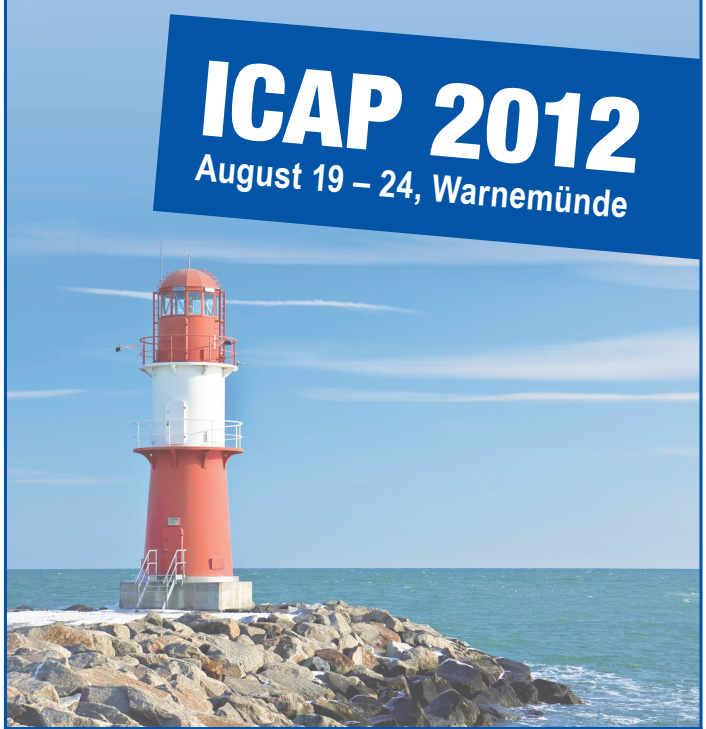
Universität Rostock

Traditio et Innovatio

August 19 – 24, 2012

11th International Computational Accelerator Physics Conference (ICAP)

Rostock-Warnemünde (Germany)
at the coast of the Baltic Sea



FACULTY OF COMPUTER SCIENCE
AND ELECTRICAL ENGINEERING

Contents

Scientific Program Committee (SPC)

Ursula van Rienen (Chair)	Universität Rostock, Germany
Thomas Weiland	Technische Universität Darmstadt, Germany
Kwok Ko	SLAC, USA
Ohmi Kazuhito	KEK, Japan

Organizing Committee (OC)

Martin Berz	Michigan State University, USA
Oliver Boine-Frankenheim	GSI, Germany
Sarah M. Cousineau	ORNL, USA
Kyoko Makino	Michigan State University, USA
Cho-Kuen Ng	SLAC, USA
Dmitri Ovsyannikov	Saint-Petersburg State University, Russia
Chris Prior	Rutherford Appleton Laboratory, UK
Stephan Russenschuck	CERN, Switzerland
Rob Ryne	Lawrence Berkeley National Laboratory, USA
Frank Schmidt	CERN, Switzerland
Tsumoro Shintake	Okinawa Institute of Science and Technology, Japan
Chuanxiang Tang	Tsinghua University, China

Local Organizing Committee

Ursula van Rienen	Universität Rostock, Germany
Dirk Hecht (Technical Editor)	Universität Rostock, Germany
Thomas Flisgen (Webmaster)	Universität Rostock, Germany

JACoW Editorial Team

Dirk Hecht	Universität Rostock, Germany
Michaela Marx	DESY, Germany
Volker RW Schaa	GSI, Germany

COMPUTATIONAL CHALLENGES IN ESS

H. Danared, M. Eshraqi, E. Laface, R. Miyamoto, S. Molloy and A. Ponton,
European Spallation Source, Box 176, SE-221 00 Lund, Sweden

Abstract

The European Spallation Source, ESS, will be based on a 2.5-GeV, 50-mA superconducting proton linac delivering 5 MW of beam power to a rotating tungsten target. ESS is a challenging project in many respects, not the least with respect to RF power and RF sources. Because of the high power, relative beam losses must be very small to avoid activation and allow hands-on maintenance of accelerator components. The beam-dynamics calculations required to ensure these low beam losses are thus another challenge. This paper gives an introduction to ESS and the ESS linac, discusses briefly computational aspects in general, and presents specific examples of computational problems from beam dynamics and RF cavity design as well as efforts initiated to benchmark beam-dynamics codes for beam parameters relevant to ESS.

ESS AND THE ESS LINAC

Lund was selected as the site for the European Spallation Source, ESS, in 2009, and a year later, the ESS Accelerator Design Update, ADU, Project was launched. This project is a collaboration between universities and institutes in five European countries with additional participation and support from accelerator laboratories in many countries inside and outside of Europe. One of the primary outputs from the ADU Collaboration and of the ESS design effort as a whole will be a Technical Design Report at the end of 2012. This will be accompanied by a cost report, time schedule and other documents needed for the final approval of the construction of ESS.

Many of the parameters of the current ESS project [1] are similar to those of the ESS Design Study from 2003/2004 [2]. A major change is that the short-pulse source has been abandoned, based on studies showing that a large fraction of the science foreseen at ESS can be done as well or better with a long-pulse source [3]. Thus, no accumulator ring is required, and the linac can accelerate protons instead of H^- ions. Significant progress has also been made in the field of superconducting RF since 2003. As a consequence, the CCL copper cavities have been replaced by superconducting spoke cavities, and the output energy of the linac has increased while the current has decreased.

Table 1: Selected ESS Top-level Parameters

Parameter	Unit	Value
Average beam power	MW	5
Proton kinetic energy	GeV	2.5
Average macro-pulse current	mA	50
Macro-pulse length	ms	2.86
Pulse repetition rate	Hz	14
Maximum cavity surface field	MV/m	40

Top-level parameters of the ESS linac are summarized in Table 1. The linac will accelerate 50 mA of protons to 2.5 GeV in pulses that are 2.86 ms long and come with a repetition rate of 14 Hz. This implies that the average beam power on the target will be 5 MW, as in the 2003 ESS study, and the peak power will be 125 MW. The linac will have a normal-conducting front-end up to 79 MeV followed by three families of superconducting cavities and a high-energy beam transport to the spallation target which will consist of a rotating tungsten wheel.

A go-ahead for construction is expected in early 2013. Then will follow an intense period of detailed design and prototyping. The most critical components are cryomodules and RF sources, where worldwide production capacity will be a limiting factor. Acceptance tests at the site in Lund and installation work in the linac tunnel and the klystron gallery will also be time-consuming. Nevertheless, the current plan calls for first neutrons from the spallation target in 2019.

LINAC DESIGN

The configuration of the current, May 2012 Baseline linac is shown schematically in Fig. 1, and selected linac parameters are listed in Table 2 [4]. The warm linac has contributions from INFN Catania, CEA Saclay, ESS-Bilbao and INFN Legnaro, the superconducting cavities and their cryomodules are designed at IPN Orsay and CEA Saclay, and the HEBT will come from ISA in Aarhus.

The 50-mA proton beam is produced in a pulsed microwave-discharge source on a platform at 75 kV. A low-energy beam transport, LEBT, with two solenoid

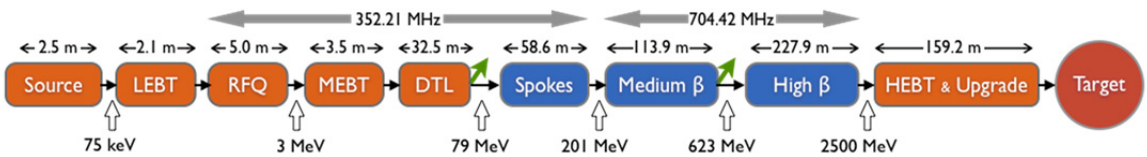


Figure 1: Schematic layout of the ESS linac [4]. Blue colour represents superconducting sections and green arrows locations where the beam could be extracted at intermediate energies.

Table 2: Selected Linac Parameters as of May 2012

Parameter	Device	Value
Length (m)	LEBT	2.1
	RFQ	5.0
	MEBT	3.5
	DTL	32.5
	Spokes	58.6
	Medium beta	113.9
	High beta	227.9
No. of cryomodules	Spokes	14
	Medium beta	15
	High beta	30
Cavities per module	Spokes	2
	Medium beta	4
	High beta	4
Optimal beta Geometrical beta	Spokes	0.50
	Medium beta	0.67
	High beta	0.92
Transition energy (MeV)	Source–RFQ	0.075
	RFQ–DTL	3
	DTL–spokes	79
	Spokes–medium beta	201
	Medium–high beta	623

magnets as focusing elements brings the beam to the entrance of the RFQ. The LEBT has a chopper that cuts away the beam while the proton pulses from the ion source stabilize, preventing a beam with off-nominal parameters from being accelerated in the RFQ and lost at high energy.

The 4-vane RFQ accelerates the beam to 3 MeV with small losses and a minimal emittance growth. It is designed specifically for ESS but it is based on the IPHI RFQ at Saclay. The RF frequency of the RFQ and the warm linac is 352.21 MHz.

After the RFQ there is a medium-energy beam transport, MEBT, with three buncher cavities and 10 quadrupole magnets. The MEBT has several different functions: it has optics to match and steer the beam from the RFQ into the drift-tube linac, it has a comprehensive set of beam-instrumentation devices, it has a chopper which acts faster than the LEBT chopper since space-charge neutralization is not an issue in the MEBT, and it allows collimation of the transverse particle distribution.

A drift-tube linac, DTL, with four tanks takes the beam from 3 MeV to 79 MeV. It has a FODO structure with permanent-magnet quadrupoles. Every second drift tube is empty or used for steering magnets and beam-position monitors.

The superconducting linac has three types of cavities: double-spoke resonators, five-cell medium-beta elliptical cavities and five-cell high-beta elliptical cavities. The May 2012 linac has 14 spoke cryomodules with two double-spoke resonators in each, and between the cryomodules there are warm quadrupole doublets. The spoke resonators operate at 352.21 MHz like the warm linac, but then there is a frequency doubling to the 704.42 MHz of the elliptical cavities. There are 15 medium-beta cryomodules with four cavities in each and quadrupole doublets between, and there are 30 high-beta cryomodules with four cavities in each and quadrupole doublets between every second cryomodule.

All accelerating structures will be powered by klystrons, except the spoke resonators where tetrodes will be used. With one klystron per elliptical cavity plus a few for the warm linac, there will be close to 200 large klystrons and almost 100 modulators since one modulator will drive two klystrons. The density of components in the klystron building would become too high if these were to be positioned linearly. Instead they will be located in groups of eight klystrons and four modulators across the klystron building.

After the last cryomodule there is 100 m of tunnel where additional cryomodules can be installed for an energy upgrade. Then the beam is brought from the tunnel to the spallation target at the surface through two vertical bends and an expansion section. Quadrupole and octupole magnets are used to blow the beam up onto the desired profile of the proton-beam window and the target window.

SIMULATIONS AND BEAM DYNAMICS

Together with the large engineering efforts required to design, prototype, manufacture and install the accelerating structures and the massive RF system, one of the great challenges with the ESS linac is to accelerate and transport the 5 MW beam without losing more than about 1 W/m. This limit is set by activation, and a higher beam loss would make hands-on maintenance of the linac components difficult and time-consuming.

The design of the ESS linac uses beam-physics laws, rules-of-thumb and experience gained from past high-current linacs. Extrapolating from earlier machines, it seems reasonable to believe that the goal of losses of at most 1 W/m can be reached. Still, it is highly desirable to be able to use computational methods to predict beam loss and to verify that the lattice and the optics are optimal with respect to beam loss. This requires computational tools where all relevant physics is included, where non-linearities are taken into account to a sufficient degree, etc. The input from the linac design must be accurate, and in particular it must include misalignments, field errors and other static and dynamic deviations from the ideal linac. Simulations have to be performed from start to end, and the starting conditions must be well understood for the final result to be realistic. Losses of course occur from particles in the tails of the distributions, so these have to be modelled with high accuracy, generally requiring a large total amount of particles in the simulations, thus needing large amounts of cpu power. All in all, realistic simulations of beam losses are extremely demanding, and it is in fact not clear to us which level of accuracy one can reach.

The layout and optics of the ESS linac has so far been designed using the TraceWin suite of codes [5], although comparisons with other codes have been initiated (see below). Care has been taken to match the different sections of cavity families well to each other and to have

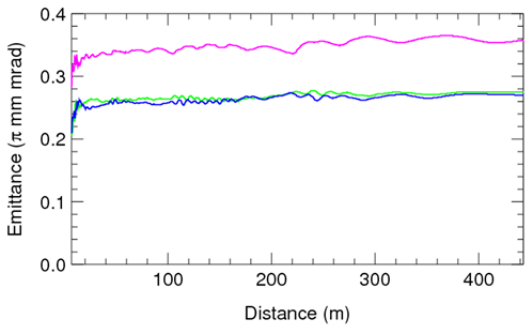


Figure 2: Normalized RMS emittances from the entrance of the MEBT to the end of the high-beta elliptical section in horizontal (green), vertical (blue) and longitudinal (violet) planes. The input distribution, at the entrance of the MEBT, is a 6D Gaussian truncated at 3 sigma with 0.21/0.21/0.28 π mm mrad.

a smooth phase advance in all planes. As an example of a result [6], the beam emittance in the three planes is plotted in Fig. 2, based on tracking of 100 000 macro particles. The RMS emittance is shown from the input of the MEBT until the end of the superconducting linac, and it is seen that the total emittance growth after the MEBT is not more than 10% in any of the planes. As already stated, however, predictions of beam loss need more aspects of the beam dynamics to be investigated, and, for instance, detailed error studies are going on at present.

Another example of beam-dynamics simulations for the ESS linac concerns the influence of unwanted cavity modes. Fundamental passband modes could prove to be dangerous due to their high *R/Q* compared to the accelerating modes at certain velocities. To explore this effect, a drift-kick-drift model was employed [7].

A pulse train of one million point-like bunches is tracked through the superconducting section of a linac, and the energy and time error generated by beam-induced modes is calculated. Figure 3 shows the resulting pulse phase space at the exit of two linacs with different velocity partitioning over the three cavity families [4, 8]. Of concern is when the growth due to passband modes (middle column) with respect to the case when no modes are acting (left column) is larger than that produced by

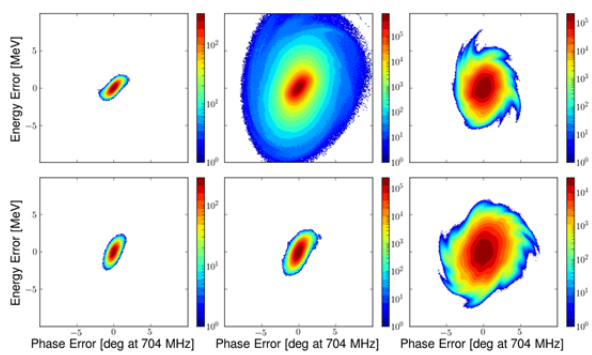


Figure 3: Pulse phase space at exit of two linacs (upper and lower rows) when no modes are acting (left), passband modes are acting (middle), there are uniformly distributed RF errors (right).

acceptable RF errors (right column). It can be seen that the linac of the top row, which represents an earlier ESS layout, is susceptible to these modes. The bottom row, corresponding to the present linac layout, shows significantly better performance.

CODE BENCHMARKING

Although TraceWin has been used so far for the ESS beam-dynamics calculations, an effort to compare the results with those of other codes has been initiated. Similar studies have been performed previously, see e.g. [9], but for different beam parameters. In the first phase, TraceWin, MADX-PTC [10, 11] and IMPACT [12] have been compared [13] with respect to acceleration and beam envelope in the transverse plane.

MADX-PTC

MADX is a widely used code for the design and study of accelerator rings as well as beam lines [10]. It does both calculation of lattice parameters and particle tracking. Complemented with the PTC library [11], MADX is capable of calculating six dimensional beam dynamics parameters and beam acceleration [14, 15]. Although the MADX and PTC already include the RF cavity as a lattice element, the included model does not have important details, such as the dependence of the longitudinal and transverse kicks on the transit time factor and its derivative as well as an offset between the electrical and mechanical centre. In our simulation with MADX-PTC, a finite length RF cavity is modelled as drift-kick-drift, where the kick includes a 4×4 thin-lens matrix modelling the transverse defocusing effect. Since a thin-lens matrix cannot be included in PTC, this simplified modelling limits us to use MADX only to study the transverse dynamics and PTC can be used to study only the longitudinal dynamics. To properly perform three-dimensional beam-dynamics calculations in a proton linac, it is ideal to develop a model of a finite length cavity with sufficient details together with the space charge effect, which is not included in the official version of MADX-PTC yet, under the framework of MADX-PTC in near future.

IMPACT

IMPACT is a particle tracking code developed at LBNL [12]. It uses a split-operator method based on a symplectic treatment of Hamilton's equations of motion and has options of both first and third order (Lorentz) integrators. In addition, it uses field maps for RF cavities, thus providing a more realistic form of tracking than codes based on the instantaneous kick-in-gap method. The space charge force is calculated using fast Fourier transform methods and a variety of boundary conditions are available. IMPACT has been regularly benchmarked with respect to other codes [9] and was also used to model the J-PARC and SNS linacs; as a result, the authors feel confident that it represents a suitable tool to check the results of the TraceWin. Comparison between the TraceWin and IMPACT is not straightforward due to

differences in conventions for the beam and lattice parameters and efforts are made to carry out a reliable comparison. The efforts include preparing a script to construct realistic field maps for drift tubes and cavities from the lattice file of the TraceWin and preparing another script to convert the relative cavity phase used by the TraceWin to the absolute cavity phase used by IMPACT based on iterative phase scanning.

Results

Figure 4 compares the kinetic energy of the beam at each longitudinal location starting from the entrance of the MEBT. The results from the TraceWin and PTC are from the lattice parameter calculations and that from the IMPACT is from tracking. The three codes are in very good agreement as it should be the case.

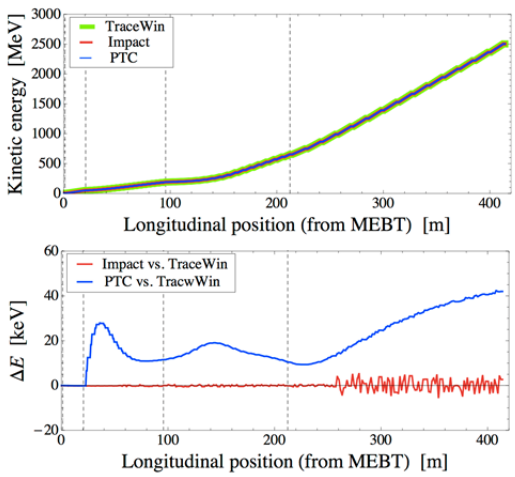


Figure 4: Comparison of kinetic energies calculated by TraceWin, IMPACT and MADX-PTC. Dotted lines represent transitions between two sections.

Figure 5 compares the transverse RMS beam size for the case of the zero current and hence with no effect of the space charge force. As the case of Fig. 2, the results from the TraceWin and MADX are from the lattice parameter calculations and that from the IMPACT is from the tracking. The third and fourth plots show the relative differences of the MADX and IMPACT with respect to the TraceWin. In TraceWin and MADX, each cell of a cavity is treated as drift-kick-drift but the kicks of the two codes are identical only up to the first order. This difference in the modelling generates the difference in the RMS sizes in the DTL. The amplitudes of $\Delta\sigma_x/\sigma_x$ and $\Delta\sigma_y/\sigma_y$ remain on the same level in sections following the DTL, suggesting the difference between TraceWin and MADX-PTC is only in the DTL and may not be significant. On the other hand, some discrepancies are evident between the TraceWin and IMPACT, considered to be due to the different orders of tracking and the use of field maps in the IMPACT but not in TraceWin.

A further comparison is in Fig. 6, which shows the transverse RMS beam sizes for the case of the full current (50 mA) based on the tracking with TraceWin and IMPACT. For the initial particle distribution, a six

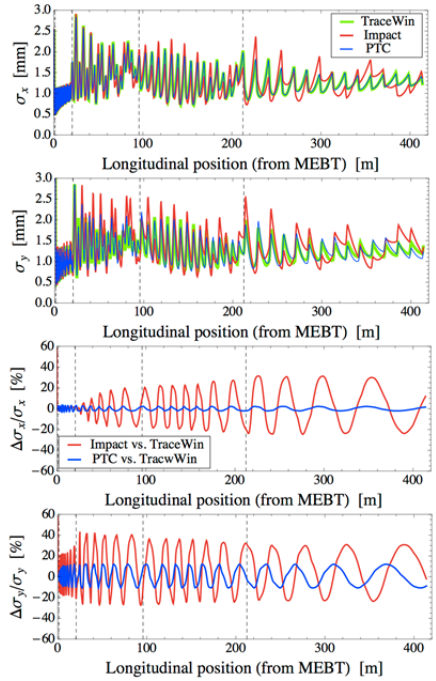


Figure 5: Comparison of RMS beam size calculated by TraceWin, IMPACT and MADX-PTC for zero current. Dotted lines represent transitions between two sections.

dimensional waterbag distribution is used. The nonuniformity of the oscillations, differences between the horizontal and vertical planes and general nonlinear growth are evident, suggesting that much needs to be done to optimize the system, not only for basic beam dynamics but also for non-linear effects.

These preliminary studies show that the three codes under consideration are in good agreement for the dynamics of the beam core in the linear regime. Nevertheless a campaign of more detailed code comparisons using the ESS linac must be performed in order to get confidence in understanding linear as well as the non-linear behaviour of the linac beam in cases both with and without errors.

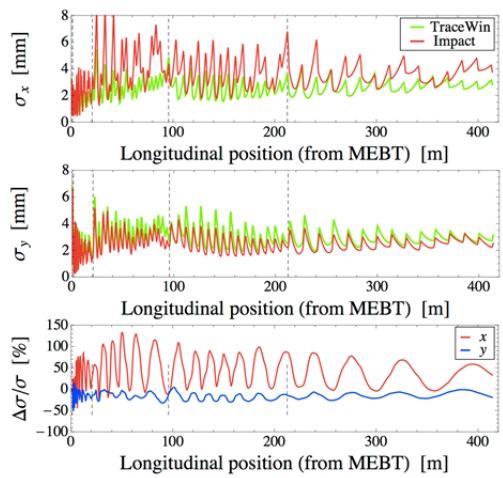


Figure 6: RMS beam envelope from IMPACT simulation with full space-charge.

COUPLER MULTIPACTOR

It is known that multipactor (MP) within higher order mode (HOM) couplers can cause considerable problems with the operation of high power superconducting cavities [16], and so it is of significant importance for the successful operation of ESS to understand any issues that may arise from installation of these couplers.

The development of a MP cascade within a coupler can have several effects, each of which may be catastrophic to the operation of the cavity:

- Absorption of the power within the accelerating mode of the cavity, thereby dropping its quality factor beyond acceptable limits.
- The increased thermal load due to absorption of the energy of the MP electrons can alter the physical geometry of the coupler, resulting in significant changes to its RF characteristics.
- A MP cascade between two points in the coupler introduces an additional conductive path in its equivalent circuit, and therefore totally changes its response to EM fields.

Note that the latter two effects result in the fundamental power being coupled out of these ports, resulting in the likely destruction of any attached electronics.

MP simulations using the ACE3P codes [17] have begun. The trajectories of electrons emitted around the inner walls of the couplers were simulated, with a postprocessing step involving scaling any resonances by a typical secondary electron yield for niobium (see Fig. 7).

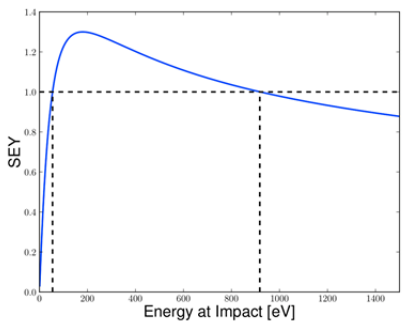


Figure 7: Secondary emission yield curve used in the postprocessing of simulation results.

Figure 8 shows a comparison of the MP behaviour of two HOM couplers proposed for use in ESS. It can be seen that one coupler displays very low-level, broad-band, MP, while the alternative has no broadband characteristics, however shows a very large, low energy, spike. Each of these effects could be catastrophic for the operation of the cavity, and so should be investigated in detail.

A remaining question relating to the computational aspects of this study is to what extent the assumptions relied on by the physics engine within this code impacts the final result. For example, the ACE3P codes always emit electrons normally to the cavity wall, and do not implement a statistical spread in the emission energy. In addition, they do not take account of space charge when tracking the emitted particles. Each of these effects may have a considerable effect on the outcome of the calculations.

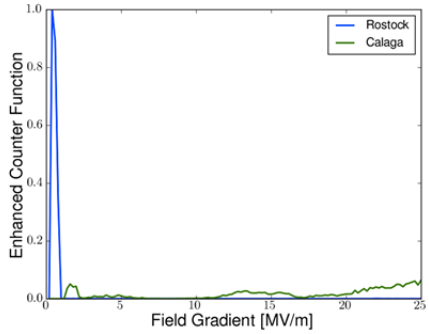


Figure 8: Comparison of the MP found in the two proposed HOM couplers [18].

MULTI-CAVITY FIELD EMISSION

As seen in [16], the statistics of the field emitted (FE) electrons impacting in the end-groups of a particular cavity are correlated with the settings of neighbouring cavities, implying that FE electrons are being transported throughout the cryomodule.

In order to investigate this effect, the ACE3P codes were used to track electrons emitted from the surface of one cavity throughout the volume enclosed by two cavities. This was done for a range of phase differences between the cavities in order to determine the effect on the impact locations of each of these electrons.

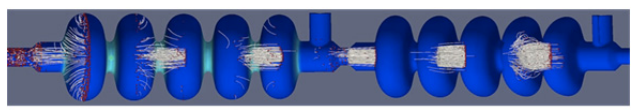


Figure 9: A simulation of FE trajectories in two cavities.

Figure 9 shows a frame from an animation of the transport of FE electrons by the accelerating field from the "most upstream" cell of one cavity into a

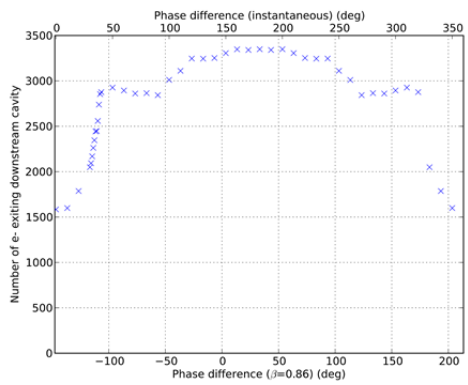


Figure 10: The integrated current emerging from the down-stream end of a two-cavity system plotted against their phase difference. The instantaneous phase difference (top axis), and that observed by a particle with $\beta = 0.86$ (bottom axis) is plotted.

neighbouring structure.

Figure 10 shows the dependence on phase difference of the current emerging from the downstream cavity. The ESS cavities will operate between approx. −80°and +10°, and so should expect to see a high degree of FE activity spread throughout the cryomodules.

REFERENCES

[1] ESS Conceptual Design Report, Ed. S. Peggs, ESS Report ESS-2012-1 (2012).
[2] The ESS Project, Vol III, Technical Report, www.esss.se/documents/VolIII.pdf and ESS SC Reference Linac, Vol III Update Report, www.esss.se/documents/Chapt_1_Vol3upd.pdf.
[3] F. Mezei, J. Neutron Res. **6** (1997) 3.
[4] H. Danared, Proc. IPAC 2012, p. 2073.
[5] R. Duperrier, N. Pichoff and D. Uriot, Proc. ICCS02, Springer-Verlag 2002, p. 411.
[6] M. Eshraqi *et al.*, Proc. IPAC 2012, p. 3935.
[7] M. Schuh *et al.*, Phys. Rev. ST Accel. Beams **14** (2011) 051001.
[8] R. Ainsworth, Thesis in preparation, Royal Holloway, University of London.
[9] L. Groening *et al.*, Proc ICAP 2009, p. 48.
[10] F. Schmidt, G. Iselin and H. Grote, MAD-X, http://mad.web.cern.ch.
[11] E. Forest, F. Schmidt and E. McIntosh, KEK-REPORT-2002-3.
[12] J. Qiang *et al.*, J. Comp. Phys. **163** (2000) 434.
[13] E. Laface *et al.*, Proc. IPAC 2012, p. 3841.
[14] F. Schmidt *et al.*, CERN-AB-2005-036.
[15] K. Skowronski *et al.*, Proc. ICAP 2006, p. 209.
[16] S. Kim *et al.*, Proc. PAC 2007, p. 2511.
[17] K. Ko *et al.*, Proc. LINAC 2010, p. 1028.
[18] S. Molloy *et al.*, Proc. IPAC 2011, p. 190.

PROJECT OVERVIEW AND COMPUTATIONAL NEEDS TO MEASURE ELECTRIC DIPOLE MOMENTS AT STORAGE RINGS

A. Lehrach# on behalf of the JEDI collaboration*, Forschungszentrum Jülich, Germany

Abstract

Different approaches to measure Electric Dipole Moments (EDMs) of proton, deuteron and light nuclei are pursued at Brookhaven National Laboratory (BNL) and Forschungszentrum Jülich (FZJ) with an ultimate goal to reach a sensitivity of 10^{-29} e·cm in a dedicated storage ring. As an intermediate step, a first direct EDM measurement of protons and deuterons at 10^{-24} e·cm sensitivity level will be carried out in a conventional storage ring, the Cooler Synchrotron COSY at FZJ [1].

Full spin-tracking simulations of the entire experiment are absolutely crucial to explore the feasibility of the planned storage ring EDM experiments and to investigate systematic limitations. For a detailed study of particle and spin dynamics during the storage and buildup of the EDM signal, one needs to track a large sample of particles for billion of turns.

INTRODUCTION

Permanent EDMs of fundamental particles violate both time invariance $\mathcal{T}$ and parity $\mathcal{P}$. Assuming the $\mathcal{CPT}$ theorem this implies $\mathcal{CP}$ violation. The Standard Model (SM) predicts non-vanishing EDMs, their magnitudes, however, are expected to be unobservably small in the near future. Hence, the discovery of a non-zero EDM would be a signal for "new physics" beyond the SM. It is mandatory to measure EDMs on different species of particles in order to disentangle various sources of $\mathcal{CP}$ violation. While neutron EDM experiments are pursued at many different places worldwide, no such direct measurements have been conducted yet for protons and other light nuclei due to special difficulties of applying electric fields on charged particles. EDM experiments with charged particles are only possible at storage rings.

As a first step towards EDM searches in storage rings we proposed R&D work to be carried out at the Cooler Synchrotron COSY [2,3], then perform a first direct EDM measurement of a charged particle in a storage ring at the Cooler Synchrotron COSY, and on a longer time scale construct a dedicated storage ring [3,4].

The COSY Infinity simulation program [5] and its updates are used to simulate beam and spin motion in storage rings. To study subtle effects and simulate the particle and spin dynamics during the storage and buildup of the EDM signal, one needs custom-tailored fast trackers capable of following up to 100 billion turns for samples of up to 10^6 particles. Given the complexity of the tasks, particle and spin dynamics simulations performed with COSY Infinity must be benchmarked with other simulation programs and experiments performed at the Cooler Synchrotron COSY, to ensure the required accuracy of the obtained simulation results.

EDM MEASUREMENTS AT STORAGE RINGS

Principle

The principle of every EDM measurement (e.g., neutral and charged particles, atom, molecule) is the interaction of an electric field with the dipole moment of the particle. In the center-of-mass system of a particle electric dipole moments $\vec{d}$ couple to electric fields, whereas magnetic dipole moments $\vec{\mu}$ couple to magnetic fields. The spin precession in the presence of both electric and magnetic fields is given by:

$$\frac{d\vec{S}}{dt} = \vec{d}\times\vec{E}^{*} + \vec{\mu}\times\vec{B}^{*}. \qquad (1)$$

Here, $\vec{E}^{*}$ and $\vec{B}^{*}$ denote the electric and magnetic fields in the particle rest frame. In case of moving particles in a circular accelerator or storage ring, the spin motion is covered by the Thomas-BMT equation and its extension for EDM:

$$\frac{d\vec{S}}{dt} = \vec{\Omega}\times\vec{S}$$

$$\vec{\Omega} = \frac{e\hbar}{mc}\left\{G\vec{B} + \left(G - \frac{1}{\gamma^2 - 1}\right)\left(\vec{v}\times\vec{E}\right) + \frac{\eta}{2}\left(\vec{E} + \vec{v}\times\vec{B}\right)\right\}. \qquad (2)$$

$\vec{E}$ and $\vec{B}$ denote the electric and magnetic fields in the laboratory system, with the constraint, that the electric and magnetic fields are perpendicular to the velocity vector $\vec{v}$ of the particle beam: $\vec{v}\cdot\vec{B} = \vec{v}\cdot\vec{E} = 0$.

The gyromagnetic anomaly G, magnetic and electric dipole moments are given by:

$$G = \frac{g-2}{2},\ \vec{\mu} = 2(G+1)\frac{e\hbar}{2mc}\vec{S},\ \text{and}\ \vec{d} = \eta\frac{e\hbar}{2mc}\vec{S}. \qquad (3)$$

Methods

Starting from equation 2, different approaches are possible to excite spin rotations via the electric dipole moment:

*JEDI spokespersons: A. Lehrach, J. Pretz, and F. Rathmann

#a.lehrach@fz-juelich.de

1. Frozen spin method [6], where the bending fields in a storage ring are adjusted according to the particle momentum in such a way that the longitudinally polarized spins of the particle beam are kept aligned ("frozen") with their momenta. If the particle has an EDM along its spin direction, the E-fields in the rest frame of the particles will precess the spin into the vertical direction. This change of the vertical component of the beam polarization from early to late storage times is the signature of the EDM signal.

   For pure electric bending fields (i.e. $\vec{B}=0$) with the additional condition that $\left(G-\frac{1}{\gamma^2-1}\right)$, equation 2 simplifies to

$$\frac{d\vec{S}}{dt}=\frac{e\hbar}{2mc}\eta\left(\vec{E}\times\vec{S}\right). \qquad (4)$$

The above condition can only be fulfilled for a positive gyromagnetic anomaly $G>0$ (i.e. for a proton) and a fixed particle momentum given by:

$$\gamma=\sqrt{\frac{1}{G}+1}\Leftrightarrow p_{magic}=\frac{m}{\sqrt{G}}. \qquad (5)$$

For protons a magic momentum of roughly p_{magic} = 0.7 GeV/c has to be chosen to freeze the spin in a pure electric bending field.

In a storage ring with combined electrostatic and magnetic bending fields the following condition has to be fulfilled to freeze the spin motion:

$$G\vec{B}+\left(G-\frac{1}{\gamma^2-1}\right)\left(\vec{v}\times\vec{E}\right)=0 \qquad (6)$$

In this case the radial E-field cancels the (g-2) spin precession in the vertical B-field, if:

$$E=\frac{GBv\gamma^2}{1-G\beta^2\gamma^2}\approx GBv\gamma^2, \qquad (7)$$

with the approximation that the denominator in equation (7) is close to one. Freezing the spin in a storage ring with combined electric and magnetic bending fields is possible for all particles with both positive and negative gyrogagnetic anomaly at any beam momentum.

Two collaborations, one located in the USA (srEDM-US) and the other one in Germany (JEDI, Jülich Electric Dipole moment Investigations), have put forward proposals for dedicated storage ring EDM (srEDM) experiments [3,4]. The aim is to advance the sensitivity level for EDMs of hadronic particles like the proton, deuteron, and ^{3}He nuclei to $10^{-29}e$·cm, utilizing the frozen spin method. The srEDM-US collaboration at BNL has proposed an all electric storage ring for protons, while the newly founded JEDI collaboration at FZJ is pursuing an approach to perform EDM measurements of proton, deuteron, and ^{3}He ions in one and the same storage ring with combined electric-magnetic deflectors [7].

2. Resonant method [8], were an RF-E/B spin flipper runs at a frequency tuned to the spin tune ($\gamma \cdot G \pm K$, K integer). The transverse electric field of the spin flipper would rotate the spin of stored particles away from the vertical or horizontal spin axis, building up a polarization perpendicular to the original spin direction, depending on the magnitude of the EDM.

As a first step towards EDM searches in storage rings, the JEDI collaboration has submitted a proposal [2] to study the proton and deuteron EDM with an intermediate goal of $10^{-24}e$·cm applying the resonance method [9]. It will use the existing but upgraded Cooler Synchrotron COSY, with the addition of RF-E or RF-E/B fields to enhance the EDM sensitivity. This would be the first EDM measurement in the world of a dedicated EDM storage ring experiment.

FIRST DIRECT EDM MEASUREMENT AT THE COOLER SYNCHROTRON COSY

The EDM signal would be a polarization component perpendicular to the original spin direction of the stored particles produced by the EDM-induced precession of the spin in a radio-frequency spin flipper with transverse electric field [9]. With accessible electric field gradients in the MV/m range, the spin will be tipped by a minuscule angle 10^{-13}–10^{-12} rad per single pass and the buildup of an observable polarization in the percent range demands a spin coherence during 10^{10}–10^{11} revolutions of the stored beam, which requires a coherent buildup of polarization for an extremely long time scales of 10^4–10^5 s. Coherent betatron oscillations induced by the transverse electric RF spin flipper field will lead to uncontrolled systematic rotations of the spin.

To suppress perturbations of the particle trajectory induced by the transverse electric RF spin flipper field, RF-E/B fields can be adjusted according to the "Wien-filter" condition: $\vec{E}+\vec{v}\times\vec{B}=0$. Since $\vec{E}^*=0$ for the reference particle, there will be no direct EDM effect induced by the RF spin flipper. The spin precession will still be perturbed in the RF-E/B spin flipper via the magnetic moment ($\vec{B}^*\neq 0$) in such way, that the symmetry between spins components parallel and anti-parallel along the momentum vector is broken. If this so called "magic RF Wien-filter" is operated at a resonance frequency of spin motion an EDM effect can be accumulated in the main magnets along the ring.

The different measures towards a first direct EDM measurement contain preparatory measurements at COSY, development of simulation tools and a dedicated hardware:

1. Experimental and theoretical studies of the spin coherence time,
2. investigation of systematic effects,
3. development of precision simulation programs for spin dynamics in storage rings,
4. development of a full-scale polarimeter,
5. and development of a high-power RF-E/B spin flipper system capable to operate at electric fields with more than 1 MV/m with a B-field of roughly 70 G in a frequency range of 0.1 to 1 MHz.

These measures are essential to perform a first direct EDM measurement of proton and deuteron at COSY with a statistical sensitivity of about 10^{-24} e·cm. They will be partially outlined below with an emphasis on the first two items.

Experimental and Theoretical Studies of the Spin Coherence Time

In a storage ring spin decoherence is caused by spin tune spread due to the momentum spread of the beam and spin kicks induced by nonlinear fields.

An important finding from the Jülich EDM Study Group is, that the phase of a particle entering the RF cavity and the phase of the RF spin flipper field are strongly connected to each other, making possible a mutual cancelation spin decoherence for properly chosen beam energies and RF spin flipper harmonics [9]. This prediction needs to be experimentally tested at COSY. An enhancement of the spin coherence time can also be achieved in COSY by operating the RF spin flipper in a special flattop mode. The possibility that operating an RF spin flipper might suppress spin decoherence induced by spin tune spread of stored particles would be an entirely new observation. A search for these decoherence-free energies has never been performed before and would have a strong impact on the whole EDM program.

The spin coherence time of the idle precession (without RF spin flipper on) will limit the observable spin coherence time with spin flipper on. To perform these studies the spin coherence time without RF spin flipper has to be optimized first by means of a phase-space cooling and multipole correction. First results are obtained from a recent experiment conducted at COSY in early 2011 [10]. The analysis of the experimental data is in progress and the preliminary data suggest spin coherence times for deuterons of larger than several hundred seconds.

Investigations of Systematic Effects Studies with RF Spin Flipper

Main sources of systematic errors are the alignment of the RF spin flipper fields with respect to the invariant spin field, opening angle of spin ensemble, field quality (fringe fields) of the RF-E/B spin flipper.

First the alignment angle of the RF spin flipper will be modified to investigate and suppress false spin rotations. After that beam cooling and heating will be applied to change the opening angle of the spin ensemble.

Investigations for a Required COSY Upgrade

It is obvious that a substantial improvement of COSY is required in order to reach the desired EDM sensitivity. False spin rotations as a function of closed-orbit excitation, quadrupole alignment and ring impedances will be studied. The aim of this part is to reduce systematic errors. The results will be the bases to specify the required COSY upgrade (orbit correction system, beam-position monitors (BPMs), power supply stability, magnet alignment and ring impedances):

- An improved closed-orbit control system for orbit correction in the micrometer range is necessary, which requires increasing the stability of correction-dipole power supplies by at least one order in magnitude. The number of correction dipoles and BPMs has to be increased significantly, since the orbit has to be controlled along the entire path length of the beam in the COSY.
- The BPM accuracy, presently limited by electronic offset and amplifier linearity, has to be substantially improved as well. Systematic errors of the closed-orbit measurement (e.g., temperature drift, beam current dependence) have to be studied in detail. In particular, a precise adjustment of the quadrupole and sextupole magnets is mandatory, and the BPMs have to be aligned with respect to the magnetic axis of these magnets. The geodetic alignment of COSY magnets has to be verified. Compensation of phase space coupling and multipole correction to high accuracy is absolutely crucial. Methods of orbit-response matrix, local orbit bumps, turn-by-turn orbit measurements, and beam-based orbit alignment have to be applied to significantly increase the precision of orbit control and knowledge of machine imperfections.
- Beam oscillations can be excited by vibrations of magnetic fields induced by the jitter of power supplies. Investigations have to be carried out with the aim to understand and suppress these beam oscillations to a sufficiently low level, where they do not interfere anymore with the design EDM sensitivity goal.
- The interaction of the circulating beam with the surrounding vacuum chamber produces longitudinal and transverse wake fields, which can lead to transverse and longitudinal beam kicks and excite instabilities. The main sources of wake fields are generally RF cavities and kickers, finite conductivity of wall material, discontinuities of the chamber geometry due to transitions, bellows and beam-position monitors. Transitions of the vacuum chamber profile can have a large impact on transverse and longitudinal beam motion. An accurate estimation of the total impedance budget of the COSY machine has to be carried out. Depending on the outcome, those sections in conflict with the sensitivity goal will be modified.

Table 1: Development Steps of the JEDI Project

Step	Aim / scientific goal	Device / Tools	Storage ring
1	Spin coherence time studies	Low-power RF-B/E spin flipper	COSY
	Systematic error studies	Low-power RF-B/E spin flipper	COSY
2	COSY upgrade	Orbit control, magnets, …	COSY
	First direct EDM measurement at 10^{-24} e·cm	High-power RF-E/B spin flipper	Modified COSY
3	Built a dedicated all-in-one ring for p, d, ^{3}He	Common magnetic-electrostatic deflectors R&D funded by ARD (Accelerator Research and Development) of HGF	Dedicated storage ring
4	EDM measurement for p, d, ^{3}He at 10^{-29} e·cm	Data analysis, systematic error simulation and estimates	Dedicated storage ring

Overall Schedule of the JEDI Project

The stepwise approach of the JEDI Project, starting with the R&D work at COSY, the first direct measurement of a charged particle EDM at COSY, and the design of a dedicated storage ring is summarized in Table 1. The studies of spin coherence time and systematic studies will be performed with a low-power RF-E/B spin flipper with an electric field gradient of tenth of MV/m. For the first direct EDM measurement with a sensitivity goal of at 10^{-24} e·cm a high-power RF-E/B spin flipper is required with an electric field gradient of more than 1 MV/m.

For this first two development steps a timeline of more than five years is estimated. For the next two development steps, that finally lead to the construction of a dedicated EDM storage ring to reach ultimate sensitivity goal of 10^{-29} e·cm, at least another five years have to be projected.

DEVELOPMENT OF PRECISION SIMULATION PROGRAMS FOR SPIN DYNAMICS

Existing spin tracking programs have to be extended to properly simulate spin motion in presence of an electric dipole moment. The appropriate EDM kicks and electric field elements (static and RF) have to be implemented and benchmarked. Furthermore, a symplectic description of fringe fields, field errors, and misalignments of magnets has to be adapted and verified.

One urgently needs spin tracking tools capable of handling up to 10^{11} particle revolutions with highest precision in a realistically modeled storage ring and for large samples of particles. The main challenge for spin tracking including EDM induced spin rotations is the fact that the spin will only be tipped by a minuscule angle due to the EDM of the particle. In the finale dedicated storage ring the EDM rotates the spin by roughly 10^{-9} radians per turn. Assuming a ring structure which contains roughly hundred elements, the EDM rotates the spin by an angle of approximately 10^{-18} radians per element on average. Thus, a simulation program based on double precision numbers has a very serious limitation to reach the required precision. With a mantissa length of 52 bits roughly sixteen significant decimal digits can only be provided.

Full spin-tracking simulations of the entire experiment are absolutely crucial to explore in a systematic way the feasibility to reach the desired spin coherence time of 10^4–10^5 s. Even if the required spin coherence times are achievable, one needs to check whether systematic errors do not limit the anticipated sensitivity to EDMs. In order to provide the required CPU time for the simulations of spin motion with a time scale larger than tens of seconds, spin tracking programs have to be migrated to powerful computer systems or clusters.

Given the complexity of the task, and in order to ensure the credibility of the results, at least two generic simulation programs using different approaches must be developed and benchmarked with the required accuracy and efficiency:

- COSY Infinity [5], based on map generation using differential algebra and the subsequent calculation of the spin-orbital motion for an arbitrary particle. COSY Infinity and its updates are used including higher-order nonlinearities, normal form analysis, and symplectic tracking [11,12]. The upgrade of COSY Infinity will be supervised by M. Berz, the principal developer of the presently available version of this powerful tracking tool. An MPI version of COSY Infinity is already running on the computer cluster at Michigan State University. A project for the Jülich supercomputer is starting in November 2012.
- Integrating program [12], where differential equations of particle and spin motion in electric and magnetic fields are solved using Runge-Kutta integration [13].

They have been shown to be accurate to sub-part per billion levels in describing the muon (g-2) spin precession frequency. The integration step size is 0.5 ps, making it rather slow with a possible maximum tracking time of about 10 ms for a particle in the ring. This simulation programs are suitable to study effects that do not require a long numerical time.

The integrating method, even though it is slow, is used for benchmarking the results of the much more efficient COSY Infinity [14]. In addition, numerical integration of the Thomas -BMT differential equations for a spin motion with smoothly approximated parameters of orbital motion can be used to benchmark all simulation programs.

Finally, benchmarking experiments will be performed at COSY to check and to further improve the simulation tools.

CONCLUSION/OUTLOOK

Dedicated storage ring EDM experiments promise to reach a sensitivity level of 10^{-29} e·cm for charged hadronic EDMs. The two collaborations, srEDM-US and JEDI, are pursuing complementary approaches to perform EDM measurements in dedicated storage rings. As an intermediate step, a first direct EDM measurement of protons and deuterons at 10^{-24} e·cm sensitivity level will be carried out in the Cooler Synchrotron COSY.

Full spin-tracking simulations of the entire experiment are absolutely crucial to explore the feasibility of the planned storage ring EDM experiments and to investigate systematic limitations. For a detailed study of particle and spin dynamics during the storage and buildup of the EDM signal, one needs to track a large sample of particles for billion of turns. Existing spin tracking programs like COSY Infinity have to be extended to properly simulate spin motion in presence of an electric dipole moment. In order to provide the required CPU time for the simulations of spin motion with a time scale larger than tens of seconds, spin tracking programs have to be migrated to powerful computer systems or clusters. At least two generic simulation programs using different approaches must be developed and benchmarked with the required accuracy and efficiency. In addition, benchmarking experiments will be performed at the Cooler Synchrotron COSY to check and to further improve the simulation tools. Finally, the layout of a dedicated storage ring has to be optimized by a full simulation of spin motion.

ACKNOWLEDGMENT

The author would like to thank all members of the JEDI and srEDM-US collaborations for fruitful discussions. I'm especially very grateful to M. Berz and K. Makino for supervised the upgrades of COSY Infinity.

REFERENCES

[1] R. Maier, "Cooler Synchrotron COSY – performance and perspectives", Nucl. Instr. Meth. A 390, 1 (1997).
[2] Proposal by EDM@COSY to COSY-PAC, "Studies of the Horizontal Spin Coherence Lifetime", Proposal No. 176.6 (2012).
[3] Proposal by JEDI to COSY-PAC, "Search for Permanent Electric Dipole Moments at COSY. Step 1: Spin coherence and systematic error studies", Proposal No. 216 (2012); http://www2.fz-juelich.de/ikp/jedi/documents/proposals/20120503_jedi_proposal_216.pdf
[4] Storage Ring EDM Collaboration, "A Proposal to Measure the Proton Electric Dipole Moment with 10^{-29} e·cm Sensitivity", Brookhaven National Laboratory, October 2011.
[5] M. Berz, "Computational Aspects of Design and Simulation: COSY Infinity", Nucl. Inst. Meth. A 298, 473 (1990).
[6] F.J.M. Farley et al., Phys. Rev. Lett. 93, 052001 (2004); hep-ex/0307006.
[7] R. Talman, "All-In-One Storage Ring Lattice for Baryon EDM Measurements", Jülich Internal Report, February 2012.
[8] Y.F. Orlov et al., Phys. Rev. Lett. 96, 214802 (2006); hep-ex/0605022.
[9] A. Lehrach, B. Lorentz, W. Morse, N. Nikolaev, F. Rathmann, "Precursor Experiments to Search for Permanent Electric Dipole Moments (EDMs) of Protons and Deuterons at COSY", Proc. of Advanced XIV Workshop on High Energy Spin Physics 2012, Dubna, Russia; http://arxiv.org/abs:1201.5773
[10] E.J. Stephenson et al., "Synchrotron oscillation effects on an RF-solenoid spin resonance", draft version, unpublished.
[11] Yu. Senichev, A. Lehrach, R. Maier, D. Zyuzin, "The Spin Aberration of Polarized Beam in Electrostatic Rings", Proc. of the 2nd International Particle Accelerator Conference IPAC'11 (2011), San Sebastian, Spain, WEPC067 (2011).
[12] Y. Senichev et al., these proceedings.
[13] W. Oevel, M. Sofroniou, Symplectic Runge-Kutta Schemes II: Classification of Symmetric Methods, preprint, University of Paderborn, Germany, 1996.
[14] D. Zyuzin, et al., "Comparison of different numerical modeling methods for beam dynamics in electrostatic rings", Proc. of the 3rd International Particle Accelerator Conference IPAC'12 (2012), New Orleans, TUPPC071 (2012).

SIMULATING THE LHC COLLIMATION SYSTEM WITH THE ACCELERATOR PHYSICS LIBRARY MERLIN, AND LOSS MAP RESULTS

J.G. Molson*, R.B. Appleby, M. Serluca, A. Toader
The University of Manchester and the Cockcroft Institute, UK
R.J. Barlow, University of Huddersfield, UK

Abstract

We present large scale simulations of the LHC collimation system using the MERLIN code for calculations of loss maps, currently using up to 1.5×10^9 halo particles. In the dispersion suppressors following the collimation regions, protons that have undergone diffractive interactions can be lost into the cold magnets. This causes radiation damage and could possibly cause a magnet quench in the future with higher stored beam energies. In order to correctly simulate the loss rates in these regions, a high statistics physics simulation must be created that includes both accurate beam physics, and an accurate description of the scattering of a 7 TeV proton in bulk materials. The current version includes the ability to simulate new possible materials for upgraded collimators, and advances to beam-collimator interactions, including proton-nucleus interactions using the Donnachie-Landshoff Regge-Pomeron scattering model. Magnet alignment and field errors are included, in addition to collimator jaw alignment errors, and their effects on the beam losses are systematically estimated. Collimator wakefield simulations are now fully parallel via MPI, and many other speed enhancements have been made.

INTRODUCTION

The LHC is a superconducting 7 TeV proton-proton collider with a high nominal stored beam energy (360MJ) and a low quench limit on the superconducting magnets $\left(4.5\mathrm{mW/cm}^3\right)$ [1]. To protect the machine from this high stored energy, the LHC is equipped with an efficient collimation system to collimate halo particles and prevent quenching, in addition to reducing the background at the experimental regions and preventing radiation damage to sensitive electronics. There exist two collimation regions - one in interaction region 7 (IR7) which contains a series of betatron collimators for transverse collimation. The primary collimators in this regions are the aperture restriction in the machine. In IR3 there is a region of beam dispersion to perform momentum collimation.

Of critical importance are regions known as the dispersion suppressors, which match the long straight section (LSS) optics to the periodic optics of the arcs. In these regions, the dispersion rises rapidly, and any protons that have undergone any interactions in the collimators that causes them to lose momentum may be lost in a localised region, see Figure 4. Due to this, one must have an accurate simulation of the accelerator optics, the machine physical aperture, and the scattering physics of a proton inside a collimator jaw.

Merlin is a C++ accelerator physics library [2] initially developed for the ILC beam delivery system [3, 4], then later extended to model the ILC damping rings. Merlin has been extended to be used for large scale proton collimation simulations, with the aim of providing an accurate simulation of the Large Hadron Collider (LHC) collimation system, and any future upgrades. In this paper we describe the developments of the Merlin code to enable study of the LHC collimation system and present beam loss maps for 2012 running.

THE MERLIN ACCELERATOR PHYSICS LIBRARY

The Merlin library consists of a large number of classes designed to simulate a particle accelerator, and any additional systems required. The classes can be split into three main categories.

The *AcceleratorModel* and associated classes deal with the creation and storage of an accelerator lattice. The lattice is stored as a series of *AcceleratorComponent* classes, which contain information about each element. Different types of accelerator component are child classes of the main *AcceleratorComponent* class. These contain pointers to classes describing specific properties of the element: *EMField* describes any electromagnetic fields inside the element, *AcceleratorGeometry* describes any geometric transforms that the element has undergone, *Aperture* describes the experimental beam pipe, and *WakePotentials* describe any wake fields that exist for this element class. Input can take place via multiple methods: the direct creation and addition of elements, via the MAD-X [5] TFS output (*MADInterface*), or tape format (*XTFFInterface*), both of which create an *AcceleratorModel* as output.

The *ParticleTracker* and associated classes deal with the transport of particles along the accelerator optical lattice, including stepping between elements, and within individual elements, whilst applying additional physics processes at appropriate locations. These create integrator sets for tracking, and individual integrators can be overridden for selected class types, e.g. crab cavities. The *ParticleTracker* takes as its input a *ParticleBunch* class, and a *Beamline*, where the *ParticleBunch* can be one of many different types, e.g. gaussian, flat and ring amongst others. The *Beamline* is a subsection of an *AcceleratorModel*, and bunches can be passed between multiple trackers, allowing

* james.molson@hep.manchester.ac.uk

situations such as the transfer between different accelerators to be simulated.

A series of *BunchProcess* classes exist to apply additional physics within elements. These can include wakefields, collimation, synchrotron radiation, and others. Templates exist for such classes, and it is straightforward for new users to add additional physics of their choice via this method, without having to adjust other parts of the library.

This modular design allows a user to use as much or as little of the library as they wish. In addition, if one wishes to investigate additional physics relevant to their accelerator system, a new tracking code does not need to be written, but simply a new *BunchProcesses* class can be created. An example simulation run is shown in Figure 1.

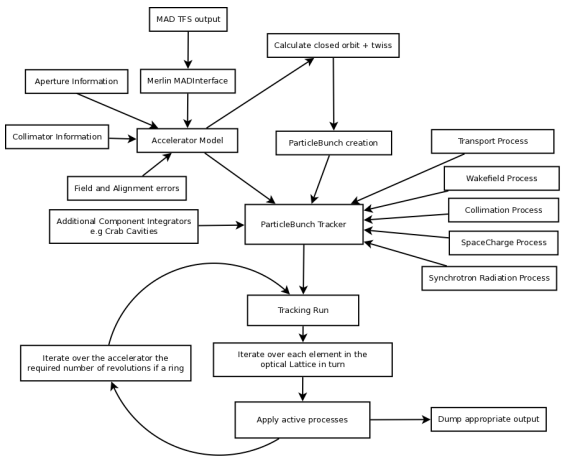


Figure 1: An example logical flow of a Merlin run.

LHC LOSS MAPS RESULTS

Merlin is used to simulate the 2012 as-built optics of the LHC in order to generate loss maps. These can be used to define the maximum possible safe beam current, and indicate areas which may need additional collimators or shielding. The first stage of this calculation is the construction of an accurate optical model of the LHC in Merlin. The machine optics are generated by MAD-X and are used as the input to Merlin. Inside Merlin, the *LatticeFunctions* class calculates beam parameters using Merlin's integrators. A comparison between Merlin and MAD-X of the β-functions and the linear dispersion is shown in Figure 2, and excellent agreement is found in all regions of the machine.

Loss maps can be generated using different optics configurations, e.g. the β-function at the interaction points (β^*), beam crossing angles, and so on. The *CollimationProcess* simulates all proton-collimator interactions and performs all aperture checking. If a proton undergoes an inelastic interaction or touches the beam pipe it is considered lost. If this takes place, the particle is removed from the bunch and the location at which this takes place is recorded. This can be done at any desired longitudinal accuracy, and by default a bin size of 10cm is used.

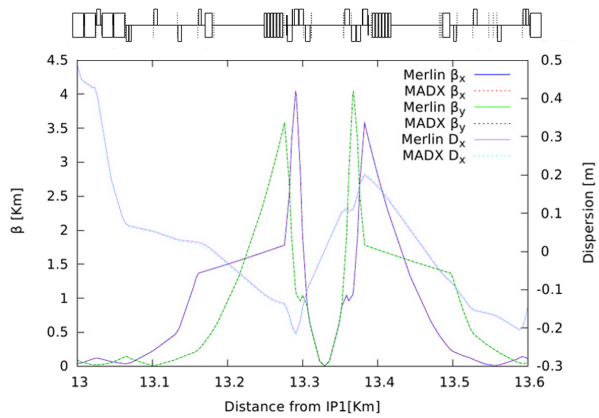


Figure 2: A comparison of optics between MAD-X and Merlin showing the β-functions and dispersion in IR5. Excellent agreement is found for both this region and the entire ring.

Figures 3 and 4 show the example loss map for 2012, 4 TeV running conditions for beam 1. A horizontal beam halo (a ring in x, x' normalized phase space, 0 in y, y') is used, which is then transformed into physical coordinate space. The initial impact parameter with the collimator can be adjusted, and in this case $1\mu m$ is used. Beam is injected at the closest (primary) collimators in IR7, and tracked for 200 turns.

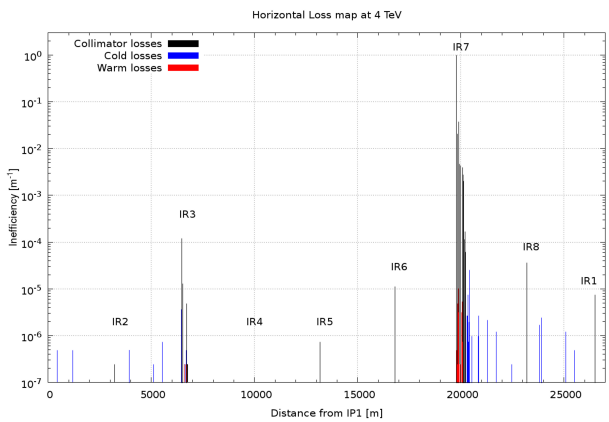


Figure 3: An example collimation loss map for 4 TeV 2012 running conditions. The initial simulated beam halo is a purely horizontal halo.

The loss map variable plotted is cleaning inefficiency around the ring, defined as:

$$\eta = \frac{n_{\text{abs}}}{\Delta s \times n_{\text{total}}}$$

where η is the inefficiency, Δs is the bin size, n_{abs} is the loss count in that bin, and n_{total} is the total number of losses.

As can be seen, the highest loss locations are in the collimation regions, specifically the IR7 betatron collimation

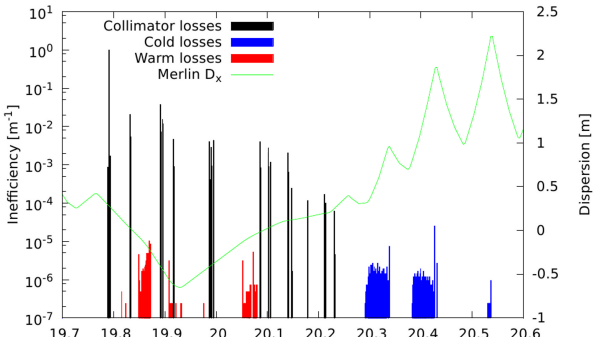


Figure 4: A zoom of the 4 TeV 2012 running conditions loss map focusing on the betatron collimation region in IR7 allowing the losses in the dispersion suppressor to be seen. The horizontal dispersion is also shown as the green line.

region. Figure 4 shows a zoom of this region. Here the effect of dispersion can be seen on the losses. Protons are lost in cold regions following the collimation region, which is minimised in the design.

Lattices have been generated in MAD-X which involve both field and alignment errors on dipole, quadrupole and sextupole magnets, to allow the effects of any lattice errors on losses to be estimated. These error configurations are then corrected using the available corrector circuits in the optical model. Collimators are aligned to the un-errored reference orbit, and then both the magnet errors are added, followed by corrections. Loss maps are generated in this configuration and it is found that as long as corrections are applied there is little quantitative difference to the loss maps generated. This includes both the locations of lost protons and the magnitude of losses.

UPDATED SCATTERING PHYSICS

In order to more accurately simulate the losses in the cold dispersion suppressor regions, more accurate simulation physics must be used over the current generation of codes. New models of proton-proton interactions have been developed, with the aim of expanding these to proton-nucleus interactions. Focus has been on two types of scattering, elastic and single diffractive scattering taking into account both theoretical considerations and experimental high energy physics data. Elastic interactions will give an angular kick to the outgoing proton, increasing the size of the beam halo, with the possibility of the proton exiting the dynamic aperture [6]. Single diffractive interactions can allow a proton to exit with an angular kick, and an energy loss [7]. This energy loss will move the outgoing proton away from the reference momentum, hence on entry to a dispersive region, they will undergo large orbit excursions, and collide with the accelerator beam pipe. Since the fits to these cross sections are mathematically highly complicated, it is faster computationally to pre-generate these distributions in an array, and interpolate them as required. The updated scattering physics will be described in a future publication.

OTHER ENHANCEMENTS

New materials, such as composites are now supported. Since Merlin is a C++ library, this is enabled via creating a new material class which inherits from the base material class, and specific access functions that are defined as virtual can be overridden. For example *GetdEdx()* will return the mean energy loss for the material mixture, whereas *GetElasticCrossSection()* will return the cross section for a randomly chosen nucleus within the material (weighted depending on the material composition). The same function calls are used for both a pure element, and material mixtures in order to model novel collimation materials.

Wakefield calculations are now fully parallel due to a parallel bunch moment calculation. Previously, particles were transferred between threads to a single node, where the wakefield calculation took place. Particles were then redistributed. This gave a speed increase over single threaded operation since standard tracking could take place in parallel. It is not the most efficient method, since a large quantity of bandwidth is required to transfer particles, and whilst this calculation takes place, CPU cores sit idle. Now the mean and standard deviation are calculated in parallel (for bunch slicing), and data is shared via a call to the *Allreduce* function.

CONCLUSION

In conclusion, the accelerator physics code Merlin has been extended in many areas to make detailed studies of the LHC collimation system and calculate halo loss maps. The loss maps have been produced for 2012 4 TeV running, and Merlin is ready to be used for studies of the LHC upgraded collimation system.

ACKNOWLEDGEMENTS

We wish to thank A. Donnachie for his assistance with the scattering physics and allowing usage of his models, and the CERN LHC collimation group for many interesting discussions.

REFERENCES

[1] O. Brüning et. al. “LHC Design Report” 2004

[2] `http://merlin-pt.sourceforge.net/`

[3] D. Krücker and F. Poirier and N. Walker, “Merlin-based start-to-end simulations of luminosity stability for the ILC” PAC 2007, THPAN023

[4] F. Poirier and D. Krücker and N. Walker, “An ILC main linac simulation package based on Merlin” EPAC 2006, MOPLS065

[5] `http://mad.web.cern.ch/mad/`

[6] A. Donnachie and P.V. Landshoff, “Elastic Scattering at the LHC” arXiv:1112.2485v1 [hep-ph]

[7] A. Donnachie and P.V. Landshoff, “Soft diffraction dissociation” arXiv:hep-ph/0305246v1

AUTOMATIC COMPUTER ALGORITHMS FOR BEAM-BASED SETUP OF THE LHC COLLIMATORS*

G. Valentino†, N. Sammut
CERN, Geneva, Switzerland and University of Malta, Msida, Malta
R.W. Assmann, R. Bruce, S. Redaelli, B. Salvachua, D. Wollmann
CERN, Geneva, Switzerland

Abstract

Beam-based setup of the LHC collimators is necessary to establish the beam centers and beam sizes at the collimator locations and determine the operational settings during various stages of the LHC machine cycle. Automatic software algorithms have been successful in reducing the costly beam time required for the alignment, as well as significantly reducing human error. In this paper, the beam-based alignment procedure is described, and the design of algorithms such as a BLM feedback loop, parallel collimator alignment, pattern recognition of BLM loss spikes, automatic loss threshold selection and coarse BPM-interpolation guided alignment is explained. A comparison on the alignment results from the 2010 to the 2012 LHC runs is presented to illustrate the improvements achieved with the automatic algorithms.

INTRODUCTION

The Large Hadron Collider (LHC) located at CERN is designed to collide two counter-rotating particle beams with an energy of 7 TeV each [1]. Machine protection systems are installed to prevent damage to the LHC in the event of beam loss scenarios. The collimation system protects the collider against unavoidable losses, which may cause quenches in the superconducting magnets, damage to beam pipe equipment or cause electronics degradation as a result of radiation effects [2].

At present, the collimation system is made up of over 100 collimators which are arranged in four levels of retraction from the beam in the form of a hierarchy. The LHC consists of 8 arcs and 8 straight sections, called insertion regions (IRs). The collimators are located mainly in IR3 and IR7 to scatter and absorb particles with large momentum and betatron offsets respectively.

Each collimator providing cleaning of normal beam losses consists of two blocks, known as jaws, of graphite or tungsten. In order to ensure maximum cleaning efficiency and protection, the jaws must be positioned symmetrically around the beam with the correct gap in units of beam standard deviations (sigmas). The jaw positioning accuracy is 5 μm, and a three-tier control system allows the upstream and downstream edges of each jaw to be moved separately via four motors [3].

The beam centers and beam sizes at the collimator locations are determined from beam-based alignments. The setup procedure relies on feedback from beam loss monitors (BLMs) [4], which consist of ionization chambers placed downstream of the collimators. The BLMs capture beam loss showers caused by primary beam particles impacting on the collimators. A collimator jaw is aligned to the beam halo when a clear spike is observed in the BLM signal.

In 2010, the setups were performed 'manually', meaning that human feedback was required to determine when the jaw is aligned to the beam. This was achieved by observing the BLM signal on a screen following a jaw movement. A disadvantage of this method is the setup time required, which is data lost for the experiments and beam time for other users. Human error results in incorrect jaw movements, causing high losses and beam dumps, therefore contributing to the setup time. In order to speed up the collimator alignment and minimize the intervention required from the operator, a set of automatic algorithms has been implemented in the top-layer of the LHC software architecture in Java.

COLLIMATOR SETUP PROCEDURE

Each collimator is aligned in a four-step procedure, as established in [5]. The setup procedure is illustrated in Fig. 1. The first step is to align the left and right jaws of a reference collimator to form a reference cut in the beam halo. This collimator is taken to be the primary collimator (TCP) in the same plane (horizontal, vertical or skew) as the collimator to be aligned, called collimator *i*. As a particular jaw can be declared to be aligned only if it was the only jaw moving when the BLM signal spike occurs, the left and right jaws are moved towards the beam separately. After aligning the reference collimator, the same procedure is performed for the collimator *i* (2), and the reference collimator is aligned once again (3). The beam center can then be determined from the final measured left and right jaw positions $x_i^{L,m}$ and $x_i^{R,m}$ of collimator *i*:

$$\Delta x_i = \frac{x_i^{L,m} + x_i^{R,m}}{2} \quad (1)$$

A full derivation of the calculation for the inferred beam size at collimator *i* is available in [6], and can be calculated from the jaw half gap and the average cut n_1 in units of sigma at the reference collimator:

* Work supported by FP7 EuCARD
† gianluca.valentino@cern.ch

$$\sigma_i^{inf} = \frac{x_i^{L,m} - x_i^{R,m}}{n_1^{k-1} + n_1^k} \quad (2)$$

where k is an index for the number of reference collimator alignments. The reference collimator is aligned both before and after the setup of collimator i to account for the halo that is scraped away. The final step is to set the left and right jaws of collimator i using the values obtained for the beam center and beam size to maintain the collimation hierarchy, where N_i is the half gap opening for a particular collimator family:

$$x_i^{L,set} = \Delta x_i + N_i \sigma_i^{inf} \quad (3)$$

$$x_i^{R,set} = \Delta x_i - N_i \sigma_i^{inf} \quad (4)$$

An alignment of all moveable 86 collimators is performed yearly during the beam re-commissioning phase. The beam centers and beam sizes must be known for various stages of the machine cycle. At 450 GeV (injection energy), all collimators are aligned. For the 4 TeV operational energy in 2012, 80 collimators are set up (the six injection protection collimators are excluded). The 16 tertiary collimators (TCTs) are aligned with both beams squeezed to the operational $\beta*$ in the experimental IPs, and are set up again when the orbit separation bumps are collapsed and the beams are brought into collisions. In addition, the tertiary collimators which protect the experimental regions are aligned following a change in the beam orbit to accommodate modifications for the experiments, such as a change in the crossing angle.

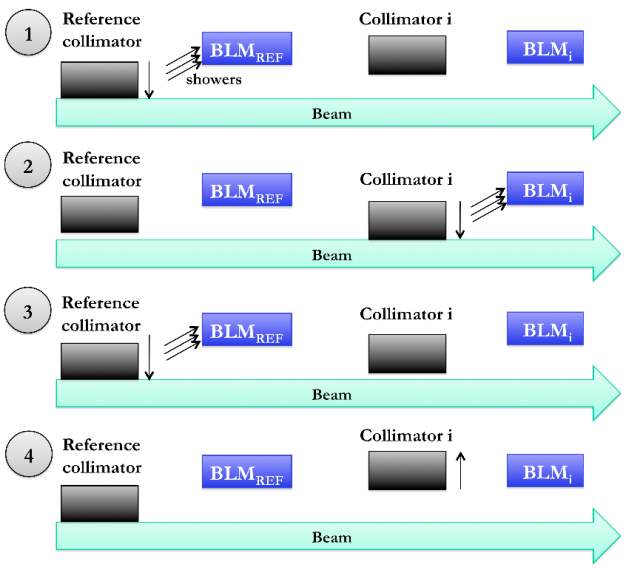


Figure 1: The four-stage beam-based alignment procedure for collimator *i*. The reference collimator is aligned to form a reference cut in the beam halo (1). Collimator *i* is aligned (2), followed by a re-alignment of the reference collimator (3). Finally, collimator *i* is opened to its position in the hierarchy (4).

BEAM LOSS FEEDBACK

The first algorithm developed to speed up and automate the alignment process makes use of a beam loss feedback loop [6]. A periodic jaw movement can be started from the top-level application, and before each jaw step the algorithm would ensure that the BLM signal is below a predefined threshold, which is several orders of magnitude lower than the beam dump loss threshold. The algorithm takes four inputs, consisting of the left and right jaw step sizes in μm, Δx_i^L and Δx_i^R, the BLM signal threshold S_i^{Thres} in Gy/s and the time interval between each step t_i^s in seconds. In the 2010 and 2011 LHC runs, the BLM data was acquired at a rate of 1 Hz, and following software upgrades as of 2012, the data is received at a rate of 12.5 Hz. The jaw moves at 2 mm/s, and steps can be sent at a rate of 1-8 Hz. When the jaw movement is stopped, human feedback is required to ascertain whether the jaw is touching the beam halo from the BLM loss spike displayed. Hence, the algorithm is semi-automatic. A flowchart of the algorithm is shown in Fig. 2.

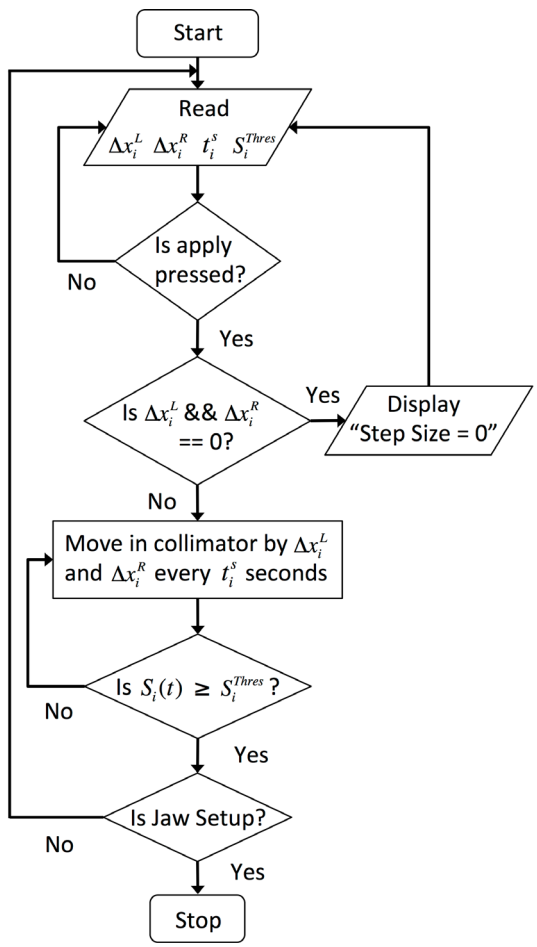


Figure 2: A flowchart of the beam loss feedback loop, which is used to align the collimator jaws in a semi-automatic manner. Human feedback is still required at the end to check whether the BLM signal is a clear loss spike, which would indicate that the jaw is aligned.

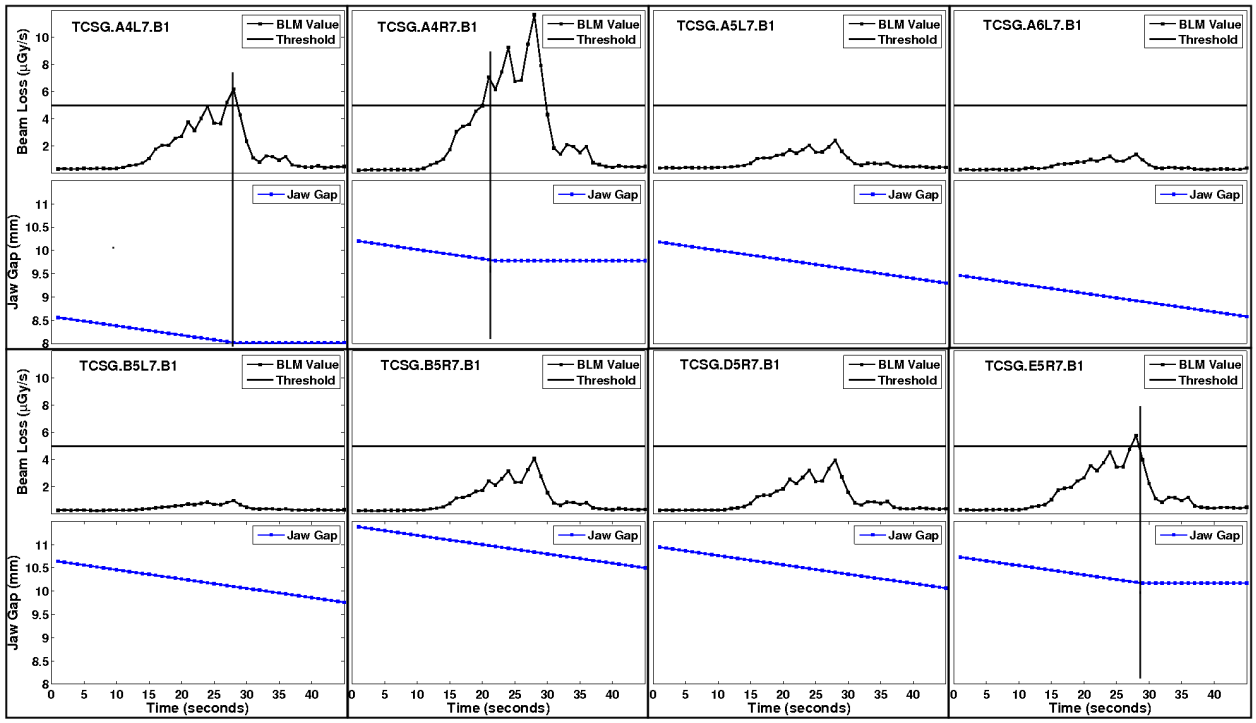


Figure 3: Both jaws of eight skew B1 collimators moving in parallel. The similarity of the BLM spike patterns and the simultaneous stopping of three collimators highlights the need to automatically identify which collimator jaw is actually aligned to the beam. The perpendicular lines indicate where the jaw stops when the losses exceed the threshold.

PARALLEL COLLIMATOR ALIGNMENT

A parallel alignment algorithm was developed to move all collimator jaws in steps towards the beam using the beam loss feedback loop, once the latter proved to be successful. It provides a coarse but quick way of positioning a set of collimator jaws around the beam, after which each jaw can be finely aligned in sequence. During initial tests of the algorithm, an expected cross-talk effect was observed, in which a loss pattern registered on the BLM of one particular collimator was also detected on on other collimator BLMs downstream. An example is illustrated in Fig. 3, where the BLM threshold was set to 5×10^{-6} Gy/s for all collimators. Three have stopped moving as the losses on their BLMs have exceeded the threshold. Cross-talk prevented the parallel setup method from functioning efficiently, and therefore the algorithm was designed to identify which collimator jaw is at the beam.

The parallel setup algorithm uses a timer task (CheckColls) to check whether any collimators have stopped moving. As soon as a collimator stops moving due to an exceeded BLM threshold, another timer task (CheckCollsT) is started to determine whether any other collimators also stop within a pre-defined time period T. If this is the case, all the other collimators moving in parallel are stopped so that the algorithm can concentrate on the collimators that stop within T. In case the BLM threshold S_i^{Thres} set during the previous movement is now below the background signal, an option allows the user to instruct the program to automatically increase the threshold in steps up to a maximum amount S_{max}^{Thres} . If the threshold is exceeded at the second step or thereafter, the jaw is declared to be aligned to the beam, and the algorithm terminates to allow the operator to start the sequential alignment.

BEAM LOSS SPIKE RECOGNITION

The alignment procedure can be automated further if a collimator expert is no longer required to visually judge whether a loss pattern is a clear indication that the jaw has touched the beam. An example of an optimal loss spike is shown in Fig. 4. The four pattern components include the steady-state losses before the spike, the loss spike, the temporal decay and the steady-state losses after the spike. Features inherent in the loss pattern were identified to discriminate between optimal and non-optimal loss spikes. A Gaussian fit can be applied to the loss spike component folded around the maximum value, while a power fit can be made to the temporal decay. Six features were considered:

- Maximum value
- Steady-state average
- Variance
- Gaussian fit correlation coefficient
- Power fit gradient
- Power fit correlation coefficient

The Support Vector Machines (SVM) algorithm is a supervised learning technique that can be used for classifica-

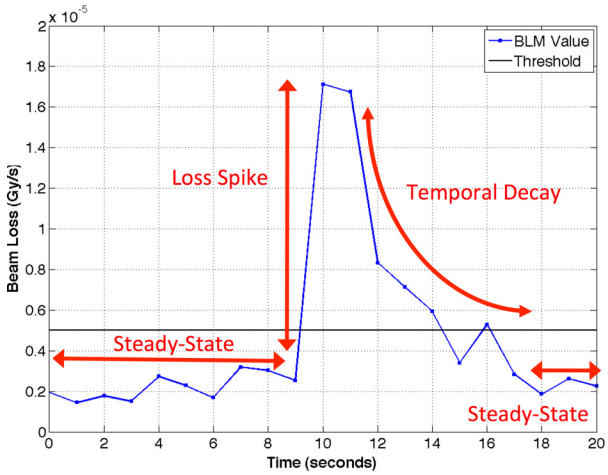


Figure 4: The components of a typical clear BLM signal when the collimator jaw touches the beam halo.

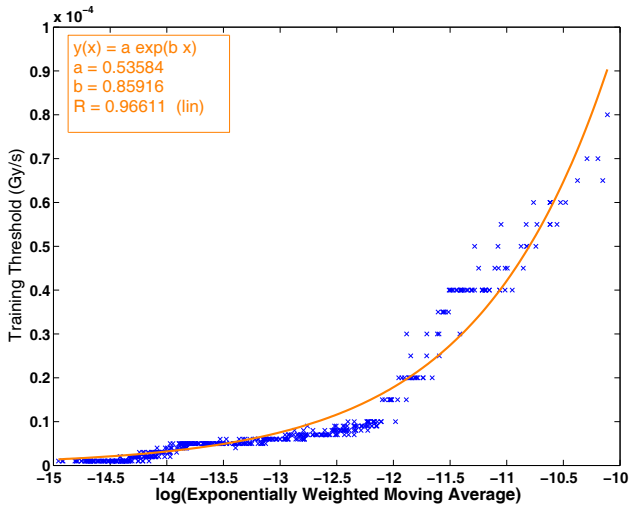


Figure 5: Loss thresholds applied before the start of a jaw movement as a function of the logarithm of the exponentially weighted moving average of the BLM signal. An exponential fit can be applied to the data.

tion of data, and was used to classify the loss spikes [7]. It operates by maximizing the margin between the training data points and the decision boundary. The LIBSVM tool [8] was used for training and testing of the SVM model. The radial basis function kernel was chosen as it has less hyperparameters, and presents fewer numerical difficulties. A total of 444 samples were available from alignments in 2011 at 3.5 TeV. The sizes of the training and the testing datasets were chosen to be equal. An accuracy rate of 97.3% was achieved for the training data, while 82.4% of the test data were classified correctly, which gives an overall prediction rate of 89.9%.

AUTOMATIC THRESHOLD SELECTION

An algorithm that could automatically set the loss threshold at the start of each jaw movement would contribute greatly to automating the alignment procedure further. Samples of the steady-state BLM signal in 20 second intervals and the subsequent threshold set by the operator were collected. The exponentially weighted moving average of each sample was determined, with the larger weights assigned to the most recent values. If the thresholds set by the operator averages are plotted as a function of the logarithm of the averages, an exponential fit can be applied to the data as shown in Fig. 5. The threshold set by the algorithm during the alignment is therefore:

$$S_i^{Thres} = 0.53584e^{0.85916x} \quad (5)$$

The maximum threshold that can be set is 1×10^{-4}, which is an order of magnitude below the BLM dump thresholds.

BPM-INTERPOLATION GUIDED COLLIMATOR ALIGNMENT

An approximation to the beam centers at the collimators can be obtained from an interpolation of the orbit measured at specific locations by Beam Position Monitors (BPMs).

The interpolation can be exploited to speed up the alignment process, if the errors between the interpolation and the beam-based measurements are not too large. Two datasets were built, one containing the beam-based alignment centers measured in 2011 and 2012, the other containing the interpolated orbit at each collimator at the time of alignment [9]. The average delta between the datasets is of $\sim$ 550μm, with maximum deltas of $\sim$ 3000μm for the tertiary collimators, where the interpolation reliability is known to be worse. The jaws can now be moved in one step at a rate of 2 mm/s from the initial positions to a safe margin around the beam based on the IR7 TCP cuts. A gain in time of a factor 200 can be achieved for this part of the alignment using this technique, instead of the standard 10 μm step every second. The left and right jaws are hence moved in to the following settings:

$$x_i^L = \Delta x_i^{int.} + (N_{TCP} + N_{margin}) \times \sigma_i^n + \frac{\Delta_{m,int.}}{2} \quad (6)$$

$$x_i^R = \Delta x_i^{int.} - (N_{TCP} + N_{margin}) \times \sigma_i^n - \frac{\Delta_{m,int.}}{2} \quad (7)$$

where $\Delta x_i^{int.}$ is the interpolated beam center at collimator i, N_{TCP} is the half-gap of the IR7 TCP in units of σ, σ_i^n is the nominal 1σ beam size and $\Delta_{m,int.}$ is the expected offset between the interpolated and the measured center from beam-based alignment, based on the empirical analysis. Once the IR7 TCP is aligned, typically at 3 - 4 σ, a further safety margin N_{margin} is applied (e.g. 2σ) over and above the TCP cut. In a LHC Machine Development (MD) study at 450 GeV [10], 27 collimators were positioned around the beam guided by the interpolated orbit, and were subsequently aligned using the parallel alignment algorithm. These collimators were aligned in 1.75

hours, which if extrapolated to a full alignment of all 80 collimators at 4 TeV flat top results in a setup time of 5.5 hours. This is a factor 5 improvement over the setup time of 28 hours achieved with manual alignment at 3.5 TeV.

RESULTS

The time taken to set up collimators is the most important indicator of the efficiency of a setup algorithm. The total time required T_{setup} is defined as follows [6]:

$$T_{setup} = T_{beam} + d \times T_{turnaround} \quad (8)$$

where T_{beam} is the beam time used for setup, C is the number of collimators set up and d is the number of beam dumps caused by collimator setup. The turnaround time $T_{turnaround}$ is the time consumed from the point of beam dump until the machine is cycled back to the setup operating point, which can reach 3 hours for flat top. Figure 6 shows the evolution in T_{setup} and the average jaw step size for collimator alignments at injection and flat top from 2010 to 2012. A larger setup time was required for 2011 due to a phased change-over between manual and semi-automatic alignment and the use of a smaller average step size, which reduces the probability of dumping the beam and improves the alignment accuracy. The setup times achieved from 2011 onwards are more impressive when one considers that the average jaw step size was reduced by a factor 4, and hence the jaw takes longer to cover a given distance in mm. The setup times continued to improve in 2012 with a higher BLM data rate of 12.5 Hz allowing the jaws to be moved at the maximum rate of 8 Hz.

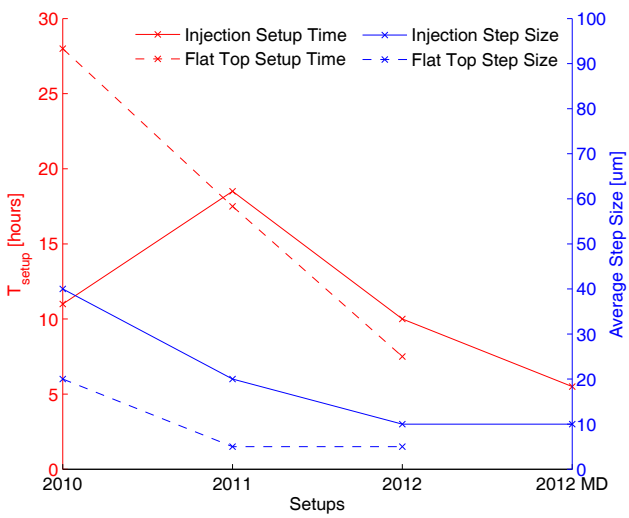


Figure 6: Evolution of the setup time for the collimator alignments and the average jaw step sizes used at injection and flat top from 2010 to 2012. The 1 hour 45 minutes required to setup 27 collimators during the 2012 MD is extrapolated to 5.5 hours for a full setup.

SUMMARY

The cleaning efficiency of the LHC collimation system is highly dependent on the careful positioning of all collimators. The positions are determined as a result of beam-based alignment. Due to the large number of collimators, the time to complete the alignment can reach over 20 hours in the absence of automatic algorithms, which results in reduced time for LHC physics operation. In this paper, several algorithms aimed at automating and speeding up the alignment procedure have been described. The algorithms have been successful in drastically reducing the need for operator intervention, and have decreased the set up time by a factor 5 at flat top.

ACKNOWLEDGEMENTS

This work is funded by the WP8 ColMat package in the FP7 EuCARD programme.

REFERENCES

[1] Report No. CERN-2004-003-V1, edited by O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, P. Proudlock (2004).

[2] R.W. Assmann et al. Requirements for the LHC Collimation System. In Proceedings of EPAC'02, Paris, France, 2002, pp. 197-199.

[3] A. Masi, R. Losito. LHC Collimators Low Level Control System. IEEE Transactions on Nuclear Science 55, 333 (2008).

[4] B. Dehning et al. The LHC Beam Loss Measurement System. In Proceedings of PAC'07, Albuquerque, USA, 2007, pp. 4192-4194.

[5] R.W. Assmann, E. Holzer, J.-B. Jeanneret, V. Kain, S. Redaelli, G. Robert-Demolaize, J. Wenninger. Expected Performance and Beam-Based Optimization of the LHC Collimation System. In Proceedings of EPAC'04, Lucerne, Switzerland, 2004, pp. 1825-1827.

[6] G. Valentino, R.W. Assmann, R. Bruce, S. Redaelli, A. Rossi, N. Sammut, D. Wollmann. Semiautomatic Beam-Based LHC Collimator Alignment. Phys. Rev. Spec. Top. Accel. Beams 15, 051002 (2012).

[7] G. Valentino, R. Assmann, R. Bruce, N. Sammut. Classification of LHC beam loss spikes using Support Vector Machines. In Proceedings of IEEE SAMI'12, Herlany, Slovakia, 2012, pp. 355-358.

[8] C.-C. Chang, C.-J. Lin, ACM Transactions on Intelligent Systems and Technology 2, 27:1 (2011).

[9] G. Valentino, R.W. Assmann, R. Bruce, L. Lari, G. J. Müller, S. Redaelli, A. Rossi, N. Sammut. Comparison of LHC Collimator Beam-Based Alignment to BPM-Interpolated Centers. In Proceedings of International Particle Accelerator Conference, New Orleans, USA, 2012, pp. 2064-2062.

[10] G. Valentino, R.W. Assmann, R. Bruce, G.J. Müller, S. Redaelli, B. Salvachua. Beam-Based Collimator Alignment MD. CERN-ATS-Note-2012-046 MD, 2012.

LINAC BEAM DYNAMICS SIMULATIONS WITH PY-ORBIT

A. Shishlo*, ORNL, Oak Ridge, TN 37830, USA

Abstract

Linac dynamics simulation capabilities of the PyORBIT code are discussed. PyORBIT is an open source code and a further development of the original ORBIT code that was created and used for design, studies, and commissioning of the SNS ring. The PyORBIT code, like the original one, has a two-layer structure. C++ is used to perform time-consuming computations, and the program flow is controlled from a Python language shell. The flexible structure makes it possible to use PyORBIT also for linac dynamics simulations. A benchmark of PyORBIT with Parmila and the XAL Online model is presented.

INTRODUCTION

The negative hydrogen ion SNS linac was designed [1] by using the Parmila accelerator simulation code [2]. In addition to envelope tracking codes, such as the XAL online model that is used routinely in the SNS control room, particle in cell (PIC) simulation codes like Parmila are valuable tools for halo growth calculations and beam loss estimation in heavy ion linear accelerators. Unfortunately, the Parmila code is not actively supported today, and a large project, such as SNS, cannot afford to lose PIC simulation capabilities for any part of the accelerator. During the search performed by the SNS accelerator physics group, we could not find an accelerator code that satisfies the necessary conditions of the full control over the source code, the underlying physical models, and possible modifications. To solve this problem we started the development of a home-grown open source linac accelerator code on an existing platform, namely the PyORBIT code.

PY-ORBIT AND ORBIT CODES

PyORBIT is a PIC code developed from the original ORBIT code [3]. ORBIT has been used for the design of the SNS ring and transfer lines, simulations of collective effects for SNS, and for other projects. PyORBIT, like the original ORBIT, has a two-language structure. Time-consuming calculations are performed at the C++ language level, and a simulation flow control is implemented in a scripting language. In PyORBIT the outdated and unsupported SuperCode is replaced by Python, an interpreted, interactive, object-oriented, extensible programming language. The PyORBIT project was started not only to replace the old programming technologies, but to perform calculations for the laser stripping experiment [4] which could not be performed by the original ORBIT. At this moment, PyORBIT does not have all features of the original code, but we are in the process of transferring the old ORBIT modules to the new code. It is not a straightforward process because of ubiquitous SuperCode dependencies in ORBIT. On the other hand, this provides the opportunity to restructure the modules in a more efficient and clear way.

The structure of the PyORBIT code was described in Ref. [3]. From the beginning the code has been developing as a loose structure capable of accommodating many weakly or completely unrelated projects. We took advantage of this feature of PyORBIT when the linac simulation part was included into the code.

LINAC PART OF PY-ORBIT CODE

The new linear accelerator lattice package in PyORBIT is a concrete implementation of abstract accelerator lattice classes described in [3]. Therefore, right now we have two types of the lattices – one for rings and transfer lines, which is similar to the lattice of the original ORBIT code, and another for the linear accelerators. The new implementation was necessary because the energy of a synchronous particle changes along the linac lattice, and the parameters of the lattice elements must be changed accordingly. In addition to this feature of the linear accelerator lattice, we implemented a more complicated structure that includes subsequences and RF cavities that in turn consist of RF gaps. The RF cavities themselves are not lattice elements. They are used to synchronize the phases of all RF gaps that belong to a particular RF cavity. Before using this type of lattice, it must be initialized by tracking a design particle to map its arrival time at each RF cavity. Only after that will changes to the cavity amplitudes or phases in the model reflect the changes in the real machine.

At present, a linac lattice can be built in two ways. First, it can be constructed right in a PyORBIT script by adding lattice elements one by one. Second, it can be built by using an input XML file and linac lattice parser. The parser assumes a certain structure of the input XML file. This structure will be standardized in the future when the list of necessary parameters is agreed upon among all users. At this moment, all classes in the linac packages are considered experimental, and they are kept in a specific SNS linac directory. All these classes are pure Python classes. They are lightweight and can be easily modified to accommodate different requests. It is also possible to implement a third type of abstract accelerator lattice if a more universal approach is required in the future.

There are several new C++ classes that have been created to simulate physics in linear accelerators. They include two types of space charge calculations and a simplified RF gap model. We plan to implement more sophisticated RF models in the near future.

*shishlo@ornl.gov

THE RF GAP MODEL

An accelerator node representing a zero length RF gap changes the energy of each macro-particle in the bunch and performs transverse focusing or defocusing according the arrival time (a relative phase) of this particle:

$$W_{out} = W_{in} + E_0 TL \cdot \cos(\varphi_{RF} + \varphi) \quad (1)$$

$$r'_{out} = \frac{(\gamma\beta)_{in}}{(\gamma\beta)_{out}} \cdot r' - k_f(\varphi_{RF} + \varphi) \cdot r \quad (2)$$

where W is the kinetic energy, $E_0 TL$ is the maximum energy gain, φ_{RF} and φ are the phases of RF and macro-particle relative to the synchronous particle, γ and β are relativistic parameters, r and r' are the transverse coordinate and angle of the macro-particle, and k_f is a transverse focusing coefficient [5].

THE SPACE CHARGE MODULES

Usually, particle bunches in rings and linear accelerators have different ratios between longitudinal and transverse sizes, and therefore methods of space charge calculation are different. At present, the PyORBIT space charge modules for linacs, where the transverse and longitudinal sizes are comparable, include two methods: 3D uniformly charged ellipsoid field and a 3D FFT Poisson solver.

Electric Field of Uniformly Charged Ellipsoid

The simplest way to calculate space charge forces for the linac bunch is to approximate its charge distribution by a uniformly charged 3D ellipsoid. The electric field inside and outside of such an object can be easily calculated [6], and the space charge force momentum kicks are applied to each macro-particle in the bunch. The parameters of the ellipsoid are found from the condition of equality of rms sizes for the real bunch and the approximation. This scheme is very simple and fast, and it will work even if the linac bunch consists of very few macro-particles (several hundreds is an acceptable number). This approach can be considered a variant of beam envelope calculations in Trace3D or in the XAL online model. If the charge density distribution is more complicated, an arbitrary number of uniformly charged ellipsoids can be used.

3D FFT Poisson Solver

If the linac bunch has an asymmetric shape, the Poisson equation for the electric potential should be solved by an exact method. For this purpose PyORBIT has a 3D FFT Poisson solver that uses the FFTW library. This solver will work in the case of parallel calculations, but it has bad parallel scalability, because all calculations related to the FFT transformations are identical and are performed on each CPU. The parallel efficiency of this module is determined only by the distribution of the macro-particles between CPUs. In the future we plan to add more efficient 3D Poisson Poisson solvers.

PY-ORBIT VS. PARMILA

A benchmark between Parmila and PyORBIT for linear accelerators was performed for the warm linac of the SNS accelerator. It includes the Medium Energy Beam Transfer (MEBT) line, Drift Tube Linac (DTL), and the Coupled Cavity Linac (CCL). All efforts were made to create identical lattices for the two codes. PyORBIT generated an initial "water-bag" distribution of 20000 macro-particles. The exact particle coordinates were translated to Parmila notation and packed into the Parmila's direct access FORTRAN file. The design initial Twiss parameters were used. The peak current of the linac bunch was set to 38 mA. The uniformly charged ellipsoid model was used for the space charge nodes in the PyORBIT lattice. For Parmila's space charge calculations the option "3D Picnic" was used. The execution time for the PyORBIT script on one CPU was about five times faster than for Parmila. The results of the simulations for all 90 meters of the warm SNS linac are shown in Fig. 1.

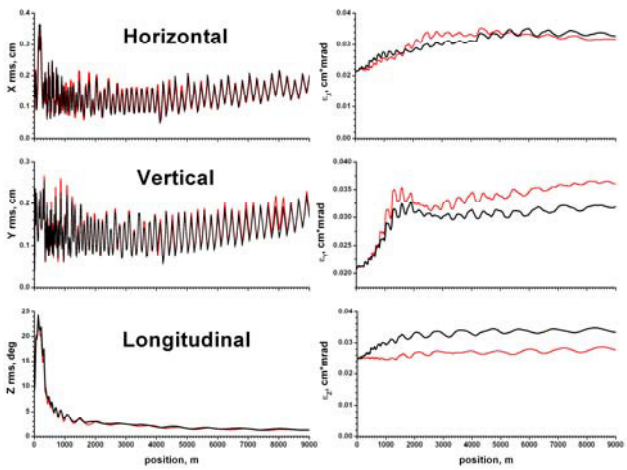


Figure 1: The calculated rms sizes of the bunch (on the left side) and emittances (on the right) in the warm part (MEBT-DTL-CCL) of the SNS linac. The black line is for Parmila, and the red one is for PyORBIT.

There is not perfect agreement between the Parmila and PyORBIT results, but we have to take into account that in the PyORBIT simulations we used a simplified RF gap model. That can explain the big differences between the longitudinal emittances for the two codes seen in Fig. 1. The usage of 3D the Poisson solver instead of the uniform ellipse model did not change the PyORBIT results much. The execution time was twice as fast for PyORBIT with the same number of grid points in the 3D solver.

Our benchmark can be compared to the benchmark between the Track code and Parmila for the DTL part of the SNS linac [7]. The quality of agreement in the Track-Parmila and PyORBIT-Parmila benchmarks are similar, especially if we take into account that here we include 90 meters (MEBT-DTL-CCL) of the SNS linac instead of 40 meters (DTL only) as in Ref. [7].

PY-ORBIT AND XAL ONLINE MODEL

Another benchmark to check the correctness of the linac part of the PyORBIT code involves the XAL online model (OM) [8]. OM is an envelope tracking code that was carefully tested against Parmila, Impact, Trace3D, and the SNS linac real measurements. The results of PyORBIT and OM for the SNS Superconducting Linac (SCL) are shown in Fig. 2. Both simulations include space charge effects for a peak current of 38 mA. The agreement between the two codes is very good, especially if we keep in mind the fundamental differences between codes: one is a PIC and the other is an envelope-tracking code.

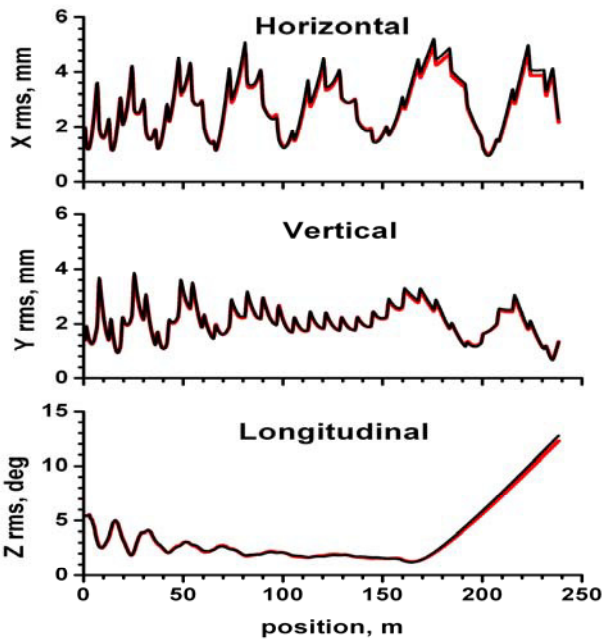


Figure 2: The transverse and longitudinal rms sizes of the bunch in the SNS SCL linac simulated with PyORBIT (red line) and OM (black line) codes.

FUTURE PY-ORBIT DEVELOPMENT

The most urgent need of the linac part of PyORBIT is a development of more comprehensive RF gap models. We are going to develop at least two new models. The first model will reproduce Parmila's approach [5] with the same simulation time, and the second one will include the 3D tracking of each particle in the time dependent electromagnetic field of the cavity. The second model will slow down the simulations, but it will allow verification of the simplified approaches.

Another addition will be a collimation module that will be moved from the original ORBIT. It will be shared with the ring related part of PyORBIT, and it will provide the loss accounting for the linac models. It also can be used for a design of collimation systems at SNS.

The development of the new space charge modules will depend on available manpower and a real necessity in parallel calculations.

CONCLUSIONS

Linac beam dynamics simulation capabilities have been successfully implemented into the PyORBIT code, which was previously restricted to rings and transport lines. The new linac part has been successfully benchmarked against Parmila and the XAL Online Model. The simplified RF cavity model and a non-scalable parallel 3D space charge solver will be supplemented with more realistic and effective models.

PyORBIT is an open source code, and it can be downloaded from the Google project hosting [9]

ACKNOWLEDGMENT

Author is grateful to A. Aleksandrov and C. Allen for help with the Parmila and the XAL online model simulations during the benchmarks.

ORNL/SNS is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725.

REFERENCES

[1] J. Stovall, et al., "Expected Beam Performance of the SNS Linac," Proc. of the 2001 Particle. Accelerator Conference, Chicago, Ill., June 18-22, 2001, p. 446.

[2] J.H. Billen and H. Takeda, PARMILA Manual, Report LAUR-98-4478, Los Alamos, 1998 (Revised 2004).

[3] A. Shishlo, J. Holmes, and T. Gorlov, "The Python Shell for the ORBIT code", Proc. of ICAP '09, THPSC052 (2009); http://www.JACoW.org

[4] T. Gorlov, A. Shishlo, "Laser stripping computing with the Python ORBIT code", Proc. of ICAP '09, TH3IOPK03 (2009); http://www.JACoW.org

[5] T. P. Wangler, "RF Linear Accelerators", (Wiley-VCH Verlag GmbhH & Co, 2008), 209

[6] O. D. Kellogg, Foundations of Potential Theory, (Dover, New York, 1953), 192.

[7] B. Mustapha, "First Track Simulation of the SNS Linac," LINAC'06, Knoxville, TN 2006, TUP076, p. 432 (2006); http://www.JACoW.org

[8] http://sourceforge.net/projects/xaldev/

[9] http://code.google.com/p/py-orbit/source/checkout

AN IMPLEMENTATION OF THE VIRTUAL ACCELERATOR IN THE TANGO CONTROL SYSTEM*

P. Goryl#, A. I. Wawrzyniak, Solaris at the Jagiellonian University, Krakow, Poland
T. Szymocha, ACK Cyfronet AGH, Krakow, Poland
M. Sjöström, MAX IV laboratory, Lund, Sweden

Abstract

Integrating physics codes into the control system gives a possibility to improve machine operation. Providing tools for making computations directly within the control system and letting exchange data between the control system and models is a way of simplifying the whole process of calculating and applying machine's operational parameters as well as keeping track of them. In addition, having a so-called on-line model could be useful for the system diagnostic and faults detection, especially when the objective approach is considered. The concept of the Virtual Accelerator as well as its implementation for the Tango control system will be presented as it is planned to be used for both facilities: the Solaris in Krakow, Poland and the MAX IV in Lund, Sweden. This includes the ModelServer tango device, the simplified C/C++ Tango API to be used with physics codes like Tracy 3 and the tango2elegant script providing an easy solution for integrating the `elegant` tool with the Tango Control System.

INTRODUCTION

The Virtual Accelerator is a concept of integrating and interfacing different simulation codes into the control system (CS). However, an implementation presented here could be used with any kind of computation codes. It has been successfully implemented at the SLS[1] using the CORBA protocol and at the Diamond Light Source for the EPICS control system[2].

The Virtual Accelerator implementation for the Tango CS presented in this paper has been developed with several goals:

- Standardize a way to starting and stopping computations from the control system
- Simplify data exchange between the control system and the simulation code
- Provide a standard interfaces to different kind of equipment (classes of devices in case of the Tango CS)
- Provide a central point for storing calculated machine parameters within the control system architecture thus making these directly available for all control system applications
- Provide a facility to debug control system scripts and high level applications before these will be run for a real machine. It is expected to make the machine commissioning more efficient.

The Tango CS

The Tango control system implements an objective approach to interface controlled devices. It provides an interface with attributes, representing process values (PVs) or state values and commands representing operations one can invoke on a particular device. Each device is in fact represented as an object of a certain class as it is in the Object Oriented Programming (OOP) [3].

The OOP introduce a mechanism called the polymorphism along with a mechanism of the classes inheritance. This provides a way to abstract from what is behind an interface. As an example, all power supplies integrated in a control system (CS) could expose the same interface to users regardless of a fieldbus the CS is using to communicate with a particular equipment. This also means that it is possible to use the same interface to a model of a controlled system as to the real one. With some assumptions actors interacting with the CS will not be able to distinguish if they work with a real or a modelled system. It does not mean that the same could not be achieved with a non-object oriented CS. Nonetheless, the OOP makes it natural.

Integrating physics codes with the CS takes advantage of the existing tools of the control system. The integration is expected to improve the system diagnostic as well as simplify applying to the machine calculated machine's parameters. The computation results could be logged with the Historical and the Temporary Databases (the HDB and the TDB). Computed settings could be remembered as a snapshot with the Tango SNAP tool. Making the model available through the control system could enable online comparison between the real machine and the virtual one. Distributing machine's parameters, simulation results and real machine's signals through the Tango control system will prevent from clutter caused by using different protocols for data exchange.

The Virtual Accelerator for the Tango Project provides the following components:

- The Model Server device. It is a device that standardizes and simplifies invoking of computation from the Tango CS.
- The Simple-Tango (STango) API. This is C/C++ library enabling access to a Tango CS in a 'one line code'. It hides some Tango complexity.
- The tango2elegant. This is a script interfacing the `elegant` tool with the Tango CS.

THE MODEL SERVER

The Model Server is a tool to run computations inside the Tango control system. It is implemented as a Tango device. Its class name is ModelServer. It is written in the Python language with the PyTango library. The Model-Server class inherits from the DynamicDS class[4].

* Work supported by the European Regional Development Fund within the frame of the Innovative Economy Operational Program: POIG.02.01.00-12-213/09
#piotr.goryl@uj.edu.pl

The Model Server device is able to run any arbitrary program, script or operating system command. All what is supposed to be run is called a Job. The Model Server device is an interface to run and to control the execution of jobs through the Tango CS.

For the security reasons jobs are placed on a disk in subdirectories of a selected directory. This is specified by a device property named *JobsPath* (device properties are the feature of the Tango CS). Those subdirectories names denote Jobs' names. Additional security measure is provided with possibility to execute the jobs with system credentials provided by two another device properties: *SystemUser* and *SystemGroup*. Details on how to deploy and configure the Model Server are found in [5].

The Model Server Interface

The Model Server provides four main commands to control jobs

- On – to initiate a job
- Off – to clean after a job
- Start – to start the initiated job for certain number of iterations or in an infinite loop.
- Stop – to stop execution of the job. It is possible to stop it immediately (kill the process) or wait for the current iteration to finish.

The model server provides reading attributes, to diagnose job progress: number of finished iterations, when the job has been started, how long it takes to do the job, the job's name.

Since the device class is based on the DynamicDS class it is possible to define so-called dynamic attributes. These may be used by a job to set and read values.

Jobs

However, while the Model Server can run any arbitrary script or code. Its primary application is to run a beam physics code using input from the control system. This kind of codes typically run iteratively in steps:

- Loading a model (lattice file)
- Reading values from the control system
- Modifying the model according to the values from the CS
- Computing
- Reading the results
- Writing the above to the CS

It could be that some of the steps are not implemented in a particular job. However, the reading from and the writing to the CS are common. It is not necessary for an operator or a beam physicist to know details of the control system programming. This is where two another components - the Simple Tango and the tango2elegant - comes into play. Those are described later in this paper.

The Model Server Limitations

The main limitations for the current implementation of the Model Server are that it can only be run on *nix operating systems and that it depends on the *sudo* tool. However, future versions will aim to improve flexibility in starting jobs. As an example it will be possible to directly use the Unicore middleware for setting the job's credentials as well as sending it to a remote computing resource like the PL-Grid infrastructure [7]. In addition, it would be of interest to trigger a job based on CS events, such as changes in certain settings or read values.

THE SIMPLE TANGO (STANGO)

Simple-tango is a C/C++ library that simplifies access to the Tango control system with C/C++ codes. It is a set of functions for reading and writing devices' attributes in a "one line" code. It hides most of the Tango complexity (even though it is not very complex) from a code writer. It is not intended to replace the Tango API. For more complex operations, one will need to use the Tango API anyway.

The STango uses the standard C/C++ data types to get and to set attributes' values. It does a real conversion from the Tango data types to the C numeric types: `double`, `float`, `short`, `unsigned short`, `int`, `unsigned int`.

There are functions for reading and for writing values of scalar, spectrum (1D array) and image (2D array) formats. There is also a function for calling argument-less tango commands. For compatibility with the C there are specific functions that deal with each of the supported C data types. Template functions are provided for use in C++ codes. Please refer to the documentation [5] and the STango source code for details.

The STango Source Code

Simple-Tango provides two header files: stango.h, stangoimpl.h and two implementation files: stangolib.cpp that implements functions for the C and stangoimpl.cpp that implements a device proxies management. These are compiled to the library libstango.so, which have to be linked with a program (i.e. the *gcc* option -lstango).

The stango.h is the header file for use with pure C programs. In C++ code one can use the stangoimpl.h. It provides generic functions as well as giving access to some internal library functionalities.

THE TANGO2ELEGANT

The `elegant` physics code is used at MAX IV and at Solaris to design and investigate the linear accelerators as well as the storage rings [8,9]. `elegant` is a code provided by the Advanced Photon Source [10]. It does a variety of beam dynamics related calculations. The main input file of `elegant` is structured as a series of Fortran namelists. These namelists are commands with their parameters. The commands setup and invoke calculations. The command parameters' values are specified directly or in specified external files. As an example, a lattice file describing a machine lattice is provided as a reference. The commands are defined and described in the elegant documentation [6]. Among the elegant commands there is a command `&alter_elements`. It allows runtime modification of a parameter value for one or more elements [6]. This feature is exploited by the tango2elegant.

The tango2elegant is written in the Python language. It acts as a pre- and a post-processor of the elegant input and output files. In pre-processing it reads a main input file and sends effects to the standard output or a file. Its output can be used by the elegant as an input. For the post-processing it uses the main input file and the elegant output files and modifies values in the control system.

It searches the elegant main input file for additional commands (not defined in the elegant) and special strings. Two commands has been arbitrarily defined - `&tango_input` and `&tango_output`. Special strings are replaced with respective values see [5].

If `&tango_input` command is found it is altered with a set of proper elegant commands in the output. Values for parameters in these commands come from the control system.

If `&tango_output` is found it is not sent to the output (hence `elegant` will not see it) but according to its parameters the tango2elegant reads elegant output files and set values in the control system respectively.

Other commands are directly passed to the elegant. Comments are omitted. Please refer to the documentation [5] to see details on how to use the commands.

Usage of the tango2elegant

The simplest way of using the tango2elegant is to call it as the following:

```
tango2elegant input.ele | elegant \
-pipe=in,out
```

Providing that *input.ele* is an elegant input file with added `&tango_input` and/or `&tango_output` commands.

The other way is suitable for running with the Pelegant. To assure that there is ready and valid output from the Pelegant, it is worth to wait until it stops before processing the output:

```
# pre-process only &tango_input:
tango2elegant --no-tango-out input.ele tmp.ele
 # run computation:
mpirun --mca btl openib,tcp,self -np 6 \
  Pelegant  tmp.ele
# process only &tango_output:
tango2elegant --no-tango-in input.ele
```

STATUS

All three modules have been implemented.

There are efforts to deploy the Virtual Accelerator on the PL-Grid infrastructure [7]. A dedicated virtual machine (vm) acting as a Tango host [4] and running several Tango device servers is up and running. The vm is run inside a cloud being a part of the PL-Grid computation infrastructure. It is possible to send jobs from the virtual machine to infrastructure clusters using the Unicore midleware. The Model Server is supposed to use it. A current setup reveal a bug in the *sudo* tool (on Scienific Linux el5), that in the certain condition wait infinitely for a subprocess to finish even if it has already ended. It makes impossible to detect the end of an iteration. Measures has been taken and the Model Server is in redevelopment to be more flexible in the way it starts and waits for iterations.

The `elegant` and the Pelegant have been installed and successfully used for the beam dynamics calculation on the ZEUS cluster at the ACK Cyfronet in Krakow, which is a part of the PL-Grid infrastructure. Using parallel elegant on the Zeus cluster for particles tracking studies, performing frequency maps analysis, etc. benefits in much shorter calculation time (hours instead of days on a PC). The Tango with its libraries and tools has been compiled on the ZEUS, too. It has been proven that it is possible to use the Tango protocol to exchange data between the cluster and the virtual machine. It means that it is possible to have the Virtual Accelerator running on a clusters infrastructure.

It is planned to deploy the Virtual Accelerator at the MAX IV Laboratory, soon. It will be used to tests high level applications and scripts for the MAX IV and the Solaris control system, before real machines will be installed. It is expected to shorten the commissioning time as well as make it smother. Later it will be run in parallel to the real machines also providing an online model.

All the source codes and documentation will be available through a publicly available SVN server, soon.

ACKNOWLEDGMENT

This research was supported in part by the PL-Grid Infrastructure [7].

Thanks for Johan Bengtsson for bringing the idea of the Virtual Accelerator and his comments on the project.

Thanks for Claudio Scafuri from the Elettra and Filip Lindau from the MAX IV Laboratory for their codes which, although now heavily modified, served as a base for tango2elegant.

REFERENCES

[1] M. Böge et al., “Commissioning of the SLS using CORBA based beam dynamics applications”, PAC’01, TOPB012 (2001).

[2] M. T. Heron et al., “The DIAMOND Light Source Control System”, EPAC’06, THPCH113 (2006); http://accelconf.web.cern.ch/Accelconf/e06/PAPERS/THPCH113.PDF

[3] The TANGO Team, *The TANGO Control System Manual*, Pink Site (2012); http://ftp.esrf.eu/pub/cs/tango/tango_80.pdf

[4] S. Rubio-Manrique et al., “Dynamic attributes and other functional flexibilities of PyTango”, ICALAPCS’09, THP079 (2009).

[5] P. Goryl, “The Virtual Accelerator for the Tango CS documentation”, Solaris (2012).

[6] M. Borland, *User’s Manual for elegant*, APS (2011).

[7] M. Bubak et al., *Building a National Distributed e-Infrastructure -- PL-Grid,* (Springer, 2012).

[8] A. I. Wawrzyniak, et al., “Injector Layout and Beam Injection into Solaris”, THPC123, Proceedings of IPAC 2011, San Sebastián, Spain.

[9] A. I. Wawrzyniak, et al; “Solaris storage ring lattice optimization with strong insertion devices”; TUPPC025; Proceedings of IPAC’12, New Orleans Louisiana, USA (2012).

[10]http://www.aps.anl.gov/Accelerator_Systems_Division/Accelerator_Operations_Physics/software.shtml

SIMULATION OF ELECTRON CLOUD INSTABILITY

K. Ohmi, KEK, 1-1 Oho, Tsukuba, 305-0801, Japan

Abstract

We discuss coupled bunch and single bunch instabilities caused by electron cloud in positron circular accelerators. Unstable mode spectrum, which characterizes the instabilities, is focused.

INTRODUCTION

Electron cloud causes coupled bunch and single bunch instabilities. The coupled bunch instability caused by electron cloud has been observed at KEK-PF, since start of positron operation in 1988. The instability was confirmed at BEPC in IHEP, China. The instability has been observed in KEKB-Low Energy Ring, which is positron storage ring. Weak solenoid coils were wound along the whole ring to protect electrons near the beam. The unstable mode of the coupled bunch instability was affected by the solenoid status, ON or OFF. Corrective electron motion reflected to beam unstable mode. The coupled bunch instability has been observed at DAFNE in Frascati, Italy. DAFNE is a small e+e- collider ring with the circumference of 98m. Bending magnets and wiggler magnets are occupied in a large part of ring. It seemed that electrons in the bending field play important role. The unstable mode was characteristic for the electron motion in the bending field.

Single bunch instability has been observed in KEKB-LER. Generally fast head-tail instability is caused by merge between 0 and -1 synchrotron sideband modes in positron ring. In the single bunch instability, clear positive side band $\nu_y+a\nu_s$ ($1<a<2$) has been observed. The similar signal has been observed at PETRA-III, DESY, Germany. The instability has been observed in Cesr-TA, Cornell, USA. The signal appears as negative side band, $\nu_y-\nu_s$.

Simulations have been performed to explain the both of single and coupled bunch instabilities. We present the simulation results with focusing unstable mode in this paper.

COUPLED BUNCH INSTABILITY DUE TO ELECTRON CLOUD

Simulation of Electron Cloud Build Up and Coupled Bunch Instability

Beam-electron cloud system in multi-bunch regime is described by following equations:

$$\frac{d^2\boldsymbol{x}_p}{ds^2} + K(s)\boldsymbol{x}_p = \frac{2N_e r_e}{\gamma}\sum_{e=1}^{N_e}\boldsymbol{F}_G(\boldsymbol{x}_p-\boldsymbol{x}_e)\delta_P(s-s_e) \quad (1)$$

$$\frac{d^2\boldsymbol{x}_e}{dt^2} = \frac{e}{m_e}\frac{d\boldsymbol{x}_e}{dt}\times\boldsymbol{B} - 2N_p r_e c\sum_{p=1}^{N_p}\boldsymbol{F}_G(\boldsymbol{x}_e-\boldsymbol{x}_p)\delta_P(t-t_p(s_e)) - r_e c^2\frac{\partial\phi(\boldsymbol{x}_e)}{\partial\boldsymbol{x}_e}, \quad (2)$$

where the electric potential of electron cloud is given by

$$\triangle\phi(\boldsymbol{x}) = \sum_{e=1}^{N_e}\delta(\boldsymbol{x}-\boldsymbol{x}_e) \quad (3)$$

Bunches, which are rigid Gaussian shape in transverse and are located equal spacing along s, are represented by their center of mass $\mathbf{x}_p$. $\mathbf{F}_G$ is expressed by the Bassetti-Erskine formula.

The electron cloud build-up is simulated by integrating the second equation for the motion of macro-electrons $\mathbf{x}_e$ under $\mathbf{x}_p$=0. The initial condition of electrons, where and when electrons are created, is sketched in Figure 1. When beam pass through the chamber cross-section, electrons are created with a energy distribution. When an electron hits the wall, secondary electrons are produced with a probability.

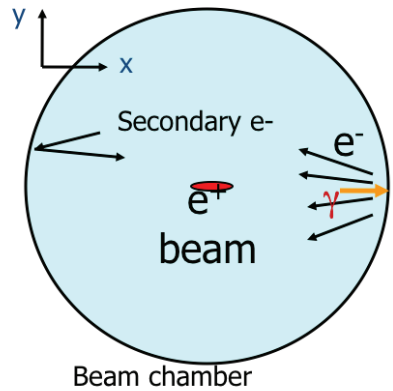


Figure 1: Electron cloud build-up model.

The coupled bunch instability is studied by solving the two coupled equations (1) and (2).

Coupled Bunch Instability in KEK-PF

Progress of electron cloud effects in these 17 years (since 1995) started from the interpretation of an instability observed in KEK-Photon Factory. Very strong coupled bunch instability had been observed since the positron storage had started. The threshold of the instability is very low 10-15mA. Either of positron or electron could be storaged by changing the polarity of the magnets. The instability is observed only in positron storage. Figure 2 shows frequency spectrum for the instability published in [1]. The first electron cloud build-up code (PEI) is developed in 1995 [2], and the coupled bunch instability was interpreted as a wake effect of the electron cloud. The wake force was estimated by perturbation of the cloud due to a passage through of a shifted bunch [2]. Figure 3 shows the simulated wake force.

Growth rate for each mode is calculated by the wake force. Figure 4 shows unstable mode spectrum estimated by the simulated wake force. The growth rate is 0.3msec for m=250-300. The corresponding frequency is (h-m-

$\nu_y)f_0$, where h and f_0 are beam harmonic number and revolution frequency, respectively. The growth rate and mode spectrum well agreed with the measurements.

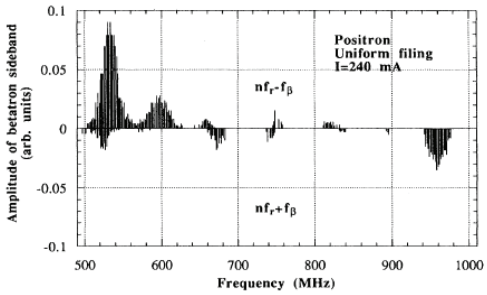


Figure 2: Unstable mode spectrum measured in KEK-PF [1].

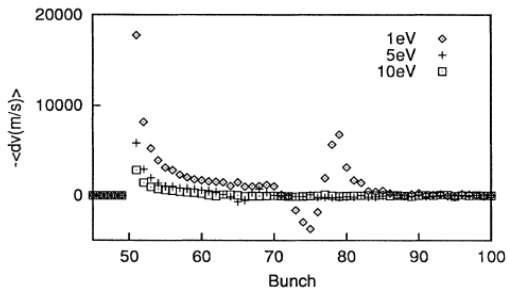


Figure 3: Simulated wake force due to electron cloud in KEK-PF [2].

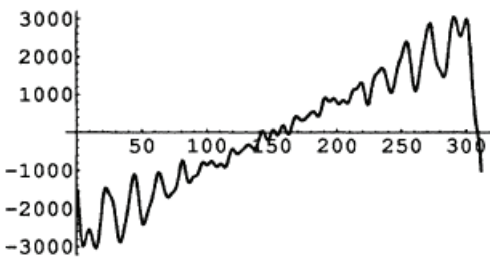


Figure 4: Unstable mode spectra due to the wake force in Figure 3 [2].

Coupled Bunch Instability in BEPC

Experiments were held to study reappearance of the electron cloud instability in BEPC since 1997. Both/either of positron and/or electron beams could be stored in BEPC; where the circulating directions are opposite.

Figure 5 shows mode spectra in positron and electron storage, respectively. Multi-mode instability similar as KEK-PF observation (Figure 2) was seen in BEPC. While single mode instability in electron storage also agreed with KEK-PF observation.

Simulation using Eqs.(1) and (2) gives beam centroid motion. FFT of the centroid positions gives unstable mode spectrum, which is compared with measurement. The spectra well agree with the measurements as shown in Figure 6.

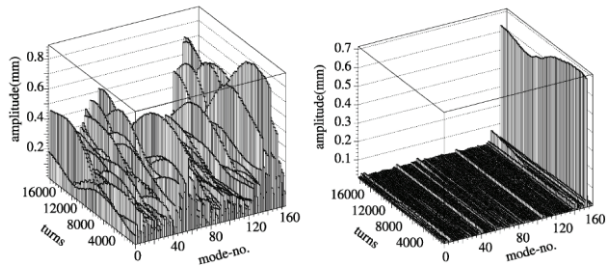


Figure 5: Mode spectra for electron cloud and ion instabilities, respectively [3]. Left and right pictures show mode variation in positron and electron storage, respectively.

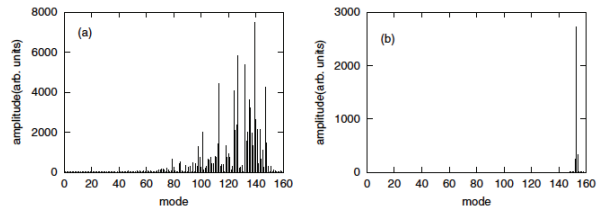


Figure 6: Mode spectra given by simulation for electron cloud and ion instabilities, respectively [3].

Coupled Bunch Instability in KEKB-LER

Coupled bunch instability due to electron cloud has been observed since the early stage of the commissioning of KEKB. For narrow bunch spacing 4-6 ns, fast beam losses was serious in the operation.

Solenoid coils were wound at the magnet free section to protect the electrons coming to the beam position. The field strange is ~<50G. The solenoid coils are covered 95% of the magnet free section finally. The solenoid coil did not work well to suppress the coupled bunch instability, while it works to suppress the single bunch instability very well. Electrons stay longer time in the vacuum chamber than that for the case of no solenoid. R/Q of the wake force induced by electron cloud was reduced, but the range of the wake field (Q) was longer. Bunch-by-bunch feedback system suppressed the coupled bunch instability for wider spacing >6ns.

Systematic measurement and simulations in KEKB were published in [4,5], respectively. Motion of electrons is different between in drift space or in the weak solenoid field (~<50 G). Electrons rotate along the chamber wall and do not approach to the beam. The electron motion reflects the instability signal.

Figure 7 shows the horizontal mode spectrum for solenoid OFF. The left picture was given by measurement. The right picture was given by simulation for electron cloud in drift space. Vertical spectra for measurement and simulation were similar as horizontal ones in Figure 7.

Figure 8 shows the horizontal mode spectrum for solenoid ON. The left and right pictures are given for measurement and simulation. Vertical spectra for measurement and simulation were similar as Figure 8.

The spectra with and without solenoid had very good agreement with simulations.

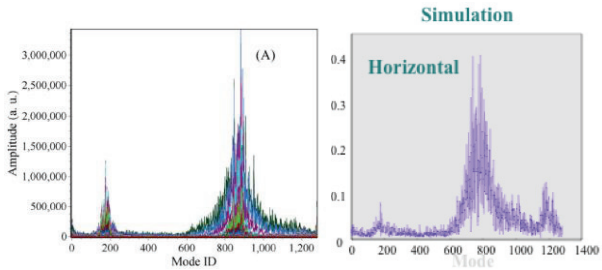


Figure 7: Horizontal mode spectrum in KEKB. Left picture is given by measurement with solenoid OFF [4,5]. Right picture is simulated by electron cloud in drift space.

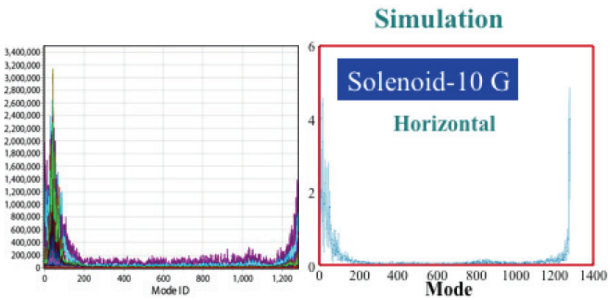


Figure 8: Horizontal mode spectrum in KEKB. Left picture is given by measurement with solenoid ON [4,5]. Right picture is simulated by electron cloud in solenoid field 10G.

Coupled Bunch Instability in DAFNE

Strong horizontal coupled bunch instability has been observed in DAFNE [6]. The unstable mode was slowest one; $f=(1-\Delta\nu_x)f_0$, where $\Delta\nu_x$ is the fractional part of the horizontal tune. DAFNE is a small ring; its circumference is 98m. Electrons in bending magnet and damping wiggler seemed dominant for the instability. Simulation was performed for electrons moving in a strong bending field. Figure 9 shows growth of transverse amplitude and horizontal mode spectrum given by the simulation.

Electrons in bending magnets form stripes, because electrons are trapped along the vertical field line. The density of the stripe increases for increasing secondary emission rate. In the instability, the stripe and bunch train experience coherent motion, thus the slowest unstable mode (m=114, $f=(120\text{-}m\text{-}\nu_x)f_0=(1-\Delta\nu_x)f_0$) is induced as shown in Figure 9.

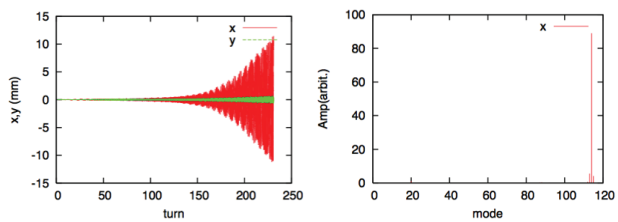


Figure 9: Growth of transverse amplitude and horizontal mode spectrum given by the simulation.

Figure 10 shows two cut of movie of beam-stripe motion. The white point is the position where bunch passes in the rectangular chamber; the wiggler chamber of DAFNE is the size 120mmx40mm. When a bunch passes though with a left deviation, the stripe deviates to right in left picture, vice versa in right picture. In the movie the stripe motion oscillates with a delay (less than π) for the beam motion.

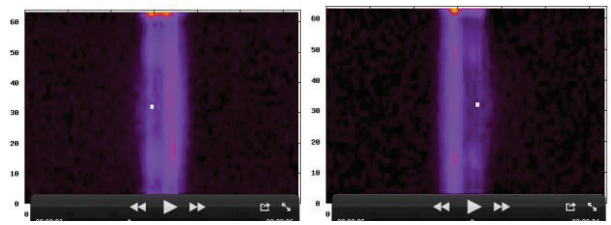


Figure 10: Two cut of movie of beam-stripe motion in bending magnet.

SINGLE BUNCH INSTABILITY DUE TO ELECTRON CLOUD

Single bunch fast head-tail instability caused by electron cloud has been observed in KEKB. When beam current exceeds a threshold value, emittance increases and synchro-beta side band signal has been observed.

The simulation of the fast head-tail instability is performed by solving following equations [7]

$$\frac{d^2\boldsymbol{x}_p}{ds^2} + K(s)\boldsymbol{x}_p = \frac{r_e}{\gamma}\frac{\partial\phi_e(\boldsymbol{x}_p)}{\partial\boldsymbol{x}_p}\delta_P(s-s_e)$$
$$\frac{d^2\boldsymbol{x}_e}{dt^2} = \frac{e}{m_e}\frac{d\boldsymbol{x}_e}{dt}\times\boldsymbol{B} - r_e c^2\frac{\partial\phi_p(\boldsymbol{x}_e)}{\partial\boldsymbol{x}_e}\delta_P(t-t_p(s_e)) \quad (3)$$

Each potential of beam and electrons are solved using PIC algorithm. Since beam (1mmx0.1mm) is localized at the chamber center, free boundary condition is employed. Electron cloud is initialized every interactions with beam with a flat distribution $\sim 40\sigma_x \times 60\sigma_y$.

The single bunch instability signal has been also observed in PETRA-III and Cesr-TA, recently.

Single Bunch Instability in KEKB-LER

A beam size blow-up had been observed since early stage of KEKB operation around ~1999 [8]. The blow-up limited the luminosity performance. Figure 11 shows beam size blow-up and luminosity limitation in 2000-2001. Solenoid coils were wound 2000-2001 in 50%-70% of the whole drift space. Left picture shows beam size blowup without (green) and with (red) the weak solenoid field. Threshold of the beam size blow-up increases 400 to 800mA. Left picture shows luminosity as function of current. Luminosity was saturated around 550mA at 2000 December. After winding solenoid additionally, 50% to 70%, luminosity was not saturated by 700mA at 2001 March. The solenoid coil was wound further after 2001, and was covered 95% of drift space at around 2005.

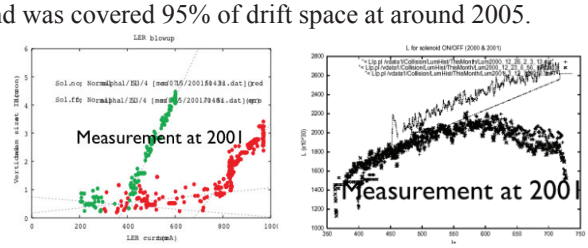


Figure 11: Beam size blow-up and luminosity limitation in 2000-2001.

Synchro-betatron sideband signal, which indicates head-tail instability caused by electron cloud, has been observed [9]. The synchro-beta signal synchronizes with the beam size blow-up: when the beam size blow-up is suppressed by the solenoid, the sideband disappears, vice versa. Figure 12 shows the betatron and synchrotron sideband spectra along the bunch train. Vertical axis is bunch train. Betatron signal, which is left white line, shift positively. It is tune shift due to electron cloud. Right white line is a positive synchrotron sideband, whose frequency is $\nu_y+a\nu_s$, where $1<a<2$. The separation of betatron and sideband is larger than $\nu_s=0.025$, because of the fast head-tail instability; perhaps mode coupling of m=1 and m=2.

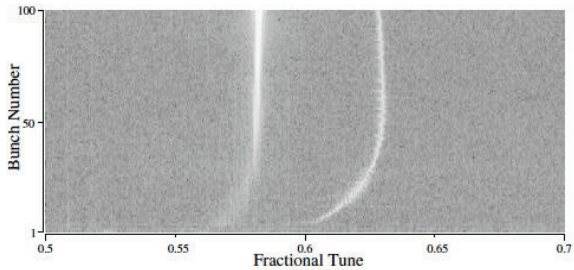


Figure 12: Measured spectrum for vertical bunch motion. Lower and upper is FFT signal of head and tail bunches [9].

Figure 13 shows simulation result of the fast head-tail instability using Eq. (3). Beam size evolution and FFT of the bunch motion are depicted in top and bottom pictures. The figure shows the threshold electron density is around 8×10^{11} m^{-3}. Upper synchrotron sideband appears above 10^{12} m^{-3}. The separation from betatron tune is larger than synchrotron tune 0.025. The result agrees well with the measurement.

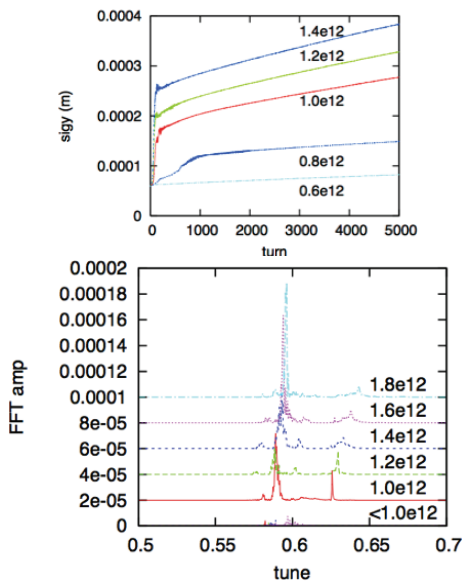


Figure 13: Simulation result of the fast head-tail instability using Eq. (3). Top picture shows beam size growth for various electron densities. Bottom picture shows FFT of the bunch motion in the top picture.

Single Bunch Instability in PETRA-III and Cesr-TA

The single bunch instability has been observed in PETRA-III and Cesr-TA. PETRA-III is synchrotron light source with very low emittance (ε_x=1nm). The electron cloud instability is observed at multi-bunch operation with a narrow bunch spacing 16 ns [11]. They have observed upper sideband signal, which separates from betatron signal $>\nu_s$. The measurement in PETRA-III agrees well with that in KEKB.

Cesr-TA is operated to study the electron cloud effect in linear collider positron damping ring with very low emittance. The sideband signal observed in Cesr-TA is negative sideband of betatron signal; $\nu_y-\nu_s$ [12]. The separation from the betatron tune is just the synchrotron tune ν_s. Simulations, which were performed for Cesr-TA, showed that negative sideband was dominant in the low emittance [14].

SUMMARY AND CONCLUSIONS

Mode spectra observed in the coupled bunch instability are clear evidence of the electron cloud effects. Corrective motion of electrons interacting with beam reflects to the unstable mode spectra. Electrons moving in drift space induces broad modes higher than m~h/2. Electrons in weak solenoid field induces slow and positive broad modes m~+0. Electrons in bending field causes instabilities slow mode m=-1.

Synchro-betatron sideband spectra in the single bunch instability appear as positive sideband in KEKB and PETRA-III, while negative sideband in Cesr-TA. Simulations can give consistent results as the measurement, but underlying physics is not clear.

ACKNOWLDGMENT

The author thanks teams of KEK-PF, KEKB, BEPC, DAFNE, PETRA-III and Cesr-TA for fruitful discussions on the experimental results.

REFERENCES

[1] M. Izawa, Y. Sato, T. Toyomasu, Phys. Rev. Lett. 74, 5044 (1995).
[2] K. Ohmi, Phys. Rev. Lett. 75, 1526 (1995).
[3] Z. Guo et al., Phys. Rev. ST-AB 5, 124403 (2002).
[4] M. Tobiyama et al., Phys. Rev. ST-AB 9, 012801 (2006).
[5] S. Win et al., Phys. Rev. ST-AB 8, 094401 (2005).
[6] T. Demma et al., Proceeding of PAC09. M. Zobov et al., Proceedings of ECLOUD12.
[7] K. Ohmi, Z. Zimmermann, Phys. Rev. Lett. 85, 3821 (2000).
[8] H. Fukuma et al., Proceedings of HEACC2001.
[9] J. Flanagan et al., Phys. Rev. Lett. 94, 054801 (2005).
[10] E. Benedetto et al., Proceedings of PAC07, p. 4033 (2007).

[11] R. Wanzenberg, Proceedings of ECLOUD'10 & 12.
[12] G. Dugan et al, Proceedings of ECLOUD'10.
[13] H. Jin et al., Japanese Journal of Applied Physics 50, 26401 (2011).

TRACKING OF A PETRA III POSITRON BUNCH WITH A PRE-COMPUTED WAKE MATRIX DUE TO ELECTRON CLOUDS *

A. Markovik, U. van Rienen, University of Rostock, Germany

Abstract

At the synchrotron radiation facility at DESY transversal tune spectra have been observed which are characteristic for an interaction of the positron beam with possible electron clouds in the ring. The filling patterns at which these incoherent tune shifts happen are favorable to the growth of the electron cloud density, i.e. long bunch trains with short intra-bunch distances. Eventually the vertical emittance growth with the originally designed equidistant filling (with 8 or 16 ns bunch spacing) has been avoided by fillings with short bunch trains and longer gaps between them and yet achieving the designed beam current of 100 mA. In this paper we examine the positron bunch stability of PETRA III for certain e-cloud densities and bunch parameters. A PIC simulation of the interaction of the bunch with an e-cloud yields the wake kick on the tail particles for an offset in the transverse centroid position of the head parts. With such a pre-computed wake matrix, we investigate the stability of a single bunch by tracking it through the linear optics of the ring while at each turn applying the kick from the e-cloud. The simulation results are in a good agreement with the measurements.

INTRODUCTION

PETRAIII at DESY is a synchrotron radiation facility running in a top up operation modus with positrons. The machine is characterized (Table 1) by an ultra low emittance and with an emittance ratio of 1% it features very flat bunches. The design beam current of 100 mA was planned to be achieved with fillings of 40 or 960 equally spaced bunches. However, for the filling scheme with 960 bunches with only 8 ns bunch-to-bunch distance a strong vertical emittance growth has been reported for currents about 50 mA [1]. The corresponding measurements of the tune spectra (Figure 1) show sidebands in the vertical tune which suggest incoherent effects. These effects are brought in connection with electron cloud effects on the bunch. Indeed the designed beam current of 100mA has been achieved by filling patterns where the e-cloud can not reach dangerous densities i. e. 60 trains of only 4 bunches with train to train distances of 80ns and bunch to bunch distance of 8ns. The e-cloud build up simulations with ECLOUD 4.0 (reported in [1]) for a train with bunch to bunch spacing of 8 ns and bunch population of $0.5 \cdot 10^{10}$ positrons (SEY δ= 2.5) show that after the first 4 bunches the e-cloud density is still below $5 \cdot 10^{11}$ 1/m^3 which is below the instability threshold computed as in [2]. In 2011 and 2012 further 100 mA runs were performed with trains of 40, 60, 240 and 480 equidistantly spaced bunches with bunch to bunch distance of 192, 128, 32 and 16ns, respectively. Only during the run with 480 bunches a significant emittance growth has been measured. The measurements are very valuable since they give the opportunity to validate the simulations. Eventually the question the instability simulation should answer is at what e-cloud density the instability of a single bunch may occur.

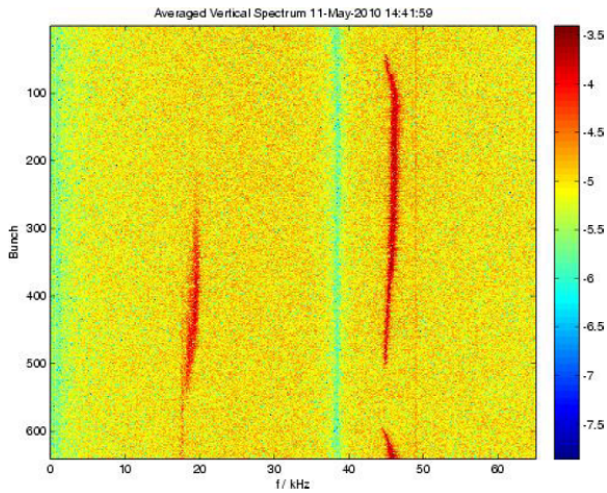


Figure 1: Vertical tune spectra of each of the 640 bunches with 8ns spacing, measured on May 11, 2010 [1]. The total beam current was 62 mA. The red color represents the sidebands and the vertical betatron frequency of 38kHz is green. (Courtesy of R. Wanzenberg.)

SIMULATION

In order to simulate the stability of a single bunch, the bunch particles are tracked through the linear optics of the machine [3]. Thereby the action of the e-cloud on the bunch is approximated by a transverse wake kick which is applied on each turn. The idea of K. Ohmi to slice the 3D bunch and compute a wake function from every longitudinal slice of the bunch backwards, leads to a triangular wake matrix. In order to apply the computed wake matrix for the bunch tracking, properties of the wake field such as time invariance, superposition and linearity are supposed.

The program MOEVE PIC Tracking [4] simulates the interaction of a single bunch with an electron cloud. The cloud is modelled as a uniform distribution of electrons in a beam pipe and is assumed to be generated by the preceding bunches. Since the perturbation of the e-cloud in the transverse plane is due to the transverse displacement

* Work supported by DFG under contract number RI 814/20-2.

of the bunch we simulate the interaction by vertically displacing each slice of the bunch by $\Delta y = \sigma_y$. The beam pipe has a small radius of 5 mm and a uniform electron distribution fills a length of 10 mm. The positron bunch is represented by 10^6 macro-particles whereas the cloud, at least for lower densities, is represented by unit charges. In each direction the bunch particles have a Gaussian distribution, longitudinally the bunch spreads from $-3\sigma_z$ to $+3\sigma_z$ and is virtually sliced in $M = 60$ slices. The thickness of the slices in the lab frame corresponds to the time which the electrons on the beam axis need to change their vertical position for one σ_y. The time step used for the interaction simulation is 1 ps. After performing M simulations of the

Table 1: PETRA III Machine Parameters

Parameter	Symbol	PETRA III
Circumference	L	2304 m
Beam energy	E_b	6 GeV
Length (rms)	σ_z	12 mm
Emittance	ϵ_x ϵ_y	1nm 0.01nm
Synchrotron tune	ν_s	0.049
Betatron tune	$\nu_{x(y)}$	36.13/30.29
Radiation Damping	horizontal vertical longitudinal	19.75 ms 19.75 ms 9.84 ms
Momentum compaction factor	α	1.2010^{-4}
RF Frequency	RF	499.564 MHz
Beam Current	I	100 mA
Beam Charge	Q	769 nC
Bunch Charge	Q_b	1.6 nC
Positrons per Bunch	N_b	10^{10}
Mean β function	$\beta_{x/y}$	15 m
Transverse beam size (rms)	σ_x σ_y	122.47 μm 17.321 μm

interaction where each of the slices $i = 1, \ldots, M$ has an off-set at a time we receive the change of the vertical momentum $\Delta p_y(i,j)$ of the bunch particles averaged for each trailing slice $j = i+1, \ldots, M$. Normalizing the dipole kick $\Delta p_y(i,j)$ by the number of particles N_i contained in the displaced slice i and the amount of the displacement Δy_i the entries of the wake matrix $W_1(z_j, z_i)$ write as:

$$W_1(z_j, z_i) = \frac{\gamma \Delta p_y(j,i)}{p_b r_e \Delta y_i N_i} [1/\mathrm{m}^2], \qquad (1)$$

where p_b is the momentum of the bunch, γ the Lorentz factor and r_e the classical electron radius. Such a computed wake field can be converted in $[\mathrm{V/Cm}]$ by multiplying it with $1/4\pi\varepsilon_0$.

RESULTS

As reported in [5] the run with 480 equidistantly spaced bunches with bunch-to-bunch distance of 16ns show a significant emittance growth. Emittance measurements were made during two days in March 2012 as the beam was used for conditioning of the vacuum chamber (beam scrubbing). Hence we wanted to see if the simulation could reproduce the measurements. The corresponding bunch parameters are given in the second part of Table 1. The starting vertical emittance was taken to be $\epsilon_y = 20$ pm. We computed the wake matrices for the following e-cloud volume densities: $\rho_e = 1 \cdot 10^{11}$ N$_e$/m^3, $\rho_e = 5 \cdot 10^{11}$ N$_e$/m^3 (Figure 2), $\rho_e = 2 \cdot 10^{12}$ N$_e$/m^3 and $\rho_e = 5 \cdot 10^{12}$ N$_e$/m^3 (Figure 3). Plugging the wake matrices and the machine description in K.Ohmi's tracking program PETHS [3] we tracked the bunch for 2048 turns.

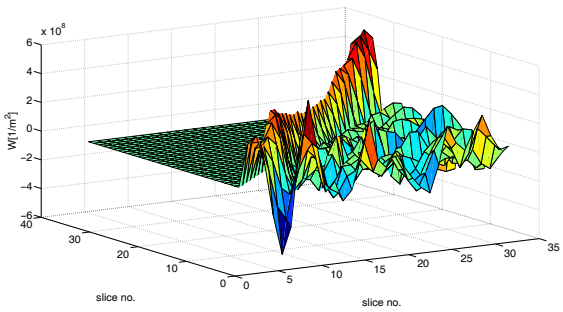


Figure 2: Wake matrix for $\rho_e = 5 \cdot 10^{11}$ N$_e$/m^3.

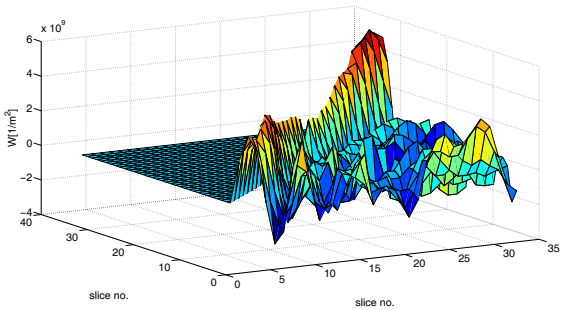


Figure 3: Wake matrix for $\rho_e = 5 \cdot 10^{12}$ N$_e$ /m^3.

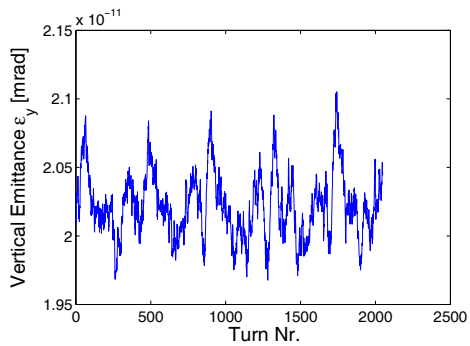


Figure 4: Almost no emittance growth after 2048 turns with $\rho_e = 5 \cdot 10^{11}$ N$_e$/m^3.

The simulation for $\rho_e = 1 \cdot 10^{11}$ N$_e$/m^3 and $\rho_e = 5 \cdot 10^{11}$ N$_e$/m^3, as shown in Figure 4 for $\rho_e = 5 \cdot 10^{11}$ N$_e$/m^3 doesn't reveal any emittance growth over

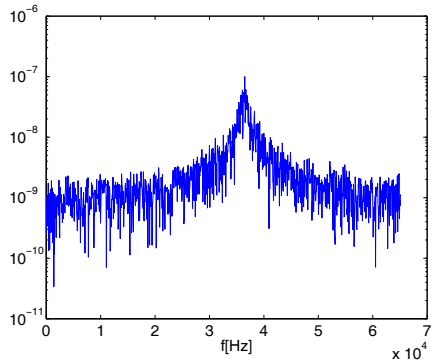


Figure 5: Vertical tune spectra for $\rho_e = 5 \cdot 10^{11}$ N$_e$/m^3

the 2048 turns. The FFT of the vertical or horizontal centroid position of the bunch during 2048 turns gives the corresponding vertical or horizontal tune spectra. The vertical tune spectra for $\rho_e = 5 \cdot 10^{11}$ N$_e$/m^3 (Figure 5) shows only the peak at the betatron frequency of 38kHz. However, for higher e-cloud densities ($\rho_e = 2 \cdot 10^{12}$ N$_e$/m^3 and $\rho_e = 5 \cdot 10^{12}$ N$_e$/m^3) the vertical tune spectra exhibits side bands (Figure 8) indicating incoherent effects on the bunch. Figure 6 shows the emittance growth for $\rho_e = 2 \cdot 10^{12}$ N$_e$/m^3 and $\rho_e = 5 \cdot 10^{12}$ N$_e$/m^3 which is clearly more moderate for the lower e-cloud density.

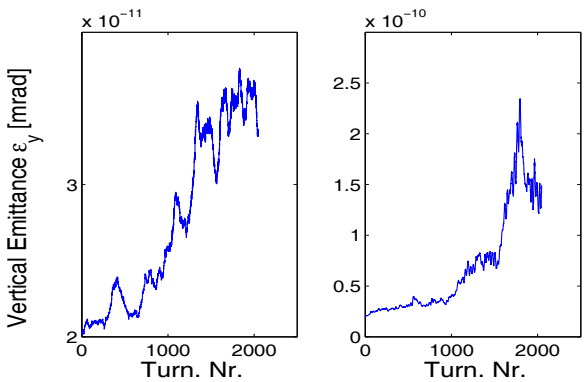


Figure 6: Emittance growth after 2048 turns for $\rho_e = 2 \cdot 10^{12}$ N$_e$/m^3 (left) and for $\rho_e = 5 \cdot 10^{12}$ N$_e$/m^3 (right).

CONCLUSION

The tracking simulation using the pre-computed wake matrix for the given bunch was able to predict the instability. As in the measurements the simulation for e-cloud densities above threshold show also sidebands in the betatron tune spectra. The simulated emittance growth seems realistic, even more the emittance from the simulation with $\rho_e = 5 \cdot 10^{12}$ N$_e$/m^3 seems to match the measured long time emittance (140 pm) from the beam scrubbing run with 480 bunches. Although further validation of the procedure is needed it seems that such a simulation may also be used to numerically estimate the threshold e-cloud density.

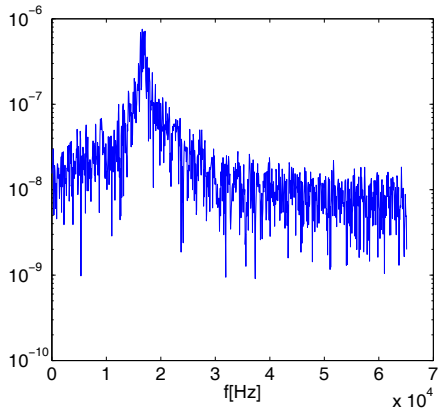


Figure 7: Horizontal tune spectra for $\rho_e = 2 \cdot 10^{12}$ N$_e$/m^3.

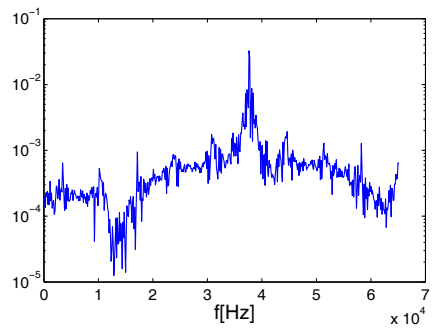


Figure 8: Sidebands in the vertical tune spectra for $\rho_e = 5 \cdot 10^{12}$ N$_e$/m^3.

ACKNOWLEDGMENT

We would like to thank K. Ohmi for letting us use his program and R. Wanzenberg for sharing his knowledge with us regarding e-cloud effects in the PETRAIII ring.

REFERENCES

[1] R. Wanzenberg. *Emittance Growth and Tune Spectra at PETRA III.* In Proceedings of ECLOUD 2010: 49th ICFA Advanced Beam Dynamics Workshop on Electron Cloud Physics. http://www.lepp.cornell.edu/Events/ECLOUD10/

[2] K. Ohmi, Electron cloud effect in damping rings of linear colliders., In *31st ICFA Beam Dynamics Workshop on Electron Cloud Effects (ECLOUD04), Napa, California, 19-23 Apr 2004*

[3] K. Ohmi. Particle-in-cell simulation of beam-electron cloud interactions. In *Particle Accelerator Conference, 2001.*, volume 3, pages 1895–1897, 2001.

[4] A. Markoviḱ, G. Pöplau, and U. van Rienen. Computation of a Two Variable Wake Field Induced by an Electron Cloud. In *Proceedings of the ICAP09, San Francisco, California, Aug 31- Sept 4.*, 2009.

[5] R. Wanzenberg. *Observations of Electron Cloud Phenomena at PETRAIII.* In Proceedings of ECLOUD 2012: Workshop on Electron Cloud Physics. http://agenda.infn.it/conferenceDisplay.py?confId=4303/

BEAM DYNAMICS STUDY CONCERNING SIS-100 PROTON OPERATION INCLUDING SPACE CHARGE EFFECTS

S. Sorge*, GSI, Darmstadt, Germany

Abstract

The projected SIS-100 synchrotron at GSI will be used for operation with intense proton and heavy ion beams. In order to avoid the crossing of the transition energy during proton operation a complicated optics scheme is proposed to provide a transition energy above the extraction energy of $E = 29$ GeV. For the purpose of optimizing the lattice, and to find a suitable working point, regime simulation scans of the dynamic aperture are performed based on MAD-X tracking. In the next step working point candidates will be used for particle tracking simulations in order to estimate beam loss due to space charge induced resonance crossing. For these studies different codes and space charge models are considered.

LATTICE PROPERTIES

Besides heavy ion operation, the SIS-100 synchrotron is also foreseen to deliver high intensity proton beams of 2.0×10^{13} protons in a short single bunch. The particles will be injected at $E_{in} = 4$ GeV and fast extracted at $E_{ex} = 29$ GeV. The working point is still under discussion. In this study, $(Q_x, Q_y) = (21.8, 17.7)$ is used.

In order to stay below transition energy, linear optics were developed that provide a transition energy corresponding to $\gamma_{tr} = 45.5$. Applying three independent families of quadrupoles, the strongly oscillating dispersion function represented by the red curve in Fig. 1 will be created. In doing so the momentum compaction factor

$$\alpha_c = \frac{1}{\gamma_{tr}^2} = \frac{1}{C} \oint \frac{D_x(s)\mathrm{d}s}{\rho(s)} \qquad (1)$$

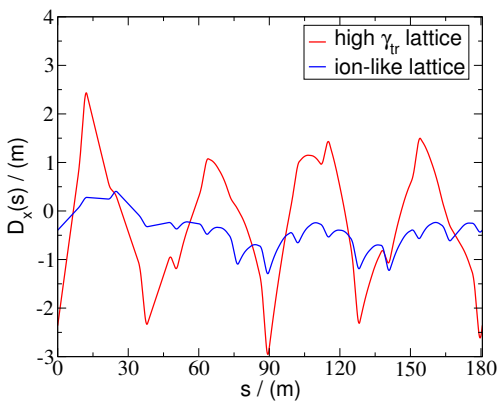


Figure 1: Dispersion function of the high γ_{tr} lattice in one of six SIS-100 sectors compared to that of an "ion-like" lattice of the same working point.

* S.Sorge@gsi.de

Table 1: Maximum values of some lattice functions of high γ_{tr} lattice and ion-like lattice, both at the working point is $(21.8, 17.7)$. 2σ emittances are $(\epsilon_{x,2\sigma}, \epsilon_{y,2\sigma}) = (13 \times 4)$ mm mrad are assumed. The natural chromaticities in row 6 correspond to Eq. (2).

	high γ_{tr}	ion-like
quadrupole families	3	2
γ_{tr}	45.5	18.4
$D_{x,max}/(\mathrm{m})$	2.9	1.3
$(\beta_{x,max}, \beta_{y,max})/(\mathrm{m})$	$(72, 29)$	$(19, 21)$
max. 2σ beam widths at $\delta = 0$, (h,v)/(mm)	$(31, 11)$	$(16, 9)$
$(\xi_{x,nat}, \xi_{y,nat})$	$(-2.4, -1.4)$	$(-0.9, -1.1)$

is minimised. The extrema of the high γ_{tr} dispersion function are much higher than those of the dispersion function generated with two quadrupole families used during heavy ion operation. In addition, the usage of the high γ_{tr} quadrupole settings increases other lattice variables, see Table 1. The chromaticities $\xi_{x,nat}, \xi_{y,nat}$ shown in that table are defined by

$$\Delta Q = \delta \xi Q. \qquad (2)$$

MAGNET IMPERFECTIONS AND RESONANCES

Real magnets are with imperfections. In particular random gradient errors in the main quadrupoles lead to an increase of the maximum horizontal beta function. Assuming them to follow a Gaussian distribution truncated at 2σ with the relative rms width of $\sigma_{rel} \approx 0.003$ of the averaged focusing strength of all quadrupole families, the maximum horizontal beta function becomes increased from 72 m to about 100 m, the exact value depends on the actual sample of random errors. In addition, there are non-linear multipole errors in main dipoles and quadrupoles each consisting of a systematic and a random contribution. The systematic components depend on $B\rho$ [1]. The random components are assumed to follow a Gaussian distribution truncated at 2σ with σ be 30 % of the corresponding systematic component. In particular the random components drive non-linear resonances which restrict the choice of the working point. To make them visible and to find a suitable working point, the dynamic aperture was determined at 100×100 working points in the range $Q_x \in [21, 22], Q_y \in [17, 18]$. The result is shown in Fig. 2. For $Q_x < 21.5$ dynamic apertures could not be determined because the lattice properties are too difficult for MAD-X even to determine the lattice functions

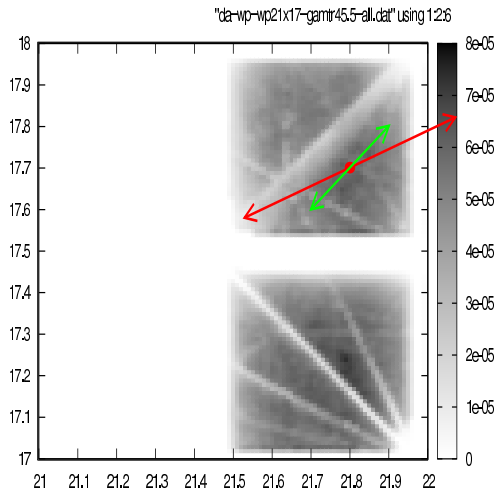


Figure 2: Resonance diagram, working point and tune spread due to $\delta = \pm 0.005$ and chromaticity without (red curve) and with (green curve) correction obtained for $\gamma_{tr} = 45.5$.

there. That is a severe restriction to the working point. So, the tune scan confirms the working point mentioned above.

The large natural chromaticities and the maximum momentum spread of $\delta_m = \pm 0.005$, reached at $E = 7$ GeV where the rf voltage reaches its maximum value, create the chromatic tune spread $(\Delta Q_x = \pm 0.27, \Delta Q_y = \pm 0.12)$, see the red arrow in Fig. 2. Because the tune spread does not fit between the half integer resonance at $Q_x = 21.5$ and the integer resonance at $Q_x = 22$ it needs to be reduced. Otherwise, particles with large δ will become lost within a synchrotron period.

CHROMATIC TUNE SPREAD AND DYNAMIC APERTURE

The reduction of the tune spread was done in two steps.

The first step consists in correcting the chromaticity with all 52 sextupoles in SIS-100 in order to achieve $\Delta Q_{x,max} = \Delta Q_{y,max} = \pm 0.1$ to make the chromatic tune spread fit in the mesh of resonances, see the green arrow in Fig. 2. The usage of 52 sextupoles to correct two variables, ξ_x, ξ_y, allowed to set the additional constraint to their focusing strengths $k_2 L$

$$\sum_{n=1}^{52} (k_2 L)_n^2 \rightarrow \text{minimum} \qquad (3)$$

in order to avoid strong sextupoles.

The second step consists in using the high-γ_{tr} settings only at high energies. At low energies the ion-like lattice described in the right column of Table 1 is applied as long as $\gamma_{tr}(E) < 18.4$, where $\gamma_{tr}(E)$ is determined by

$$\eta(E) = \frac{1}{\gamma^2} - \frac{1}{\gamma_{tr}^2(E)} = \text{const.} \qquad (4)$$

with $\gamma_{tr}(E = 29 \text{ GeV}) = 45.5$. The simplest indicator that shows the advantage of that way to proceed is the so

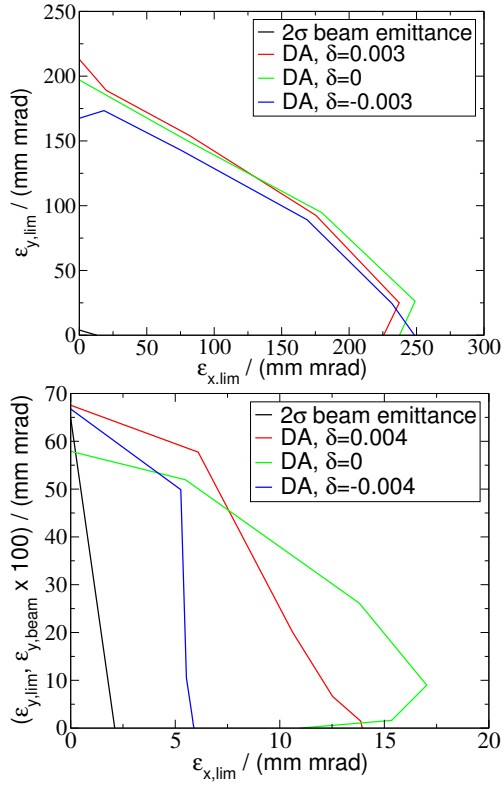


Figure 3: Dynamic apertures calculated with ion-like optics at $E = 4$ GeV (graph above) and high-γ_{tr} optics for $E = 29$ GeV (graph below). Both graphs contain curves for $\delta = 0, \pm\delta_m$ where $\delta_m = 0.003$ at 4 GeV and $\delta_m = 0.004$ at 29 GeV. The 2σ beam emittances are (13×4) mm mrad and (2.1×0.7) mm mrad, respectively. In the graph below, the vertical beam emittance is multiplied by 10.

called diagonal dynamic aperture

$$\epsilon_{lim,x}(\phi) = \epsilon_{lim}(\phi) \cos^2 \phi, \ \epsilon_{lim,y}(\phi) = \epsilon_{lim}(\phi) \sin^2 \phi \qquad (5)$$

with the angles $\phi = (0, 0.1 \cdot \pi, 0.2 \cdot \pi, 0.3 \cdot \pi, 0.4 \cdot \pi, 0.5 \cdot \pi)$. $\epsilon_{lim}(\phi)$ was determined by means of single particle tracking for $\delta = 0, \pm\delta_m$ over a period of 500 turns. Figure 3 shows the dynamic apertures for $E = 4$ GeV, $\gamma_{tr} = 18.4$, and for $E = 29$ GeV, $\gamma_{tr} = 45.5$. At 4 GeV the dynamic aperture is much larger than the 2σ beam emittance, whereas at 29 GeV and for $\delta = -0.004$, the horizontal dynamic aperture is smaller than four times the horizontal beam emittance. In other words, the spatial horizontal dynamic aperture is smaller than twice the horizontal beam width.

MULTI-PARTICLE SIMULATION

The important quantity is the number of lost particles. To estimate it, multi-particle tracking simulations were performed using the thin-lens tracking tool of MAD-X. The used lattice contained an rf cavity in order to introduce synchrotron oscillations.

At the present stage, 100 test particles were tracked over

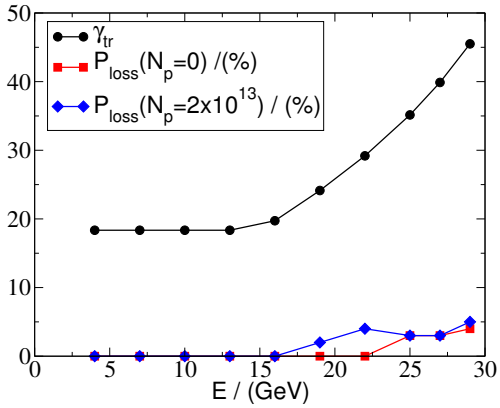


Figure 4: Beam loss without space charge and with space charge arising from 2×10^{23} protons as well as γ_{tr} as functions of the particle energy.

a period of of 16000 turns what is long enough to cover at least one synchrotron period for all energies considered The transverse initial coordinates of the test particles were chosen according to a bi-Gaussian distribution truncated at 2σ. The corresponding energy dependent rms emittances were $(\epsilon_{x,rms}, \epsilon_{y,rms})(E) = (3.25, 1)$ mm mrad $\beta_{in}\gamma_{in}/(\beta_E\gamma_E)$, where β_{in}, γ_{in} are the relativistic factors at injection energy and β_E, γ_E those of the energy considered. The momentum distribution was a Gaussian distribution truncated at two sigma with $\sigma = \delta_m/2$. According to the rf cycle of the proton ramp in SIS-100, the maximum momentum deviations were $\delta_m = 0.003$ at $E = 4$ GeV, $\delta_m = 0.005$ at $E = 7$ GeV, and $\delta_m = 0.004$ at $E = 29$ GeV [2]. Between 7 GeV and 29 GeV, δ_m was linearly interpolated. Beam loss up to 4 % was found, see red curve in in Fig. 4. Note, that these losses correspond to the sample of random magnet errors used in the simulations up to now. If choosing a different sample, the beam losses can change. In agreement with the γ_{tr} dependence of the dynamic aperture, also particle loss was found to occur only for high γ_{tr} or, more precisely, when $\gamma_{tr} > 25$ see the black curve in Fig. 4.

INCLUSION OF SPACE CHARGE

In this study only incoherent space charge effects were taken into account. The thin lens tracking tool of MAD-X was utilized to perform the simulations. Space charge could included only within the frozen space charge model. It was implemented by means of so called beambeam elements being elements implemented in MAD-X. They provide a transverse momentum kick according to a Gaussian beam shape truncated at two times the rms width. The rms width of each beambeam element was adapted to the rms width of the beam at its longitudinal position. Their implementation consists of two steps [3]. In the first step, marker elements are put at the positions of the beambeam kicks. Now the beta functions and, so the rms beam widths

$$z_{rms}(s) = \sqrt{\beta_z(s)\epsilon_{rms,z}} \text{ with } z = x, y \qquad (6)$$

can be determined. In the second step, the markers become replaced with the beambeam elements. In doing so, the oscillation of the space charge induced contribution to the betatron tune that arises from longitudinal particle motion through region of different space charge in the bunch was neglected.

Nevertheless, the results suggest the occurrence of space charge induced beam loss around $E = 20$ GeV, see blue curve in Fig. 4. The appearance of space charge induce beam loss below maximum energy seems reasonable because the influence of space charge is usually larger at lower energies. On the other hand, the appearance of a minimum energy is a hint for the necessity of complicated lattice functions arising from the high γ_{tr} optics for space charge induced beam loss to occur.

CONCLUSION

Aim of this paper was to present results from a tracking simulation study concerning the proton cycle in the SIS-100 synchrotron in order to do the first step in estimating beam loss. Key ingredient of the proton cycle is the usage of a high γ_{tr} lattice to keep the beam energy below the transition energy. This lattice causes lattice functions of strange shape. For that reason, the optics will be changed during the ramp in order to use the high γ_{tr} lattice only at high energies and to minimise the occurrence of difficulties due to the lattice functions. The tracking simulation were performed with MAD-X assuming constant energy. Effects due to synchrotron motion were regarded with respect to the tune change arising from the chromaticity. Space charge effects were included within the 2D frozen space charge model.

The results indicate that beam loss is generated by the complicated optics of the high γ_{tr} lattice. Space charge plays an important role at energies below the maximum energy. On the other hand, there is a lower energy limit for beam loss. That suggests that beam loss at medium energies is driven by the interconnection of space charge effects and the complicated high γ_{tr} optics.

Nevertheless, the present results are uncertain because the particle numbers in the simulation were small and only one sample of random errors was applied. Hence, the next step will be to repeat the simulations with larger particle numbers and for different random error samples to consolidate the results. Additionally, it would be desirable to perform simulations using codes which can model the influence of synchrotron motion on the space charge fields. These effects play a key role in the description of space charge induced beam loss in the presence of resonances.

REFERENCES

[1] V. Kapin and G. Franchetti, "The SIS100 Dipole and Quadrupole integrated kicks for nonlinear dynamics", GSI internal report ACC-note-2010-004, Darmstadt 2010.

[2] O. Chorniy, private communication.

[3] V. Kapin, private communication.

SIMULATION OF SPACE EFFECTS DURING MULTITURN INJECTION INTO THE GSI SIS18 SYNCHROTRON

S. Appel, GSI, Darmstadt, Germany
Oliver Boine-Frankenheim, TEMF, TU Darmstadt, Darmstadt, Germany

Abstract

The optimization of the Multiturn Injection (MTI) from the UNILAC into the SIS18 is crucial in order to reach the FAIR beam intensities required for heavy ions. In order to achieve the design intensities, the efficiency of the multiturn injection from the UNILAC has to be optimized for high beam currents. We developed a simulation model for the MTI including the closed orbit bump, lattice errors, the parameters of the injected UNILAC beam, the position of the septum and other aperture limiting components, and finally the space charge force and other high-intensity effects. The model is also used to estimate the required proton and heavy-ion beam emittances from the UNILAC and from the projected p-linac. For the accurate prediction of the MTI efficiency a careful validation of the simulation model is necessary. We will present first results of the comparison between experiments and simulation for low and high uranium beam currents.

INTRODUCTION

The GSI SIS18 synchrotron and the linac UNILAC are being upgraded in order to increase the beam intensity to the FAIR design parameter. For FAIR the SIS18 has to work as booster for the new SIS100 synchrotron. One crucial point in the upgrade program is the optimization of the Multiturn Injection (MTI) from the UNILAC into the SIS18. The beam loss during the MTI into the SIS18 must be minimized to avoid an intolerable increase of the dynamic vacuum pressure, which in turn leads to a reduced life-time of intermediate charge state heavy-ions [1]. The main beam loss is expected to occur on the injection septum. For FAIR intensities collective effects are expected to affect the MTI. The impact of space charge and image currents on the injection efficiency are therefore being investigated.

The aim of the present study is the development of a detailed simulation model for the MTI including the closed orbit bump and errors, the parameters of the injected UNILAC beam, the position of the septum and other aperture limiting components, and finally the space charge force and other high-intensity effects. The model can also be used to indicate the required proton and heavy-ion beam emittances from the UNILAC and from the projected p-linac. Before the model can be applied to predict and optimize the MTI for high currents a careful validation with MTI experiments is necessary.

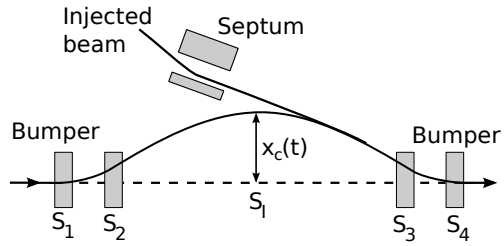


Figure 1: Layout of the multiturn injection.

MULTITURN INJECTION

In the SIS18 the beam is stacked in the horizontal betatron phase space using a closed orbit bump to bring the stacked beam close to the injection septum (See Fig. 1). The incoming beam centre will have a linear x and an angular x' displacements with respect to the undeformed closed orbit. After injection the beam will undergo betatron oscillations. One turn later the beam will come again to the injection point. Due to the betatron oscillation around the closed orbit the beam will avoid the septum. Meanwhile a new beam will be injected. This beam will have a larger amplitude of the betatron oscillation as the orbit bump is reduced. The process goes on until the maximum number of injection is reached. The beam emittance after the injection process is considered as area of the smallest ellipse that contains all injected particles. The dilation during the injection is defined as [2]

$$D = \frac{\epsilon_f}{n_{MTI}\epsilon_i} \tag{1}$$

where ϵ_i is the emittance of the injected beamlet, ϵ_j the emittance of the final beam and n_{MTI} the number of injected beamlets. The final beam emittance muss be smaller than the machine acceptance. This means, the best injection schemes have the smallest dilation and the lowest loss at the septum.

The injection bump is produced by four bumper magnets located in the injection region with positions

$$s_1 < s_2 < s_I < s_3 < s_4 \tag{2}$$

where s_I is the point of injection (See Fig. 1). If we require that the four bumper magnets produce no closed orbit distortion outside the injection region and the horizontal position and angle of the closed orbit x_c, x_c' at the injection position s_I are the degrees of freedom, than the angular kick produced by the bumper magnets are [3]

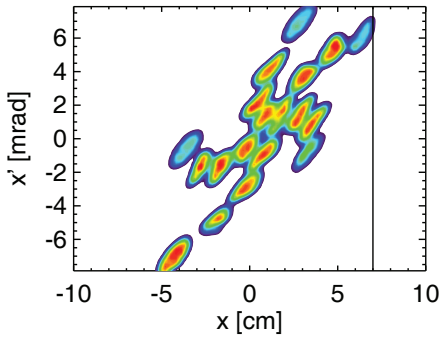


Figure 2: Phase space after 21 turns without collective effects.

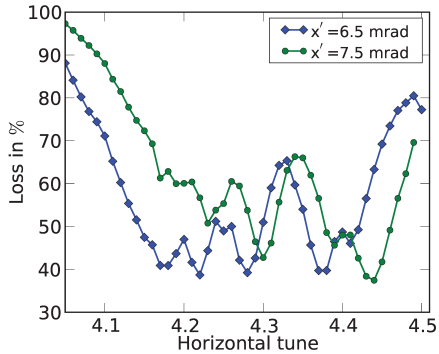


Figure 3: MTI loss as a function of the horizontal tune for a mated and a mismatched beam.

$$\varphi_1 = \frac{-b_{41}\phi_2 - b_{42}\phi_4}{b_{43}}, \quad \varphi_2 = \frac{b_{31}\phi_3 + b_{32}\phi_4}{b_{43}}, \tag{3}$$

$$\varphi_3 = \frac{d_{I2}x_c - b_{I2}x'_c}{b_{21}}, \quad \varphi_4 = \frac{-d_{I1}x_c + b_{I1}x'_c}{b_{21}} \tag{4}$$

with

$$b_{ji} = \sqrt{\beta_j \beta_i}\sin(\phi_j - \phi_i), \tag{5}$$

$$d_{ji} = \sqrt{\frac{\beta_i}{\beta_j}}\sin(\phi_j - \phi_i) - \frac{\alpha_j}{\beta_j} b_{ji}. \tag{6}$$

Here α_j, β_j, and ϕ_i are the horizontal lattice parameters at the point s_j.

For simplicity and technical reasons the SIS18 operation control program SISMODI uses an approximation of the described analytic solution for the calculation of the angular kick during injection. With this approximation the two degrees of freedom are being limited to one. For the new control system such limitations are not planned. For fixed horizontal position and angle of the closed orbit at the injection position and several horizontal tunes the angular kicks were calculated and quadratic functions were fitted on these results. By normalizing the angular kick functions with the fixed horizontal position and by multiplying with the desired bump amplitude one can adjust the four functions on each bump amplitude [4].

MTI SIMULATION CODE (PATRIC)

At the GSI accelerator physics department the PATRIC simulation code has been developed over many years for numerical collective effects. In order to investigate the MTI the sources of the code were modified such that a time dependent local orbit bump can be adjusted to the incoming beam [5].

Figure 2 shows a snapshot of a MTI simulation of the horizontal phase space without collective effects at the septum after 21 turns are injected. In the simulation the bump is added by four horizontal kicks at the bumper position on the SIS18 lattice. The deflection angles are adapted turn by turn until the injection orbit bump has disappeared given by Eq. 3-6 or calculated by the SIS18 control approximation. Collimators count turn by turn the loss on the septum and on the SIS acceptance. Figure 2 shows that some of the outside beamlets (light-blue) loose particles during the injection at this collimators. If space-charge effects are to be included the Poisson's equation is solved on a 2D transverse grid and momentum kicks corresponding to the local space charge field strength are applied.

The modified version of PATRIC can be employed to study losses, particle accumulation, emittance growth and the phase-space distribution for varying tune, bump settings, injection duration and initial particle distribution, emittance and intensity. Also the effects of a linear or nonlinear ramp, the effect of measured local close orbits deformation caused by lattice errors and the effect of the approximate SIS18 model can be studied.

MTI SIMULATION STUDIES

The MTI efficiency depends on various machine and beam parameters. Some important parameters, like the injected beam parameters, are not precisely known from measurements. In the measurements with the correction devices in the transfer channel and in the injection area in the SIS18 the injected beam slope (divergence angle x') can be modified by $\pm 1 - 3$ mrad and the beam position (x) by $\pm 10 - 20$ mm in respect to the septum position and slope. Unfortunately, it is not possible to measure the beam slope and position during injection [4]. Therefore we study the effect of the mismatched beam slope within the SIS18 injection model.

Figure 3 shows the dependency of the MTI efficiency on the horizontal tune for a mated beam slope of 6.5 mrad and for a mismatched beam of 7.5 mrad. For the mated beam (blue circles) the simulation shown the well known maxima losses at the fractional tunes related to the resonance condition $q_x n = m$ [6]. The smallest loss $\sim 40\%$ is at 4.17. For the mismatched beam (green diamonds) the maxima are shifted to the right. For tunes between 4.0-4.3 the losses are increased by more than 15% and for tunes between 4.4-4.5 the losses is 15% smaller related to the mated beam losses.

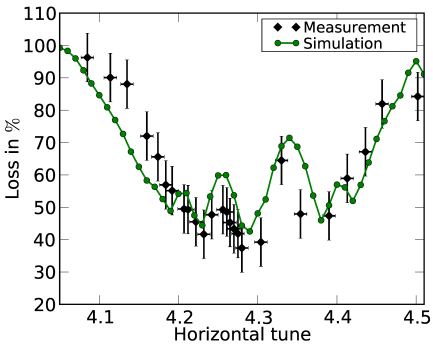


Figure 4: Simulated and measured MTI loss as a function of the horizontal tune. Measurement results provided by [7].

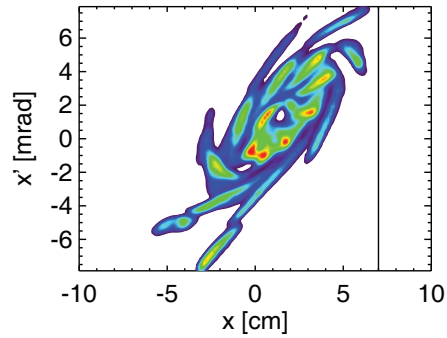


Figure 5: Phase space after 21 turns with collective effects.

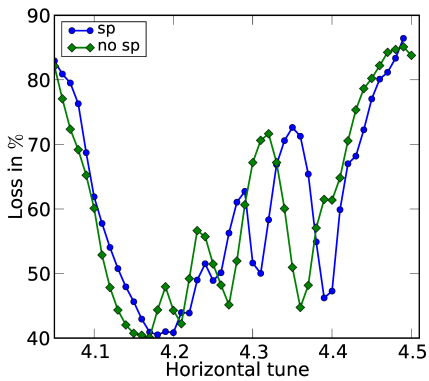


Figure 6: MTI loss as a function of the horizontal tune with and without collective effects.

For the comparison between experiments and simulation we used the machine experiments bump fall (orbit amplitude of x_c = 90 mm and bump fall of Δx_c = 2.5 mm per turn), the SISMODI angular kick calculation, measured orbit errors and beamlet emittance. The emittance was measured to $\epsilon_{rms} = 1.625$ mm mrad in the transfer channel to the SIS18 a few meters before the injection point [7]. [7] provided also the loss measurement results after 21 turns of U^{28+} beam with 1.35 mA have been injected. In Fig. 6 the simulated and measured dependence of the MTI beam loss on the tune is shown for low currents. The Figure shows a good agreement between measurements and simulations. Both show that the local beam loss maxima are located at the same fractional tunes. This was possible by setting the beamlet parameters to x = 90 mm (position of septum plus beamlet radius) and x' = 7.9 mrad (variable between 3-9 mrad) in the simulation. The reason for the good agreement is especially the choice of the divergence angle of the injected beam.

The effect of space charge and image currents on the MTI efficiency and particle distribution are considered. Figure 5 shows a snapshot of an MTI simulation with collective effects at the same moment and simulation parameters as Fig. 2. The smearing out of the particle distribution due to space charge is obvious. Close to the center individual beamlets even cannot be distinguished. The outer beamlets differ by position compared to Fig. 2 though all injection settings were equal. This observation is attributed to the tune shift.

Figure 6 shows the injection efficiency as a function of the horizontal tune with and without collective effects. The maxima and minima of the efficiency are shifted. For the injection efficiency the SIS18 high working point $q_x = 4.17$ is a good choice for high and low currents.

CONCLUSIONS AND OUTLOOK

For low beam currents a good agreement between a simulation model and MTI measurements in the SIS18 is obtained. For the present default settings in the SIS18 the injection efficiency depends very sensitively on the horizontal tune. Further well-controlled measurements at low and high beam currents are required in order to fully validate the model. Other injection schemes for lower loss like a non-linear ramp are considered. Full 3D space charge simulations a planned to understand better the space charge effects and validate the model.

ACKNOWLEDGEMENT

The authors thank D. Ondreka for useful discussions about the injection into the SIS18 and Y. El-Hayek for the measurements.

REFERENCES

[1] P. Puppel et al., IPAC11, WEPS094, p. 2724

[2] S. Fenster et al, IEEE, Vol. NS-28, No.3, June 1981

[3] C.J. Gardner, Booster Technical Note No. 197, Brookhaven National Laboratory, 1991

[4] D. Ondreka, GSI, private communication

[5] S. Paret, O. Boine-Frankenheim, HB2010, MOPD11, p. 72

[6] G. Franchetti et al, EPAC00, p. 1292

[7] Y. El-Hayek, GSI, private communication

LOW-ENERGY p-He AND mu-He SIMULATION IN GEANT4

Yu Bao
Paul Scherrer Institute, 5232 Villigen, PSI, Switzerland

Abstract

The frictional cooling method is one of the most promising methods on cooling a muon beam. Several frictional cooling schemes have been simulated in Geant4 to be efficient to produce intense muon beams. Frictional cooling works at a low energy range, where the energy loss (momentum transfer) from elastic collision is not negligible. In this paper, the p-He collision process is implemented into Geant4 and the simulation results are compared to the literature data. The process is then scaled for mu-He interaction, which will provide more accurate Geant4 simulations at low energies.

INTRODUCTION

Frictional cooling is one promising method to produce a "cold" beam. It balances the energy loss to a material with energy gain from an external electric field, so that the beam reaches an equilibrium energy and the energy dispersion is reduced. Several cooling schemes based on frictional cooling were outlined for various experiments [1–4]. In most of these schemes, low density helium gas is chosen as the retarding material for its high effective charge [5]. For accurately simulating the transport of the particles in the frictional cooling energy range in the helium gas, the low energy physics processes are needed.

Geant4 is a powerful toolkit for simulating the particle-material interations. The energy loss of the particle is handled by the ionisation process according to the stopping power from the NIST table down to 1 keV. In the energy range between 10 eV and 1 keV the model of a free electron gas [6] is used, in which the energy loss is calculated proportional to the velocity of the particle. When the particle energy gets lower than 10 eV, it's treated as "stopped" and if there is no "AtRest" process the tracking of the particle will be terminated. The scattering of the particles are simulated in Geant4 by the multiple-scattering method, which has been proved to have the same accuracy as the single-scattering simulations, and the energy loss due to elastic scattering can be neglected at high energies.

At energies lower than 1 keV, the cross sections of the elastic processes are much larger than the inelastic ones. The energy loss due to the elastic scattering plays the dominant role in the particle transport. These processes were investigated in the plasma physics decades ago and were summarized by P. S. Krstic and D. R. Schultz in the reference [7].

In this work the elastic scattering process of the p-He interaction is implemented into Geant4 and is scaled for the μ^+-He interaction. The multiple scattering and the ionisation processes in Geant4 are turned off at energies lower than 1 keV, and the elastic scattering process is used in the range between 1 keV and 0.1 eV. When the energy goes lower than 0.1 eV, the tracking is terminated.

THEORY

The scattering is usually described by the differential cross section $d\sigma(\theta,\phi)/d\Omega$, defined as the ratio of the number of particles scattered per unit time into an element of solid angle $d\Omega = sin\theta d\theta d\phi$, per unit solid angle, to the flux of incoming particles.

The total (elastic) scattering cross section is the flux of particles scattered in all directions, defined as:

$$\sigma_{el} = \int d\Omega \frac{d\sigma(\theta,\phi)}{d\Omega} = 2\pi \int_0^\pi sin\theta |f(\theta)|^2 d\theta \quad (1)$$

and the momentum transfer cross section is defined as:

$$\begin{aligned}\sigma_{mt} &= \int d\Omega \frac{d\sigma(\theta,\phi)}{d\Omega}(1-cos\theta) \\ &= 2\pi \int_0^\pi sin\theta |f(\theta)|^2 (1-cos\theta) d\theta \end{aligned} \quad (2)$$

in which $f(\theta)$ is the amplitude of the scattered wave. The differential cross section is only a function of the scattering angle θ for a certain particle velocity:

$$\sigma_d(\theta, v) = \frac{d\sigma(\theta,\phi)}{d\Omega} = |f(\theta)|^2 \quad (3)$$

$f(\theta)$ and all these cross sections are computed in the center-of-mass (CM) reference frame in Ref. [7] for ten points per energy decade at $E_{CM} = 10^{0.1j-1} eV, j = 0, 30$ (Data available for $j \leq 50$ online[1]). Both the differential and the total cross sections are obtained from extensive quantum-mechanical calculations and can be regarded as having very high accuracy.

To see how the momentum transfer cross section relates to the particle transport, consider the elastic scattering of a particle labled a from a material atom. In the CM frame the momentum of the particle is simply μv_a, where μ is the reduced mass of the ion-atom pair and v_a is the drift velocity of particle a. Hence the momentum loss is $\mu v_a(1-cos\theta_{CM})$ and σ_{mt} is the average momentum transfer in a

[1] http://www-cfadc.phy.ornl.gov/elastic/homeh.html

collision. So σ_{mt} can be used to define the energy loss per collision and the corresponding mean free path.

The average number of scatters into an angle θ per unit time is $N\overline{v}2\pi\sigma_d(\theta,\overline{v})d\theta$, where $\overline{v}$ is the mean relative particle-atom speed and N is the number of atoms per volume in the material. Multiplying it by the momentum loss in one collision and integrating over all angles, the average momentum loss per unit time is $\mu v_a N\overline{v}\sigma_{mt}$. v_a and $\overline{v}$ can be related by the partition of the total energy:

$$\frac{m\overline{v}^2}{2} = \frac{mv_a^2}{2} + \frac{Mv_a^2}{2} \tag{4}$$

in which m is the mass of the sample particle and M is the mass of the material atom or molecular. Then the average momentum loss can be written as: $2N[m/(m+M)]^{1/2}\epsilon\sigma_{mt}(\epsilon)$, where ϵ is the relative energy of the collision: $\epsilon = \mu\overline{v}^2/2$.

If an electric field E is applied, an equilibrium between the acceleration and the collision deceleration can be reached: $eE = 2N[m/(m+M)]^{1/2}\epsilon\sigma_{mt}(\epsilon)$ and the kinetic energy of the particle should be fixed at the equilibrium energy.

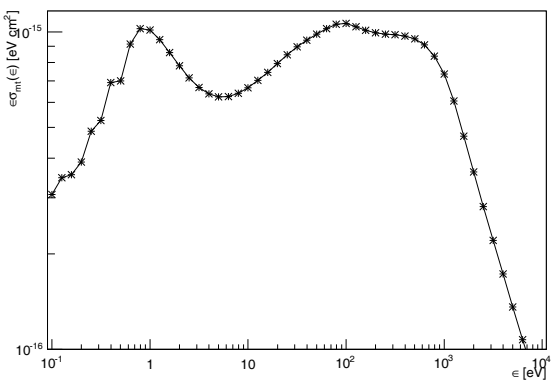


Figure 1: Collisional momentum loss as a function of relative collision energy for p-He collision.

This is true in most of the cases. Normally the parameter $\epsilon\sigma_{mt}$ continues to rise with increasing ϵ and does not have an absolute maximum in the energy range lower than the Bragg peak. This confirms the free electron gas model in Geant4. However, this is not the case for the proton-helium interaction. Figure 1 shows the curve of $\epsilon\sigma_{mt}$ versus ϵ for proton in the helium gas. The curve never rises higher than about 110×10^{-17} eVcm^2 at the energies lower than 1 keV, so if the electric field strength is higher than the energy loss the particle will eventually runaway from the low drift velocity and reach a higher kinetic energy. This "runaway" effect is first predicted by S. L. Lin [8] and experimentally observed by F. Howorka [9].

In order to compare the collisional energy loss and the ionisation energy loss, the $\epsilon\sigma_{mt}$ curve is converted to the energy loss per distance (the stopping power) in the laboratory reference frame. Figure 2 shows the energy loss

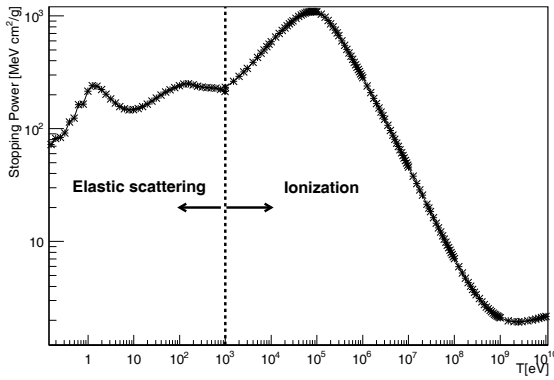


Figure 2: Stopping power of proton in helium. Left side of the dash line is the energy loss from elastic scattering, and the right side of the line is the ionisation energy loss according to the NIST data.

of the two processes matches good at the energy of 1 keV. From Fig. 1 we know that at energies higher than 1 keV the collisional energy loss rapidly gets down to zero at 6 keV. It's much smaller compared to the ionisation energy loss. In the energy range between 1 eV and 1 keV the stopping power is roughly the same. In case an electric field is applied, the equilibrium energy is hard to reach in this energy range. The particle will run away from the low energy to above 1 keV as electric field raises.

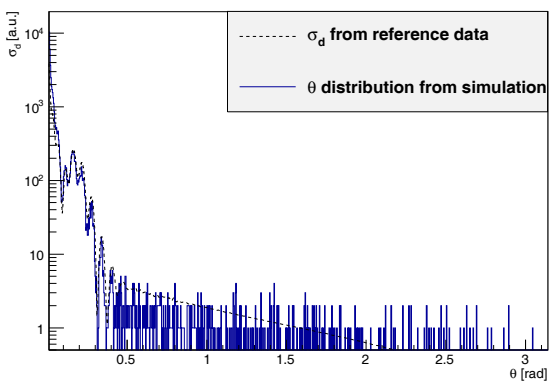


Figure 3: Simulated θ distribution (solid curve) and the differential cross section from reference data (dash curve). $T = 10eV$

Except for the average calculation, the energy loss for each single scattering can be accurately calculated from the differential cross section. The scattering angle θ is firstly randomly sampled according to the particle kinetic energy T and the differential cross section σ_d. Figure 3 shows an example of simulated θ distribution at $T = 10$ eV compared with the differential cross section data from reference [7]. Most of the time the particles are scattered in very small angles, whereas the large angle scatterings are

important for the energy loss. Considering the kinetic energy and momentum conservation in the elastic process, the energy loss dT of the particle and the scattering angle Φ in the laboratory frame are then obtained:

$$dT = \frac{2mM}{(m+M)^2}(1 - cos\theta)T \tag{5}$$

$$\tan\Phi = \frac{sin\theta}{cos\theta + m/M} \tag{6}$$

Equation (5) shows that the energy loss of each collision dT is strongly correlated to the scattering angle θ by the factor $(1 - cos\theta)$. The small angle scatterings have minimum effect on the energy loss. The energy loss is mostly from the large angle scattering. The randomly generated θ causes a large fluctuation in dT.

Instead of using momentum transfer cross section σ_{mt} as in the average calculation, the transport cross section σ_{el} is used for determin the mean free path in the accurate simulations. Because σ_{el} is much larger than σ_{mt}, the computing time is much longer in the accurate simulation.

SIMULATED "RUNAWAY" EFFECT

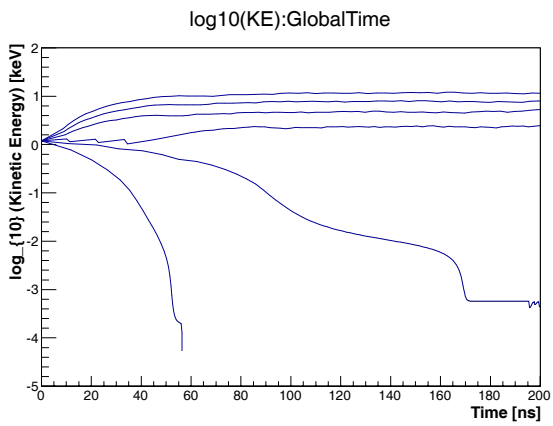


Figure 4: Run away of protons in helium with a density of $0.01\ \mathrm{mg/cm^3}$. Under various electric fields the protons reach different equilibrium energy. As the electric field increases linearly from $100\ \mathrm{kV/m}$ to $600\ \mathrm{kV/m}$, the equilibrium energy does not raise linearly but jumps from 1 eV to 1.5 keV, indicating the run away effect.

By applying various electric fields, different equilibrium energies are expected. Because of the "runaway" effect the equilibrium energy should have a jump as the electric field changes linearly. Figure 4 shows this jump. The protons are injected to a helium gas (with a density of $0.01\ \mathrm{mg/cm^3}$) tube with the same initial energy. The kinetic energy changes with time under 6 different electric fields from $100\ \mathrm{kV/m}$ to $600\ \mathrm{kV/m}$. A field strength of $100\ \mathrm{kV/m}$ is not enough to compensate the energy loss. The kinetic energy goes lower than the low limit and the particle is killed by the program. With $200\ \mathrm{kV/m}$ the particle is slowly reaching the equilibrium energy around 1 eV. When the electric field reaches $300\ \mathrm{kV/m}$, the equilibrium energy jumps to 1.5 keV, and as the field strength raises linearly to $600\ \mathrm{kV/m}$, the equilibrium energy goes up linearly to 10 keV.

CONCLUSION AND THE CODE STATUS

The p-He elastic scattering process has been implemented into Geant4 at energies lower than 1 keV. Based on this process, the "runaway" effect is successfully simulated.

Currently the simulation for muons is still under testing. Several scaling methods are considered. Because of the lack of the muon scattering study in literature, we still need to check the muon simulation with our experiment at the Paul Scherrer Institut.

REFERENCES

[1] D. Taqqu, "Compression and extraction of stopped muons," *Phys. Rev. Lett.*, vol. 97, p. 194801, Nov 2006.

[2] H. Abramowicz *et al.*, "A muon collider scheme based on frictional cooling," *Nucl. Instr. and Meth. A*, vol. 546, no. 3, pp. 356 – 375, 2005.

[3] T. Hart, T. Luo, D. Summers, and K. Paul, "End-to-End G4Beamline Simulation of an Inverse Cyclotron for Muon Cooling," *Proceeding of IPAC-2012-MOPPC046*, 2012.

[4] M. Zolotorev, A. Sessler, G. Penn, J. S. Wurtele, and A. E. Charman, "Enhancing trappable antiproton populations through deceleration and frictional cooling," *Phys. Rev. ST Accel. Beams*, vol. 15, p. 033502, Mar 2012.

[5] Y. Bao, A. Caldwell, D. Greenwald, and G. Xia, "Low-energy [mu]+ via frictional cooling," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 622, no. 1, pp. 28 – 34, 2010.

[6] J. Linhard and A. Winther *Mat. Fys. Medd. Dan. Vid. Selsk*, vol. 34, no. 10, 1963.

[7] P. Krstic and D. Schultz, "Elastic and related transport cross sections for collisions among isotopomers of $h^+ + h$, $h^+ + h_2$, $h^+ + he$, h + h, and $h + h_2$," *Atomic and plasma-material interaction data for fusion*, vol. 8, 1998.

[8] I. R. G. S L Lin and E. A. Mason, "Mobility and diffusion of protons and deuterons in helium-a runaway effect," *J. Phys. B: At. Mol. Phys. 12*, vol. 12, no. 24, p. 4179, 1979.

[9] F. C. F. F Howorka and D. L. Albritton, "h^+ and d^+ ions in he: observations of a runaway mobility," *Journal of Physics B: Atomic and Molecular Physics*, vol. 12, no. 24, p. 4189, 1979.

HIGH PRECISION CAVITY SIMULATIONS*

W. Ackermann**, T. Weiland, Technische Universität Darmstadt,
Institut für Theorie Elektromagnetischer Felder (TEMF),
Schlossgartenstraße 8, 64289 Darmstadt, Germany

Abstract

The design of resonant radio frequency cavities used in particle accelerator machines to accelerate charged particles of various species is heavily based on proper computer simulations. While the determination of the resulting field distribution can be obtained by analytical means only for a limited number of cavity shapes it is essential to apply appropriate computer programs to find sufficiently accurate approximate solutions. The achievable quality first depends on the underlying mathematical model which then has to be solved accurately on a discrete level. The precise knowledge of the distribution of the electromagnetic fields both within the cavities as well as on the surface of the resonators is essential for appropriate cavity-shape optimizations and accurate beam dynamics studies.

INTRODUCTION

In the context of highly resonating cavities a promising method to determine the electromagnetic field distribution inside the structures is given by the eigenmode analysis where a limited number of eigensolutions is used to characterize the devices within a specified frequency range. Selected eigenmodes can be determined with the help of suitable eigenmode solvers once the underlying continuous mathematical model is properly transformed into a convenient matrix formulation. Under actual operating conditions, the fields in the resonators have to be coupled to the fields in the external devices either to enable energy transfer from e.g. the sources to the beam or, conversely, to dump beam-driven parasitic modes to internal or external loads. Compared to lossless standing-wave structures where real-valued variables are sufficient to describe the entire field distribution this is no longer true in a dissipative environment where a resulting net energy transfer has to be supported. On the other hand, a complex-valued formulation completely enables to describe the physical space-dependent phase variation and additionally allows to characterize the oscillation by simultaneously extracting the corresponding quality factor next to the resonance frequency. Unfortunately, the solution of complex-valued eigenvalue systems is much more demanding compared to the widespread real-valued formulations and special care has to be put in the implementation of the computer programs to achieve good performances for large-scale applications.

* Work partially supported by DESY, Hamburg
**ackermann@temf.tu-darmstadt.de

MATHEMATICAL MODELING

A proper mathematical model to describe the electromagnetic field distribution within highly resonating structures with an eigenmode analysis in mind is obviously based on Maxwell's equations in frequency domain. We use the differential notation

$$\operatorname{curl} \vec{H} = \vec{J} + j\omega\vec{D} \tag{1a}$$

$$\operatorname{curl} \vec{E} = -j\omega\vec{B} \tag{1b}$$

$$\operatorname{div} \vec{B} = 0 \tag{1c}$$

$$\operatorname{div} \vec{D} = \varrho \tag{1d}$$

to point out the interdependence of the applied electromagnetic field components. In this context, $\vec{E}$ and $\vec{H}$ represent the phasors of the electric and magnetic field strength while $\vec{D}$ and $\vec{B}$ specify the electric and magnetic flux densities, respectivly. The symbol ω represents the angular frequency while $\vec{J}$ and ϱ describe the sources of the electromagnetic fields. In the following we concentrate on linear isotropic materials which may be inhomogeneous if required. This limitation simplifies the material relations to $\vec{D} = \varepsilon\vec{E}$ and $\vec{B} = \mu\vec{H}$ with space-depending scalar proportionality factors. Finally, the electric conductivity σ give rise to the electric current density $\vec{J} = \sigma\vec{E}$ which is responsible for the specific bulk-related losses.

Discretization

To transform the continuous formulation into a suitable matrix equation we discretize (1) with the help of the finite element method (FEM) [1]. Eliminating either the electric or the magnetic field combines the two first-order differential equations into one second-order equation. This procedure simplifies the overall solution process because only one type of field either with tangential or normal continuity conditions has to be properly represented on the discrete level. In the end, the initially eliminated field can be reconstructed in a postprocessing step if one of the curl relations in (1) is applied to the selected quantity.

With the well-known Nédélec elements in mind we set the focus of the formulation on the electric field only and combine (1a) and (1b) to the double-curl equation

$$\operatorname{curl}\left(\frac{1}{\mu}\operatorname{curl}\vec{E}\right) + j\omega\vec{J} - \omega^2\varepsilon\vec{E} = 0 \tag{2}$$

which finally will be discretized following Galerkin's approach. The electric field strength $\vec{E} = \sum_i x_i \vec{\omega}_i^{3D}$ is expanded in terms of locally defined real-valued vector basis

functions such that the introduced degrees of freedom x_i have to be of complex type [2].

Collecting the unknown weighting coefficients x_i in the vector variable $\vec{x}$ allows to formulate the entire problem in terms of the eigenvalue problem

$$A^{3D}\vec{x} + j\omega\mu_0\, C^{3D}\vec{x} - \omega^2\mu_0\varepsilon_0\, B^{3D}\vec{x} = 0 \qquad (3)$$

which has to be solved accurately to fulfill the full set of Maxwell's equations also on the discrete level. This is especially important because the fundamental charge relation (1d) is not explicitly stated in the problem formulation but will be implicitly satisfied in the dynamic case once the curl equations are properly treated. The matrices A^{3D}, B^{3D} and C^{3D} in (3) are specified according to their components

$$A_{ij}^{3D} = \iiint_\Omega \frac{1}{\mu_r}\,\mathrm{curl}\,\vec{\omega}_i^{3D}\cdot\mathrm{curl}\,\vec{\omega}_j^{3D}\,\mathrm{d}\Omega \qquad (4a)$$

$$B_{ij}^{3D} = \iiint_\Omega \varepsilon_r\,\vec{\omega}_i^{3D}\cdot\vec{\omega}_j^{3D}\,\mathrm{d}\Omega \qquad (4b)$$

$$C_{ij}^{3D} = \iiint_\Omega \sigma\,\vec{\omega}_i^{3D}\cdot\vec{\omega}_j^{3D}\,\mathrm{d}\Omega \qquad (4c)$$

and are sparsely populated on principle because of the local support of the applied basis functions.

If losses are considered in the simulation the resulting damping behavior of the oscillation is naturally described by an occuring imaginary part of the angular frequency. Conversely, if neither conductive material is available within the computational domain nor any boundary-related loss mechanism has to be considered the matrix system (3) reduces to a classical generalized eigenvalue problem which can be solved following standard techniques.

Boundary Conditions

The important boundary conditions are incorporated into the system with the help of the surface contribution

$$-j\omega\mu_0 \oiint_A (\vec{n}\times\vec{H})\cdot\vec{\omega}_j^{3D}\,\mathrm{d}A \qquad (5)$$

which originates from a partial integration of the fundamental double-curl contribution in combination with Faraday's law (1b) to eliminate the emerging curl of the electric field strength. There is a large variety of possible boundary conditions available to encapsulate the computational domain from the remaining structure. Regarding cavity simulations the most important and often idealized boundary conditions are specified as follows:

- Perfect magnetic conductive material (PMC):

  The interface condition to perfect magnetic conductive material is characterized by vanishing tangential magnetic field components in the interface plane. With $\vec{n}$ representing the local surface normal vector an appropriate relation is given by $\vec{n}\times\vec{H} = 0$ which prevents the boundary integral (5) to contribute to the overall FEM formulation and is therefore known as the natural boundary condition of the method. In the context of eigenmode analysis the magnetic boundary condition is used to realize symmetry planes.

- Perfect electric conductive material (PEC):

  The interface condition to perfect electric conductive material is characterized by vanishing tangential electric field components in the interface plane. This widespread boundary condition is not represented with the help of the boundary integral (5) but will be incorporated into the final formulation by explicitly forcing the corresponding weighting coefficients of the applied Nédélec-type vector basis functions to be zero. The integral (5) finally does not contribute to the formulation because the adjacent vector basis functions do not show any tangential component to the surface. An efficient implementation will eliminate those contributions a priori and therefore has to distinguish between volume and surface related basis functions. The electric boundary condition is applied to realize symmetry planes and to approximate the metallic surfaces of superconducting cavities.

- Impedance boundary condition:

  One of the local acting lossy boundary conditions which can be used to formulate the necessary energy exchange in the interface plane is given by the impedance relation $Z\,(\vec{n}\times\vec{H}) = \vec{n}\times(\vec{n}\times\vec{E})$ where Z represents the ratio of the tangential electric to the magnetic field strength. Incorporating this representation into the boundary formulation (5) results in

  $$j\omega\mu_0 \oiint_A \frac{1}{Z}\,(\vec{n}\times\vec{\omega}_i^{3D})\cdot(\vec{n}\times\vec{\omega}_j^{3D})\,\mathrm{d}A \qquad (6)$$

  while the integral part can be immediately added to the system matrix (4c) if the specified impedance is modeled independent of the frequency. If a frequency dependent representation has to be resolved in the system (3) the matrix equation turns from quadratic to higher-order polynomial or even nonlinear in general and radically increases the requirements on the underlying eigenvalue solver. With respect to the cavity eigenmode calculation the concept of impedance boundary condition leads to the most simple approach to enable wave propagation in either coaxial lines attached to the resonators with the help of carefully designed coupler units or the beam tubes. Unfortunately, this efficient formulation is restricted to single-mode propagation only so that the usable frequency range is limited to the interval of fundamental mode transport in the particular waveguides. Compared to more generally applicable boundary conditions the formulation in (6) advantageously retains the sparsity pattern of the original system and does not introduce additional matrix elements. This favorable property is of particular importance if efficient eigenvalue solvers have to be implemented where preconditioners are applied to solve linear systems of equations.

- Port boundary condition:

  Accelerating cavities are necessarily equipped with fundamental couplers to enable the energy transfer from the radio-frequency sources to the beam. In the same way intentionally installed highly-specialized couplers allow the extraction of parasitic modes to prevent performance degradation in high-current applications. While a single-mode wave propagation can be efficiently modeled with the help of the impedance boundary condition this is no longer true for the multi-mode case. The accurate modeling of a true port interface can be realized with the help of a two-dimensional (2-D) modal expansion of the resulting electromagnetic field in the specified boundary plane. This procedure enables the correct treatment of any kind of wave propagation in the attached line including even the excitation below cut-off. This ability is especially interesting even for the simple mono-mode case because the port-interface plane can then be located near to the cavity where higher-order modes still have to be considered. Following this approach a large amount of degrees of freedom can be saved because otherwise the interface plane has to be moved far away from the cavity to guarantee that all higher-order modes are sufficiently decayed. The modeling of the coaxial lines is particularly expensive because it naturally requires a fine grid resolution to resolve the small geometric details. The numerical treatment of the port boundary condition is established again with the help of the boundary integral representation (5) in combination with Faraday's law (1b) because the presented three-dimensional (3-D) formulation is based on the electric field strength only. A full modal approach has to consider all modes traveling in both directions of the attached waveguide

$$\vec{E} = \sum_{\nu} \left(A_{\nu}^{(-)} \vec{E}_{\nu}^{(-)} + A_{\nu}^{(+)} \vec{E}_{\nu}^{(+)} \right) \tag{7}$$

  with the incoming waves representing the desired excitation. Keeping the aspired eigenvalue formulation in mind we do not explicitly excite the resonators at the port interfaces and concentrate the modeling on the outgoing waves. The electric field strength can then be composed of individual mode contributions according to

$$\vec{E} = \sum_{\nu} A_{\nu} \left(\vec{E}_{t,\nu} + \vec{e}_z \, E_{z,\nu} \right) e^{-jk_{\nu} z} \tag{8}$$

  where the full field is decomposed in terms of the transverse (t) and longitudinal (z) components. A local coordinate system is introduced such that $\vec{n} = \vec{e}_z$ represents the orientation of the outward normal vector of the relevant port plane. Following this approach the TEM, TE and TM modes in homogeneous waveguides can be described with only one formulation which is in addition even applicable for inhomogeneous port planes [3].

  To evaluate the resulting boundary integral representation

$$\iint_{A_{\text{port}}} \frac{1}{\mu_r} \left(\vec{n} \times \operatorname{curl} \vec{E} \right) \cdot \vec{\omega}_j^{\text{3D}} \, \mathrm{d}A \tag{9}$$

  the curl of the electric field can be rephrased analytically while special care has to be put in the analysis because of the fundamental differences in the handling of the transverse and longitudinal coordinates. The resulting expression

$$\vec{n} \times \operatorname{curl} \vec{E} = \sum_{\nu} A_{\nu} \left(\operatorname{grad} E_{z,\nu} + jk_{\nu} \vec{E}_{t,\nu} \right) \tag{10}$$

  has already been evaluated in the port plane to simplify the notation. According to the orthogonality relation of the waveguide modes the missing scaling factors are specified according to the expression

$$A_{\nu} = \iint_{A_{\text{port}}} \varepsilon_r \, \vec{E} \cdot \vec{E}_{t,\nu}^{*} \, \mathrm{d}A \tag{11}$$

  which finally enables to set up the entire port boundary condition once the desired two-dimensional mode patterns are available for the specified port interfaces.

Port Description

Setting up a 3-D eigenvalue formulation where electromagnetic energy can dissipate through matched waveguides necessitates the calculation of the modal field pattern in the boundary plane together with the corresponding propagation constants. The desired modal data can be derived from a two-dimensional eigenvalue formulation once the evaluation frequency is fixed. In an iterative solution process the ultimate resonance frequency is then obtained by successive evaluation steps where the overall convergence rate depends on the coupling strength of the 2-D and 3-D fields. For practical applications regarding cavity simulations a few fixed-point iterations already lead to accurate results up to solver-precision.

The required field pattern $\vec{E}_{\nu} = \sum_i x_{i,\nu} \, \vec{\omega}_i^{\text{2D}}$ in the port plane can be calculated with the help of a 2-D eigenvalue formulation which is derived from the 3-D problem (3) with a necessary restriction to the port plane. Concentrating on the lossless case results in the notation

$$A^{\text{2D}} \vec{x}_{\nu} = \omega_{\nu}^2 \mu_0 \varepsilon_0 \, B^{\text{2D}} \vec{x}_{\nu} \tag{12}$$

with matrices A^{2D} and B^{2D} defined by the elements

$$A_{ij}^{\text{2D}} = \iint_{A_{\text{port}}} \frac{1}{\mu_r} \operatorname{curl} \vec{\omega}_i^{\text{2D}} \cdot \operatorname{curl} \vec{\omega}_j^{\text{2D}} \, \mathrm{d}\Omega \tag{13a}$$

$$B_{ij}^{\text{2D}} = \iint_{A_{\text{port}}} \varepsilon_r \, \vec{\omega}_i^{\text{2D}} \cdot \vec{\omega}_j^{\text{2D}} \, \mathrm{d}\Omega \tag{13b}$$

where the important propagation constants k_{ν} are hidden in the longitudinal dependence of the basis functions. Explicitly expanding the separated variables and simultaneously applying a partitioning of the weighting coefficients

according to the transverse and longitudinal components results in a block-oriented quadratic eigenvalue problem which can be also stated using the two equations

$$A_{11}^{2D}\,\vec{x}_t - jk_z B_{12}^{2D}\,\vec{x}_z + \; k_z^2 B_{11}^{2D}\,\vec{x}_t = 0 \qquad (14a)$$

$$B_{22}^{2D}\,\vec{x}_z + jk_z B_{21}^{2D}\,\vec{x}_t = 0 \qquad (14b)$$

and allows to determine the propagation constants once the evaluation frequency is fixed. The necessary submatrices are defined according to the expressions

$$A_{11,ij}^{2D} = \iint_{A_{\text{port}}} \frac{1}{\mu_r}\,\text{curl}_t\,\vec{\omega}_{t,i}^{2D}\cdot\text{curl}_t\,\vec{\omega}_{t,j}^{2D}\,\text{d}\Omega + \\ - \omega_{\text{port}}^2\,\mu_0\varepsilon_0 \iint_{A_{\text{port}}} \varepsilon_r\,\vec{\omega}_{t,i}^{2D}\cdot\vec{\omega}_{t,j}^{2D}\,\text{d}\Omega \qquad (15a)$$

$$B_{11,ij}^{2D} = \iint_{A_{\text{port}}} \frac{1}{\mu_r}\,\vec{\omega}_{t,i}^{2D}\cdot\vec{\omega}_{t,j}^{2D}\,\text{d}\Omega \qquad (15b)$$

$$B_{12,ij}^{2D} = \iint_{A_{\text{port}}} \frac{1}{\mu_r}\,\text{grad}_t\,\omega_{z,i}^{2D}\cdot\vec{\omega}_{t,j}^{2D}\,\text{d}\Omega \qquad (15c)$$

$$B_{21,ij}^{2D} = \iint_{A_{\text{port}}} \frac{1}{\mu_r}\,\vec{\omega}_{t,i}^{2D}\cdot\text{grad}_t\,\omega_{z,j}^{2D}\,\text{d}\Omega \qquad (15d)$$

$$B_{22,ij}^{2D} = \iint_{A_{\text{port}}} \frac{1}{\mu_r}\,\text{grad}_t\,\omega_{z,i}^{2D}\cdot\text{grad}_t\,\omega_{z,j}^{2D}\,\text{d}\Omega + \\ - \omega_{\text{port}}^2\,\mu_0\varepsilon_0 \iint_{A_{\text{port}}} \varepsilon_r\,\omega_{z,i}^{2D}\,\omega_{z,j}^{2D}\,\text{d}\Omega \qquad (15e)$$

and have to be evaluated with the help of the 2-D basis functions in the specified port planes. The unfavorable quadratic formulation can be rephrased into the classical generalized eigenvalue problem

$$\begin{pmatrix} A_{11} & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \vec{y}_t \\ \vec{y}_z \end{pmatrix} = -k_z^2\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}\begin{pmatrix} \vec{y}_t \\ \vec{y}_z \end{pmatrix} \qquad (16)$$

using the substitutions $\vec{y}_t = k_z\,\vec{x}_t$ and $\vec{y}_z = -j\,\vec{x}_z$ [4]. The last equation in (16) is responsible for the source-free condition of the port modes although the derived formulation originates from Maxwell's curl equations only. The formulation is of particular interest because all eigenvalues and eigenvectors can be calculated using real-valued arithmetic even for the evanescent fields.

Once the 2-D eigenvalue problem (16) is solved for the lowest eigenvalues the extracted eigensolutions have to be incorporated into the originally stated 3-D formulation using the boundary representation (9). Because of the tangential nature of the vector expression $\vec{n}\times\text{curl}\,\vec{E}$ the essential equation can also be written in the form

$$\iint_{A_{\text{port}}} \frac{1}{\mu_r}\,(\vec{n}\times\text{curl}\,\vec{E}\,)\cdot\vec{\omega}_j^{2D}\,\text{d}A \qquad (17)$$

where the 3-D basis functions have been interchanged with the corresponding 2-D variants without any approximation. Consequently, if the port-mode calculation has been performed on a 2-D subset of the 3-D discretization the specified integration does not have to be completed again but can be copied instead from the 2-D calculations, where they have been carried out anyhow. This favorable feature is readily available if the applied 2-D basis functions are originating from the 3-D counterparts by projection on the specified port plane. All necessary basis functions automatically meet this fundamental requirement if the same construction scheme is used to set up the initially unknown vector functions from the known scalar counterparts.

Supplementary to the propagation constants the knowledge of the cut-off frequencies of the considered modes is advantageous to characterize the wave propagation. The necessary values do not have to be calculated from scratch but can be extracted from (16) with the help of a Rayleigh quotient simply by setting k_z to zero.

Splitting the submatrix in (15a) according to the definition $A_{11}^{2D} = A'_{11} - \omega_{\text{port}}^2\,\mu_0\varepsilon_0\,A''_{11}$ into the specified parts A'_{11} and A''_{11} finally enables a simple incorporation of the boundary contribution (17) into the original eigenvalue formulation (3). Special care has been put in the conversions to demonstrate the symmetry of the resulting matrix block so that the entire formulation retains the favorable property.

IMPLEMENTATION

Based on the presented eigenvalue formulation a reliable computer program to enable high precision cavity simulations has been set up. The inherently high computational demands welcomes a parallel implementation utilizing distributed memory machines. The geometrical modeling of the structures is performed with the CST Studio Suite® [5]. A flexible tetrahedral mesh is used to perform the underlying FEM calculations. The electric field strength is described with the help of Nédélec-type basis functions up to the second order which are formulated on curved elements to retain the high approximation order even for non-flat material interfaces.

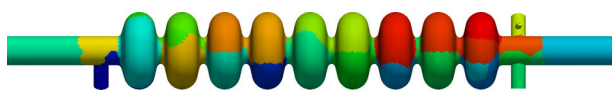


Figure 1: Partitioning of the computational domain among 32 processes. A unique color is assigned to each of the extracted sub domains to model e.g. a TESLA 9-cell cavity.

Because of the immanent coupling of excited modes to any kind of carefully designed loads the applied computer program explicitly has to enable a suitable handling of local loss mechanisms. The underlying algebraic eigenvalue solver naturally has to cope with a complex formulation which is not straightforwardly applicable to available solver packages. A robust realization of a Jacobi-Davidson-type eigenvalue solver has been implemented to extract the complex-valued eigenvectors and corresponding eigenvalues for the large 3-D problem while the 2-D port modes are determined as a comparable small problem with the help of standard LAPACK routines [6].

APPLICATION

The available algorithms have been applied to the computational model of a TESLA 1.3 GHz accelerating cavity to determine the complex resonance frequency of various configurations next to the corresponding field distributions for all modes e.g. in the first monopole passband and in the mixed first and second dipole passband [7]. The computational model consists of the nine-cell cavity, the fundamental input coupler as well as the up- and downstream higher-order modes couplers. Port boundary conditions are placed at the coaxial lines of the three couplers as well as on the two beam tubes.

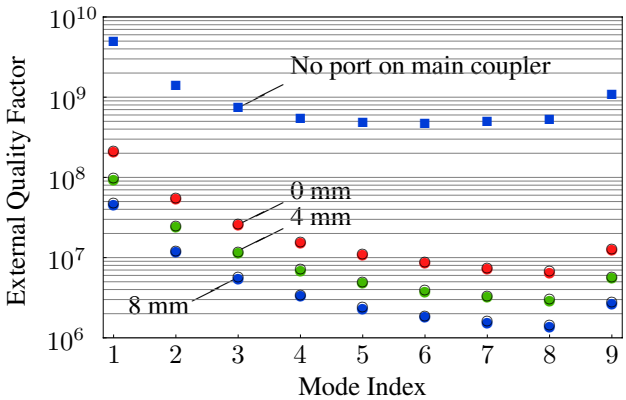


Figure 2: Quality factors for the monopole modes for various penetration depth of the main input coupler. The colored data points are obtained on a mesh using 1.3 million tetrahedrons (8.1 million complex DOF) while the black circles are included for comparison reasons and correspond to 283 thousand tetrahedrons (1.7 million complex DOF).

The implemented formulation can cover a wide range of dynamical systems reaching from strongly coupled configurations to nearly isolated resonators. A distinct feature of the complex formulation is given by the possibility to simultaneously determine the real part as well as the imaginary part of the angular resonance frequency. While the real part immediately characterizes the oscillation the imaginary part is used to describe the damping. Instead of directly specifying the damping coefficient the widespread quality factor $Q = \omega_{\text{real}}/(2\,\omega_{\text{imag}})$ is preferably used instead. Because the coupling to the external devices is the only loss mechanism considered in this work the quality factor is specified by the external quality factor only. In Figs. 2 and 3 selected simulation results are summarized to demonstrate a proper handling of the notoriously difficult large-scale numerical problem.

CONCLUSION

A robust FEM implementation of a dissipative eigenvalue formulation has been set up which enables to simulate lossy structures based on complex-valued variables without the necessity to precalculate real-valued systems in a preprocessing step. The available parallel computer program utilizes hierarchical Nédélec-type basis functions

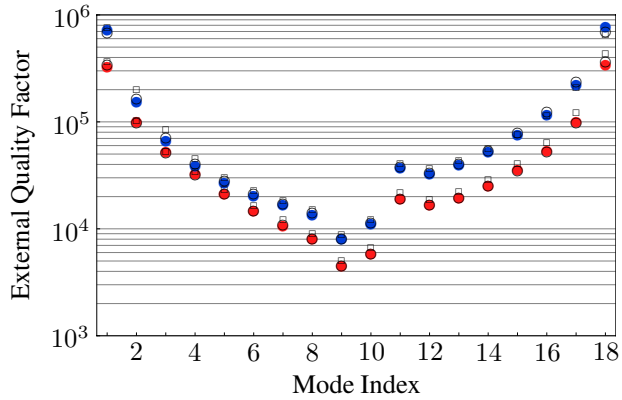


Figure 3: Quality factors for the 36 modes in the 1 st and 2nd dipole band considering both polarizations. The calculations are performed on meshes with 0.354, 0.991 and 2.090 million tetrahedrons indicated by squared, circled and colored data points which translates to 2.1, 6.1 and 12.9 million complex degrees of freedom (DOF).

up to the second order on curved tetrahedral elements and can handle besides losses in bulk materials in particular an impedance boundary condition and a port mode representation for arbitrary number of ports. The practical applicability of the combined 2-D and 3-D formulation to real-life problems has been successfully demonstrated with the precise electromagnetic modeling of a TESLA-type accelerating cavity where the influence of the attached coupling units has been incorporated consistently in the formulation.

ACKNOWLEDGMENT

The Authors thank Martin Dohlus from DESY Hamburg, Germany, as well as Wolfgang F. O. Müller from the Technische Universität Darmstadt, Germany, for many valuable discussions on the physical problem formulations and the support for using the high performance computer.

REFERENCES

[1] P. P. Silvester, R. L. Ferrari, *Finite Elements for Electrical Engineers*, Cambridge University Press, 1996

[2] P. Ingelstrm, "A New Set of H(curl)-Conforming Hierarchical Basis Functions for Tetrahedral Meshes", IEEE Trans. on Microwave Theory and Techniques, 54(1): 106-114, 2006

[3] L.-Q. Lee, Z. Li, C. Ng, K. Ko, "Omega3P: A Parallel Finite-Element Eigenmode Analysis Code for Accelerator Cavities", SLAC-PUB-13529, February 2009

[4] C. J. Reddy, M. D. Deshpande, C. R. Cockrell, F. B. Beck, "Finite Element Method for Eigenvalue Problems in Electromagnetics", NASA Technical Paper 3485, December 1994

[5] CST AG - Computer Simulation Technology, CST Studio Suite 2012, `http://www.cst.com`

[6] Intel Corporation, Intel Math Kernel Library 2011, LAPACK Math Functions, `http://www.intel.com`

[7] M. Dohlus, V. Kaljuzhny, S. G. Wipf, "Higher Order Mode Absorption in TTF Modules in the Frequency Range of the 3rd Dipole Band", TESLA Report 2002-05, 2002

IMPLEMENTATIONAL ASPECTS OF EIGENMODE COMPUTATION BASED ON PERTURBATION THEORY*

K. Brackebusch†, U. van Rienen, Institute of General Electrical Engineering,
University of Rostock, Germany

Abstract

Geometry perturbations affect the eigenmodes of a resonant cavity and thereby can improve but also impair the performance characteristics of the cavity. To investigate the effects of both, intentional and inevitable geometry variations parameter studies are to be undertaken. Using common eigenmode solvers involves to perform a full eigenmode computation for each variation step, even if the geometry is only slightly altered. Therefore, such investigations tend to be computationally extensive and inefficient. Yet, the computational effort for parameter studies may be significantly reduced by using perturbative computation methods. Knowing a set of initial eigenmodes of the unperturbed geometry these allow for the expansion of the eigenmodes of the perturbed geometry in terms of the unperturbed modes. In this paper, we study the complexity of a numerical implementation of perturbative methods. An essential aspect is the computation and analysis of the unperturbed modes since the number and order of these modes determine the accuracy of the results.

INTRODUCTION

Aiming to design a cavity with best possible performance it is necessary to optimize miscellaneous of its characteristics that all depend on the cavity's eigenmodes. The cavity shape mainly influences these characteristics and has to be changed numerous times during the optimization process. Any geometry modification entails a full recomputation of the eigenmodes causing an immense total computational effort. Perturbative methods allow for avoiding this repetitive procedure. The methods (discussed here) base on the approach of computing the eigenmodes (denoted as unperturbed) of exclusively one initial geometry using a common computation method and deriving the eigenmodes (denoted as perturbed) of a modified geometry directly from the unperturbed eigenmodes with substantially less effort.

To do so, the interaction of each unperturbed mode i with every other unperturbed mode k has to be determined by forming an expression that includes their resonant frequencies f and a volume integration over the scalar product of their electric / magnetic fields $\mathbf{E}(\mathbf{r})$, $\mathbf{H}(\mathbf{r})$

$$\mathrm{IT}_{E(\mathrm{ik})} = \iiint_{\Delta V} \mathbf{E}_\mathrm{i}(\mathbf{r}) \cdot \mathbf{E}_\mathrm{k}(\mathbf{r}) \, \mathrm{d}V \tag{1}$$

$$\mathrm{IT}_{H(\mathrm{ik})} = \iiint_{\Delta V} \mathbf{H}_\mathrm{i}(\mathbf{r}) \cdot \mathbf{H}_\mathrm{k}(\mathbf{r}) \, \mathrm{d}V \tag{2}$$

Here ΔV is the volume that is removed by the modification from the unperturbed volume V. Applying further arithmetic operations to the resulting interaction term (IT) matrix finally yields the perturbed resonant frequencies and weighting factors that allow for expanding each perturbed electromagnetic field as a series of the unperturbed fields. In [1] and [2] two perturbative methods differing in the composition of their ITs and arithmetic operations are described in detail.

In [2] and [3] their applicability was proved by means of analytically evaluable cavity geometries providing very accurate results. Using analytically computed unperturbed eigenmodes all operations can be implemented with very high precision and very low effort. However, for most cavity structures the unperturbed modes have to be numerically computed. This is not only more complex and affected by additional numerical errors but also involves some difficulties that are discussed in the following. The numerical implementation was investigated using the example of a cylindrical cavity subject to one-dimensional perturbations to compare its outcomes with the analytically computed ones.

NUMERICAL IMPLEMENTATION

The computation of the unperturbed eigenmodes is the first and most expensive operation but has to be done only once for a certain cavity structure. It is required for the ITs and the series expansion and therefore particularly important for the accuracy of the results. The simulations are done with CST MWS Eigenmode Solver [4]. The integrals IT_E and IT_H solely depend on the fields in the boundary region (where the perturbation occurs). Besides an appropriate mesh density, hence a precise discretization of the boundary is significant. Since a discrepancy between discretized and actual boundary leads to an abrupt transition of the fields to zero (Fig. 1) an insufficient discretization may seriously impair the accuracy of the IT_E and IT_H. The Finite Element Method (FEM) combined with a tetrahedral mesh and curved elements proved to reproduce the boundary much more precisely than the hexahedral dual grid of the Finite Integration Theory (FIT). Therefore, here, FEM should be used for the eigenmode computation of cavities with a curved shape.

For the further processing the computed fields have to be exported from CST MWS as discrete field points. But despite the achieved precise boundary discretization using FEM some of these field values are still affected by the de-

* Work supported by Federal Ministry for Research and Education BMBF under contracts 05H09HR5 and 05K10HRC

† korinna.brackebusch@uni-rostock.de

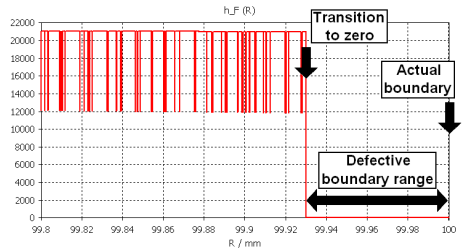


Figure 1: $\mathbf{H}_{\varphi}(r)$, cavity radius R: 100 mm, method: FIT. FEM-computed fields have a significantly smaller defective boundary range (0.3 μm) than the shown FIT-computed ones (70 μm) which are allocated on the dual grid in a certain distance to the boundary.

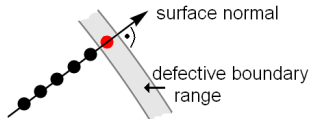


Figure 2: One-dimensional extrapolation of defective value (red) from correct values on the boundary surface normal.

fective boundary range. To minimize this impact a simple but effective procedure was developed. For it, each discrete field value that lies inside the defective region is extrapolated from a set of correct values located on the surface normal vector of the boundary (Fig. 2). By doing so, the accuracy of the ITs can be further improved.

For the computation of the ITs (done with Wolfram Mathematica [5]) two approaches were pursued. For the first approach, all discrete fields are interpolated to create continuous functions for the volume integration in Eqs. 1 and 2 in order to achieve a high accuracy. However, the computation of the interpolation functions and integrals turned out to cost an enormous effort. The main reason is that very accurate results can only be obtained by an interpolation of higher degree. But this can only be done on a structured grid which requires to interpolate the fields inside a larger volume than ΔV (in many cases the complete volume V) while the grid density must be kept appropriate for ΔV. Such a three-dimensional higher degree interpolation for a huge number of grid points is very inefficient. Consequently, an alternative approach was needed. Its main idea is to partition the volume ΔV into elements, determine the discrete field values in their centers and then substitute the integration by a summation of the products

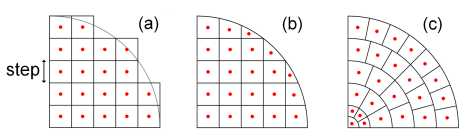


Figure 3: Volume partitioning of a cylinder (cross cut): (a) Pure cubic elements. (b) Cubic / (c) cylindrical elements with analytically determined boundary elements.

of the discrete integrands and volumes. These basic operations need only little effort and solely the fields inside ΔV have to be processed. But the commonly used partitioning into cubic elements (Fig. 3(a)) is not adequate for a summation over ΔV because of the required precise boundary discretization. Therefore, an improved partitioning algorithm was developed. If the boundary of ΔV and an element intersect volume and center of the element change. But both can be analytically computed due to the fact that usually all cavity shapes are described by (piecewise) analytical functions (Fig. 3(b),(c)). Correcting these intersection elements can essentially improve the accuracy of the IT_E and IT_H.

After the computation of all ITs, the arithmetic operations of the respective method (mostly matrix operations) have to be performed. These finally yield the perturbed resonant frequencies and the weighting factors of the series expansion. The operations are equal for analytically and numerically computed ITs and easy to implement with a low effort and very high accuracy. In case that besides the frequencies also the perturbed fields or certain cavity characteristics (e.g. E_{peak}/E_{acc} or Q_{ext}) are desired the complete or at least parts of unperturbed fields have to be previously exported to do the series expansion.

RESULTS

In the following, the implementational outcomes for a radial perturbation of a cylindrical cavity (R=100 mm, L=100 mm, fundamental mode of 1.15 GHz) investigating $\mathrm{TM}_{2.n.2}$ modes (n: radial mode index) are exemplarily discussed. The frequencies of the numerically computed unperturbed mode are highly accurate. The relative deviation is less than $1.6 \cdot 10^{-4}$ inside a very large frequency range (up to 35.71 GHz). Thereby an essential condition for the following operations is fulfilled since all intermediate and final results highly depend on the computable frequency range. However, it has to be considered that the accuracy degrades with increasing frequency due to the fact that the number of simulated mesh cells is limited by the available memory (Fig. 4).

To obtain accurate IT_E and IT_H the discrete volume elements have to be chosen small enough to properly discretize the fields of all unperturbed modes. While the IT_E and IT_H of modes with a low frequency can be computed

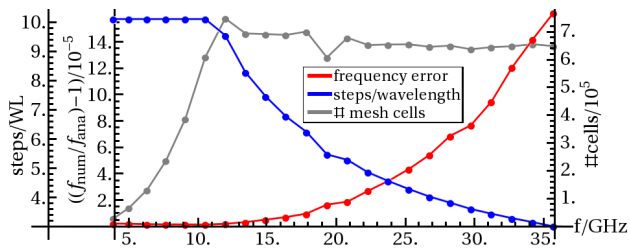


Figure 4: Parameters of CST MWS eigenmode computation of $\mathrm{TM}_{2.1.2}$ to $\mathrm{TM}_{2.23.2}$. The intended steps per wavelength reduce from 12 GHz on due to full memory usage.

with very coarse elements their accuracy for higher modes largely depends on the element step size (Fig. 5(a)). However, a large perturbation does not require elements as fine as a small perturbation does (Fig. 5(b)). A very precise step size is only needed if the perturbation is also very small. Considering this, the efficiency can be increased by avoiding the use of an unnecessarily large number of elements.

Choosing volume elements with an appropriate discretization and effort, the relative deviation of the integral IT_E of the electrical fields is largely below $2 \cdot 10^{-3}$ and hence very accurate (Fig. 6). Only for comparatively small IT_E it rises up to $7.42 \cdot 10^{-3}$. The deviation of the IT_H is equal up to the 13th mode. The IT_H of higher modes are relatively small so that they are increasingly impaired by the numerical simulation error. As a consequence the relative deviation rises up to $3 \cdot 10^{-2}$. The accuracy of the final ITs depends on the used perturbative method. For [2] solely the IT_E are required so that the so computed ITs remain as accurate as the IT_E. But for the ITs in [1] the integrals IT_E and IT_H are multiplied by the further quantities and subtracted from each other. Thereby the relaitve error contribution of IT_E and IT_H changes. This may increase the deviation for some of the resulting ITs. Due to the complexity of the mathematical relations the details are not discussed here.

Compairing the final results, the perturbed frequencies $\tilde{f}$

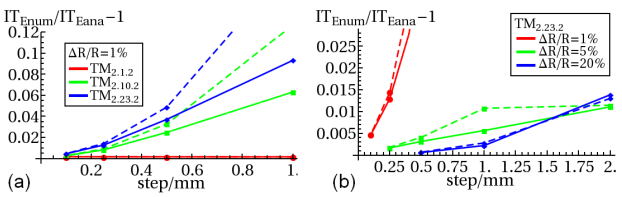


Figure 5: Relative error of selected $IT_{E(ii)}$ depending on the step size of the volume elements (— cubic, - - - cylindridal) for different modes (a) and radial perturbations (b).

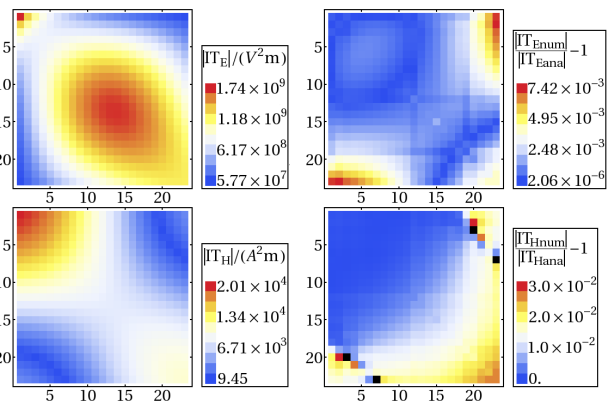


Figure 6: Absolute values and relative error of the IT_E and IT_H for radial perturbation of 5% using $3.4 \cdot 10^5$ cubic volume elements with 0.5 mm step size. For two IT_H that are almost zero the error is not shown (black areas).

based on a numerically implementation highly coincide with the ones based on an analytically implementation (Fig. 7). This proves the practicability and accuracy of a numerical implementation. Figure 7 also shows that the results using [1] deviate more from the expected ones than the results using [2] due to the previous described different composition of the ITs.

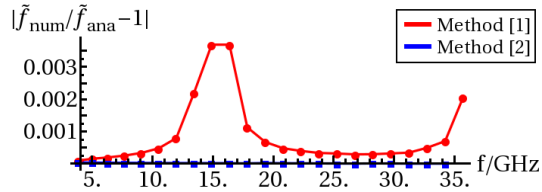


Figure 7: Relative deviation between perturbed frequencies based on numerically computed unperturbed eigenmodes and the ones based on analytically computed eigenmodes.

CONCLUSIONS

The investigations showed that a completely numerical implementation of perturbative methods using standard software is feasible. The development of special procedures for the processing of numerically computed unperturbed eigenmodes allows for a very accurate and efficient computation of the necessary basic parameters (unperturbed frequencies and volume integrals of the electromagnetic fields) over a large frequency range. The perturbed eigenmodes determined from these basic parameters very exactly match the results of analytically comparative calculations. But it also became apparent that there are differences in the error propagation depending on the algorithm of the used method which should be further examined.

The most important conclusion of the investigations is that the implementation algorithm described in this paper allows for the application of perturbative methods to any desired cavity geometry providing reasonable results.

REFERENCES

[1] D. Meidlinger, *A General Perturbation Theory for Cavity Mode Field Patterns*, Proceedings of SRF2009, Berlin, Germany, THPPO005, 2009.

[2] K. Brackebusch, U. van Rienen, *Eigenmode Computation for Cavities with Perturbed Geometry based on a Series Expansion of Unperturbed Eigenmodes*, Proceedings of 3rd IPAC, New Orleans, USA, May 20-25, pp. 277-279, 2012.

[3] K. Brackebusch, H.-W. Glock, U. van Rienen, *Calculation of High Frequency Fields in Resonant Cavities based on Perturbation Theory*, Proceedings of 2nd IPAC, San Sebastian, Spain, September 4-9, pp. 2235-2237, 2011.

[4] Computer Simulation Technology AG, *Microwave Studio®*, 64289 Darmstadt, Germany.

[5] Wolfram Research, Inc., *Mathematica®Edition: Version 8.0*, Champaign, 2010

AN APPLICATION OF THE NON-CONFORMING CROUZEIX-RAVIART FINITE ELEMENT METHOD TO SPACE CHARGE CALCULATIONS

C. Bahls*, U. van Rienen, University of Rostock, Rostock, Germany

Abstract

The calculation of space charge effects in linear accelerators is an important prerequisite to understand the interaction between charged particles and the surrounding environment. These calculations should be as efficient as possible. In this work we explore the suitability of the Crouzeix-Raviart Finite Element Method for the computation of the self-field of an electron bunch.

INTRODUCTION

Current and future accelerator design requires efficient 3D space charge calculations. One possible approach to Space Charge Calculations is the Particle-in-Cell (PIC) method, especially the Particle-Mesh method which calculates the potential in the rest-frame of the bunch.

This computation usually is done by solving Poisson's equation on the domain Ω, using a charge weighting $\mathsf{f}(x)$:

$$-\Delta\,\mathsf{u}(x) \quad = \quad \mathsf{f}(x), \qquad \forall x \in \Omega.$$

This equation is subject to some boundary conditions:

$$\begin{aligned} \mathsf{u}(x) &= g_D(x), && \forall x \in \partial\Omega_D, \\ \operatorname{grad}\mathsf{u}(x)\cdot n(x) &= g_N(x), && \forall x \in \partial\Omega_N. \end{aligned}$$

These computations should be as efficient as possible.

SPACE CHARGE CALCULATIONS

We are aiming at computing the self-field of the bunch. Denoting with $\mathbf{D}$ the dielectric flux and with ρ the charge density we are estimating a solution to Gauss' law:

$$\operatorname{div}\mathbf{D} = \rho.$$

Usually there are infinitely many solutions to that equation, as in fact a very large subspace of all vectorial functions on the domain fulfils the equation. One can add any divergence-free field (for example a rotational field) to a known solution without changing the divergence, therefore recovering additional solutions to the field equation.

For the moment we are only interested in curl-free solutions of Gauss' law, so we will only try to estimate fields $\mathbf{\Psi}(x)$ which are gradients of a scalar potential $\mathsf{u}(x)$, as: $\mathbf{\Psi}(x) = -\operatorname{grad}\mathsf{u}(x)$, so our equations become:

$$\begin{aligned} \operatorname{grad}\mathsf{u}(x) + \mathbf{\Psi}(x) &= \mathbf{0} && \forall x \in \Omega, \\ -\operatorname{div}\boldsymbol{\varepsilon}(x)\,\mathbf{\Psi}(x) &= \rho(x) && \forall x \in \Omega. \end{aligned}$$

* christian.bahls@gmx.de

Removing the auxiliary vector function $\mathbf{\Psi}(x)$ this usually gets shortened to the following equation:

$$-\operatorname{div}\boldsymbol{\varepsilon}(x)\operatorname{grad}\mathsf{u}(x) = \rho(x) \qquad \forall x \in \Omega.$$

If the permittivity tensor $\boldsymbol{\varepsilon}(x)$ isotropic and can be replaced by a divergence free scalar function $\varepsilon(x)$ this becomes:

$$-\Delta\mathsf{u}(x) = \varepsilon(x)^{-1}\rho(x) \qquad \forall x \in \Omega.$$

Our currently used numerical scheme [5] - solving Poissons equation $-\Delta\mathsf{u}(x) = \rho(x)/\varepsilon_0$ in vacuum using a finite difference scheme - seems to less than optimal for estimating the electric field.

We are loosing one order of convergence ($O(h^2) \rightarrow O(h^1)$) by the numerically computation of the electric field from the potential u (even if using the exact derivative on the underlying function space).

The discretized solution u_h on an equidistant structured mesh approximates the solution u with an order $O(h^2)$:

$$\mathsf{u}_h(x) = \mathsf{u}(x) + O(h^2).$$

The gradient $\mathbf{\Psi}$ (which corresponds to our accelerating field) will than be approximated with an order of $O(h^1)$:

$$\mathbf{\Psi}_h(x) = \mathbf{\Psi}(x) + O(h^1).$$

As we are mainly interested in the electric field we would like to approximate it with the same order of accuracy as the potential.

So we want to discretize and solve for the vector field $\mathbf{\Psi}$ directly. The discretization used has to be curl-free and should somehow allow for a reasonable definition of the divergence of the field (e.g. be conformal).

Raviart-Thomas Mixed Finite Elements

One suitable ansatz-space is the lowest order Raviart-Thomas space $\mathbf{RT}_0$ whose linear vector functions have following element-wise linear expression:

$$\mathbf{\Psi}_h(\mathbf{x}) \quad = \quad \mathbf{a}_k + b_k\mathbf{x},$$

where $\mathbf{x}$ is in the element T_k of the triangulation $\mathbf{T}$ of $\mathbf{\Omega}$.

For the discretisation to be conformal the normal components of the field have to be continuous at every interface (edges in 2D or faces in 3D).

To include this continuity constraint in the construction of the finite element space, $\mathbf{RT}_0$ usually is represented by an interface-based discretization using following definition:

$$\psi_{E_j}(\mathbf{x}) = \sigma_j \frac{|E_j|}{2|T_k|}(\mathbf{x} - \mathbf{P}_j).$$

Where $|T_k|$ is the area/volume of the simplex T_k, $|E_j|$ is the length/area of the edge/face E_j, σ_j indicates the orientation of E_j and $\mathbf{P}_j$ is the vertex opposite to E_j.

Mixed and Hybrid Formulation

We are now using the canonical Galerkin-approach for Mixed Finite Elements to compute approximate solutions for the field $\mathbf{\Psi}_h$ and the potential u_h:

$$\begin{aligned} \int_\Omega \boldsymbol{\tau} \cdot \mathbf{\Psi}_h + \int_\Omega \boldsymbol{\tau}\, \varepsilon \,\mathrm{grad}\, u_h &= 0 \quad \forall \boldsymbol{\tau} \in \mathbf{RT}_0, \\ \int_\Omega v \,\mathrm{div}\, \mathbf{\Psi}_h &= \int_\Omega v\, f \quad \forall v \in \mathbf{P}_1^{-1}. \end{aligned}$$

To later remove the flux variable from the system, one can relax the continuity requirement on the ansatz-space and use the flux $\tilde{\mathbf{\Psi}}_h$ from the space $\mathbf{RT}_0^{-1}$ of piecewise linear vector and discontinuous vector functions.

One could then enforce the continuity of the normal component of the flux on the interfaces by the use of piecewise discontinuous Lagrange multipliers $\lambda_h \in \mathbf{M}_1^{-1}$, leading to following system of equations:

$$\begin{aligned} \int_\Omega \tilde{\boldsymbol{\tau}} \cdot \tilde{\mathbf{\Psi}}_h + \int_\Omega \tilde{\boldsymbol{\tau}}\, \varepsilon \,\mathrm{grad}\, u_h + \int_{\delta\Omega} \lambda_h\, \mathbf{n}_T \cdot \tilde{\boldsymbol{\tau}} &= 0, \\ \int_\Omega v \,\mathrm{div}\, \tilde{\mathbf{\Psi}}_h &= \int_\Omega v\, f, \\ \int_{\delta\Omega} \mu\, \mathbf{n}_T \cdot \tilde{\mathbf{\Psi}}_h &= 0. \end{aligned}$$

for all $\tilde{\boldsymbol{\tau}}$, v and μ in $\mathbf{RT}_0^{-1}$, $\mathbf{P}_1^{-1}$ and $\mathbf{M}_1^{-1}$ respectively.

This is equivalent to a system of linear equations such as:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ \mathbf{B}^T & & \\ \mathbf{C}^T & & \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{\Psi}}_h \\ u_h \\ \lambda_h \end{pmatrix} = \begin{pmatrix} 0 \\ f_h \\ 0 \end{pmatrix}.$$

Because $\mathbf{A}$ is block-diagonal it is element-wise invertible, the local $\tilde{\mathbf{\Psi}}_h$ can easily be computed by:

$$\tilde{\mathbf{\Psi}}_h = -\mathbf{A}^{-1}(\mathbf{B}\, u_h + \mathbf{C}\, \lambda_h).$$

Using block Gauss elimination this leads to the following system of linear equations:

$$\begin{pmatrix} \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B} & \mathbf{B}^T\mathbf{A}^{-1}\mathbf{C} \\ \mathbf{C}^T\mathbf{A}^{-1}\mathbf{B} & \mathbf{C}^T\mathbf{A}^{-1}\mathbf{C} \end{pmatrix} \begin{pmatrix} u_h \\ \lambda_h \end{pmatrix} = \begin{pmatrix} -f_h \\ 0 \end{pmatrix}.$$

This method is called static condensation. Also eliminating the u_h using a Schur complement one arrives at the Crouzeix-Raviart Finite Element Method, which can also be derived differently.

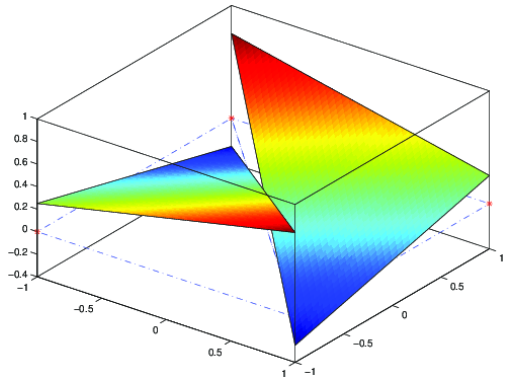


Figure 1: The ansatz-functions are only continuous at the midpoints of interfaces

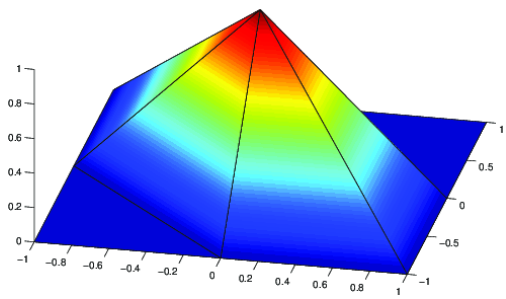


Figure 2: The space of continuous $\mathbf{P}_1$ is contained in $\mathbf{P}_1^{\mathsf{NC}}$.

Crouzeix-Raviart Non-conformal FEM

In [1] Marini suggested to directly use the locally $\mathbf{P}_1$ but nonconforming finite element spaces $\mathbf{P}_1^{\mathsf{NC}}$, these are also called Crouzeix-Raviart or loof finite element spaces. (This approach can also be found in [2])

These finite elements have their degrees of freedom allocated at the barycenters of their interfaces, rather than at their vertices. The function space $\mathbf{P}_1^{\mathsf{NC}}$ actually contains the contains the space $\mathbf{P}_1$ of Lagrange finite elements, so it can at least represent the continuous solutions from nodal $\mathbf{P}_1$ discretizations, but it is larger.

The direct way to arrive at the Crouzeix-Raviart Finite Element Method is to apply the usual Galerkin approach to the nonconforming ansatz space $\mathbf{P}_1^{\mathsf{NC}}$ directly:

$$\sum_{T_k \in \mathbf{T}} \int_{T_k} \varepsilon \,\mathrm{grad}_h\, \mathsf{u}_h \cdot \mathrm{grad}_h\, v = \int_\Omega f_h v \quad \forall v \in \mathbf{P}_1^{\mathsf{NC}}.$$

With decreasing mesh size h the numerical solution u_h converges to u with $O(h^2)$. More interestingly a special post-processing can recover a linear flux $\mathbf{\Psi}_h(x)$ of second order accuracy $O(h^2)$ from $\mathsf{u}_h(x)$ by locally choosing:

$$\mathbf{\Psi}_h(x) = \varepsilon \,\mathrm{grad}_h\, \mathsf{u}_h - f_h\, (\mathbf{x} - \mathbf{x}_{T_k})/n$$

on the element T_k. (with n denoting the dimension of the problem setting and $\mathbf{x}_{T_k}$ the barycenter of the simplex T_k)

Table 1: Tabulated convergence of successive refinements of the square domain $[0,1]\times[0,1]$; $\|e_{\mathsf{u}_h}\|_2$ and $\|e_{\mathsf{u}_h}\|_\infty$ are the L_2 and the maximum error of the potential u_h, while $\|e_{\mathbf{\Psi}_h}\|_\infty$ indicates the maximum error of the approximated gradient $\mathbf{\Psi}_h$ at interface midpoints.

K	t(sec)	$\|e_{\mathsf{u}_h}\|_2$	$\|e_{\mathsf{u}_h}\|_\infty$	$\|e_{\mathbf{\Psi}_h}\|_\infty$
40	0.002	6.20e-2	1.20e-1	2.85e+0
176	0.004	1.49e-2	3.69e-2	1.28e+0
736	0.009	3.70e-3	9.59e-3	3.65e-1
3008	0.035	9.24e-4	2.40e-3	9.47e-2
12160	0.204	2.31e-4	6.01e-4	2.39e-2
48896	1.433	5.77e-5	1.50e-4	5.97e-3

Table 2: Tabulated convergence of successive refinements of the cube $[0,1]\times[0,1]\times[0,1]$; $\|e_{\mathsf{u}_h}\|_2$, $\|e_{\mathsf{u}_h}\|_\infty$ and $\|e_{\mathbf{\Psi}_h}\|_\infty$ retain the meaning indicated in Table 1.

K	t(sec)	$\|e_{\mathsf{u}_h}\|_2$	$\|e_{\mathsf{u}_h}\|_\infty$	$\|e_{\mathbf{\Psi}_h}\|_\infty$
6	0.002	1.42e+0	9.74e-1	1.44e+1
72	0.002	7.14e-1	9.20e-1	1.32e+1
672	0.006	1.49e-1	3.20e-1	6.87e+0
5760	0.044	4.02e-2	1.13e-1	2.93e+0
47616	0.597	1.07e-2	3.21e-2	8.63e-1
387072	9.373	2.74e-3	8.58e-3	2.23e-1

RESULTS

For convergence studies we have implemented Crouzeix-Raviart Finite Elements in MATLAB [6] for two and three dimensional simplical meshes.

For benchmarking the robustness and for accessing the efficiency of this approach we are solving:

$$\begin{aligned} -\Delta \mathsf{u}(x) &= 3\pi^2\, sin(\pi x_1)\, sin(\pi x_2)\, sin(\pi x_3) \text{ in 3D,} \\ -\Delta \mathsf{u}(x) &= 2\pi^2\, sin(\pi x_1)\, sin(\pi x_2) \text{ in 2D.} \end{aligned}$$

The analytic solutions for the scalar potential u is given by:

$$\begin{aligned} \mathsf{u}(x) &= sin(\pi x_1)\, sin(\pi x_2)\, sin(\pi x_3) \quad \text{in 3D} \\ \mathsf{u}(x) &= sin(\pi x_1)\, sin(\pi x_2) \quad \text{in 2D respectively.} \end{aligned}$$

The convergence of the solution for successively refined meshes (halving the minimum element diameter in each step) is tabulated for two dimensions in table 1 and for three dimensions in table 2.

The convergence of the potential shows the expected behaviour (order $O(h^2)$ implying a reduction of the error by a factor of 4 when the step-size is halved).

The error in the approximated field distribution seems to be dominated by the error at the boundary (especially the corners of the domain) first - only approaching order $O(h^2)$ on highly refined grids.

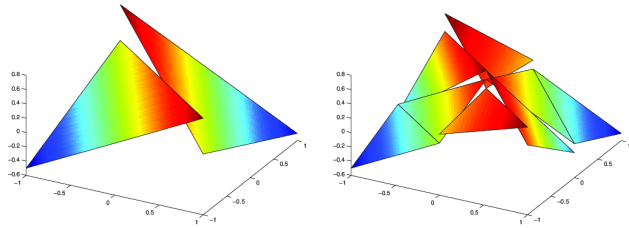


Figure 3: Coarse grid function is not contained in fine grid

CONCLUSIONS

Though in both cases the computation time using the conjugate gradient algorithm could be reduced by a factor of 2 by using SSOR as a preconditioner for our current applications and the accuracy sought the computational cost seems to be prohibitive.

The algorithm still does not have the optimal complexity. Besides using an efficient preconditioner we want to explore the use of geometric multigrid for Crouzeix Raviart finite elements.

Using the right prolongation and restriction operators is not completely straight-forward for the finite element spaces involved as the successively refined function spaces are not nested (see Figure 3). Exploring the approaches in [4] and geometric multigrid as described in [3] seems to be most promising at the moment.

ACKNOWLEDGMENT

We would like to thank Dr. Pöplau for her input during the preparation of this paper.

REFERENCES

[1] L. Marini. An inexpensive method for the evaluation of the solution of the lowest order Raviart–Thomas mixed method. *SIAM Journal on Numerical Analysis*, 22(3):493–496, 1985.

[2] D.N. Arnold and F. Brezzi. Mixed and nonconforming finite element methods: implementation, postprocessing and error estimates. *Math. Anal. Numér*, 19(1), 1985.

[3] D. Braess, M. Dryja, and W. Hackbusch. A multigrid method for nonconforming FE-discretisations with application to non-matching grids. *Computing*, 63(1):1–25, July 1999.

[4] J. Kraus, S. Margenov, and J. Synka. On the multilevel preconditioning of Crouzeix–Raviart elliptic problems. *Numerical Linear Algebra with Applications*, 15(5):395–416, 2008.

[5] G. Pöplau and U. van Rienen. An efficient 3D space charge routine with self-adaptive discretization. In *Proceedings of ICAP 2009 (Proceedings of the 10th International Computational Accelerator Physics Conference), San Francisco, USA*, pages 23–26, 2010.

[6] MATLAB. *Version 7.10.0 (R2010a)*. The MathWorks Inc.

ANALYZING MULTIPACTING PROBLEMS IN ACCELERATORS USING ACE3P ON HIGH PERFORMANCE COMPUTERS*

Lixin Ge[1], Kwok Ko, Kihwan Lee, Zenghai Li, Cho Ng, Liling Xiao
SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

Abstract

Track3P is the particle tracking module of ACE3P, a 3D parallel finite element electromagnetic code suite developed at SLAC which has been implemented on the US DOE supercomputers at NERSC to simulate large-scale complex accelerator designs [1-3]. Using the higher-order cavity fields [4] generated by ACE3P codes, Track3P has been used for analysing multipacting (MP) in accelerator cavities [5]. The prediction of the MP barriers in the ICHIRO cavity at KEK was the first Track3P benchmark against measurements. Using a large number of processors, Track3P can scan through the field gradient and cavity surface efficiently, and its comprehensive post-processing tool allows the identifications of both the hard and soft MP barriers and the locations of MP activities. Results from applications of this high performance simulation capability to accelerators such as the Half Wave Resonator (HWR) [6], Quarter Wave Resonator (QWR) for FRIB [7], 704 MHz SRF gun cavity for BNL ERL [8] and the Muon cooling cavity for Muon Collider [9] will be presented.

INTRODUCTION

Multipacting (MP) is an undesired, resonant built-up of electrons inside RF structures. It can cause wall heating and high power RF components like couplers, windows, etc. breakdown. There are also other bad effects, such as significant power loss, low achievable field gradient and thermal breakdown in superconducting structures.

Due to the critical role of multipacting effects on accelerator design, much work has been done on multipacting studies to identify potential MP activities and their locations and to mitigate MP effects using different methods such as modifying geometry to eliminate MP barriers, changing surface conditions to reduce secondary emission yield (SEY) and imposing external DC biasing field.

In recent years, with more computing power, first 2D, then also full 3D simulation tools have been developed to investigate potential multipacting activities in RF-structures [10]. There are several requirements for realistic multipacting simulation, namely, high resolution EM field, correct representation of particle emission from curved surface, realistic SEY curve for surface material and comprehensive post-processing of particle data to identify MP events.

* The work was supported by the U.S. DOE contract DE-AC02-76SF00515 and used the resources of NERSC at LBNL under US DOE Contract No. DE-AC03-76SF00098.
[1]lge@slac.stanford.edu

ACE3P CODE SUITE

For more than a decade, SLAC has been developing the conformal, higher-order, C++/MPI based parallel finite element suite of electromagnetic codes [1-3]. ACE3P consists of the following modules: Omega3P for calculating cavity modes and damping, and S3P for transmission in open structures in frequency domain; T3P for calculating wakefields and transients in time domain; Track3P for multipacting and dark current studies using particle tracking; Pic3P for RF gun design with particle-in-cell (PIC) method; and TEM3P for multi-physics analysis including EM, thermal and mechanical effects.

There are several strengths of the parallel finite-element method used in ACE3P. First, high-fidelity geometry modelling can be achieved using curved quadratic tetrahedral elements. Second, higher-order field interpolation functions (p = 1-6) improve field accuracy as shown in Fig. 2 when using linear, quadratic and cubic basis functions for solving the eigenmode of the cavity shown in Fig. 1. Third, parallel processing speeds up the computation by taking advantages of the code scalability. The calculation took less than 1 minute to obtain the mode frequency within 0.001% using a mesh with 67k quadratic elements on 16 CPUs with 6GB memory. All these factors are important for multipacting simulation as will be seen later in this paper.

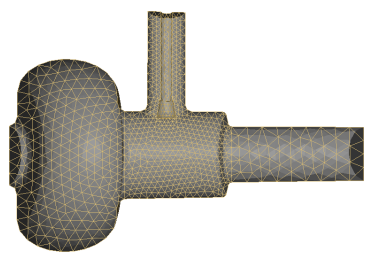


Figure 1: End cell cavity with input coupler.

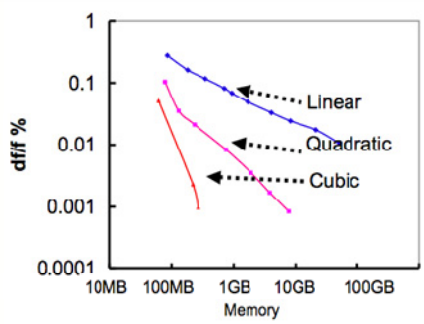


Figure 2: Convergence study for using linear, quadratic and cubic basis functions.

ACE3P has been installed and runs on the supercomputers at NERSC (National Energy Research Centre) of Lawrence Berkeley National Laboratory. One of the NERSC machines, hopper, has 153,216 compute cores, 217 terabytes of memory and 2 petabytes of disk storage. Its peak performance is 1.28 Petaflops/sec. More information can be found at [11].

TRACK3P

Track3P is a 3D parallel high-order finite element particle tracking code for multipacting and dark current simulations. For multipacting simulation, Track3P can load EM fields calculated by Omega3P, S3P and T3P, which are high resolution EM solvers to calculate standing wave, traveling wave and transient field solutions as input for particle tracking. High-fidelity geometry representation built inside ACE3P allows realistic modeling of particle emission on cavity surface. Realistic SEY curves provided by experiments can be used to construct multipacting maps. The versatile post-processing capabilities help users to identify onset of multipacting through various parameter scans.

Parallelization of Track3P

In Track3P, the whole geometry mesh and EM fields are loaded in individual processors. Particles are divided evenly among the processors, so no communication is required between processors and thus excellent load balancing is achieved. For a typical MP simulation, for example, the FRIB Half Wave Resonator (HWR), the total number of field level scans is 300 and the estimated number of particles per field scan is 8 millions (with initial particles distributed on all exterior surfaces). It took about four hours with 9600 CPUs to finish the simulation.

Multipacting Simulation in Track3P

In a typical MP simulation, electrons are launched from specific surfaces of a cavity at different phases over a full RF period. The initial launched electrons follow the electromagnetic fields in the structure and eventually hit the boundary, where secondary electrons are emitted based on the secondary emission yield (SEY) of the surface material. The tracing of electrons will continue for a specified number of RF cycles. The process described above is repeated for a set of combined launching parameters: RF field level and phase, launching location energy and angle. After all the calculations for various launching conditions are finished, resonant particle trajectories are identified. Not all of these resonant trajectories contribute to multipacting, and only those with successive impact energies within the right range for secondary emission yield bigger than unity will be treated as multipacting events. Finally a MP susceptible zone is constructed. A post-processing tool is also included in Track3P for the effective extraction of MP events. The tool not only enables to determine the multipacting order and type, which are defined as the number of cycles per impact and the number of impacts per multipacting cycle, respectively, but also includes enhancement counter functions to crosscheck the MP barriers determined using the information from the SEY curve.

Benchmark of Track3P

Track3P has been extensively benchmarked with theories and measurements. During high power RF processing, the KEK ICHRIO cavity experienced low achievable field gradient and long RF processing time. To understand these processing limitations, Track3P was used to study potential MP barriers in the cavity cell and beam pipe step region as shown in Fig 3.

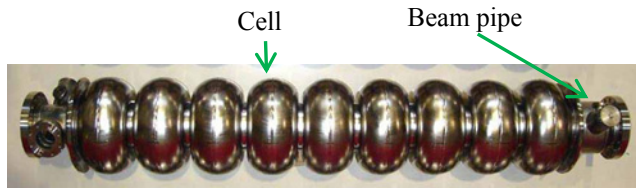


Figure 3: KEK ICHIRO cavity.

Track3P simulation on the ICHIRO cavity cell shows that there are resonant trajectories near the cavity cell equator. Fig. 4 shows examples of the resonant trajectories at three different field levels. The impact energy ranges from 20 eV to 55 eV, based on the SEY curve provided by KEK. This MP barrier is soft, which agrees with RF tests as some radiation signals were observes at these field levels but the tests can be processed through them.

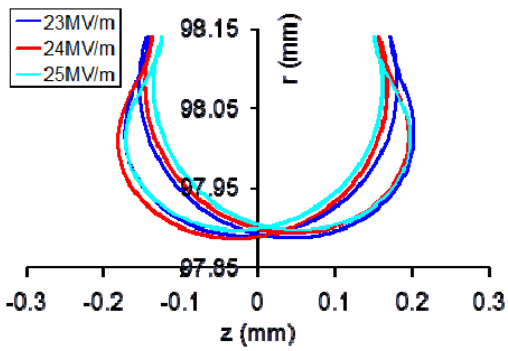


Figure 4: Resonant particle trajectories at three different field gradients: 23 MV/m, 24MV/m and 25MV/m.

Table 1: MP Barriers in the Beam Pipe Step Region of the KEK Ichiro Cavity

Track3P MP simulation		ICHIRO #0 (K. Saito,KEK)
Impact Energy (eV)	Gradient (MV/m)	X-ray Barriers (MV/m)
300-400	12	11-29.3 12-18
200-500	14	13, 14, 14-18, 13-27
300-500	17	(17, 18)
300-900	21.2	20.8
600-1000	29.4	28.7, 29.0, 29.3, 29.4

Multipacting simulation on the beampipe step shows there are several MP barriers. Table 1 shows the comparison of the simulated and measured MP bands for field levels up to 30 MV/m. Track3P predicted all the MP bands observed in measurements, and, in particular, the hard barrier at 29.4 MV/m at which the cavity could not be processed through.

MULTIPACTING STUDIES FOR ACCELERATOR STRUCTURES

Track3P has been used to study multipacting in many accelerator structures. Its applications on the Half Wave Resonator (HWR), the Quarter Wave Resonator (QWR) for FRIB, the Muon cooling cavity for the Muon Collider and the 704 MHz SRF gun cavity for BNL ERL will be presented in this paper.

Multipacting Study for HWR Cavity of FRIB

The driver linac for the Facility for Rare Isotope Beams (FRIB) will use superconducting cavities to accelerate heavy-ion beams [12]. There are two kinds of superconducting accelerating cavities, the Half Wave Resonators and Quarter Wave Resonators. Multipacting is an issue of concern for the design of these superconducting resonators.

The HWR is being developed at Michigan State University (MSU)[13]. Fig. 5 is the CAD model of the beta = 0.53 HWR cavity. Four ports have been added to the top shorting plate for ease of cavity processing and a superconducting plunger is used for fine-tuning of the frequency. This cavity is designed to provide 3.7 MV of accelerating voltage at an optimum beta = v/c =0.539. The detailed physical parameters of the HWR can be found in [12].

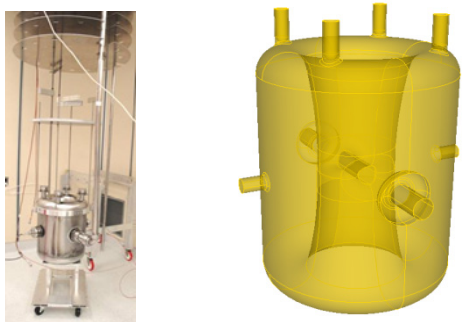


Figure 5: Beta=0.53 superconducting half-wave resonator cavity for the FRIB. Left: setup of the HWR cavity; Right: simulation CAD model.

Simulation results show that there are different kinds of resonant particles occurred in different regions, according to typical Niobium SEY curve shown in Fig. 6 and the distribution of resonant particle impact energies. Potential MP was found to be located at the tip region [6]. Detailed MP studies were performed at the tip region by changing the plunger location and modifying the tip shape. The particle impact energies decrease while the plunger is in an extracted position, and the resonant particle impact area shrinks with intruding plunger. The multipacting barrier is hardest at zero intrusion (Fig. 7).

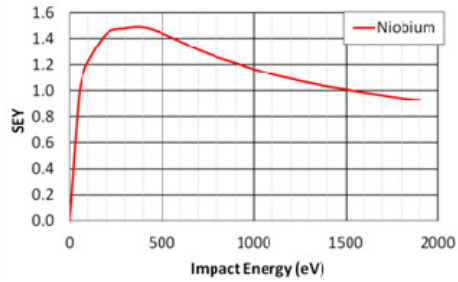


Figure 6: Impact energy dependence of SEY for Niobium. The peak SEY is around 150-700eV. Resonant particles with impact energies within this range likely contribute to MP.

The impact energy of the resonant trajectories on the tip of the plunger is well around the peak of Niobium SEY. It is desirable in the design to minimize such resonant conditions to avoid potential strong multipacting. Based on Omega3P and Track3P results, two ways have been suggested to mitigate MP at the rinse port. One method is to round the plunger tip with full radius and the other to remove the plunger completely from the cavity. Both are very successful in mitigating MP effects. Details can be found in [6].

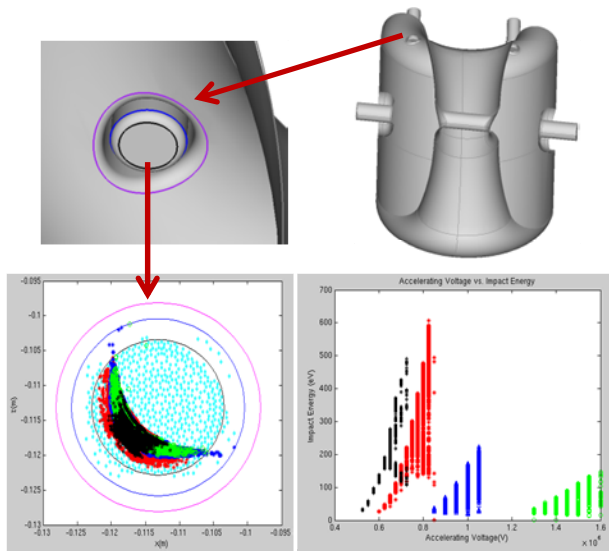


Figure 7: Left: Resonant trajectory distribution; Right: Impact energy vs field level. Cyan: initial electron distribution; Red 0 mm plunger insertion; Black: 3 mm insertion; Blue: 5 mm retraction; Green: 10 mm retraction.

Multipacting Study for QWR Cavity of FRIB

Figure 8 shows the CAD model and electric and magnetic field profiles of the 80.1 MHz β = 0.085 Quarter Wave Resonator [12][14]. The cavity will operate at a

maximum peak electric field of 30 MV/m. The corresponding accelerating voltage for a β = 0.085 beam is 1.5MV. Omega3P [4] was used to obtain the RF parameters and field maps required for multipacting (MP) simulation. Track3P [7] was used to track the particles and identify resonant trajectories. Calculations were performed on the NERSC Franklin machine [11], which has 38,128 Opteron compute cores. Typical Niobium SEY curve shown in Fig. 6 was used to estimate the MP strength.

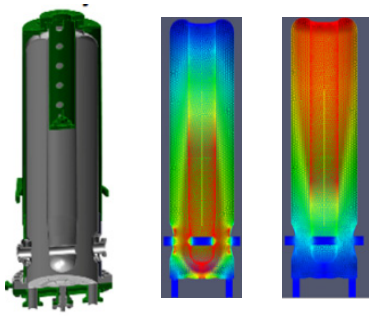


Figure 8: CAD model and field profiles of the β=0.085 superconducting quarter-wave resonator for the FRIB.

The distribution of resonant particles is shown is Fig. 9. There are two potential MP bands, one at low field levels with impact energies from tens of eV to about keV, and the other at high field levels from 360kV to 600kV with low impact energies around 100 eV, which are below the peak SEY energy (Fig. 6). The latter MP band is expected to be a soft barrier and could be processed through without many difficulties.

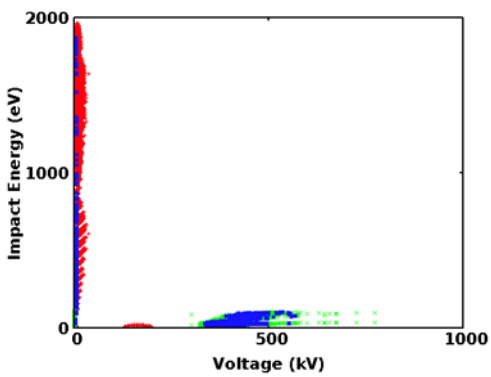


Figure 9: Impact energy vs. accelerating voltage for particles with resonant trajectories.

Stable resonant particles are observed at voltage levels from 800V to 7.5kV. Fig. 10 shows the impact energies of the resonant particles within this MP band. It appears to be a relatively hard barrier as the impact energies of the particles are around the peak of the SEY curve (Fig. 6). This simulated hard MP band agrees with the high power test data which showed processing barriers from 1.2kV - 7.2kV accelerating voltage [7].

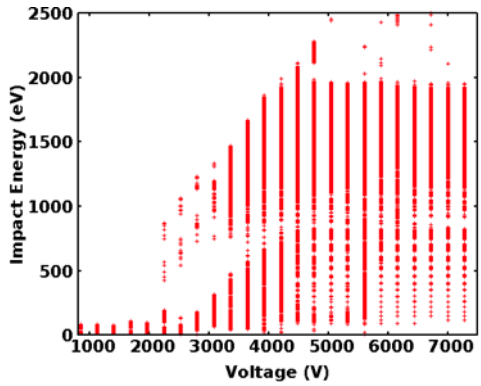


Figure 10: Resonant particle distributions for accelerating voltage between 800V - 7.5kV.

Four snapshots of particle tracking are shown in Fig. 11 to illustrate MP resonances at 23kV accelerating voltage.

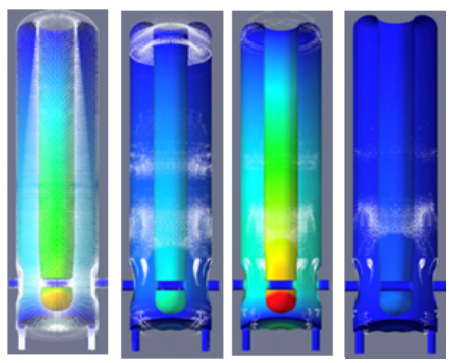


Figure 11: a-d) Evolution of particles survived at increasing RF cycles. Particles survived a large number of RF cycles are considered potential MP particles.

Multipacting Study for Muon Cooling Cavity

Experimental studies using an 805 MHz pillbox cavity at Fermilab's MuCool Test Area (MTA) have shown that its gradient significantly degrades and rf breakdown related damage occurs in high rf field regions when the cavity is operated in a DC solenoidal magnetic field up to 4 Tesla. These effects are believed to be related to the dark current and/or multipacting activities in the presence of the external magnetic field. The 805 MHz cavity under investigation has a pillbox geometry and was the first such cavity to be tested. ACE3P was used to optimize the cavity shape to minimize peak surface fields and to study the dark current and multipacting dependences on cavity RF and geometry parameters under a strong external magnetic field [9]. Using these simulations tools, we analyzed the characteristics of the dark current energy deposition at various cavity lengths; identified high field enhancement areas that could have resulted in high dark current damage; optimized the surface profiles in those areas to minimize the field enhancement; and identified potential multipacting bands. An improved 805 MHz

cavity design was obtained with significantly lower surface field, and strong MP zones were minimized by locally modifying the cavity geometry (Fig. 12).

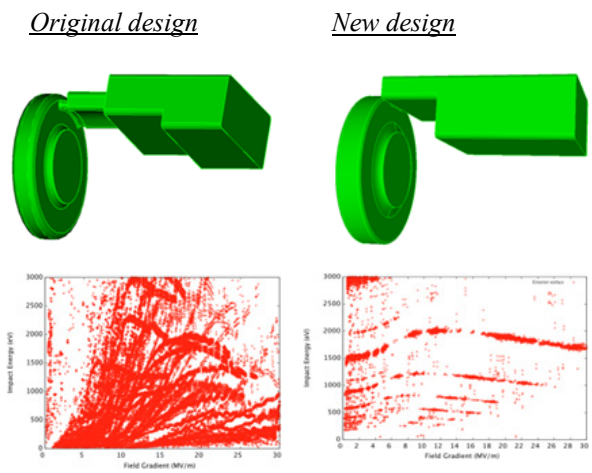


Figure 12: Original and new cavity design with corresponding resonant particles distribution on coupling slot region.

Multipacting Study for BNL SRF Gun Coupler

Scientists at BNL have successfully applied ACE3P to studying MP bands of the fundamental power coupler of an SRF gun. A paper [8] has been published based on the simulations using ACE3P and experiments performed at BNL. Simulations and test results match reasonably well as shown from Fig. 13, in which the shaded areas indicate the observed multipacting zones. It should be noted that Track3P simulations predicted multipacting zones that were later found during conditioning.

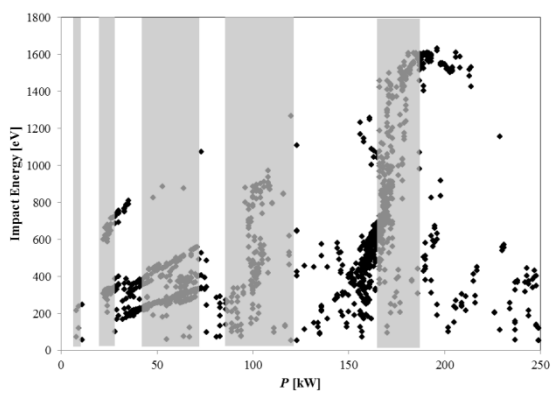


Figure 13: Comparison of multipacting zones from FPC conditioning test (grey-shaded areas) and simulation results at 703.9MHz with 0 degree (black dots).

SUMMARY

Track3P is the particle tracking module of the parallel finite-element code suite ACE3P and has been applied to study multipacting in accelerator cavities. Running on massively parallel computers, Track3P provides an efficient multipacting simulation tool in identifying multipacting barriers, locations and types of trajectories for the accelerator community. Track3P has been extensively used to study multipacting activities in many accelerator cavities and help scientists successfully mitigate multipacting activities and improve cavity design.

REFERENCES

[1] C.-K. Ng, et al., "State of the Art in EM Field Computation," Proc. EPAC06, Edinburg, Scotland.

[2] Z. Li, et al., "Towards Simulation of Electromagnetic and Beam Physics at the Petascale," Proc. PAC07, Albuquerque, New Mexico.

[3] K. Ko, et al, "Advances in Parallel Computing Codes for Accelerator Science and Development," Proc. LINAC2010, Tsukuba, Japan, 2010.

[4] Lie-Quan Lee, et al, "Omega3P: A Parallel Finite-Element Eigenmode Analysis Code for Accelerator Cavities," SLAC-PUB-13529, 2009.

[5] L. Ge, et al., "Multipacting Simulations of TTF-III Power Coupler Components," Proc. PAC07, Albuquerque, New Mexico.

[6] L. Ge et al., "Multipacting Simulation and Analysis for the FRIB Superconducting Resonators Using Track3P," Proc. LINAC2010, Tsukuba, Japan, 2010.

[7] L. Ge, et al., "Multipacting Simulations and Analysis for the FRIB β=**0.085** Quarter Waver Resonators Using Track3P," Proc. IPAC12, New Orleans, LA, USA. 2012.

[8] W. Xu, et al., "Design, Simulations and Conditioning of 500 kW Fundamental Power Couplers for a SRF Gun," Phys. Rev. ST Accel. Beams 15, 072001 (2012).

[9] Z. Li, et al., "RF Optimization and Analysis of the 805-MHz Cavity for the US MuCool Program Using ACE3P," Proc. AAC12, Austin, TX, USA. 2012.

[10] F. L. Krawczyk, "Status of Multipacting Simulation Capabilities for SCRF Applications," Proc. 10th workshop on RF Superconductivity, Tsukuba, Japan, 2001.

[11] http://www.nersc.gov/users/computational-systems/hopper/

[12] http://www.frib.msu.edu/

[13] J. Popielarski, et al., "Development of a Superconducting Half Wave Resonator for Beta 0.53". Proceedings of PAC09, Vancouver, Canada.

[14] C. Compton, et al., "Superconducting Resonator and Cryomodule Production for Ion Linacs at Michigan State University".

GPGPU IMPLEMENTATION OF MATRIX FORMALISM FOR BEAM DYNAMICS SIMULATION

N. Kulabukhova*, Saint-Petersburg State University, Russia

Abstract

Matrix formalism is a map integration method for ODE solving. It allows to present solution of the system as sums and multiplications of 2-indexes numeric matrix. This approach can be easy implement in parallel codes. As the most natural for matrix operation GPU architecture has been chosen. The set of the methods for beam dynamics has been implemented. Particles and envelope dynamics are supported. The computing facilities are located in St. Petersburg State University and presented by the NVIDIA Tesla-based clusters.

INTRODUCTION

The performance available on graphics processing units (GPUs) has led to interest in using GPUs for general-purpose programming [1]. It is difficult, however, for most programmers to program GPUs for general-purpose uses.

The raw computational power of a GPU dwarfs that of the most powerful CPU, and the gap is steadily widening. Furthermore, GPUs have moved away from the traditional fixed-function 3D graphics pipeline toward a flexible general-purpose computational engine. Today, GPUs can implement many parallel algorithms directly using graphics hardware. Well-suited algorithms that leverage all the underlying computational horsepower often achieve tremendous speedups[2]. Truly, the GPU is the first widely deployed commodity desktop parallel computer.

Here, we provide a comparison of different sort of parallel technologies of the classic graphics pipeline; our goal is to highlight those aspects of the real-time rendering calculation that allow graphics application developers to exploit modern GPUs as general-purpose parallel computation engines.

COMPARISON OF GPU AND CPU

The highly parallel workload of real-time computer graphics demands extremely high arithmetic throughput and streaming memory bandwidth but tolerates considerable latency in an individual computation since final images are only displayed every 16 milliseconds. These workload characteristics have shaped the underlying GPU architecture: Whereas CPUs are optimized for low latency, GPUs are optimized for high throughput. The GPUs specialized architecture is not well suited to every algorithm. Many applications are inherently serial and are characterized by incoherent and unpredictable memory access. Nonetheless, many important problems require significant computational resources, mapping well to the GPUs many-core arithmetic intensity, or they require streaming through large quantities of data, mapping well to the GPUs streaming memory subsystem. Porting a judiciously chosen algorithm to the GPU often produces speedups of five to 20 times over mature, optimized CPU codes running on state-of-the-art CPUs, and speedups of more than 100 times have been reported for some algorithms that map especially well. Some successful examples of using GPUs are described in [2].

In contrast, most CPU programs use a sequential programming model and are not benefiting from the continued increase in transistors due to Moores law. They will need to be rewritten or modified substantially to obtain increased performance from new CPUs with multiple cores on a chip. Furthermore, CPU clock speeds have plateaued due to power concerns. On the other hand, GPU architectures have a number of disadvantages for parallel programming. First, GPUs have a SIMD programming model. GPUs are moving toward a SPMD model, although use of loops and control-flow instructions in GPU pixel shaders may currently degrade performance, not improve it. Second, the actual architectures of GPUs are hidden behind device drivers that support APIs that implement virtual machines. This abstraction and lack of detail can hamper obtaining the most performance out of a graphics processor. For example, the virtual machines have no model of caches and no easy way for programmers to indicate how to traverse memory to facilitate memory reuse. This can make it difficult to write parallel programs where memory locality is crucial to performance. In addition, we must target APIs designed to support graphics that introduce extra complexity and overhead. Third, the programmable parts of GPUs have had limited support for primitive types typically found on CPUs.

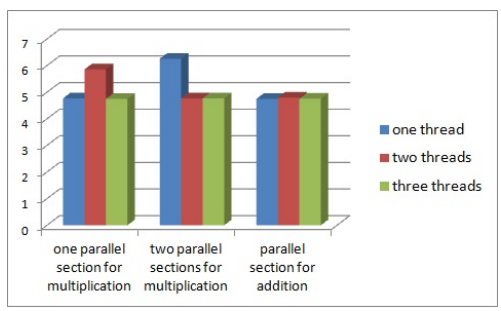


Figure 1: Diagram of comparative performance.

* kulabukhova.nv@gmail.com

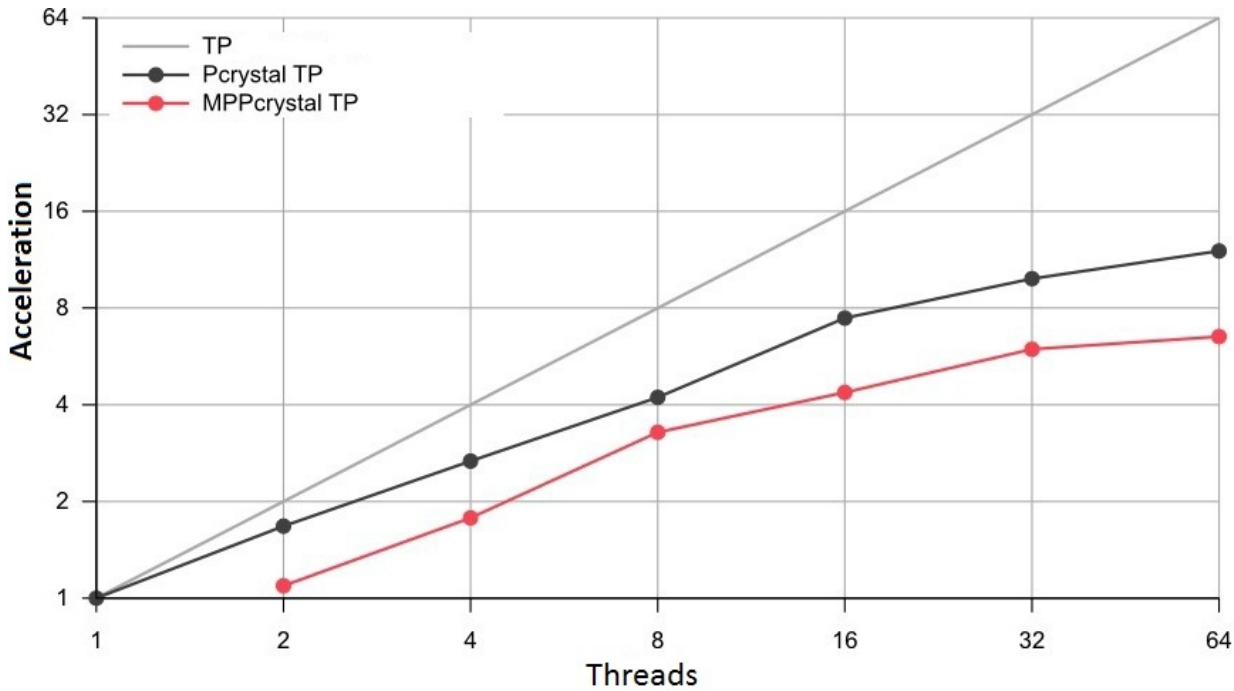


Figure 2: Comparison of SPbSU T-platform.

PARALLEL TECHNOLOGIES

Methods, libraries, interfaces to parallel a program are a lot to choose. There are different groups of technologies, those who use:

- a high-level communication libraries and interfaces (API) (MPI, MPL, OOMPI, OpenMP);
- special "parallelizing" structures in the programming language (MPC++, mpC, Ada,MC#, Cray MPP Fortran);
- automatic parallelization of sequential programs (FORSE, KAP, PIPS, VAST, V-Ray);
- parallelized procedures of the specialized libraries (ATLAS, PLAPACK, PIM, PARPACK);
- specialized software packages (ANSYS, ABAQUS, CFX, FLOWVISION, LMS Virtual Lab., GDT).

We evaluate the effectiveness of using data parallelism to program GPUs by providing results for a set of compute-intensive benchmarks. All calculations were performed on a hybrid cluster of SPbSU (see Fig. 2) computing center. Its nodes contain a NVIDIA Tesla S2050 system that was developed specifically as a GPGPU unit. For our goal we choose OpenMP technology.

OPENMP

When we speak about parallel computations, some problems must be solved. First, the time is spent on writing the program. Code must be effective and removable. It is important to view how this program will be used in future. The OpenMP API using the fork-join model of parallel execution is suitable for these aims.

OpenMP parallel program is constructed on sequential code by adding directive, procedures, process variables. This technology bases on the concept of shared memory, thats why Symmetric Multiprocessing is used (SMP computers). For this architecture threads (flows), running on different processors, easy to support. For classic UNIX-processes is more expensive to do this[4].

Scheme FORK/JOIN is used to support the parallel code. Entering the parallel environment thread-master generates complementary threads (operation FORK is running). After that each thread has its own number, thread-master has zero number. All threads run the same code of the parallel environment. After work thread-master waits ending of all other threads and continues to do next step (operation JOIN is running).

SIMULATION

Matrix formalism [5, 6] is a high-performance mapping approach for ODE solving. It allows to present solution of the system in following form

$$X = \sum_{i=0}^{k} R^{1i}(t) X_0^{[i]}, \qquad (1)$$

where R^{1i} are numerical matrices. So this approach can be easy implement in parallel code. Due to the fact that only matrix multiplication and addition are used, GPU programming is especially suitable for this purpose.

There are exist two way for beam dynamics simulation:

- based on particle simulation;
- based on envelope description.

Particles Simulation

The research have shown that there is no great benefits via parallelization of computational code for one particle by using GPU (see Fig. 1) In Table 1 performance of different parallelization modes are presented, where

- **1 mult** means one parallel section for multiplication;
- **2 mult** means two parallel section for multiplication;
- **add** means one parallel section for addition.

In this case overhead on data sending is significant. On the other hand matrix formalism allows to process a set of the initial points, where parallelization is more preferably.

Let's introduce a set of initial particle

$$M = (X_0^1 X_0^2 \dots X_0^p).$$

In according to the equation (1) the resulting points can be calculated

$$M = \sum_{i=0}^{k} R^{1i}(t)((X_0^1)^{[i]}(X_0^2)^{[i]} \dots (X_0^p)^{[i]}). \qquad (2)$$

Note that the sizes of matrices in the equation (2) is much greater than in (1) when a set of initial particles is quite large.

Table 1: Performance of Different Parallelization [sec]

	1 mult	2 mult	add
1	4.77	6.26	4.75
2	5.86	4.77	4.8
3	4.75	4.77	4.76

Envelope Simulation

Beam dynamics description based on envelope provides an efficient approach to modeling. Let's consider the envelope simulation is linear case (Fig. 3). In nonlinear case the equations are quit difficult, but the concept is based on linearization by introducing an extended space.

In linear case equation (1) is wrote in following form

$$X = R \cdot X_0.$$

The elliptical envelope can be described by a quadratic form

$$X^* A X < 1,$$

where X^* means transpose of vector X.

By these equations a new envelope can be obtained:

$$X^*(R^* A R)X < 1,$$

where $R^* A R$ is a matrix of new quadratic form.

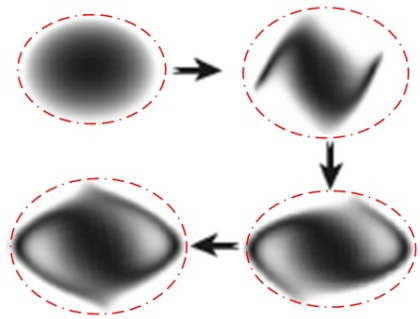


Figure 3: Envelope simulation.

CONCLUSION

Matrix formalism is a high-performance approach for beam dynamic modeling. The method can be implemented in parallel codes on GPU. It allows simulate both long-term evolution of a set of particles, and evaluating based on envelope description.

The future development of the research can be based on other parallel techniques investigation and complete implementation of the described approaches.

ACKNOWLEDGMENT

Computations were partly carried out on cluster HPC-0011654-001 of Saint-Petersburg State University, Faculty of Applied Mathematics and Control Processes. Special thanks for my scientific supervisor S. Andrianov.

REFERENCES

[1] D. Tarditi, S. Puri, J. Oglesby, "Accelerator: Using Data Parallelism to Program GPUs for General-Purpose Uses," proceedings of ASPLOS, San Jose, California, USA, 2006, pp. 325–335.

[2] D. Luebke, G. Humphreys, "How GPUs Work," How Things Work, Computer, February, 2007, pp. 126-130.

[3] A. Degtyarev, I. Gankevich, "Efficiency Comparison of Wave Surface Generation Using OpenCL, OpenMP and MPI," Proceedings of 8th International Conf. Computer Science and Information Tech., Yerevan, Armenia, 2011, p. 248–251.

[4] V. V. Voevodin, "Mathematical foundations of parallel computing," World Scientific Publishing Co., Series in computer science, 1992, pp. 33–343.

[5] S. Andrianov, "The Convergence and Accuracy of the Matrix Formalism Approximation," TUSCC2, ICAP2012.

[6] A. Ivanov, S. Andrianov, "Matrix formalism for long-term evolution of charged particle and spin dynamics in electrostatic fields," WEACC3, ICAP2012.

A MASSIVELY PARALLEL GENERAL PURPOSE MULTI-OBJECTIVE OPTIMIZATION FRAMEWORK, APPLIED TO BEAM DYNAMIC STUDIES

Y. Ineichen, A. Adelmann*, PSI, Villigen, Switzerland
C. Bekas, A. Curioni, IBM Research – Zurich, Switzerland
P. Arbenz, Department of Computer Science, ETH Zurich, Switzerland

Abstract

Particle accelerators are invaluable tools for research in the basic and applied sciences, in fields such as materials science, chemistry, the biosciences, particle physics, nuclear physics and medicine. The design, commissioning, and operation of accelerator facilities is a non-trivial task, due to the large number of control parameters and the complex interplay of several conflicting design goals.

We propose to tackle this problem by means of multi-objective optimization algorithms which also facilitate massively parallel deployment. In order to compute solutions in a meaningful time frame, that can even admit online optimization, we require a fast and scalable software framework.

In this paper, we present an implementation of such a framework and report first results of multi-objective optimization problems in the domain of beam dynamics.

INTRODUCTION

In contemporary scientific research, particle accelerators play a significant role. Fields, such as material science, chemistry, the biosciences, particle physics, nuclear physics and medicine rely on reliable and effective particle accelerators as research tools. Achieving the required performance is a complex and multifaceted problem in the design, commissioning, and operation of accelerator facilities. Today, tuning machine parameters, e.g., bunch charge, emission time and various parameters of beamline elements, is most commonly done manually by running simulation codes to scan the parameter space. This approach is tedious, time consuming and can be error prone. In order to be able to reliably identify optimal configurations of accelerators we propose to solve large multi-objective design optimization problems to automate the investigation for an optimal set of tuning parameters. Observe that multiple and conflicting optimality criteria call for a multi-objective approach.

We developed a modular multi-objective software framework (see Fig. 1) where the core functionality is decoupled from the "forward solver" and optimizer (master/slave). This allows to easily interchange optimizer algorithms, forward solvers and optimization problems. A "pilot" coordinates all efforts between the optimization algorithm and the forward solver. This forms a robust and general framework

* andreas.adelmann@psi.ch

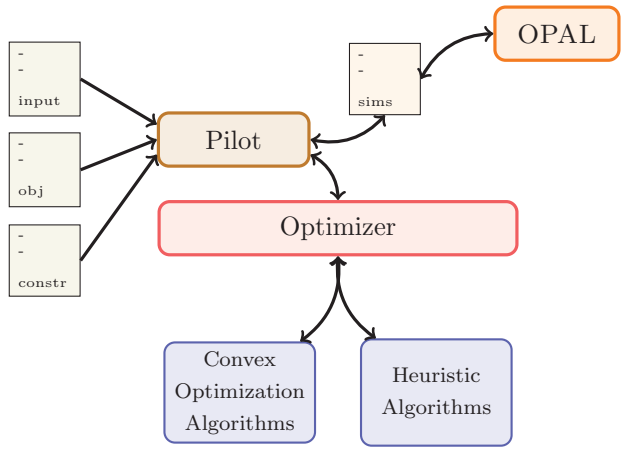


Figure 1: Multi-objective framework: the pilot (master) solves the optimization problem specified in the input file by coordinating optimizer algorithm and workers running forward solves.

for massively parallel multi-objective optimization. Currently the framework offers one concrete optimization algorithm, an evolutionary algorithm employing a NSGAII selector [1]. Normally simulation based approaches are plagued by the trade-of between level of detail and time to solution. We address this problem by using forward solvers with different time and detail complexity.

The first section covers a brief introduction to multi-objective optimization theory and describes the available optimizer. Next we discuss the implementation of the framework and present a proof of concept application of a beam dynamics problem.

MULTI-OBJECTIVE OPTIMIZATION

Optimization problems deal with finding one or more feasible solutions corresponding to extreme values of objectives. If more than one objective is present in the optimization problem we call this a multi-objective optimization problems (MOOP). A MOOP is defined as

$$\begin{aligned} \min \quad & f_m(\mathbf{x}), && m = 1 \dots M && (1)\\ \text{s.t.} \quad & g_j(\mathbf{x}) \geq 0, && j = 0 \dots J && (2)\\ & x_i^L \leq \mathbf{x} = x_i \leq x_i^U, && i = 0 \dots n, && (3) \end{aligned}$$

where we denote f as the objectives (1), g the constraints (2) and x the design variables (3).

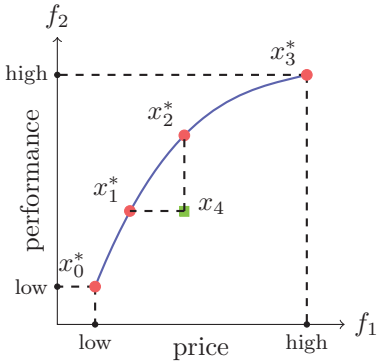


Figure 2: Two competing objectives can not be optimal at the same time. Red points represent Pareto optimal points, while x_4 is dominated (exhibits a worse price per performance ratio than e.g. x_2^*) by all points on the blue curve (Pareto front).

Often, we encounter conflicting objectives complicating the concept of optimality. To illustrate this, let us consider the problem of buying a car. Naturally, we want to get the best performance for the lowest price. This can be formulated as MOOP (4).

$$\begin{array}{ll} \min & \text{price} \\ \max & \text{performance} \\ \text{s.t.} & \ldots \end{array} \tag{4}$$

Obviously it is not possible to get the maximal performance for the lowest price and a trade-off decision between performance and price has to be reached (see Figure 2). Since not every choice is equally profitable for the buyer (car x_4 costs as much as x_2^* but offers less performance), we pick trade-offs (red points) that are essentially "equally optimal" in both conflicting objectives, meaning we cannot improve one point without hurting at least one other solution. This is known as the notion of Pareto optimality.The set of Pareto optimal points (blue curve) form the Pareto front or surface. All points on this surface are optimal.

Once the shape of the Pareto front has been determined the buyer can *specify preference*, balancing which features are more important. This is called *a-posteriori* preference specification since we select a solution after all possible trade-offs have been presented to us. The other alternative is to specify preference *a-priori*, e.g., by weighting (specifying preference before solving the problem) and combining all objectives into one and applying a single-objective method to solve the problem (yielding only one solution). In many situations preference is not known *a-priori* and an *a-posteriori* preference specification helps conveying a deeper understanding of the solution space. The Pareto front can be explored and the impact of trade-off decision become visible.

Sampling Pareto fronts is far from trivial. A number of different approaches have been proposed, e.g. evolutionary algorithms, sampling, simulated annealing, swarm methods and many more. The next section describes how evolutionary algorithms (available in our framework) are used to approximate the Pareto front of an multi-optimization problem.

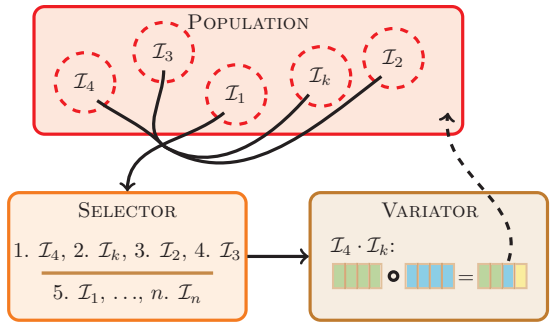


Figure 3: Schematic view of interplay between selector and variator.

Evolutionary Algorithms

Evolutionary algorithms (EA) are loosely based on nature's evolutionary principles to guide individuals towards an optimal solution (survival of the fittest). This "simulated" evolutionary process preserves entropy/diversity by using mutation and crossover to remix the fittest individuals in a population. Maintaining diversity in a population is critical for the success of all evolutionary algorithms.

In general, any evolutionary algorithm consists of the following components:

- *Genes*: properties/traits of an individual
- *Fitness*: a mapping from genes to fitness of individuals
- *Selector*: ordering relation to select k fittest individuals
- *Variator*: recombination (mutations and crossover) schemes for offspring generation.

In the context of our framework, genes correspond to specified design variables and individuals are ranked by their objective values (fitness). Evolutionary algorithms schematically work as depicted in Figure 3. Since there already exist plenty of implementations of evolutionary algorithms, we decided to incorporated the existing PISA library [1] into our framework. One of the advantages of PISA is that it separates variator from selector, rendering the library expendable and configurable. Implementing a variator was enough to use PISA in our framework and retain access to all available PISA selectors. As shown in Figure 3, the selector is in charge of ordering a set of d-dimensional vectors and selecting the k fittest individuals currently in the population. The performance of a selector depends on the number of objectives and the surface of the search space. So far, we only used an NSGA-II selector [2] exhibiting satisfactory convergence performance.

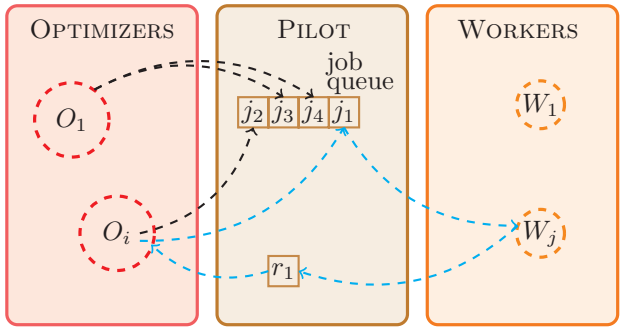


Figure 4: Schematic view of messages passed within the network. The cyan path describes a request sent from O_i to the Pilot being handled by W_j. Subsequently the result is returned to the requesting Optimizer (O_i).

The task of the variator is to generate offspring and ensure diversity in the population. The variator can start generating offspring once the fitness of every individual of the population has been evaluated. This explicit synchronization point defines an obvious parallel bottleneck of evolutionary algorithms. In the worst case one group of processors is taking a long time to compute the fitness of the last individual in the population. During this time all other processors are idle and wait for the result of this one individual in order to continue to generate offspring. To counteract this effect we call the selector already when 2 individuals have finished evaluating their fitness, lifting the boundaries between generations and evaluating the performance of individuals. New offspring will be generated and processors can immediately go back to work on the next fitness evaluation.

Our variator implementation uses the master/slave architecture, presented in the next section, to run as many function evaluations as possible in parallel. Additionally various crossover and mutation policies are available for tuning the algorithm to the optimization problem.

THE FRAMEWORK

The fact that multi-objective optimization problems are omnipresent in research and industry calls for a general purpose framework. The following requirements are necessary to be able to apply the framework independently of field and optimization problem:

- supporting any multi-objective optimization method,
- supporting any forward solver (simulation code or measurements),
- general description/specification of objectives, constraints and design variables.

The reminder of this section discusses each requirement and show how this is implemented in the opt-pilot framework.

Every processor will take up one of three available roles (see Figure 1): one core will act as Pilot, the remaining cores are divided amongst Workers and Optimizers. As shown in Figure 4 the Pilot is used to coordinate all "information requests" between Optimizer and Worker. An information request job consists of a set of design variables (e.g. the genes of an individual) and a type of information it requests (e.g. function evaluation or derivative). The Pilot keeps checking for idle Workers and assigns jobs in the queue to any free Worker groups. Once the Worker has computed and evaluated the request it is routed back to the Optimizer that originally requested the information.

Using template parameters the composition of the optimizer can be specified:

Code Listing 1: Calling the Pilot

```
typedef OpalInputFileParser Input_t;
typedef PisaVariator Opt_t;
typedef OpalSimulation Sim_t;
typedef Pilot< Input_t, Opt_t, Sim_t,
               /* ... */ > pilot_t;
scoped_ptr< pilot_t > pi(new pilot_t(args, comm));
```

After a processor gets appointed a role it starts a polling loop asynchronously listening for incoming requests. To that end a Poller base class has been introduced. Classes implementing the Poller interface enter an infinite loop and upon receiving MPI messages the appropriate handler is called. A concrete implementation of the Poller is allowed to implement special methods acting as hooks in the polling process, e.g. for actions that need to be taken after a message has been handled. The core of the Poller is the `onMessage()` method. The method is called with the `MPI_Status` of the received message and a `size_t` value specifying different values depending on the value of the `MPI_Tag`. Every Poller terminates the loop upon receiving a special MPI tag.

Implementing an Optimizer

All Optimizer implementations have to respect the API shown in Listing 2.

Code Listing 2: Optimizer API

```
virtual void initialize() = 0;

// Poller hooks
virtual void setupPoll() = 0;
virtual void prePoll() = 0;
virtual void postPoll() = 0;
virtual void onStop() = 0;
virtual bool onMessage(MPI_Status status,
                       size_t length) = 0;
```

The optimizer entry point corresponds to the `initialize()` method. Since an optimizer derives from the `Poller` interface, predefined hooks can be used to influence the polling procedure. Using a special "external" communicator group (passed to the Optimizer in the constructor) serves as interface to the Pilot to queue job requests. Every set of optimizers working on one optimization problem have their own "internal" MPI

communicator for handling the optimization part of the framework.

Implementing a Forward Solver

Forward solvers can easily be added by implementing a wrapper to run the simulation using a set of design variables. Listing 3 describes the API a forward solver has to implement.

CODE LISTING 3: SIMULATION API

```
virtual void run() = 0;
virtual void collectResults() = 0;
virtual reqVarContainer_t getResults() = 0;
```

The `run()` method will be called by workers and should execute the simulation in a *blocking* fashion. Subsequently, the worker will call `collectResults` to parse data from output files to the result data structures and ultimately use `getResults()` to get the resulting objectives and pass the data back to the optimizer. As before, every set of workers handling the same problem has access to a shared MPI communicator.

Optimization Problem Specification

Generally, any combination of simulation output variables (during any point of the simulation) can be used to compute objectives, constraints or act as design variables. Motivated by the principle of keeping meta-data (optimization and simulation input data) together, we decided to embed the optimization problem in the simulation input file (e.g. using comments). In some cases it might not be possible to annotate the simulation input file. By using another input file parser the optimization problems from stand-alone files.

Expression strings are parsed using Boost Spirit[1]. In the process an annotated expression tree is constructed and upon evaluation, all unknown variables are replaced with values from simulation results. To improve the expressive power of objectives and constraints we added a simple mechanism to define and call custom functions in expressions. This enables the user to define custom functions (e.g. over data produced by a simulation or measurement files) with ease using functors as shown in Listing 4.

CODE LISTING 4: SIMPLE AVERAGE FUNCTOR

```
struct avg {

    double operator()(
      client::function::arguments_t args) const {

        double limit = boost::get<double>(args[0]);
        std::string filename =
          boost::get<std::string>(args[1]);

        double sum = 0.0;
        for(int i = 0; i < limit; i++)
            sum += getDataFromFile(filename, i);

        return sum / limit;
    }
};
```

[1] http://boost-spirit.com/

All custom functions have to be registered with the expression to ensure the expression knows how to resolve function calls in its tree.

Parallelization

Parallelization is defined by a mapping of roles to available cores. Command-line options allow the user to steer the number of processors used in worker and optimizer groups. Here, we mainly use the command-line options to steer the number of processors running a forward solver.

FORWARD SOLVER

The framework contains a wrapper implementing the API mentioned in Listing 3 for OPAL [3]. OPAL provides different trackers for cyclotrons and linear accelerators with satisfactory parallel performance [4]. Recently we introduced a reduced envelope model [5] into OPAL reducing time to solution by several orders of magnitude.

Access to the OPAL forward solver enables the optimizer to solve a multitude of optimization problems arising in the domain of particle accelerators.

EXPERIMENTS

Experiments were executed on the PSI FELSIM cluster, running the framework using the components described above. The FELSIM cluster consists of 8 dual quad-core Intel Xeon processors at 3.0 GHz and has 2 GB memory per core with a total of 128 cores. The nodes are connected via Infiniband network with a total bandwidth of 16 GB/s.

A first benchmark tries to reproduce the Ferrario matching point [6] using the optimization problem given in equations (5) to (13).

$$
\begin{aligned}
\min \quad & \varepsilon_x && (5)\\
& \Delta\mathrm{rms}_{x,\mathrm{peak}} && (6)\\
& \Delta\varepsilon_{x,\mathrm{peak}} && (7)\\
\text{s.t.} \quad & q = 200\,[\mathrm{pC}] && (8)\\
& \mathrm{Volt}_{\mathrm{RF}} = 100\,[\mathrm{MV/m}] && (9)\\
& \sigma_L \leq \sigma_x = \sigma_y \leq \sigma_U && (10)\\
& \mathrm{KS}_L \leq \mathrm{KS}_{\mathrm{RF}} \leq \mathrm{KS}_U && (11)\\
& \mathrm{LAG}_L \leq \mathrm{LAG}_{\mathrm{RF}} \leq \mathrm{LAG}_U && (12)\\
& \Delta z_{L\mathrm{KS}} \leq \Delta z_{\mathrm{KS}} \leq \Delta z_{U\mathrm{KS}} && (13)\\
& && (14)
\end{aligned}
$$

The first objective minimizes the emittance at the end of the first traveling wave structure. The remaining two objectives minimize the distance from the position of the current minimum peak to the expected peak location at 3.025 m for transverse bunch size (beam waist) and emittance. Equations (8) and (9) define constraints for initial conditions and design variables given in (10) to (13) correspond to fieldstrength, and displacement of the solenoid.

Visualization

For visualization purposes we implemented a simple Pareto front explorer in Python. This simple tool provides a simple mapping from solutions on the Pareto front to the correct design variable values and helps investigate effects of the trade-off decision on the design variables.

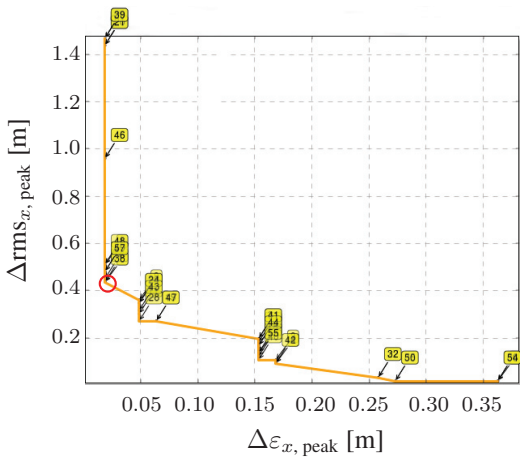


Figure 5: Approximation of Pareto front for the Ferrario matching point optimization problem. Red circle marks selected trade-off solution (38).

The result of a 1000 generation run (taking approximately 20 minutes with 16 cores) for the Ferrario matching problem mentioned above is show in Figure 5. For the trade-off solution 38 we show the complete time evolution of the slice beam in Figure 6. The minimum of beam waist and emittance peak are coinciding and this point falls close to the expected 3.025 m.

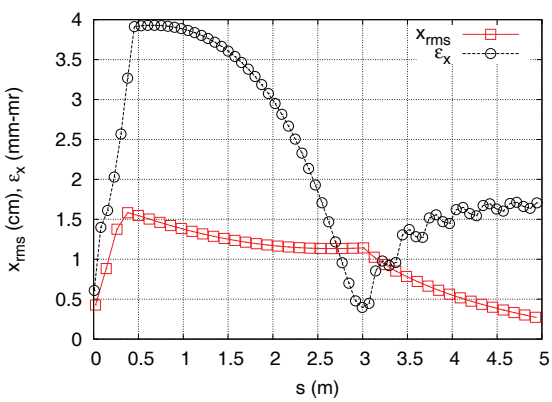


Figure 6: Simulation results for individual 38 of Pareto front shown in Figure 5.

CONCLUSIONS

We presented a framework to solve general multi-objective optimization framework. Its modularity allows to cover simulation-based optimization of a wide range of problems.

A first study on a simple benchmark shows that the framework is ready to tackle problems arising in the domain of beam dynamics. Even tough the stability analysis of the presented results is still work in progress the results are very convincing. With help of the master/slave parallelization the framework and using only a small amount of processors shows that a good approximation of the Pareto front can be computed in a matter of minutes.

Improving parallel efficiency on massively parallel systems is currently work in progress.

ACKNOWLEDGMENT

The authors thank the SwissFEL team for contributing to the formulation of optimization problems.

REFERENCES

[1] PISA — a platform and programming language independent interface for search algorithms. In Carlos M. Fonseca, Peter J. Fleming, Eckart Zitzler, Kalyanmoy Deb, and Lothar Thiele, editors, *Evolutionary Multi-Criterion Optimization (EMO 2003)*, Lecture Notes in Computer Science, pages 494 – 508, Berlin, 2003. Springer.

[2] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, April 2002.

[3] A. Adelmann, Ch. Kraus, Y. Ineichen, and J. J. Yang. The OPAL (Object Oriented Parallel Accelerator Library) Framework. Technical Report PSI-PR-08-02, Paul Scherrer Institut, 2008-2010. `http://amas.web.psi.ch/docs/opal/opal_user_guide.pdf`.

[4] A. Adelmann, Ch. Kraus, Y. Ineichen, S. Russell, Y. Bi, and J.J. Yang. The object oriented parallel accelerator library (opal), design, implementation and application. In *Proceedings ICAP09*, 2009.

[5] Yves Ineichen, Andreas Adelmann, Costas Bekas, Alessandro Curioni, and Peter Arbenz. A fast and scalable low dimensional solver for charged particle dynamics in large particle accelerators. *Computer Science - Research and Development*, pages 1–8, May 2012.

[6] M. Ferrario, J.E. Clendenin, D.T. Palmer, J.B. Rosenzweig, and L. Serafini. HOMDYN study for the LCLS RF photoinjector. pages 534–563, 2000.

THE TRIUMF OPTIMIZATION PLATFORM AND APPLICATION TO THE E-LINAC INJECTOR*

C. Gong and Y.C. Chao
TRIUMF, 4004 Wesbrook Mall, Vancouver V6T 2A3, Canada

Abstract

Multi-objective genetic algorithms (MOGA) have demonstrated their usefulness for the global optimization of accelerator design using Elegant [1] and Astra [2]. A MOGA platform developed at TRIUMF seeks to expand the capabilities of such tools by allowing multiple simulation engines to be used. The TRIUMF optimization software platform was applied to the transport design of an injection line leading from a cryomodule to the beam dump. The optimization involves two simulation engines, Astra and MAD-X, and demonstrates the ability for the platform to handle multi-engine optimization for a realistic problem. Results of the optimization are shown.

INTRODUCTION

A software platform for global optimization was created at TRIUMF [3]. The platform uses Multi-Objective Genetic Algorithms (MOGA) [4] as a wrapper around simulation engines. MOGA creates an initial population, with each member of the population some combination of randomly chosen variables. For each of a fixed number of iterations, the members are assigned a density, based on how well they satisfy the optimization constraints and objectives. Members are then randomly chosen, with a bias based on the density function, to 'breed' new members (by mutation, crossing variables, etc). Such algorithms are suitable for global optimization because they ignore the geometry of the search space, and given a large enough population and long enough running time, will always find the global optimum. In accelerator design, MOGA codes are wrappers for simulation engines such as Astra [5]. Each member of the population is an individual Astra run with different input variables such as magnet strengths and RF phase. Constraints and objectives are set on the Astra outputs, such as emittance and bunch length.

Previous single-engine MOGA codes [1, 2] have demonstrated their usefulness for accelerator design. However, since different engines are more useful in different situations, as listed in Table 2, modern accelerators require multi-engine simulations to encompass the wide range of design parameters. This motivates the TRIUMF code platform (Fig. 1), designed to handle multi-engine problems with arbitrary topology descriptions. The code was tested and performed satisfactorily in trial problems in a parallel-capable Linux environment [3].

The ARIEL/e-linac [6] project currently underway at

* Work supported by Natural Sciences and Engineering Research Council of Canada and National Research Council of Canada

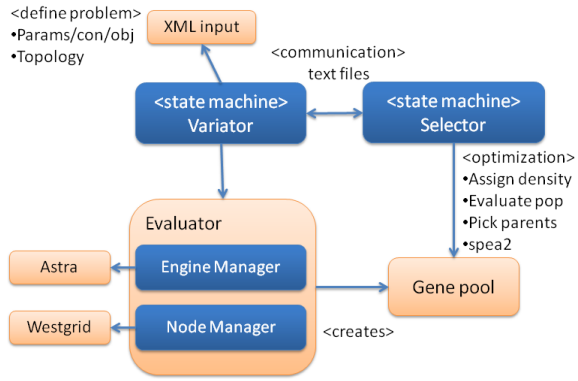


Figure 1: The architecture is based on A Platform and Programming Language Independent Interface for Search Algorithms [7]. The main components of the design are Variator, Selector, and Evaluator. Variator is a state machine which reads from the input file and defines the optimization problem. Evaluator handles the direct evaluation of all individuals in the population, including assigning and managing jobs to network nodes [8], and running the engine executables. Selector is a state machine which calculates the density, i.e. fitness, of the individuals and selects those for reproduction. The benefits of such a framework is decoupling between the algorithm, the problem definition, and the wrapper.

TRIUMF sees the possibility of operating a machine in dual rare isotope (RIB) and light source operation (FEL) (Fig. 2). The assortment of problems present in such a machine and the coupling between parameters, especially the machine settings shared by the low intensity RIB and high intensity FEL beams, means it would be difficult to separate the simulation into individual modules. The pre-mentioned problems leads to a design requiring the use of multi-engine simulations. The optimization platform was created in preparation as a wrapper for such simulations. A sample of anticipated problems is listed in Table 1.

TRIUMF E-LINAC: INJECTOR CRYOMODULE TO DUMP

We demonstrate the platform capabilities in a realistic and practical problem for TRIUMF's e-linac project [6]. The injector consists of 300 keV beam from a thermionic gun going through a buncher and accelerated to 10 MeV through the injection cryomodule (ICM), which houses a 9-cell superconducting cavity. For the current stage test plan, the beam after the ICM terminates on a beam dump. Three

Table 1: Different aspects of a dual RIB/ERL (Fig. 2) machine requires different and specific treatment. Information courtesy of Y.C. Chao.

Section	Issues	Objectives	Constraints
Injector Complex - gun, cryomodule, simultaneous transport of low intensity RIB and high intensity FEL CW beams	Space charge, phase space preservation, RF bunching and capture	Emittance (longitudinal, transverse), beam profile, momentum spread, transmission, robustness - both beams	Hardware limitation, acceptance (longitudinal, transverse), cost
Injection Merger - injected and recirculated beams, collimation	Longitudinal space charge, phase space preservation, momentum tail collimation	Emittance (longitudinal, transverse), transmission, tunability	Hardware limitation, acceptance (longitudinal, transverse)
Linac - acceleration of low and high intensity beams, energy recovery, transport of 3 different beams	Energy gain, phase space preservation, energy recovery, beam breakup	Beam properties, RF efficiency, instability threshold, transmission, robustness - all 3 beams	Hardware limitation, acceptance (longitudinal, transverse), cost (construction, operation)
RF Separation	Loss less transport	Beam properties, phase space preservation, robustness, tunability	Hardware limitation, acceptance
Arcs - transport, longitudinal manipulation, chromatic control	Phase space manipulation and preservation	Beam properties, transmission, robustness, tunability	Hardware limitation, acceptance (longitudinal, transverse)
Chicane	Phase space manipulation	Beam properties, CSR, tunability	Hardware limitation, acceptance (longitudinal, transverse)
FEL - wiggler, optical cavities	Beam-FEL interaction	Beam properties, FEL efficiency and performance, tunability	Hardware limitation, acceptance (longitudinal, transverse)
Global			Real estate, cost

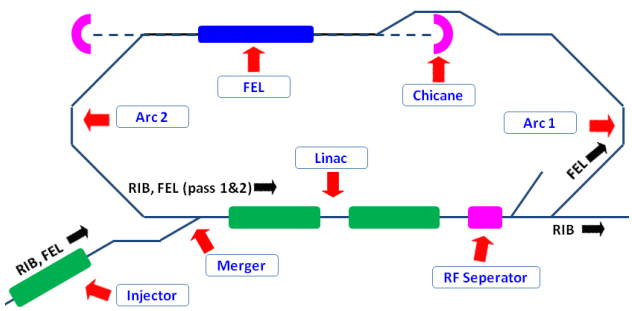


Figure 2: Layout of the upgraded e-linac including dual low intensity rare isotope beam and high intensity beam for light source operation. Many issues need simulations, and is unlikely to be done by a single engine.

solenoids (S1, S2, S3) are placed upstream of the ICM, and one solenoid and two quadrupoles (S4, Q1, Q2) are placed downstream (Fig. 3). A space of 2.5 m is allocated for the dump line from the end of the cryomodule and the entrance of the dump. The objectives are

- want the beam size to be large (≈1 cm) at the dump to minimize heating,
- between S4 and Q1, want the beam size to be minimized, so the beam pipe can be narrowed at this 'neck' to prevent backscatter at the dump from damaging equipment upstream,
- undemanding magnet settings in order to achieve the beam size requirements, and
- feasibility study to determine whether the 2.5 m space reserved from the ICM end to the dump is sufficient to accomplish the above goals.

The variables comprising the search space, i.e. to be optimized, are

- S3 strength,
- S4 strength and position,
- Q1 strength and position, and
- Q2 strength and position.

Table 2: Comparison of a few common simulation engines and their capabilities [9]. Astra is traditionally used at TRIUMF for modeling the e-linac injector, but its dipole modeling is not optimal, therefore another engine should be used for the merger and dogleg modeling. With the future upgrade of the e-linac into an ERL, a combined MAD/Genesis can be used to model the transport and also radiative properties through the FEL.

Engine	Field Map Required	Space Charge	Radiation Field Interaction	3D	Matrix/Tracking
Astra	1D	yes		limited	tracking
GPT	1D-3D	yes		yes	tracking
MADX	no	yes			both
Elegant		no			both
CSRTrack	multipole	yes	yes		tracking
Genesis			yes		tracking

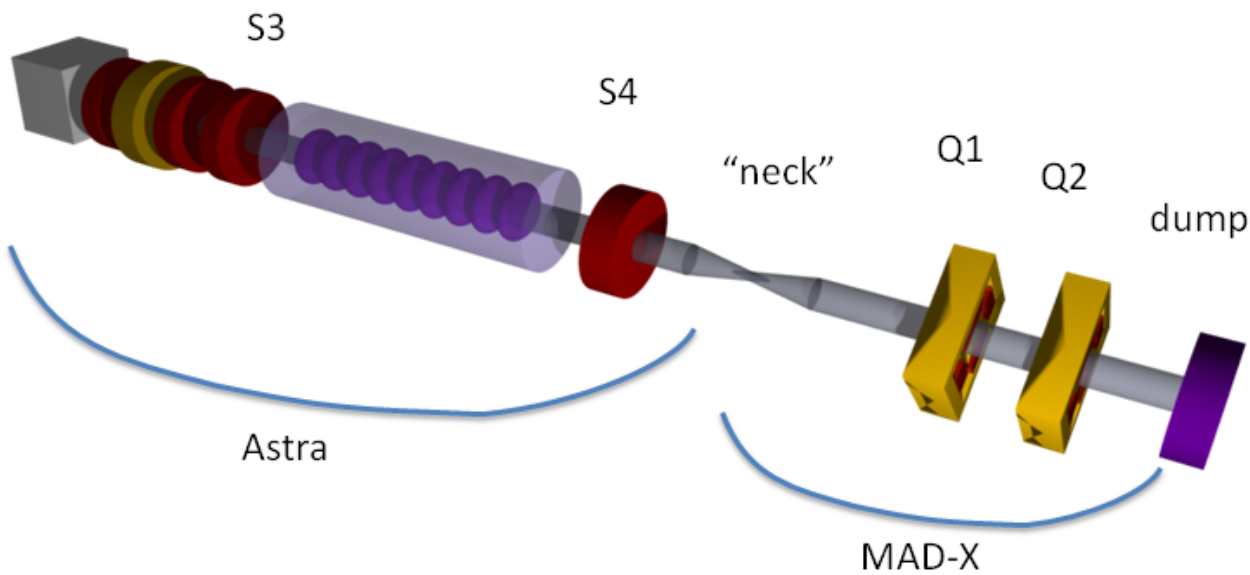


Figure 3: Astra is used to track the 300 keV beam from the exit of the gun to the neck point. The reasons for choosing Astra is 1) at the low 300 keV section, Astra is very easy to use to model space charge effects, and 2) Astra was the traditional engine used for modeling this part of the injector. After the beam accelerates to 10 MeV through the cryomodule and passes the neck point, tracking switches to MAD-X. MAD-X's tracking by Twiss parameters provided a significant boost to the running time of the optimization.

The majority of the magnets and buncher parameters upstream of the ICM are fixed for optimal transport and minimal beam loss through the ICM. Solenoid 3, although upstream of the ICM, was allowed to vary to determine whether it helps with focusing at the neck point.

Optimization was performed to determine if the layout is feasible. Concerns were raised with the preliminary design. In particular, the end of the ICM wall to the dump is 2.5 m. One goal of the optimization was to determine whether objectives can be satisfied within this distance, or if the dump line should be lengthened [10]. The main limitation is solenoid 4, a custom Niowave 20 cm magnet, which has limited focusing ability. Whether a neck can be achieved so close to S4 is a major question. In addition to the optics, space on the transport line must be made for diagnostics.

SIMULATION SETUP

Tracking starts from the gun exit with a beam distribution generated from GPT [11]. Astra is used to track the beam through the ICM and S4 up to the neck point, and MAD-X [12] for the 10 MeV section from the neck, through the quads, to the dump. MAD-X is used for the tracking through the quads because, 1) after the ICM at 10 MeV, the beam is not as sensitive to collective space charge effects, therefore using MAD-X's Twiss parameter tracking provides a significant decrease in tracking time, and 2) MADX provides convenient tools for the extraction of beam parameters and transfer matrices for analysis.

For optimization, the problem requires a two-vertex topology, with each member of the population consisting of the Astra tracking and the MAD-X tracking (Fig. 4). The optimization platform takes care of joining the two

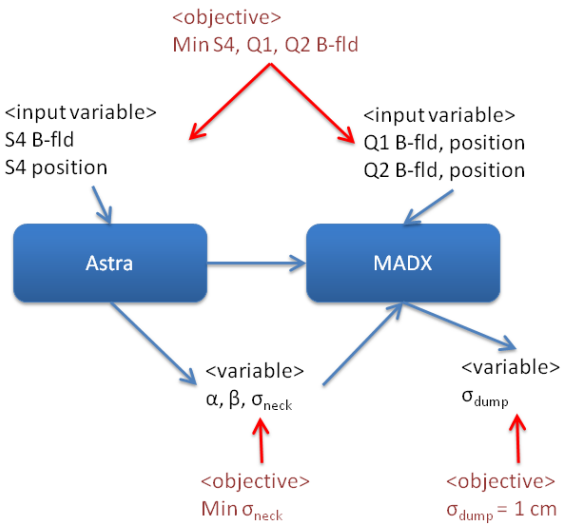


Figure 4: The two-vertex topology of the problem. MAD-X executes in sequence after Astra.

vertices, extracting the beam parameters at the end of the Astra tracking to be used as the input for MAD-X tracking. Twiss parameters are tracked for the MAD-X portion rather than a particle distribution. Only transverse properties are critical for the transport to the dump, therefore the longitudinal properties are ignored. The code also oversees unit conversions, as Astra uses [eV] while MAD-X uses [GeV]. All input for the optimization program is defined through XML (Fig. 5). The extraction of variables from engine outputs is done through python code (Fig. 6). The user can modify the python code, allowing mathematical manipulation of any combinations of variables. Other optimization settings are listed in Table 3.

Table 3: List of optimization settings for the ICM to dump line. Total running time is 15 hours. From previous experience, running time on Westgrid can vary depending on how busy it is.

Setting	Value
Population size	2000
Parent size	20
Generations	1600
Environment	Westgrid
Number of nodes	20

SIMULATION RESULTS

The objectives were easily satisfied within the search space, therefore optimal designs were selected from the solutions that required the least demanding magnet settings. The greatest demands are on the quadrupoles (Fig. 7), as no solutions exist with small field values for both quads. A list of selected individuals are shown in Table 4. A sample plot of the beam size is shown in Fig. 8, which easily satisfies the objectives. The results, particularly of the quadrupole strength and position, is consistent with results derived from an independent study [10], and lends credibility to the optimization engine.

```
<ParamList>
        <Param name="B_S4"      min="0"         max="0.10"      unit="T" />
        <Param name="s_S4"      min="4.2"       max="5.0"       unit="m" />
        <Param name="K1_Q1"     min="0"         max="150"       unit="dimensionl
        <Param name="K1_Q2"     min="-150"      max="0"         unit="dimensionl
        <Param name="s_Q1"      min="5.65"      max="6.05"      unit="m" />
        <Param name="ds_Q1Q2"   min="0.0"       max="0.5"       unit="m" />
</ParamList>

<NDParamList>
        <NDParam name="E_n"     unit="MeV"          comment="" />
        <NDParam name="betx_n"  unit="m"            comment="" />
        <NDParam name="bety_n"  unit="m"            comment="" />
        <NDParam name="alfx_n"  unit="dimensionless"            comment="" />
        <NDParam name="alfy_n"  unit="dimensionless"            comment="" />
        <NDParam name="emitx_n"         unit="m"            comment="" />
        <NDParam name="emity_n"         unit="m"            comment="" />
        <NDParam name="sigmax_n"        unit="m"            comment="" />
        <NDParam name="sigmay_n"        unit="m"            comment="" />
        <NDParam name="sigmax_2"        unit="m"            comment="" />
        <NDParam name="sigmay_2"        unit="m"            comment="" />
</NDParamList>

<ConstraintList>
        <Constraint param="sigmax_2" direction="GT" bound="0.007" unit="m" />
        <Constraint param="sigmax_2" direction="LT" bound="0.015" unit="m" />
        <Constraint param="sigmay_2" direction="GT" bound="0.007" unit="m" />
        <Constraint param="sigmay_2" direction="LT" bound="0.015" unit="m" />
</ConstraintList>

<ObjectiveList>
        <Objective direction="minimize" param="sigmax_n" />
        <Objective direction="minimize" param="sigmay_n" />
        <Objective direction="equals" value="0.01" param="sigmax_2" />
        <Objective direction="equals" value="0.01" param="sigmay_2" />
        <Objective direction="minimize" param="B_S4" />
        <Objective direction="minimize" param="K1_Q1" />
        <Objective direction="maximize" param="K1_Q2" comment="defocusing, so 0
</ObjectiveList>

<Topology>
        <Vertex name="a1" type="ASTRA" inputdir="2069X" prereqs="" timeout="600"
        <Vertex name="mad" type="MADX" inputdir="mad" prereqs="a1" timeout="60"
</Topology>
```

Figure 5: XML input for the ICM to dump problem. In addition to defining the decision variables, constraints, and objectives, the input file also specifies connecting variables (variables extracted from the output of Astra and used as the input to MAD-X), and the topology. Note in the bottom Topology tag, the MAD-X vertex lists the Astra vertex as a prerequisite, specifying that the two engines execute in sequential order. Other parameters, including the number of nodes to use, number of generations, file save/load, and unit conversions, are also defined in this file (not shown).

```
# neck, also transition between astra and mad
file1 = "%s/input.0559.001" % sys.path[0]
if os.path.isfile(file1):
        d1  = astra.ReadAstraFile(file1)
        common.SetValue("betx_n", astra.GetBetax(d1))
        common.SetValue("bety_n", astra.GetBetay(d1))
        common.SetValue("alfx_n", astra.GetAlphax(d1))
        common.SetValue("alfy_n", astra.GetAlphay(d1))
        common.SetValue("E_n", astra.GetE(d1))
        common.SetValue("emitx_n", astra.GetXemitrms(d1))
        common.SetValue("emity_n", astra.GetYemitrms(d1))
        common.SetValue("sigmax_n", astra.GetSigmaxrms(d1))
        common.SetValue("sigmay_n", astra.GetSigmayrms(d1))
```

Figure 6: Custom python code for the Astra vertex, which runs after the Astra executable finishes. The code extracts variables that are used as connecting variables to MAD-X, e.g. $betx\_n$ (beta at neck), or in objective functions, e.g. $sigmax\_n$ (beam size at neck). Library functions are given for extracting common properties, e.g. emittance.

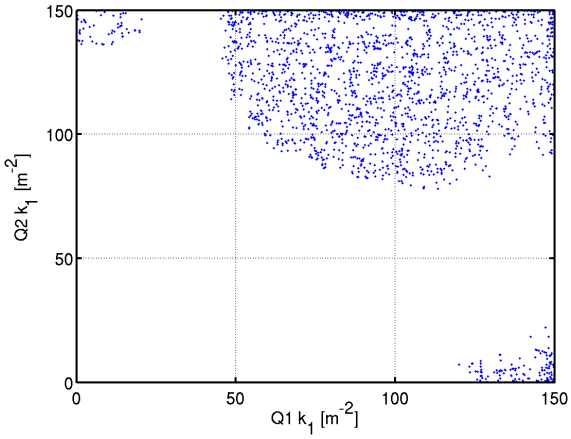


Figure 7: Solution space for quadrupole strengths. No combination of small values can minimize both quads. A k-value of 70 translates to KL = 2300 G. The island in the bottom right with large Q1 and small Q2 corresponds to solutions with Q1 close to the neck point. Here the beam size is very small and the quad does not have much effect. Effects of both quads are therefore minimized and the beam drifts to the dump with $\sigma_x \approx \sigma_y$. The island to the upper left is similar. Why the two islands survived selection is not clear, as they do not satisfy the requirement of large σ_{dump}, but is likely due to minimizing quad strengths as objectives. Table 4 lists designs selected from individuals in the center region of the plot.

CONCLUSIONS

The TRIUMF optimization platform was applied to a multi-engine simulation of the e-linac ICM to dump transport line. Preliminary results are encouraging to the usefulness of the platform. In the near future, the platform will be used on CSR problems and other sections of the e-linac.

REFERENCES

[1] M. Borland et al., "Application of Direct Methods of Optimizing Storage Ring Dynamic and Momentum Apertures," Sep 2009.

[2] APISA — LINAC08, Victoria, 2008 http://www.lepp.cornell.edu/~ib38/apisa/

[3] C. Gong and Y.C. Chao, "A New Platform for Global Optimization," IPAC12, New Orleans, 2012.

[4] E. Zitzler and M. Laumanns and L. Thiele, "SPEA2: Improving the Strength Pareto Evolutionary Algorithm." Technical report, 2001.

[5] ASTRA — http://tesla.desy.de/~lfroehli/astra/

[6] S. Koscielniak et al., "ARIEL and the TRIUMF E-Linac Initiative, a 1/2-MW Electron Linac for Rare Isotope Beam Production,"

[7] PISA — http://www.tik.ee.ethz.ch/pisa/

[8] WestGrid — http://www.westgrid.ca/

Table 4: Individuals selected from the optimized gene pool. For reference, the neck point is defined as z = 5.59 m. End of cryomodule is at 4.09 m. Many individuals satisified the problem objectives. The above were chosen based on magnet restrictions: Sol4<100G, Q1<3000G, Q2<3300G. The quad restrictions causes $\sigma_{dump,y}$ to be smaller smaller than $\sigma_{dump,x}$.

Individual	**1**	**2**	**3**	**4**
Sol4 [G]	49	10	27	6
Sol4 location [m]	4.655	4.630	4.640	4.662
KL, Q1 [G]	2855	2533	2566	2763
KL, Q2 [G]	3029	2895	3204	3167
Q1 location, after neck [m]	0.091	0.086	0.095	0.092
Q2 location, after Q1 [m]	0.319	0.416	0.452	0.312
emit_neck (geometric) [um]	0.411	0.411	0.411	0.411
σ_{neck} **[mm]**	0.669	0.670	0.670	0.670
$\sigma_{dump,x}$ **[mm]**	11.09	10.56	11.11	10.73
$\sigma_{dump,y}$ **[mm]**	7.84	7.02	7.25	8.23

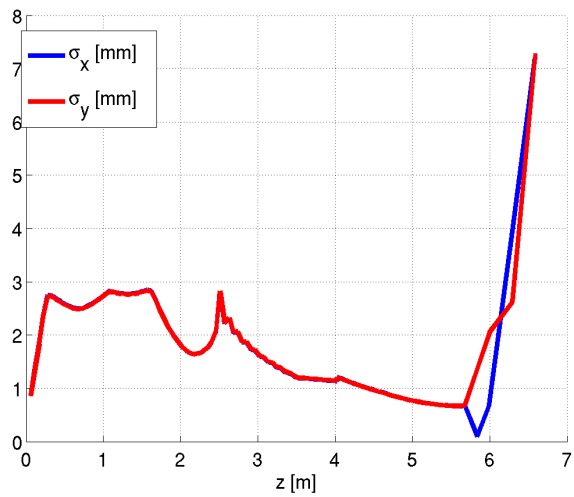


Figure 8: Beam size of one individual in solution space. It satisfies the requirements of small σ_{neck} (z=5.6 m) and large σ_{dump}. The beam pipe can be narrowed at the neck point to prevent backscattering at the dump from damaging the cryomodule upstream. Longitudinal profile of the beam was not studied, as it is not critical to the design objectives.

[9] Y.C. Chao, "Optimization Platform Concept," TRIUMF internal report, Oct 2010.

[10] Y.C. Chao, "VECC 10 MeV Optics – Parameter Feasibility," TRIUMF internal report, May 2012.

[11] General Particle Tracer - http://www.pulsar.nl/gpt/

[12] MAD-X — http://madx.web.cern.ch/madx/

GLOBAL OPTIMIZATION OF THE ANKA LATTICE USING MULTIOBJECTIVE GENETIC ALGORITHMS (MOGA)*

M. Streichert†, N. Hiller, E. Huttel, V. Judin, B. Kehrer, M. Klein‡,
S. Marsching, C. Meuter, M.J. Nasse, A.-S. Müller, M. Schuh, N.J. Smale
Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Abstract

The optimization of a storage ring lattice is a multiobjective problem, since the parameter space of possible solutions can be very large and a high number of constraints have to be taken into account during the optimization process. In this paper we used Genetic Algorithms (GA) and MultiObjective Genetic Algorithms (MOGA), which can solve such problems very efficiently and rapidly, to find the optimized settings for the ANKA storage ring lattice.

INTRODUCTION

ANKA is the synchrotron light source of the Karlsruhe Institute of Technology (KIT)[1]. It consists of 4 super periods with two double bend achromats (DBA) structures each. Each DBA structure contains 2 bending magnets, 5 quadrupole families and two chromatic sextupole families to control the vertical and horizontal chromaticity. We implemented a Global Scan of All Stable Settings (GLASS)[2], based on linear optics to find all possible quadrupole settings of the ANKA lattice. However, the GLASS technique requires a lot of computational power. To decrease the computation time, we considered a symmetric super period by reducing the number of quadrupole families from 5 to 3. Furthermore, we implemented a parallel code to make use of multicore computers. However, the whole scan took up to 28 hours. For optimizations with all five quadrupole families, the GLASS technique is no longer an effective approach. Hence we employed much faster genetic algorithms to find optimal quadrupole settings. We used the GLASS scan to benchmark GA and MOGA for the three quadrupole families and performed global optimizations with GA for all five quadrupole families.

GENETIC ALGORITHMS

GA optimization is a promising technique to find the optimal solution in a multi-dimensional, non-continuous parameter space. This technique can easily be implemented with the existing simulation models for ANKA and does not require knowledge or gradient information of the response function. We compared the GA results with the GLASS results to benchmark and to explore the optimum settings for the GA, for instance the population size, the maximum number of generations etc. We performed GA optimization to find the optimal low-emittance optics. Since not all settings of quadrupole strengths lead to stable or feasible solutions, we implemented the following constraints in GA:

- $|\mathrm{tr}(M_{\mathrm{x,y}})| < 2, \quad \beta_{x,y} < 40\,\mathrm{m},$
- $|\eta_x| < 2\,\mathrm{m}, \quad J_x, J_s > 0,$
- no tune resonance up to the 2^{nd} order,

where $M_{\mathrm{x,y}}$ is the transversal one turn transfer matrix, $\beta_{x,y}$ transversal beta function, $|\eta_x|$ the horizontal dispersion function and J_x, J_s are the damping partition numbers. Beam energy dependent fringe field integrals and quadrupole components in the ANKA bending magnets[3] were taken into account in our optimizations to get a realistic model for the ANKA lattice at the beam energy of 2.5 GeV. The optimization was performed in the range of -2.4 to 2.4 m^{-2}, corresponding to the maximum possible current for the ANKA quadrupole magnets. For a fair comparison of both techniques, we discretized the parameter space for the GA optimization with the same grid spacing (0.02 m^{-2}) as for the GLASS scan, reduced the number of quadrupole families from 5 down to 3, and turned off the sextupole magnets.

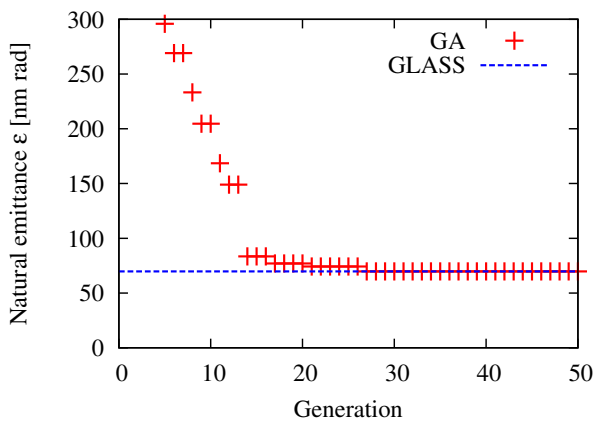


Figure 1: Comparison of the GA with the GLASS results. Minimization of the emittance at 2.5 GeV with minimal chromaticity values of -10 as additional constraint.

After 50 generations the GA converged to the same value of 69.8 nm rad, which we found by using the GLASS (see Fig. 1). However, for the whole GLASS scan, 28 hours

* This work has been supported by the Initiative and Networking Fund of the Helmholtz Association under contract number VH-NG-320.
† max.streichert@student.kit.edu
‡ now at SOLEIL, Gif-sur-Yvette, France

of computation were needed, whereas the GA optimization took only 36 minutes. The difference between both techniques is that the GLASS scan provided all stable settings with their properties, while GA only found the setting with the lowest possible emittance satisfying all above mentioned constraints.

Low-emittance Optics

Since the result of the GA optimization with 3 quadrupole families showed a good agreement with the GLASS result, we increased the number of quadrupole families to 5 and relaxed the constraint for the natural chromaticity to -16, which is the lowest possible value that can be compensated by ANKA's sextupole magnets.
The GA found the optimal stable setting with an emittance of 38.3 nm rad within 200 minutes on a single core. The corresponding optical functions are shown in Figure 2. The emittance of the low-emittance optics currently used at ANKA is 50 nm rad[4], which is 23% higher than the value found by the GA.

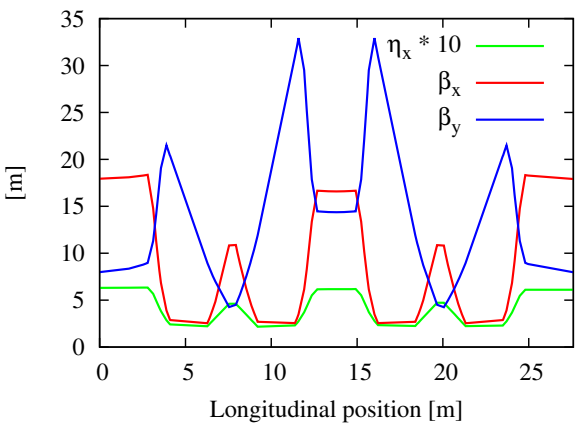


Figure 2: Low-emittance optics with a natural emittance of 38.3 nm rad, tune values of $\nu_{x,y}$ = (7.13, 1.67) and chromaticity values of $\nu'_{x,y}$ = (-15.90, -7.16). One super period of the ANKA lattice is shown.

Low-β_y Optics

We were also interested in low beta values in the straight sections of ANKA as this is required by the three insertion devices installed there. We wanted to improve the currently used low-β_y optics[5], which has a with a β_y of 1.9 m in the straight sections and an emittance value of 53 nm rad. The optimized optics (see Fig. 3) has a vertical beta of 0.7 m in the straight sections and a natural emittance of 51 nm rad, which is an improvement in both quantities.

Low-α_c Optics

A low absolute value of the momentum compaction factor is also an interesting parameter, since it directly determines the bunch length. At ANKA we use a low-α_c optics to decrease the bunch length and produce coherent synchrotron radiation (CSR)[6]. With the GLASS scan (see

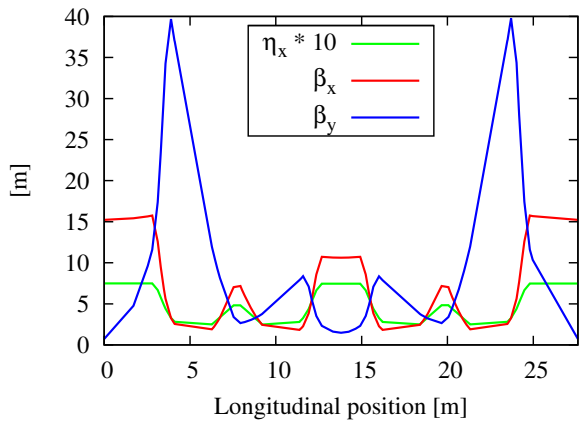


Figure 3: Low-β_y optics with a vertical beta of 0.7 m in the straight sections, an emittance of 51 nm rad, tune values of $\nu_{x,y}$ = (6.44, 4.53), and chromaticity values of $\nu'_{x,y}$ = (-11.10, -12.69).

Fig. 4) we found a stable optics with bunch lengths shorter than 5 mm, which can be considered for production of CSR at the beam energy of 2.5 GeV. For the GA optimization we added a maximum emittance value of 300 nm rad as an additional constraint. With this, we found a stable optics (see Fig. 5) with an RMS bunch length of 0.60 mm (2.0 ps).

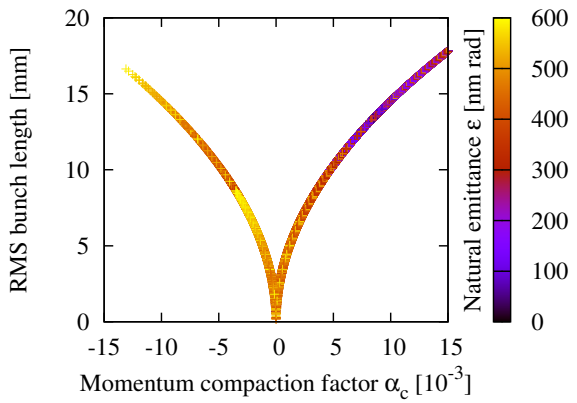


Figure 4: Relationship between momentum compaction factor α_c, natural emittance, and natural RMS bunch length for an infinitesimal current at 2.5 GeV as calculated by the GLASS method.

MULTIOBJECTIVE OPTIMIZATION

However, in most situations multiple parameters have to be optimized simultaneously. For instance, to optimize the brightness of an insertion device not only the local beta function has to be minimized but also the emittance. In this case, the optimum solution is no longer a single point but a front of optimum solutions, the so-called Pareto optimum (PO). For our simulations we used a controlled elitist

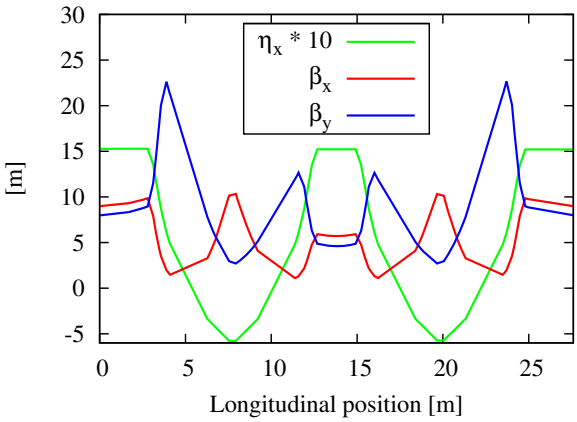


Figure 5: Low-α_c-optics with a momentum compaction factor of $2 \cdot 10^{-5}$, an RMS bunch length of 0.60 mm (2.0 ps), a natural emittance of 280 nm rad, tune values of $\nu_{x,y} = (6.45, 2.96)$ and chromaticity values of $\nu'_{x,y}$ = (-8.84, -8.31).

genetic algorithm (a variant of NSGA-II [7]), which is already implemented in MATLAB's Optimization Toolbox.

β_y – *Emittance Optimization*

We performed a MOGA optimization to find the PO between the vertical beta function in the straight sections and the natural emittance. Figure 6 shows that MOGA found two separated regions with optimum solutions, one with particularly low beta values (A), and the other with relatively low emittance values (B). This demonstrates clearly the advantage of MOGA, since optimum quadrupole settings were found in a non-continuous solution space, as can be seen from the two disjunct regions in Figure 6. The optimization took 220 minutes on a single core.

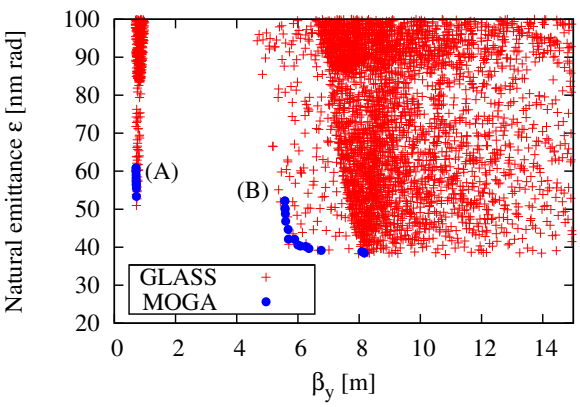


Figure 6: Vertical beta values in the straight sections and natural emittance values calculated with GLASS and MOGA for ANKA at 2.5 GeV.

Bunch Length – Emittance Optimization

As Figure 4 demonstrates a low absolute value of the momentum compaction factor leads to an increase in the natural emittance. We were interested to find the optimum between short bunch lengths and low emittances. The MOGA optimization yields a PO which, as shown in Figure 7, is in good agreement with the GLASS results.

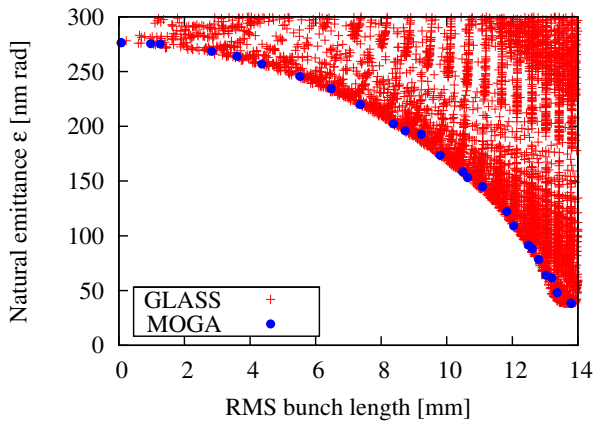


Figure 7: RMS bunch lengths and natural emittance calculated with GLASS and MOGA for the 2.5 GeV ANKA lattice.

SUMMARY AND OUTLOOK

We performed the global optimizations using genetic algorithms for the ANKA storage ring at 2.5 GeV. This technique is a valuable tool to explore the optimal linear stable optics of the ANKA ring. We plan to implement the obtained optimized optics at ANKA in an up-coming machine development shift.
Furthermore, the GA or MOGA can be also used to study non-linear effects considering sextupole, octupole magnets etc. to perform high order optimizations.

REFERENCES

[1] `http://ankaweb.fzk.de`

[2] M. Streichert et al. "Global Scan of All Stable Settings (GLASS) for the ANKA Storage Ring", ICAP 2012

[3] M. Streichert et al."Simulations of Fringe Fields and Multipoles for the ANKA Storage Ring Bending Magnets", IPAC 2012, p.1626

[4] E. Huttel et al., "Operation with a Low Emittance Optics at ANKA", PAC 2005, p.2467

[5] E. Huttel et al., "Low Beta Structure for the ANKA Storage Ring", EPAC 2008, p.2034

[6] A.-S. Müller et al., "Far Infrared Coherent Synchrotron Edge Radiation at ANKA", PAC 2005, p.2518

[7] Deb, Kalyanmoy. Multi-Objective Optimization Using Evolutionary Algorithms. John Wiley & Sons, 2001

MULTI-OBJECTIVE GENETIC OPTIMIZATION OF LINAC BEAM PARAMETERS FOR A SEEDED FEL

Marco Apollonio*, Riccardo Bartolini, Ian Martin,
Diamond Light Source, Oxfordshire, OX11 0DE, United Kingdom

Abstract

The optimization of the beam dynamics in a Linac for free electron lasers (FELs) can be a very time consuming process, in which several parameters of the acceleration and compression sections need to be varied simultaneously. The optimization procedure is required to tackle different and often opposing goals at a time, depending on the adopted FEL scheme. As such, multi-objective genetic algorithms are an interesting choice, given their ability to target several, often conflicting objectives. We have studied an optimization strategy based on a combination of multi-objective optimization with a fast parallel computation of the FEL performance and, for the specific case of the proposed UKs New Light Source (NLS), we illustrate the benefits of this method for the optimization of the average gain length and its variation along the beam pulse. The method can be extended to other sets of objectives, such as power and bandwidth of the FEL.

INTRODUCTION

X-ray free electron lasers are the natural evolution of 3rd generation light sources towards brighter, shorter and fully coherent photon pulses. The presently operating machines (see *e.g.* [1]) represent not only a proof of principle but established tools for new science. A typical structure adopted in these projects consists of a high brightness electron gun, a linear accelerator used to reach the final energy, and are equipped with few compression stages to increase the bunch peak current. A properly optimized linac section should maintain a high brightness throughout. The required linac tuning depends on the kind of lasing we want to achieve. In the simplest scheme, the self-amplified spontaneous emission (SASE), a large peak current, small energy spread and small normalized emittance are key elements. Any portion of the bunch with a length equal to the FEL cooperation length and with the aforementioned beam qualities, will contribute to an independent SASE pulse with its saturation length and limited time coherence. In order to improve the performance of a SASE scheme a seeding laser can be used. In the seeded mode of operation, the temporal coherence can in principle be extended to the length of the seed pulse. However this mode demands a higher beam quality control over a wider region of the pulse length. Proposed schemes like high gain harmonic generation (HGHG), cascaded HGHG or ECHO enabled harmonic generation (EEHG) all require a careful control over the energy spread. Beam uniformity beyond the full seed length both in terms of current and emittance is also an important asset. In general the length of the uniform portion of the beam should be larger than the seed laser pulse, in order to take into account the unavoidable arrival time jitter between the electron bunch and the laser pulse. In this way one makes sure that the laser seed is always overlapping a uniform region of the electron bunch.

In order to reach this, beam dynamics needs to be optmized by means of start-to-end simulations, taking into account collective effects like coherent synchrotron radiation (CSR), longitudinal space charge (LSC) and transverse and longitudinal wakefields present in the accelerating stages. Usually a large number of often correlated machine parameters need to be varied while the resulting objective functions can be mutually conflicting (*e.g.* high peak current and current uniformity in a bunch). Multi-objective genetic algorithms (MOGA) represent an interesting approach to this class of optimization problems, and as such have been already used to characterize accelerators in many different cases (see *e.g.* [2]).

In this paper the stress is on the use of a MOGA for the tuning of a seeded FEL linac. Even though focused on the specific design for the New Light Source, the strategy adopted here can be easily ported to another machine.

OPTIMIZATION FOR THE NLS LINAC

The NLS was a project for a 4th generation light source comprising three seeded FELs driven by a single 2.25 GeV superconducting linac [3]. The baseline set-up diagram is shown on on Fig. 1 together with the most relevant parameters of the machine. More details can be found in [4]. Downstream of the main linac a collimation section and a spreader section are used to remove the beam halo and off-momentum electrons respectively and to bring the beam to the three FELs.

Objectives and Optimization Procedure

The universal figure of merit describing the exponential amplification in a high gain FEL is the non-dimensional Pierce parameter ρ [5]. Another important figure related to ρ is the gain length, which describes the power growth in the undulator section and can be expressed in the 1-D model, by:

$$L_g = \frac{\lambda_u}{4\pi\sqrt{3}\rho} \quad (1)$$

where λ_u is the undulator period. A more realistic expression for the gain length, which takes into account the average size, emittance and energy spread of the beam is given by the Xie parametrization [6]. Both SASE and seeded modes require good electron bunch quality to deliver the shortest possible gain length. In the seeded case

* marco.apollonio@diamond.ac.uk

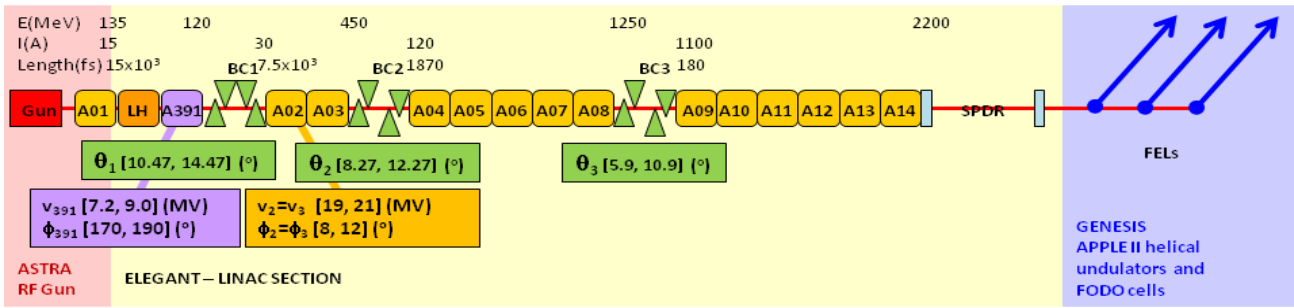


Figure 1: Layout of the NLS linac with three bunch compressors, naming convention and scope for the optimization parameters. The main sections of the system are highlighted together with the settings of every section and the relevant codes used for their description.

it is also preferable to reduce the gain length rms spread over the electron bunch. For our computing purposes we split the machine in three sections: the injector, where the beam dynamics is dominated by low energy transverse space charge forces, the linac, where acceleration and compression need to take into account CSR, LSC and wakefields, and the undulator section, where the lasing takes place. ASTRA is used for the injector simulation, ELEGANT for the linac section and GENESIS for the FEL. In the optimization the knobs used are the strength of the three bunch compressors(θ_1, θ_2, θ_3), the amplitude and phase of the 3^{rd} harmonic cavity (V_{391}, ϕ_{391}) and of the first accelerating cavities ($V_1 = V_2$,$\phi_1 = \phi_2$) as summarized in Fig. 1. The objectives are the gain length and its flatness over a time window of 100 fs around the bunch centre of charge. This value has been chosen after studies have shown that a laser seed pulse with a relative jitter of 20 ns FWHM can be accommodated withing such a time span [3]. The aforementioned objectives appear to be inversely correlated, in that a low gain length is usually associated with a large variation of the same quantity along the bunch length. This feature suits well with a MOGA approach, such as the Non-dominated Sorting Genetic Algorithm (NSGA-II) adopted for our study [7]. NSGA-II finds Pareto-optimal solutions none of which is dominated by the others. The population is given a Pareto rank, that is a fitness order by the number of strictly dominating solutions. After this sorting procedure the population goes through tournament selection, polynomial mutation, and binary crossover, generating new solutions by combining the characters of the fittest parents. The least fit offsprings are discarded. The zero-ranked Pareto front is the result of the optimization after N iterations. The NSGA-II algorithm has been implemented in PYTHON and relies on message passing interface (MPI) to distribute population members across the computing cores of the Accelerator Physics (AP) cluster. The AP cluster at Diamond runs Sun Grid Engine and has 30 nodes, each with 2 quad core Xeon E5430 processors and 16 GB of RAM. Twenty-four nodes are 4xDDR Infiniband enabled to improve the performance of MPI jobs such as GENESIS. The cluster has shared access to a 200TB Lustre parallel file system. The beam dynamics simulation in a linac requires a large number of macroparticles to overcome an instability effect originated by numerical noise. Also, a complete characterization of the FEL interaction would demand a time-dependent simulation. Both requests result in a very challenging scheme for a MOGA optmization, with an overall prohibitive computation time. We therefore adopted the following strategy: (a) track 10^5 particles through the linac, (b) adopt a bunch slice analysis with time-independent simulations. For part (a) we verified that the quality of solution is preserved when increasing the number of macroparticles up to 2×10^6. In part (b) we felt the intrinsic limitations of the usually adopted Xie approach, neglecting beam size variation along the undulator train and beam angles and offsets, could be overcome by computing the FEL gain length by using GENESIS in time-independent mode. We divide the bunch into forty slices

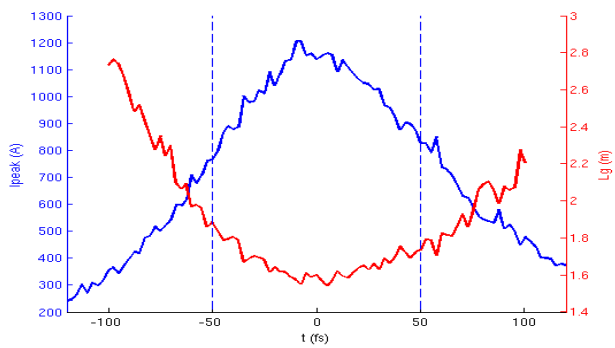


Figure 2: (blue-left vertical axis) bunch current as a function of time for a MOGA solution. (red-right vertical axis) gain length per slice. The vertical dashed lines define the r.o.i. where the average L_g and its RMS are computed.

of 2.5 fs [1] around the centre of charge and treat each slice as a single GENESIS run, where slice emittance, peak current, relative energy spread, Twiss parameters, offset and angles are taken into account. The resulting average on forty gain lengths and its rms are our objectives: ($\langle L_g \rangle$, σ_{L_g}). Fig. 2 shows a typical peak current shape as a func-

[1] comparable to the cooperation length

tion of time (left vertical axis) compared to the gain length per slice (right vertical axis) for a MOGA solution. With such a choice of objectives and knobs the AP cluster can compute a full front with a population of 100 individuals in 23 minutes.

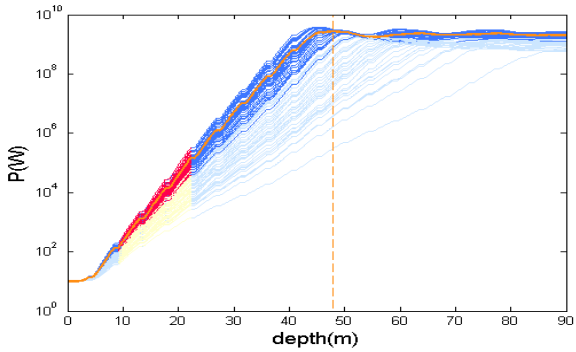


Figure 3: Determination of $(L_g,\ \sigma_{L_g})$ for a 100 iteration front MOGA solution. Power as a function of the undulator length as computed by GENESIS for each bunch slice. Dark blue curves refer to slices within the 100 fs time window around the bunch centroid of charge; cyan curves refer to slices lying out of the aforementioned region. The slope of each power curve is calculated between 9.18 and 22.17 m (red band) and the average curve is plotted in orange.

A further aspect of the optimization is represented by the control of the energy chirp at the end of the linac section. Instead of introducing another objective we opted for the following approach: each gain length L_g is penalized as,

$$\begin{cases} L_{g,chirp} = L_g \cdot (1 + \mathcal{P}) \\ \mathcal{P} = \exp\left(\frac{\delta_n - \delta_{\mathrm{MAX}}}{\sigma_\delta}\right) \text{ penalty function} \\ \delta_n = \frac{|\gamma_n - \gamma_0|}{\gamma_0} \text{ relative energy deviation (n}^{th}\text{ slice)} \end{cases} \tag{2}$$

where δ_{MAX} is the maximum tolerable energy deviation (usually chosen equal to the Pierce parameter) and σ_δ describes the rapidity with which the penalty function rises above unity. In this way every solution with high energy chirp is highly disfavoured in the optimization process.

Results

The MOGA optimization over $(\langle L_{g,chirp}\rangle,\ \sigma_{L_g})$ has been run for the NLS set-up, both for the baseline three bunch compressor case and for a version with two bunch compressors (where only six parameters are varied) [4]. In both cases we use an initial 200 pC beam. The results of the optimizations are shown in Fig. 4 where the two Pareto fronts are compared and the initial manually tuned (SOL_0) solutions are displayed too.

This diagram suggests that the additional flexibility of the third bunch compressor leads to significantly better solutions, since the Pareto-optimal front is completely dominated by the three bunch compressor case. In the two bunch compressor case it proves indeed rather more difficult to reduce the gain length without compromising the flatness, as shown by the steep rise of the Pareto-optimal front as small gain lengths are reached.

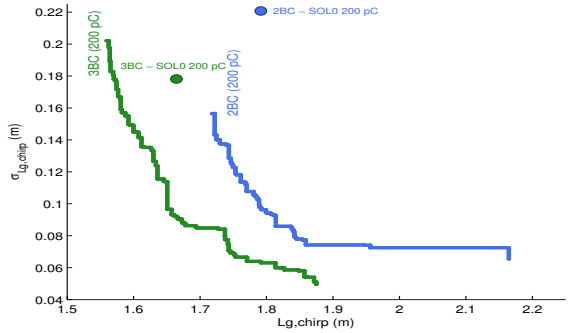


Figure 4: Comparison of the Pareto-optimal fronts for the linac with two bunch compressors (blue) and three bunch compressors (green). The large dots correspond to the starting points of a manual optimization.

CONCLUSIONS AND FUTURE WORK

A MOGA algorithm is shown to be well suited to the complex problem of the optimization of the linac parameters for a seeded FEL. Even though the final validation needs a full numerical simulation with a larger number of macroparticles this method eases the task of designing the machine or defining new operating regimes. For the NLS set of parameters, the MOGA analysis clearly singles out a three bunch compressor design with respect to a two bunch compressor. While this result is likely to be machine dependent, both in terms of layout and initial gun parameters, we believe the strategy adopted would be of benefit to other machines.

ACKNOWLEDGMENT

We wish to thank J. Rowland for the invaluable programming support in setting up the MOGA.

REFERENCES

[1] P. Emma et al., Nature Photon. 4, 641 (2010).

[2] I.V. Bazarov and H. S. Padamsee, in Proceedings of the 21st Particle Accelerator Conference, Knoxville, 2005 (Ref. [12]), p. 1736.

[3] NLS Conceptual Design Report (2010) in `http://www.newlightsource.org`.

[4] R. Bartolini, M. Apollonio and I. P. S. Martin, Phys. Rev. Special Topics - Accelerators and Beams 15, 030701 (2012)

[5] R. Bonifacio, C. Pellegrini, and L. M. Narducci, Opt. Commun. 50, 373 (1984).

[6] M. Xie, Proceedings of the 1995 Particle Accelerator Conference, Dallas (1995).

[7] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, IEEE Trans. Evol. Comput. 6, 182 (2002).

LOW NOISE PARTICLE-IN-CELL SIMULATIONS OF LASER PLASMA ACCELERATOR 10 GeV STAGES*

E. Cormier-Michel†, D. L. Bruhwiler‡ , J. R. Cary, B. M. Cowan, E. J. Hallman,
Tech-X Corporation, Boulder, CO 80303, USA
E. Esarey, C. G. R. Geddes, W. P. Leemans, C. B. Schroeder, J.-L. Vay,
LBNL, Berkeley, CA 94720, USA

Abstract

Because of their ultra-high accelerating gradient, laser plasma based accelerators (LPA) are contemplated for the next generation of high-energy colliders and light sources. The upcoming BELLA project will explore acceleration of electron bunches to 10 GeV in a 1 meter long plasma, where a wakefield is driven by a PW-class laser. Particle-in-cell (PIC) simulations are used to design LPA stages relevant to the upcoming experiments. Simulations in a Lorentz boosted frame are used to gain significant speed up and to simulate the 10 GeV stages at full scale parameters, which are otherwise impractical. As criteria on energy spread and beam emittance become more stringent, PIC simulations become more challenging as high frequency noise artificially increases those quantities. To reduce numerical noise, we consider using a Poisson solve to calculate the beam self-fields. This method allows correct cancelation of the beam transverse self-forces and prevents artificial emittance growth.

INTRODUCTION

Laser-plasma accelerators (LPA) can reach accelerating gradients several orders of magnitude larger than conventional accelerating structures [1], opening the path to more compact light sources and particle colliders. Recently, mono-energetic electron beams have been accelerated to 100 MeV in a 3 mm long plasma [2, 3, 4] and 1 GeV in a 1 cm plasma [5, 6], using 10 and 40 TW lasers respectively. Energy gain is increased by using higher laser power and lower density for longer acceleration length. The BELLA project at LBNL will explore increasing the energy gain to 10 GeV, using a 1 meter long plasma and a PW-class laser [7]. A succession of this type of stages can be used efficiently to build a TeV linear collider [8].

Particle-in-Cell (PIC) simulations are an essential tool to understand the physics inherent to laser-plasma acceleration and to design future devices. As longer stages are used to reach higher energy, simulations become more computationally intensive, since grid size and time step are limited by the laser wavelength, typically of the order of 1 micron. To make simulations more affordable, reduced models are used such as the Ponderomotive Guiding Center method [9, 10, 11, 12, 13], or envelope model, where the laser wavelength is no longer resolved and the smallest scale length is the plasma wavelength which scales with the stage length and energy gain. TThe latter method allows order of magnitude speed-up but is limited because it can not model the laser all the way into depletion. Simulations in a Lorentz boosted frame reduce the number of time steps needed for the simulations while still resolving the laser wavelength [14, 15]. This method has proven successful in simulating 10 GeV LPA stages with orders of magnitude speedup [16].

Accurate representation of the accelerated beam is limited by numerical noise in PIC simulations. This becomes more of an issue as the condition on beam energy spread and emittance becomes more stringent. Introduction of higher order particle shapes has allowed significant reduction of numerical noise in LPA simulations with PIC [17], representing beam properties within a few % of what is obtained in experiments, in contrast to results using linear interpolation which differ by an order of magnitude [18]. To more accurately represent beam evolution with % level energy spread and a fraction of mm mrad emittance, as required by applications, we explore using a method commonly used in tracking codes [19, 20, 21] but that has never been used in the context of LPA, in which the beam self-fields are calculated using a Poisson solve in the beam rest frame. This method provides correct relativistic cancelation of the beam transverse self-forces and prevents high frequency numerical noise responsible for artificial emittance growth of the beam.

SIMULATIONS IN A BOOSTED FRAME

Methods presented by J.-L. Vay and collaborators [16] are used in the Vorpal framework [22] to perform simulations in a relativistic boosted frame of 10 GeV LPA stages, relevant to the BELLA project, with parameters similar to those presented in [23]. The boosted frame technique allows the simulation to be performed at full scale, i.e., 1 m long stage at the nominal plasma density $n_0 = 10^{17}\ \mathrm{cm}^{-3}$. An externally injected 2 pC beam is used, with normalized emittance $\epsilon_n = 0.5\ \mathrm{mm\,mrad}$ and initial energy $E = 1\ \mathrm{GeV}$. The evolution of the beam properties is consistent for different values of the simulation frame relativis-

* Work supported by the US DOE office of science, office of High-Energy Physics under grants numbers DE-FC02-07ER41499 (SciDAC-2; ComPASS Project), DE-SC0004441 (SBIR), and DE-AC02-05CH11231 (LBNL). This research used resources of the National Energy Research Scientific Computing Center, which is supported by US DOE Office of Science under Contract No. DE-AC02-05CH11231.

† ecormier@txcorp.com
‡ Current address: 1348 Redwood Ave., Boulder CO 80304

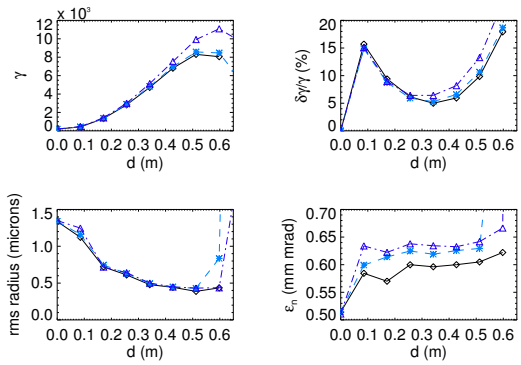


Figure 1: Evolution of electron beam mean γ, rms energy spread, rms radius, and normalized emittance as a function of propagation distance for $\gamma_{\rm boost} = 25$ (diamonds), $\gamma_{\rm boost} = 50$ (stars) and $\gamma_{\rm boost} = 75$ (triangles).

tic factor $\gamma_{\rm boost}$ as shown in Fig. 1. A factor of $3,500$ times speedup is achieved when performing the simulation with $\gamma_{\rm boost} = 75$. Numerical instabilities arise when performing the simulation with a high $\gamma_{\rm boost}$, which are mitigated by using smoothing with large stride on the current deposited by the particles [24]. Although beam energy gain, energy spread and radius evolution are converged, higher resolution is needed at high $\gamma_{\rm boost}$ to accurately represent the beam emittance, limiting the effective speedup to $550\_$ fold in this case, the simulation being performed in 4,500 processor hours.

BEAM FRAME POISSON SOLVE

Because of numerical noise in PIC simulations, high resolution is often needed to accurately represent the evolution of the accelerated electron beam. Beam emittance can particularly be affected by the high frequency noise inherent to the point like representation of the macro-particles. In simulations of conventional accelerator with tracking codes, it is common to use a Poisson solve in the beam rest frame to calculate the beam self-fields, however this method has never been employed in PIC simulations of LPA. The beam charge density is Lorentz transformed in the frame where the beam is at rest, and a Poisson solve is used to calculate the corresponding electrostatic field, which is then Lorentz transformed back into the laboratory frame to obtain the beam electro-magnetic beam self-fields. This is shown in the blue boxes of the diagram in Fig. 2. This method assumes that simultaneity in the laboratory frame implies simultaneity in the beam frame (the particle time transformation is neglected), and that all the particles of the beam are non-relativistic in the beam rest frame, and hence can only be used in the case of compact, low energy spread, low divergence beams. This is usually the case when designing LPA for future applications. In the following, we refer to this method as *Beam Frame Poisson Solve* (BFPS).

This method allows the beam fields in the laboratory frame to be calculated at the same position on the grid, and hence allows correct relativistic cancelation of the beam transverse self-forces. In the PIC algorithm, where a Yee update of the electro-magnetic fields is generally used, the $\mathbf{E}$ and $\mathbf{B}$ fields are staggered on the grid, leading to different interpolation errors to the macro-particle position. This translates into an inaccurate cancellation of the $\mathbf{E}$ and $\mathbf{v} \times \mathbf{B}$ transverse forces, which in turn leads to increase of the radius of an otherwise matched beam.

The BFPS also mitigates artificial beam emittance growth of a matched beam in the presence of a linear focusing force. Numerical emittance growth, when using the Yee algorithm, can be reduced by using higher resolution and several pass of smoothing, with larger strides, on the current and the forces applied to the beam, but is sometimes not sufficient to converge to the right answer, especially for very low emittance beams (~ 0.1 mm mrad).

The BFPS method can also be used in the plasma wakefield, thanks to the linearity of Maxwell's equations, at the condition that the beam is initially in the vacuum region. The method, implemented from the Vorpal input file, is shown in Fig. 2. The beam self-fields are calculated as explained above, using a Poisson solve in the beam frame. In parallel, the fields of the plasma are calculated self-consistently using the Yee advance normally used in PIC simulations. The fields are then combined to push particles of both beam and plasma, which then deposit charge and current densities, respectively, to perform the next field advance.

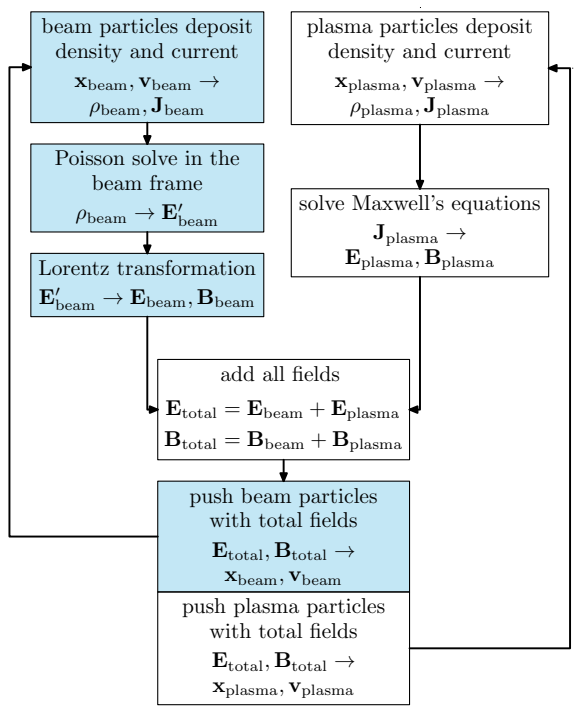


Figure 2: Diagram of the BFPS algorithm when the beam is used inside a plasma wakefield. The blue boxes indicate the algorithm for the beam alone.

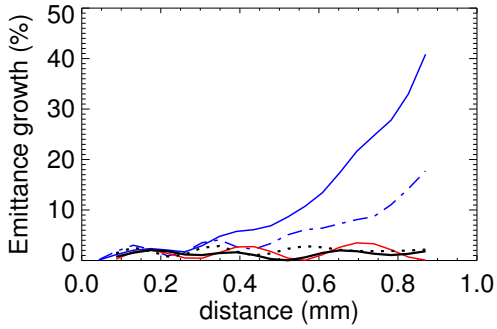


Figure 3: Percentage emittance growth of a quasi-matched electron beam in a plasma wakefield, using the BFPS (black) at the resolution $dx = \lambda_0/24$ (solid line) and $dx = \lambda_0/48$ (dotted line), and using the Yee algorithm (blue) at the resolution $dx = \lambda_0/24$ (solid line) and $dx = \lambda_0/48$ (dash-doted line), using the Yee algorithm with an additional 4-pass (1,2,1) filter with compensator of stride 2 at the resolution $dx = \lambda_0/24$ (red line).

Fig. 3 shows the beam emittance evolution of a quasi-matched Gaussian beam in a plasma wakefield. The wakefield is driven by a laser pulse of normalized intensity $a_0 = 1.4$ with a Gaussian profile both longitudinally and transversely with dimensions satisfying $k_p L = 1$ and $k_p w_0 = 5.3$, similar to what is used in [23]. Here, $k_p = \sqrt{4\pi n_0 e^2/mc^2}$ is the plasma wave number, a density $n_0 = 10^{19}$ cm^{-3} on axis is used with a parabolic profile transversely. A 10 pC electron beam with a Gaussian profile is loaded behind the laser pulse, with rms dimensions verifying $k_p\sigma_L = 0.1$ and $k_p\sigma_r = 0.35$, initial normalized emittance $\epsilon_n = 0.5$ mm mrad and energy $E = 100$ MeV $\pm$ 1%. The acceleration of the beam is turned off by setting $E_x = 0$, where x is longitudinal direction. The solution with the BFPS is shown in black for $dx = \lambda_0/24$ and $dx = \lambda_0/48$ (dotted line), where λ_0 is the laser wavelength, showing similar evolution of the beam emittance, which oscillates but does not grow overall. The solution using the standard Yee algorithm is shown in blue at the same resolution of the BFPS runs, showing obvious emittance growth, which is reduced when using higher resolution. Using an additional 4-pass smoothing filter with compensator with stride 2 on the current converges toward the BFPS response.

The BFPS algorithm can be used in the boosted frame. We verified that the beam evolution is similar in the previous case for $\gamma_{\rm boost} = 13$. This allows for simulation full-scale 10 GeV stages with accurate emittance of the beam.

CONCLUSION

Boosted frame simulations are used in the Vorpal framework to simulate full-scale meter-long 10 GeV gain LPA stages. Evolution of an externally injected beam is consistent for different values of $\gamma_{\rm boost}$.

Numerical noise is reduced by calculating the beam self-fields using a Poisson solve in the beam rest frame. This allows correct relativistic cancelation of the beam transverse self-forces and prevents artificial emittance growth of the beam. This method can be used inside the plasma wakefield if the beam is initialized in the vacuum region. A quasi-matched beam in the focusing field of the wake shows no spurious emittance growth using the BFPS, contrary to the standard Yee algorithm in the same conditions.

REFERENCES

[1] E. Esarey *et al.*, *Rev. Mod. Phys.*, **81**, 1229, (2009).

[2] S. P. D. Mangles *et al.*, *Nature*, **431**, 535, (2004).

[3] C. G. R. Geddes *et al.*, *Nature*, **431**, 538, (2004).

[4] J. Faure *et al.*, *Nature*, **431**, 541, (2004).

[5] W. P. Leemans *et al.*, *Nat. Phys.*, **2**, 696, (2006).

[6] N. A. M. Hafz *et al.*, *Nat. Photonics* **2**, 571, (2008).

[7] W. P. Leemans *et al.*, *AIP Conf. Proc.*, **1299**, 3, (2010).

[8] C. B. Schroeder *et al.*, *Phys. Rev. ST Accel. Beams*, **13**, 101301, (2010).

[9] J. Krall *et al.*, *Phys. Rev. E*, **48**, 2157, (1993).

[10] P. Mora and Jr. T. M. Antonsen, *Phys. of Plasmas*, **4**, 217, (1997).

[11] D. F. Grodon *et al. IEEE Trans. Plas. Sc.*, **28**, 1135, (2000).

[12] C. Huang *et al.*, *J. Comp. Phys.*, **217**, 658, (2006).

[13] B. M. Cowan *et al. J. Comp. Phys.*, **230**, 61, (2011).

[14] J.-L. Vay., *Phys. Rev. Lett.*, **98**, 130405, (2007).

[15] S. F. Martins *et al.*, *Nat Phys*, **6**, 311, (2010).

[16] J.-L. Vay *et al.*, *Phys. of Plasmas*, **18**, 123103, (2011).

[17] E. Cormier-Michel *et al.*, *Phys. Rev. E*, **78**, 016404, (2008).

[18] C. G. R. Geddes *et al.*, *SciDAC review*, **13**, 13, (2009).

[19] F. J. Sacherer., *Proc. of the Part. Accel. Conf.*, 1105, (1971).

[20] D. L. Bruhwiler and Michael F. Reusch, *Proc. of the Part. Accel. Conf.*, 3624, (1993).

[21] J. Qiang *et al.*, *J. of Comp. Phys.*, **163**, 434, (2000).

[22] Chet Nieter and John R. Cary, *J. of Comp. Phys.*, **196**, 448, (2004).

[23] E. Cormier-Michel *et al.*, *AIP Conf. Proc.*, **1086**, 297, (2009).

[24] J.-L. Vay *et al.*, *J. of Comp. Phys.*, **230**, 5908, (2011).

NUMERICAL MODELING OF COLLECTIVE EFFECTS IN FREE ELECTRON LASER

Igor Zagorodnov, DESY, Hamburg, Germany

Abstract

In order to have a free electron laser (FEL) of high performance we need to design and optimize it taking into account the dynamics of electrons and their interactions with each other and with their surroundings. An accurate self-consistent simulation of collective effects in the charged beams remains a challenging problem for numerical analysis. In this paper we consider only the modeling of FEL process in an undulator section. We give a short overview of the numerical methods adopted in different FEL codes. Advantages and drawbacks of these methods will be discussed. Some approaches to improve the accuracy and efficiency of the codes will be presented and the remaining challenges in FEL modeling will be highlighted.

INTRODUCTION

An accurate self consistent simulation of collective effects in the charged beams remains a challenging problem for numerical analysis. During the last decades several numerical codes have been developed to model the non-linear process in a self-amplified spontaneous emission (SASE) free electron lasers (FEL). In this paper we review the mathematical FEL model used in these codes [1-4]. To illustrate the numerical methods we use those implemented in code ALICE [5].

Modelling of the FEL is challenging due to different scales of the process [6]. The disparity of scales up to 12 orders of magnitude imposes several limitations on the size of the systems that can be modelled on the basis of classical Maxwell-Vlasov set of equations.

In order to be able to study the FEL process two main approximations are used: wiggler-period averaging of the equations of motion and slowly-varying envelope approximation of the Maxwell equations.

FEL MATHEMATICAL MODEL

We consider the case of helical undulator with the magnetic field on the axis given by

$$\vec{B}_\perp^w = (\nabla\times\vec{A}_\perp^w)_\perp = \frac{\partial}{\partial z}\begin{pmatrix}-A_y^w\\ A_x^w\end{pmatrix} = \frac{mc}{e}Kk_w\begin{pmatrix}\cos(k_w z)\\ -\sin(k_w z)\end{pmatrix}.$$

Following the approach of [1] the equations of motion for helical undulator can be derived from Hamiltonian

$$h(\vec{r}_\perp, ct; \vec{P}_\perp, P_t; z) = -\gamma\left[1-\frac{1+\left|\vec{P}_\perp\right|^2+\left|\vec{a}_\perp\right|^2+2(\vec{P}_\perp,\vec{a}_\perp)}{\gamma^2}\right]^{0.5} + a_z .$$

In order to avoid the necessity to resolve the fast "slalom" motion we carry out *wiggler-period averaging* of the Hamiltonian. The scalar product disappears and we obtain the averaged Hamiltonian

$$h(\vec{r}_\perp, ct; \vec{P}_\perp, P_t; z) = -\gamma\left[1-\frac{1+\left|\vec{P}_\perp\right|^2+\left|\vec{a}_\perp\right|^2}{\gamma^2}\right]^{0.5} + a_z .$$

The squared module of the transverse part of the vector potential $\vec{a} = e\vec{A}/(mc)$ can be approximated as

$$\left|\vec{a}_\perp\right|^2 = K^2 + 2a_s K\sin(\psi+\varphi_s).$$

Here $\psi = (k+k_w)z - \omega t$ is a particle phase, K is an undulator parameter, $a_s\exp(i\varphi_s)$ is a normalized complex amplitude of the amplified wave

$$A_x + iA_y = -imce^{-1}a_s e^{i(kz-\omega t+\varphi_s)},$$

$$\vec{E}_\perp = -\frac{\partial}{\partial t}\vec{A}.$$

The canonical moments are defined by relations

$$\vec{P}_\perp = \gamma\frac{d\vec{r}_\perp}{dz}\beta_z - \vec{a}_\perp,\ P_t = -\gamma + \frac{e}{mc^2}\varphi .$$

The equations of motion derived from the averaged Hamiltonian read

$$\frac{d\psi}{dz} = k_w - k(2\gamma^2)^{-1}\left(1+\left|\vec{P}_\perp\right|^2+K^2\right),$$

$$\frac{d\gamma}{dz} = \frac{eK}{mc^2\gamma^2}\Re(\tilde{E}e^{i\psi}) - \frac{e}{mc^2}E_z,$$

$$\frac{d\vec{r}_\perp}{dz} = \frac{\vec{P}_\perp}{\gamma\beta_z},$$

$$\frac{dP_x}{dz} = -\left(\frac{K^2k_x^2}{\gamma\beta_z}+\frac{eg}{mc}\right)x,\quad \frac{dP_y}{dz} = -\left(\frac{K^2k_y^2}{\gamma\beta_z}-\frac{eg}{mc}\right)y,$$

where $g = \partial_y B_x$ describes the gradients of the focusing quadrupole lattice and we have used the following representations of the undulator parameter and the transverse EM field

$$K(x,y,z) = K_0(x,y,z)\left(1+\frac{k_x^2}{2}x^2+\frac{k_y^2}{2}y^2\right),$$

$$E_\perp = E_x + iE_y = \tilde{E}(\vec{r},t)\,\mathrm{e}^{i(kz-\omega t)} + c.c..$$

We split the electromagnetic field in the transverse and the longitudinal components. The longitudinal electrostatic field comes from the bunching and can be suggested to be a nearly periodic one. Then the Fourier components of the longitudinal field can be found from the equation [1]

$$\left(\nabla_\perp^2 - \frac{n^2(k+k_w)^2}{\gamma_z^2}\right)E_z^{(n)} = \frac{in(k+k_w)}{\varepsilon_0\gamma_z^2}\rho^{(n)},$$

$$\rho^{(n)}(\vec{r}_\perp) = \frac{1}{2\pi}\int_0^{2\pi}\rho(\vec{r}_\perp,\psi)e^{-in\psi}d\psi = -\frac{I}{Nv_z}\sum_{i=1}^{N}e^{-in\psi_i}\delta(\vec{r}_\perp-\vec{r}_{i\perp}).$$

The transverse field is written in the complex form with amplitude $\tilde{E}(\vec{r},t)$ which fulfils the parabolic equation (paraxial or *slowly-varying envelope approximation*)

$$\left[\nabla_\perp^2 + 2ik\left(\frac{\partial}{\partial z}+\frac{1}{c}\frac{\partial}{\partial t}\right)\right]\tilde{E}(z,t) = ikc\mu_0\frac{K}{\gamma}v_z\rho^{(1)}(z,t).$$

Along the trajectory $t = t_0 + z/c$ this equation can be rewritten in a simpler form

$$\left[\nabla_\perp^2 + 2ik\frac{d}{dz}\right]\tilde{E}\left(z,t_0+\frac{z}{c}\right) = ikc\mu_0\frac{K}{\gamma}v_z\rho^{(1)}\left(z,t_0+\frac{z}{c}\right).$$

If the Pierce parameter [7] $\rho = c\gamma_z^{\,2}\Gamma\omega^{-1}$ is small and the transverse variation of the longitudinal field can be neglected than the following set of normalized and simplified equations can be considered

$$\hat{x}'' = -\left(\hat{k}_x^{\,2}+\hat{g}\right)\hat{x},\ \hat{y}'' = -\left(\hat{k}_y^{\,2}-\hat{g}\right)\hat{y},$$

$$\frac{d\psi}{d\hat{z}} = \hat{C}+\hat{\eta}-\frac{B}{2}(\hat{x}'^2+\hat{y}'^2),\ \frac{d\hat{\eta}}{d\hat{z}} = |\hat{u}|\cos(\psi+\psi_r)-\hat{E}_z,$$

$$\hat{E}_z = \hat{E}_z^{(0)} - \hat{\Lambda}_p^{\,2}\frac{1}{N_{jk}}\sum_{i=1}^{N_{jk}}\left[\pi\,\mathrm{sgn}(\psi-\psi_i)-(\psi-\psi_i)\right],$$

$$\left[\frac{1}{2iB}\hat{\Delta}_\perp + \frac{d}{d\hat{z}}\right]\hat{u}\left(\vec{\hat{r}}_\perp,\hat{z},\tau_0+\frac{\hat{z}}{c}\right) = -2a^{(1)}\left(\vec{\hat{r}}_\perp,\hat{z},\tau_0+\frac{\hat{z}}{c}\right),$$

where $\hat{x} = xr_0^{-1}$, $\hat{y} = yr_0^{-1}$, ($r_0 = \sqrt{2\sigma_x\sigma_y}$) are transverse particle coordinates, $\hat{\eta} = (\gamma-\gamma_0)(\gamma_0\rho)^{-1}$ is an energy deviation, $\hat{C}$ is a detuning parameter, $\hat{u}$ is the normalized (in complex notation) electric field and $\hat{E}_z$ is the longitudinal electric field. The other parameters in Eq. (1) are defined as in [7, Chapter I].

NUMERICAL METHODS

In last decades several FEL codes have been developed around the world. In our studies at DESY we use mainly three codes whose basic features are listed in Table 1.

Table 1: Numerical Methods in Different Codes

	FAST[2]	Genesis 1.3[3]	ALICE[5]
Equations of motion	Runge-Kutta		Leap-frog
EM field solver	Integral representation	Finite-difference, ADI	Finite-difference, Neumann
Boundary condition	Free space	Dirichlet	Free space with PML

These codes are based on the mathematical model described in the previous section. In the following we consider the numerical methods implemented in code ALICE [5]. In this code the equations of motion of the particles are integrated with symplectic "leap-frog" scheme. The parabolic field equation is resolved with implicit Neumann finite difference scheme based on azimuthal expansion. Additionally we have implemented the open boundary condition with the help of perfectly matched layer (PML) for parabolic equation. The last feature allows for a mesh only in the bunch vicinity. The implemented field solver is accurate and fast. The code is parallelized and allows to use one dimensional, rotationally symmetric or fully three dimensional models.

Particle Distribution

At the beginning we divide the bunch longitudinally in N_s slices (numerated from the tail) with the length equal to the radiation wavelength. The initial particle distribution in the slice is generated with the "quiet start" method [8]. For this purpose we use the Sobol sequences [9]. The uniform distribution is converted to the Gaussian one with the help of the inverse error function

$$y = \sigma\sqrt{2}\,\mathrm{erf}^{-1}(2x-1)+\mu.$$

Figure 1 presents a comparison with the code Genesis 1.3 [3]. We carry out a simulation with only one slice in amplifier model. The space of parameters corresponds to SASE2 undulator at wavelength of 0.1 nm as described in [10]. The left plot compares the radiation power at saturation. The first comparison was done in the January 2008 [11] (version 1.0 of Genesis 1.3). The disagreement in the saturation power at the level of 20 % was obtained. After data analysis we have found that the used in Genesis 1.3 the "Box-Mueller" algorithm [9] (to convert the uniform distribution to Gaussian one) has spoiled the "quiet-start" property and the statistics of the particle distribution. The new version 2.0 of Genesis 1.3 released in April 2008 allows to use the inverse error function transformation (parameter inverfc=1). With this option results obtained by both codes converge together.

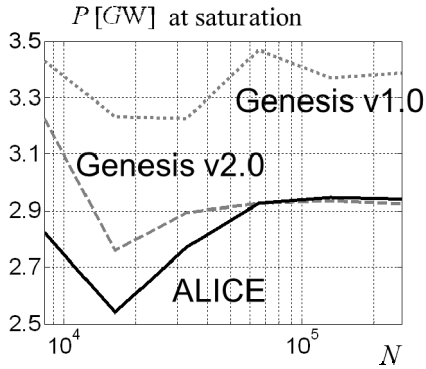


Figure 1: Convergence of the different codes.

The noise statistics in the slice is imposed following Fawley [12]. The n_p macropartiucles in each beam slice are subdivided into an ensemble of n_b beamlets, each composed of $2M$ macroparticles (M is the order of the highest harmonic of interest). The $2M$ macroparticles in each beamlet k are loaded uniformly

$$\theta_{jk} = \theta_{0k} + (j-1)\frac{\pi}{M},\qquad j = 1..2M,\ k = 1..n_b,$$

where θ_{0k} is uniformly increased with index k over the interval $\left[0,\frac{\pi}{M}\right]$. To model the effects of shot noise a small random deviations,

$$\delta\theta_{jk} = \sum_{m=1}^{m=M} a_{mk}\cos(m\theta_{jk}) + b_{mk}\sin(m\theta_{jk}),$$

are added to the macroparticle positions. Each coefficient a_{mk}, b_{mk} are independently picked from Gaussian distributions with rms width

$$a_{mk,rms} = b_{mk,rms} = \sigma_{mk} = \frac{1}{m\sqrt{N_b}} \times \begin{cases} \sqrt{2}, & m < M, \\ 1, & m = M. \end{cases}$$

Note that for $m = M$ the rms width σ_{mk} is different from that given in [12]. Indeed, instead of Eq (5a) from [12] we obtain

$$\frac{1}{2M}\sum_{j=1}^{2M}\cos(m(\theta_{jk}+\delta\theta_{jk})) =$$

$$= \begin{cases} -\frac{m}{2}b_{mk}, & m < M, \\ -\frac{M}{2}\left(2b_{Mk}\sin^2(M\theta_{0k}) + a_{Mk}\sin(2M\theta_{0k})\right), & m = M. \end{cases}$$

For $m < M$ we have

$$\left\langle \left| \sum_{j=1}^{2M}\cos(m(\theta_j+\delta\theta_j)) \right|^2 \right\rangle = m^2M^2\sigma_{mk}^2.$$

And for $m = M$ the equation reads

$$\left\langle \left| \sum_{j=1}^{2M}\cos(M(\theta_j+\delta\theta_j)) \right|^2 \right\rangle =$$

$$\left\langle \left| M^2\left(2b_{Mk}\sin^2(M\theta_{0k}) + a_{Mk}\sin(2M\theta_{0k})\right) \right|^2 \right\rangle =$$

$$= M^4\left(4b_{Mk}^2\frac{3}{8} + a_{Mk}^2\frac{1}{2}\right) = 2M^4\sigma_{Mk}^2.$$

The described "shot noise" algorithm is based on small variations in the longitudinal positions of the macroparticles. Alternatively a variation in the charge can be used as it is described in [13].

Integration of Equations of Motion

The longitudinal equations are discretized with "leap-frog" scheme

$$\frac{\psi_{i,j+0.5} - \psi_{i,j-0.5}}{\Delta\hat{z}} = \hat{\eta}_{ij} + \hat{C}_j - \frac{B}{2}\left(\hat{x}'^2_{ij} + \hat{y}'^2_{ij}\right),$$

$$\frac{\hat{\eta}_{i,j+1} - \hat{\eta}_{i,j}}{\Delta z} = \frac{\hat{u}_{j+1} + \hat{u}_j}{2} +$$

$$+ \left[\psi_{i,j+0.5} + \frac{\varphi^s_{j+1} + \varphi^s_j}{2}\right] + \hat{E}_{z,j+0.5},$$

and for the transverse equations a matrix formalism of the linear beam optics is applied

$$\begin{pmatrix} \hat{x}_{i,j+1} \\ \hat{x}'_{i,j+1} \end{pmatrix} = M_x \begin{pmatrix} \hat{x}_{i,j} \\ \hat{x}'_{i,j} \end{pmatrix}, \qquad i = 1:N_p,\ j = 1:N_z,$$

$$\begin{pmatrix} \hat{y}_{i,j+1} \\ \hat{y}'_{i,j+1} \end{pmatrix} = M_y \begin{pmatrix} \hat{y}_{i,j} \\ \hat{y}'_{i,j} \end{pmatrix}.$$

Solution of the Parabolic Field Equation

The transverse mesh is constructed in polar coordinates with N_r divisions along the radius and N_φ divisions along the angle. Each slice is tracked through the undulator with N_z periods. In the following we consider time-independent case and only later on we will discuss the general case.

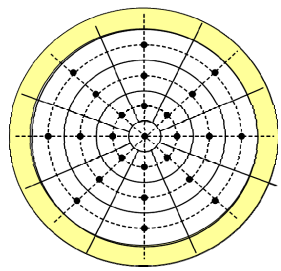


Figure 2: The transverse mesh with PML.

To find the transverse field we use the Fourier transform in azimuthal coordinate. For each azimuthal mode m we have to solve the parabolic equation (in a simplified notation)

$$\left[\frac{1}{2iB}\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial}{\partial r} - \frac{m^2}{r^2} + \frac{d}{dz}\right]u^{(m)} = -2a^{(1)(m)}, r \in (0,\infty),$$

with the condition on the axis

$$\frac{\partial}{\partial r}u^{(0)}(0) = 0,\ u^{(m)}(0) = 0,\ m > 0.$$

In order to truncate the mesh at radius r_0 we use an absorbing layer called Perfectly Matched Layer (PML) [14], which possesses the desired property of generating very low numerical reflection. In order to construct a mathematical model of the PML we introduce the complex variable

$$\tilde{r} = r + \frac{i}{B}\int_0^r \sigma(\xi)d\xi, \qquad \sigma(r) = \begin{cases} 0, & r \le r_0 \\ \ge 0, & r > r_0 \end{cases}.$$

The change of the variable r to $\tilde{r}$ (and the partial derivative $\partial/\partial r$ to $\partial/\partial\tilde{r}$) in the parabolic wave equation will give us the required equation. This change of variable does not alter the solution in the area of interest ($r < r_0$), but it extends the solution by a fast exponentially decaying part in the absorbing layer $r_0 < r < r_{PML}$.

Let us introduce the radial mesh

$$\tilde{r}_j = \tilde{r}_{j-1} + \frac{i}{B}\sigma_{j-0.5}\Delta r,\ \ j = 1:n_r,\ n_r = r_0/\Delta r,$$

$$\sigma_{j-0.5} = \begin{cases} 0, & j \le n_r, \\ B\dfrac{10}{\Delta r}\left(\dfrac{r_{j-0.5}}{L_{PML}}\right)^2, \end{cases}$$

$$r_{j-0.5} = (j - n_r - 0.5)\Delta r\,, \quad L_{PML} = r_{PML} - r_0\,.$$

The implicit Neumann numerical scheme reads

$$c_q u_{q+1}^{n+1} + b_q u_q^{n+1} + a_q u_{q-1}^{n+1} = f_q^n\,,$$

$$c_q = \frac{\Delta z}{4iB}\frac{1}{\tilde r_j}\frac{\tilde r_{j+0.5}}{(\tilde r_{j+0.5} - \tilde r_{j-0.5})(\tilde r_{j+1} - \tilde r_j)}\,,$$

$$a_q = \frac{\Delta z}{4iB}\frac{1}{\tilde r_j}\frac{\tilde r_{j-0.5}}{(\tilde r_{j+0.5} - \tilde r_{j-0.5})(\tilde r_j - \tilde r_{j-1})}\,,$$

$$b_q = (1 - a_q - c_q) - \frac{\Delta z}{iB}\frac{m^2}{\tilde r_j^{\,2}}\,,$$

$$f_q^n = -c_q u_{q+1}^n + \left(2 - b_q\right)u_q^n - a_q u_{q-1}^n - 2\Delta z_f a^{(1)}(q)\,.$$

We supply this scheme with the discrete boundary condition at r_{PML}

$$a_{n_r+n_{PML}} = c_{n_r+n_{PML}} = f^n_{n_r+n_{PML}} = 0;\ b_{n_r+n_{PML}} = 1; \qquad m \ge 0\,,$$

and a discrete boundary condition at $r = 0$. The last condition for the monopole mode reduces to

$$c_0 = \frac{\Delta z}{iB}\frac{1}{\Delta r^2}\,,\ a_0 = 0\,,\ b_0 = 1 - c_0\,,$$

$$f_0^n = -c_0 u_1^n + \left(2 - b_0\right)u_0^n - 2\Delta z a^{(1)}(0)\,,$$

and for higher order modes it reads

$$a_0 = c_0 = f_0^n = 0;\ b_0 = 1; \qquad m > 0\,.$$

Figure 2 sketches the used transverse mesh. The black points present the location of sample points for the field.

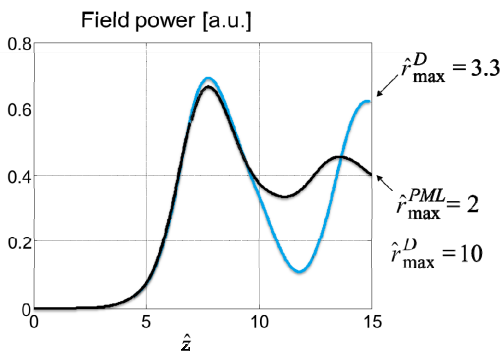


Figure 3: The radiation power with and without PML.

In FEL simulations the Dirichlet boundary condition works satisfactory in the exponential growth regime (linear regime), but it could spoil the correct solution after the saturation (non-linear regime). Figure 3 shows a 3D simulation for a round beam with radius $\hat r_b = 1$. It can be seen that for the Direchlet condition the mesh should be truncated very far from the beam (at $\hat r_0 = 10$). On the contrary, the quite thin perfectly matched layer (only 7 mesh points) allows to truncate the mesh accurately already at radius $\hat r_0 = 2$.

Time-dependent Simulations

Let us consider the time-dependent equation along the trajectory $t = t_0 + z/c$

$$\left[\frac{\nabla_\perp^2}{2ik} + \frac{d}{dz}\right]\tilde E\left(z, t_0 + \frac{z}{c}\right) = F\left(z, t_0 + \frac{z}{c}\right).$$

We divide the bunch in N_S slices of the length Δs and numerate them from the tail by index j. The slice with index j cross $z = 0$ at the moment of time

$$t_{j0} = -\frac{j\Delta s}{c}$$

and it will reach the position z at the time

$$t_j(z) = -\frac{j\Delta s}{c} + \frac{z}{v_z}\,.$$

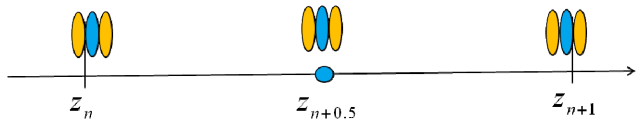


Figure 4: Time-dependent case.

The derivative in z can be approximated as

$$\frac{d}{dz}\tilde E\left(z_{n+0.5}, t_j(z_{n+0.5})\right) =$$

$$= \frac{\tilde E\left(z_{n+1}, t_j(z_{n+0.5}) + \dfrac{0.5\Delta z}{c}\right) - \tilde E\left(z_n, t_j(z_{n+0.5}) - \dfrac{0.5\Delta z}{c}\right)}{\Delta z} + O(2)$$

Let us consider Figure 4 and derive the following relations

$$t_j(z_{n+0.5}) + \frac{\Delta z}{2c} = -\frac{j\Delta s}{c} + \frac{z_{n+0.5}}{v} + \frac{\Delta z}{2c} =$$

$$= -\frac{(j+0.5)\lambda}{c} + \frac{z_{n+1}}{v} = t_{j+0.5}(z_{n+1}),$$

$$t_j(z_{n+0.5}) - \frac{\Delta z}{2c} = t_{j-0.5}(z_n)\,.$$

Hence we can write

$$\frac{\nabla_\perp^2}{2ik}\left(\frac{\tilde E_{n+1}^j + \tilde E_n^{j-1}}{2}\right) + \frac{\tilde E_{n+1}^j - \tilde E_n^{j-1}}{\Delta z} = F(z_{n+0.5}, t_j(z_{n+0.5}))\,,$$

where

$$\tilde E_n^j = \tilde E\left(z_n, t_{j+0.5}(z_n)\right)$$

means the EM field of slice j at position z_n.

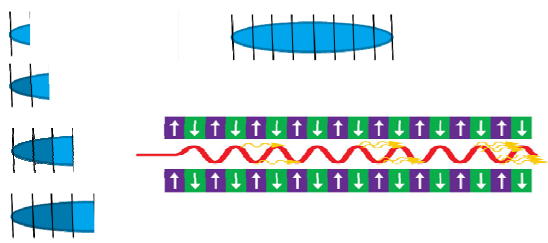


Figure 5: Parallelization.

It follows from the last equation that the field of slice j at position z_n can be found easily if we know the field of the previous slice $j-1$ at the previous position z_{n-1}. Hence the above described algorithm can be parallelized as sketched in Figure 5. We start from the last slice and track the particles of this slice through the whole undulator; the radiated EM field is saved. Then we track the next slice in the radiation field of the previous slice and so on. The three codes listed in Table 1 are parallelized and show the linear scaling of performance in proportion to the number of processes used.

NUMERICAL EXAMPLE

Expected Radiation in the FLASH with 3rd Harmonic Module

In order to linearize the energy chirp before the first bunch compressor the third harmonic module is installed in FLASH at the end of 2009.

To find working points, to define the tolerances and to characterize the parameters of the bunch at the undulator entrance we have done series of "start-to-end" simulations for different bunch charges [15].

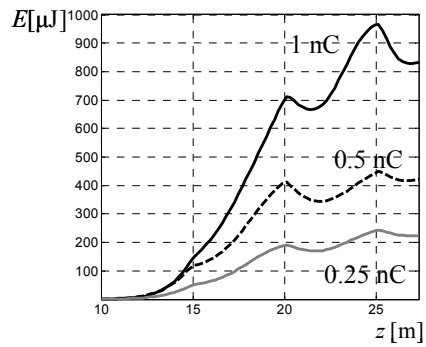


Figure 6: Radiation energy at FLASH.

Figure 6 presents the evolution along the undulator of the radiation energy in SASE mode for different bunch charges. The full set of plots for different characteristics of the obtained radiation can be found in [16].

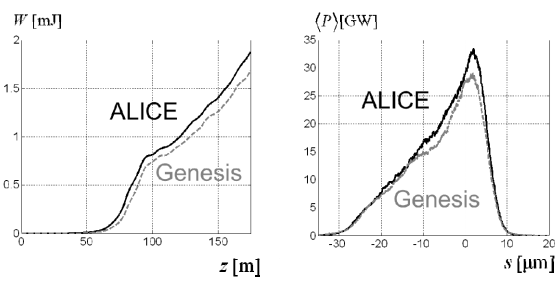


Figure 7: Radiation energy at the European XFEL.

Figure 7 presents comparison of two codes for the set of parameters of the European XFEL [16] as well.

OUTLOOK

The approximations used in the described FEL model allow studying of many features of the physical processes in undulators. The existing FEL codes are benchmarked by comparison with experiments [17] and with analytical results [7]. However for certain problems such as space charge dominated beams, ultra short electron beam pulses, high diffraction cases the used approximations begin to fail and numerical solution of the original set of Maxwell-Vlasov set of equations is required. There are several codes [18, 19] where non-averaged equations of motion are used, but the field is found from the paraxial approximation in the same manner as described in this paper. Abandoning of the paraxial approximation (slowly-varying envelope approximation) will result in necessity to use quite time consuming models as, for, example PIC model. A promising approach for the PIC simulations could be the approach based on Lorentz-boosted frame transformation [20, 21].

REFERENCES

[1] T.M. Tran, J.S.Wurtele, Phys. Reports 195 (1990) 1
[2] E.L.Saldin et al, NIM A 429 (1999) 233.
[3] S. Reiche, NIM A 429 (1999) 243.
[4] W.M. Fawley, LBNL CBP Tech Note-104 (1995).
[5] I.Zagorodnov, M.Dohlus, Numerical Studies with a New Code ALICE, FEL 2009, Malmö, 2009, MOPC16.
[6] S.Reiche, FEL Simulations: History, Status and Outlook, FEL 2009, Malmö, 2009, MOOCI1.
[7] E.L. Saldin et al, The Physics of Free Electron Lasers, Springer (1999).
[8] J.M.Dawson, Rev. of Modern Physics 55 (1983) 403.
[9] W.H.Press et al, Numerical Recipes in C, Cambridge University Press (1988).
[10] The European X-Ray Free Electron Laser TDR, DESY 2006-097 (2006).
[11] I.Zagorodnov, Genesis/ALICE benchmarking (2008), http://www.desy.de/xfel-beam/talks_a.html#
[12] W.M. Fawley, Phys. Rev. STAB 5 (2002) 070701.
[13] B.W.J. McNeil et al, PR STAB 6, (2003) 070701.
[14] F. Collino, J. Comp. Phys. 131 (1997) 164.
[15] Zagorodnov I., Dohlus M., PR STAB 14 (2010) 014403.
[16] Zagorodnov I., Ultra-short low charge operation at FLASH and the European XFEL, FEL 2010, Malmö, Sweden, 2010, WEOBI2.
[17] L. Giannesi et al, PR STAB 14 (2011) 060712.
[18] H.P. Freund et al, IEEE J.Quantum Electron. 36 (2000) 275.
[19] C.K.W.Nam et al, Unaveraged Three-Dimensional Modelling of the FEL, FEL 2008, Gyeogju, Korea, MOPPH025.
[20] J.-L. Vay, PRL 98 (2007) 130405.
[21] W.M.Fawley, J.-L-Vay, Full Electromagnetic FEL Simulation via the Lorentz-Boosted Frame Transformation, FEL 2010, Malmö, Sweden, 2010, MOPB01.

WAVE - A COMPUTER CODE FOR THE TRACKING OF ELECTRONS THROUGH MAGNETIC FIELDS AND THE CALCULATION OF SPONTANEOUS SYNCHROTRON RADIATION

M. Scheer*, Helmholtz-Zentrum Berlin für Materialien und Energie, Berlin, Germany

Abstract

WAVE is developed at BESSY - now Helmholtz-Zentrum Berlin (HZB) - since 1990 to calculate spontaneous synchrotron radiation for arbitrary magnetic fields. A variety of field models for dipoles, wavelength shifters, and undulators is available. Field maps can be read and written. Many routines to handle magnetic fields are implemented, including simulations of field error e.g. due to misalignment. Coherent radiation of electrons in a bunch and energy losses due to radiation are taken into account. Phase-space distributions of the beam electrons are taken into account by various algorithms. Generating functions and linear transfer matrices for particle tracking purposes can be calculated. Subroutines to calculate the effects of insertion devices on the storage ring are included. The program runs in batch mode, controlled by input files, but a simple GUI is also provided. The results are given as ASCII data or binary formats of the programs PAW, ROOT, and HDF5. Parallel runs of WAVE on a cluster are supported. WAVE has been checked and validated with the synchrotron radiation code of the German National Bureau of Standards (PTB) based on Schwinger's formula.

INTRODUCTION

The development of WAVE started in 1990 within the framework of a study for the PTB. Subject of this study was the question, whether a superconducting wave-length shifter (WLS) can be used at the planned storage ring BESSY II for radiometry in the X-ray regime. The code was developed as a tool to calculate the synchrotron radiation of a WLS with very high precision and also to investigate the impact of the WLS on the storage ring with respect to emittance change, beam polarization and dynamical aperture. This study has become the starting point for the PhD thesis of the author [1], where more details are given. From 1994 the calculation of undulator radiation become more important and many routines to model and handle magnetic fields of undulators has been implemented. WAVE is written in FORTRAN 90 and can be compiled and used on a variety of platforms and operating systems. However, nowadays it is developed mainly for Linux.

* Michael.Scheer@Helmholtz-Berlin.de

MAGNETIC FIELDS

The periodic magnetic fields of undulators can be approximately described by analytical formulas. Although these formulas are available in WAVE, many problems, e.g. the calculations of effects of field errors or endpole configurations, require more sophisticated methods. Complex setups like helical undulators coupled by a modulator build of rotatable, cylindrical magnets [2] are treated in WAVE as an arrangement of permanent magnets. Undulator parameters as gap, shift, dimensions and magnetization of the magnets as well as taper, misalignments, and field errors can be set by the user. The magnets can be read from an external file or defined as complete devices in the input file of WAVE. The fields of the individual magnets are calculated using the current sheet method. Another very powerful method is to model the fields by a Fourier series of analytical 3D functions [1, 3]. This includes algorithms to expand the vertical field given on the longitudinal axis of an insertion device to a Maxwell conform 2D or 3D field. However, sometimes the usage of field maps is more flexible or necessary. WAVE can handle 3D field maps with different interpolation schemes. Another group of features include symmetry operations and options to make fields periodic along the longitudinal device axis. For applications not covered by WAVE, the user can also provide his own magnetic field routine.

TRACKING

The tracking algorithm of WAVE considers the magnetic field as constant for each time step. Thus, the trajectory consists of pieces of circles with orientation and bending radii according to the local field. An option allows to calculate the energy loss due to the synchrotron radiation for each tracking step and to adapt the energy of the electron accordingly. This can be done continuously or by a Monte-Carlo technique which generates discrete photons in order to take quantum fluctuations into account. For special purposes, the user may provide an own routine called for each tracking step.

SYNCHROTRON RADIATION

The spectrum calculation is mainly a numerical integration of well-known formulas given e.g. in [4]. WAVE uses a modified version:

$$\int_0^{\Delta T} \frac{1}{R(t)} \frac{\vec{n}(t)\times[(\vec{n}(t)-\vec{\beta}(t))\times\dot{\vec{\beta}}(t)]}{(1-\vec{\beta}(t)\vec{n}(t))^2} e^{i\omega(t+R(t)/c)} dt \approx$$

$$\sum_{j=1}^{N} \frac{1}{R(t_j)} \frac{\vec{n}(t_j)\times[(\vec{n}(t_j)-\vec{\beta}(t_j))\times\dot{\vec{\beta}}(t_j)]}{(1-\vec{\beta}(t_j)\vec{n}(t_j))^2} e^{i\omega(t_j+R(t_j)/c)} \times$$

$$\frac{1-e^{i\omega(1-\vec{\beta}(t_j)\vec{n}(t_j))\Delta t_j}}{(1-\vec{\beta}(t_j)\vec{n}(t_j))\omega} \quad (1)$$

Here, $\vec{R} = R \times \vec{n}$ denotes the vector from the electron to the observer, $\vec{\beta}$ the velocity divided by c, and ω the frequency of the radiation. For the numerical integration over the time-interval ΔT, the integral of eq. 1 is approximated by a sum of N time steps. The phase is expanded up to first order, while the ω-independent factor is treated as constant within the time step. For the step the integration is then done analytically. This procedure ensures a proper integration for very rapidly varying phases with many oscillations within a step as it occurs for wigglers or dipoles. The accuracy is further increase by a recursive calculation of the phase:

$$e^{i\omega(t_{j+1}+R(t_{j+1})/c)} \approx e^{i\omega(t_j+R(t_j)/c)} \times e^{i\omega(1-\vec{\beta}(t_j)\vec{n}(t_j))\Delta t_j} \quad (2)$$

The integration yields the field amplitude for a given observation point. From this amplitude the spectral flux-density and the brilliance is calculated in terms of Stokes parameters. In addition, a similar integral for the velocity term of the field is evaluated as well. The irradiated total power and the power passing a given aperture are also determined. To calculate the flux through a rectangular or circular aperture, the calculations are carried out for each point on a rectangular or radial grid and integrated over the grid points via spline techniques. For a proper calculation of the brilliance, the size of the source is needed. It can be determined by WAVE, via a backward propagation of the field amplitudes from the grid of observation points to the device center making use of Huygens's principle.

WAVE offers two main modes for the treatment of insertion devices. In the wiggler mode all poles are treated as separate sources and added incoherently, while in the undulator mode the contributions of all steps of the trajectory are summed up coherently. Beside the normal undulator mode, a quick but less accurate mode is available, where the complex field amplitude of only one period is calculated and added up N times with the appropriate phase-shift and $1/R$-correction for N periods of the undulator. For the wiggler mode, the synchrotron radiation can also be calculated making use of Schwinger's formula [5], which is very fast and precise.

To take the emittance and energy spread of the electron beam into account two very different approaches are available. For a single electron folding procedures are applied. Alternatively, many electrons can be tracked starting from a Gaussian or user defined phase-space distribution. The radiation of all electrons of a bunch is added coherently, while the contributions of different bunches are summed up incoherently. This allows to calculate coherent radiation of bunched electrons.

If the flux through an aperture has to be calculated for a large number of electrons, a Monte-Carlo mode is available. In this mode not only the phase-space variables of the electrons are randomly generated according to their distribution, but also the coordinates of the observation point are randomly chosen within the considered aperture. Thus the flux through the aperture is calculated as fast as the flux-density for a single observation point. However, to track thousands of electrons can still be very time consuming. Therefore, WAVE can run in parallel on a cluster.

Many other features like dose rate calculations or application of filters and yield functions are available but beyond the scope of this paper.

ACCELERATOR PHYSICS

From its history WAVE was written not only for the calculation of synchrotron radiation, but also to investigate the influence of insertion devices on the storage ring. By tracking many electron through a given magnetic field, WAVE can fit the coefficients of a Taylor series of the Generating Function F_2 from the phase-space variables at the entrance and exit of the device [6, 1]. This Generating Function (GF) can be implemented in tracking codes to track a particle with a single step through a whole device. From the linear terms of the GF, the linear transfer-matrix (TM) is derived. Another method implemented to determine the TM, is to track principle trajectories and to solve a system of linear equations. As explained in details in [6], also the emittance change of the storage ring due to e.g. a superconducting wavelength-shifter as well as the degree and build-up time of the beam polarization can be estimated from the tracking through the insertion device and the parameters of the storage ring. WAVE can also calculate the beta functions of magnet configurations and determine the periodic solution, which is interesting for very long setups with many undulators and focusing elements in between.

INTERFACES

Two different kinds of interfaces are available: User defined subroutines and output data for external codes. The interface routines are provided by the user to be compiled and linked to WAVE. These interfaces exist for the magnetic field routine, the input routine and the output routine for WAVE as well as for the tracking routine on step level. Output interfaces exist for the ray-tracing and field propagation codes RAY [7] and PHASE [8] to make use of sources calculated by WAVE.

EXAMPLE

WAVE has become a standard tool for the calculation of synchrotron radiation at BESSY (now HZB), and the results are confirmed by a variety of measurements. However, with the FEL, new challenges came up with the need of new features to calculate the spontaneous synchrotron radiation of very long sections of undulators including focusing elements. These includes energy-loss of the electrons, quantum fluctuations, and tracking of many electrons with a realistic phase-space distribution. As an example, the following figure shows spectra of a 215.25 meter long straight section build of 35 cells. Each cell consists of an undulator with a period-length of 4 cm and 125 periods, phase-shifters, and quadruples. WAVE offers an option to find the periodic beta-functions for the cells. It is used here to tune the quadrupoles such that the maxima and minima of the horizontal and vertical beta-functions are 23 and 9.5 meter.

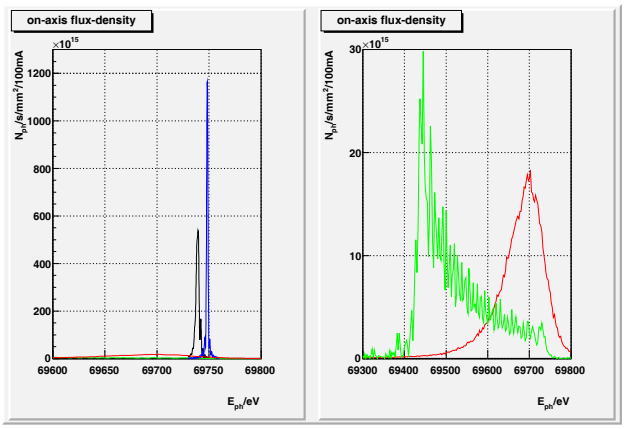


Figure 1: Spectra of the 9th harmonic of 35 undulator sections: Blue: Pencil beam, no energy loss, no taper. Green: Pencil beam, continuous energy loss, no taper. Black: Pencil beam, continuous energy loss, with compensating taper. Red: Broad beam including all effects, i.e. energy loss with quantum fluctuations, beam emittance, energy spread, and taper.

The main problem is, that the harmonics of so many coupled undulators are so narrow, that a minor change of the parameters of the considered particle changes the harmonics dramatically. As a consequence, folding procedures to take the beam emittance and energy spread into account are not applicably. Thus, a representative sample of many electrons has to be tracked. About one million integration steps per electron are require making the calculations very time consuming. An analysis of the performance of WAVE has shown that the calculation of the magnetic field is a crucial point, depending on the field model used. To speed up calculations, the complex magnetic field of the undulators (current sheet method), of the quadrupoles, and of the phase-shifters is written to a 3D field map in a first WAVE run. This takes some minutes, and the size of the map file for this examples is 235 MBytes. In the following runs, the spectra are calculated using this field map. This is much faster, and allows to calculate the spectrum of 1000 electrons for 201 photon energies within 5 hours on a conventional PC with a single processor.

Figure 1 shows the importance of the new features of WAVE, especially the treatment of the energy loss during the passage of the electrons through the undulators and the capability of tracking many electrons. The blue curve shows the pure on-axis spectrum of an electron without energy loss. The green curve shows the same situation, but with a continuous energy loss. Due to this effect, the 9th harmonic is completely spoiled. The black curve shows the effect with a continuous energy loss but compensated by a taper. The tapering is done such that the field of each undulator is scaled to keep the position of the harmonic. This restores the harmonic to about half of the original peak value. However, when all effects are included, the harmonic goes down by a factor of about 60 (red curve) compared to the pencil beam case without energy loss. This examples illustrates that WAVE is ready (although not yet cross-checked) for the calculation of spontaneous synchrotron radiation of FEL beamlines.

REFERENCES

[1] M. Scheer, "Beschleunigerphysik und radiometrische Eigenschaften supraleitender Wellenlängenschieber", Dissertation, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät I, 2008, urn:nbn:de:kobv:11-10096299

[2] J. Bahrdt, W. Frentrup, A. Gaupp, M. Scheer, W. Gudat, G. Ingold, S. Sasaki, "Elliptically polarizing insertion devices at BESSY II", NIM A 467-468 (2001) 21-29

[3] J. Bahrdt, W. Wüstefeld, "Symplectic Tracking and Compensation of Dynamic Field Integrals in Complex Undulator Structures", Proceedings of IPAC2012, New Orleans, Louisiana, USA, FRXAB01, p. 4165 (2012), http://www.JACoW.org

[4] J.D. Jackson, Classical Electrodynamics, Second Edition, (New York: Wiley, 1975), ISBN 0-471-43132-X

[5] J. Schwinger, On the Classical Radiation of Accelerated Electrons, Phys. Rev. 75 (1949), p. 1912-1925

[6] M. Scheer, G. Wüstefeld, Part I: "Canonical Tracking Calculations" , Part II: "Emittance and Beam Polarization Effects", Technischer Bericht, TB 169, BESSY mbH (now Helmholtz-Zentrum Berlin), Berlin, 1992

[7] F. Schäfers, "Ray - the BESSY raytrace program", In: Springer Series in Modern Optical Sciences: Modern Developments in X-Ray and Neutron Optics, eds. A. Erko, M. Idir, Th. Krist, A.G. Michette, Springer Berlin/Heidelberg, Vol. 137, 9-41 (2008)

[8] J. Bahrdt, U. Flechsig, S. Gerhardt, I. Schneider, "PHASE: a universal software package for the propagation of time-dependent coherent light pulses along grazing incidence optics", Proceedings of SPIE, San Diego, California, USA, 2011, Vol. 8141 p. 81410E-1 - 81410E-10

A FAST INTEGRATED GREEN FUNCTION METHOD FOR COMPUTING 1D CSR WAKEFIELDS INCLUDING UPSTREAM TRANSIENTS

C.E. Mitchell, Ji Qiang, and R.D. Ryne, Lawrence Berkeley National Laboratory, USA

Abstract

An efficient numerical method for computing wakefields due to coherent synchrotron radiation (CSR) has been implemented using a one-dimensional integrated Green function approach. The contribution from CSR that is generated upstream and propagates across one or more lattice elements before interacting with the bunch is included. This method does not require computing the derivative of the longitudinal charge density, and accurately includes the short-range behavior of the CSR interaction. As an application of this method, we examine the importance of upstream transient wakefields within several bending elements of a proposed Next Generation Light Source.

BACKGROUND

The accurate modeling of coherent synchrotron radiation is a numerical challenge of key importance to the development of future light sources [1]. Recent 3D simulations of the CSR generated by Gaussian bunches of various shapes [2] confirm that a 1D model of the longitudinal CSR wakefield is accurate provided that the transverse rms beam size $\sigma_\perp$ satisfies $\sigma_\perp << R(\sigma_z/R)^{2/3}$, where R is the bending radius and σ_z is the longitudinal rms beam size. A number of such 1D models appear in the literature [3]–[7], many of which have been implemented in existing beam dynamics codes.

In these models, the energy loss per unit length at a longitudinal location z within the bunch is given by:

$$W(z) = \int_{-\infty}^{z} \lambda(z') K_{CSR}(z, z') dz', \qquad (1)$$

where λ is the longitudinal number density of the bunch, and K_{CSR} is related to the longitudinal component of the Liénard-Wiechert field $\mathbf{E}^{LW}$ of a single particle [6]. The integration kernel K_{CSR} varies rapidly near $z \approx z'$ on the scale $R/\gamma^3 << \sigma_z$, and the integral (1) is therefore difficult to resolve numerically. This problem is typically avoided by computing the equivalent integral

$$W(z) = \int_{-\infty}^{z} \frac{d\lambda(z')}{dz'} I_{CSR}(z, z') dz', \qquad (2)$$

where

$$I_{CSR}(z, z') = -\int_{-\infty}^{z'} K_{CSR}(z, z'') dz''. \qquad (3)$$

The kernel I_{CSR} exhibits less singular behavior than K_{CSR} near $z \approx z'$. However, this method requires that one evaluate the numerical derivative of the longitudinal charge density, which in general contains significant numerical noise. It is also typical to approximate I_{CSR} by an asymptotic form. (See, for example, equation (19) of [4].)

In the following section, we describe an efficient method for evaluating (1) that makes use of the longitudinal charge density λ directly. The short-range behavior of the CSR kernel is treated analytically, so that only longitudinal variations in the charge density need to be resolved [8, 9]. In addition, this method can be used to treat the case of entry and exit transient fields [4], as well as the case of transient fields in a general lattice due to upstream elements [6].

INTEGRATED GREEN FUNCTION METHODS

Let $\lambda_j = \lambda(z_j)$, $j = 1, \ldots, N$ denote the values of the longitudinal number density of the bunch at a set of equidistant sample points z_j, $j = 1, \ldots, N$, and let $\{P_j : j = 1, \ldots, N\}$ denote a basis of piecewise polynomials of given degree. This basis can always be chosen such that $P_j(z_k) = \delta_{j,k}$. We can then write the interpolated longitudinal density $\lambda_{appx} \approx \lambda$ in the form

$$\lambda_{appx} = \sum_{j=1}^{N} \lambda_j P_j(z). \qquad (4)$$

Using (4) in (1) gives an approximate longitudinal wakefield $W_{approx} \approx W$ of the form

$$W_{approx}(z_k) = \sum_{k'=1}^{N} \lambda_{k'} w_{k,k'}, \quad \text{where} \qquad (5)$$

$$w_{k,k'} = \int_{-\infty}^{z_k} P_{k'}(z') K_{CSR}(z_k, z') dz'.$$

For the models described in [3]–[6], the weights $w_{k,k'}$ may be determined analytically in terms of rational functions, log, arctan, and polynomial roots. In particular, the quantities $w_{k,k'}$ have been determined explicitly for the 1D models described in [4, 6] in the cases of both piecewise constant and piecewise linear basis functions.

For example, a set of piecewise-constant basis functions is given, for stepsize h, by:

$$P_{k'}(z) = \begin{cases} 1, & z_{k'} - h/2 \le z \le z_{k'} + h/2, \\ 0, & \text{else} \end{cases} \qquad (6)$$

when $1 < k' < k$, with similar expressions for P_1 and P_k.

We then have for $1 \le k \le N$:

$$w_{k,k'} = I_{CSR}\left(z_k, z_{k'} - \frac{h}{2}\right) - I_{CSR}\left(z_k, z_{k'} + \frac{h}{2}\right), \qquad (7)$$

for $1 < k' < k$ and

$$w_{k,1} = I_{CSR}\left(z_k, z_1\right) - I_{CSR}\left(z_k, z_1 + \frac{h}{2}\right), \quad (8)$$

$$w_{k,k} = I_{CSR}\left(z_k, z_k - \frac{h}{2}\right) - I_{CSR}\left(z_k, z_k\right). \quad (9)$$

For the four cases considered in [4], the function I_{CSR} is given in Cases A–D, respectively, by the expressions:

$$\frac{RI_{CSR}}{\gamma r_c mc^2} = -\frac{2(\hat{\phi}+\hat{y}) + \hat{\phi}^3}{(\hat{\phi}+\hat{y})^2 + \hat{\phi}^4/4} + \frac{1}{\hat{s}}, \quad (10)$$

where $\hat{s} = \frac{\hat{\phi}+\hat{y}}{2} + \frac{\hat{\phi}^3}{24}\frac{\hat{\phi}+4\hat{y}}{\hat{\phi}+\hat{y}}$.

$$\frac{RI_{CSR}}{\gamma r_c mc^2} = -\frac{4\hat{u}(\hat{u}^2+8)}{(\hat{u}^2+4)(\hat{u}^2+12)}, \quad (11)$$

where $\hat{s} = \frac{\hat{u}^3}{24} + \frac{\hat{u}}{2}$.

$$\frac{RI_{CSR}}{\gamma r_c mc^2} = -\frac{2(\hat{\phi}_m + \hat{x} + \hat{y} + \hat{\phi}_m^3/2 + \hat{\phi}_m^2\hat{x})}{(\hat{x}+\hat{y}+\hat{\phi}_m)^2 + (\hat{\phi}\hat{x} + \hat{\phi}_m^2/2)^2} + \frac{1}{\hat{s}}, \quad (12)$$

where $\hat{s} = \frac{\hat{\phi}+\hat{x}+\hat{y}}{2} + \frac{\hat{\phi}_m^2}{24}\frac{\hat{\phi}_m^2 + 4\hat{\phi}_m(\hat{x}+\hat{y}) + 12\hat{x}\hat{y}}{\hat{\phi}+\hat{x}+\hat{y}}$.

$$\frac{RI_{CSR}}{\gamma r_c mc^2} = -\frac{2(\hat{\psi}+\hat{x}+\hat{\psi}^3/2 + \hat{\psi}^2\hat{x})}{(\hat{x}+\hat{\psi})^2 + (\hat{\psi}\hat{x} + \hat{\psi}^2/2)^2} + \frac{1}{\hat{s}}, \quad (13)$$

where $\hat{s} = \frac{\hat{\psi}+\hat{x}}{2} + \frac{\hat{\psi}^2}{24}\frac{\hat{\psi}^2 + 4\hat{x}\hat{\psi}}{\hat{\psi}+\hat{x}}$.

The notation appearing on the right-hand side in (10–13) is the same as the notation used in [4].

It is possible to express the sum (5) in the form of a discrete convolution, which can then be evaluated using an FFT in $O(N \log N)$ operations. This technique has been implemented in the code IMPACT. A routine using the more general expression for I_{CSR} provided in [6] has also been developed, and may also be used when we wish to include transient effects due to CSR from upstream lattice elements.

NUMERICAL ANALYSIS

Several methods for computing the longitudinal wake integral (1) were compared against the integrated Green function method described in the previous section. These include:

- Direct evaluation of (1) using an extended trapezoidal rule.
- Evaluation using an integrated Green function method with piecewise constant basis functions.
- Evaluation using an integrated Green function method with piecewise linear basis functions.
- Evaluation of (2) using an extended trapezoidal rule, when $d\lambda/dz$ is exactly known.
- Evaluation of (2) using an extended trapezoidal rule, when $d\lambda/dz$ is approximated using a central difference formula.

These methods share the property that the local error on each subinterval scales as $O(h^3)$. Figure 1 illustrates the relative error in the value of the CSR wake computed at the centroid of a Gaussian bunch as a function of stepsize, for the five algorithms above. Note that the convergence behavior of the direct integration method changes near $h = R/\gamma^3$ (indicated by the vertical dashed line). When this scale is not resolved, this method has an error that is a factor of $10^5 - 10^6$ worse than the other methods considered here. The figure below shows the error in the region $h \geq R/\gamma^3$, indicating that the IGF method with a piecewise constant basis results in the smallest error over a large range of stepsizes.

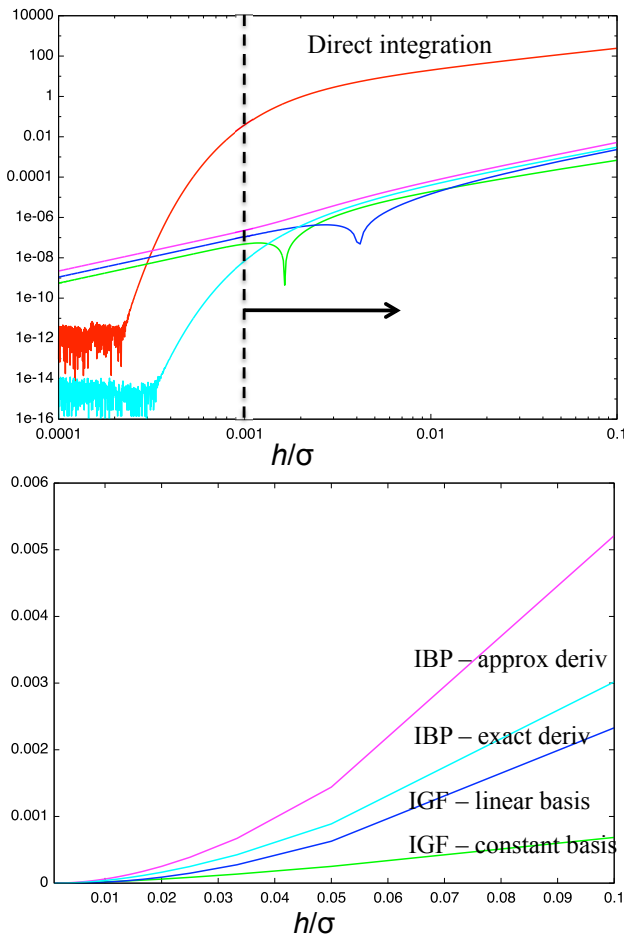


Figure 1: (Upper) Relative error in the computed longitudinal CSR wake at $z = 0$ for a Gaussian bunch with $\sigma = 0.1$ mm at 200 MeV in a bend with $R = 10$ m. The result is shown for 5 distinct quadrature formulas as a function of stepsize (log-log scale). (Lower) The same quantities shown on a linear scale for $h \geq R/\gamma^3$.

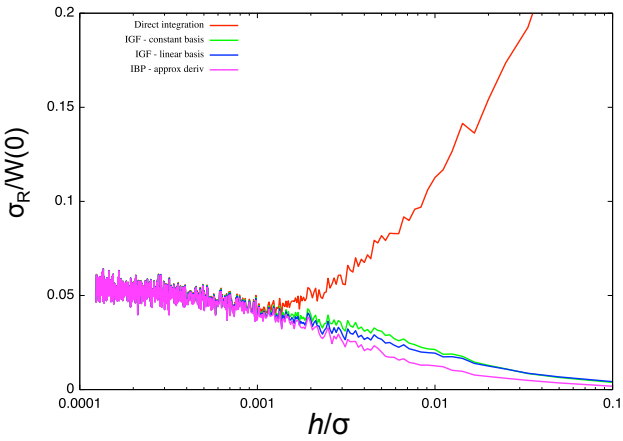


Figure 2: Sensitivity to noise in the CSR wake at $z = 0$ for a Gaussian bunch for 5 distinct quadrature formulas, shown as a function of stepsize (log-linear scale). Colors have the same meaning as in the previous figure. The shot noise (14) at a given step size h is determined by the two parameters $N = 20\sigma/h$ and $N_p = 10^6$.

The values λ_j are in general obtained by projecting a distribution of N_p macroparticles into N longitudinal bins, therefore introducing shot noise to the longitudinal density profile of the beam. We model this noise by setting

$$\lambda_j^r = \lambda_j(1 + \epsilon_j), \quad j = 1, \ldots, N, \tag{14}$$

where each ϵ_j is a normal random variable with $\langle \epsilon_j \rangle = 0$ and $\langle \epsilon_j^2 \rangle = N/N_p$. The CSR wake may now be computed using either the smooth values λ_j or the noisy values λ_j^r. We refer to the difference between these two results as the *sensitivity to noise*. The rms value of the sensitivity to noise at the centroid of a Gaussian bunch is shown in Fig. 2 for the five algorithms described above. Each value was obtained by averaging over 100 distinct random seeds. Note that the direct integration method (red) is very sensitive to the presence of noise, while the other methods exhibit sensitivities that are comparable to one another.

APPLICATION TO NGLS

The integrated Green function method of (5) was used to study longitudinal CSR wakefields in the spreader section of a proposed Next Generation Light Source [1]. The 2.4 GeV spreader section directs the beam bunches into several parallel FEL lines using a series of six 10° dipoles. The CSR wake was computed inside Bend 6 (Fig. 3) using the model of [6] with piecewise constant basis functions. Figure 4 illustrates an analytical model of the current pulse at the spreader entry, together with the upstream contribution of Bends 4 and 5 to the longitudinal wakefield at several locations within Bend 6. After computing the net energy kick experienced by particles in the bunch as they propagate through the spreader dipoles, the CSR from upstream bends was found to contribute 12% of the total energy kick in the spreader.

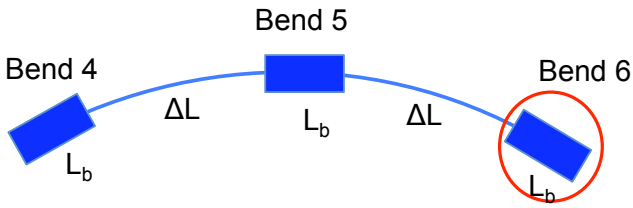


Figure 3: The final 3 dipoles appearing in the spreader section of a Next Generation Light Source. The longitudinal CSR wake was computed in Bend 6 both with and without the effects of radiation from Bends 4 and 5 included.

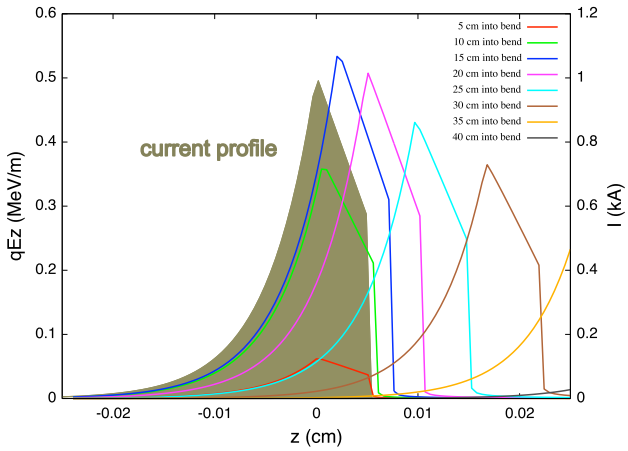


Figure 4: (Brown) Model of the current pulse entering the NGLS spreader. (Curves) The difference between the longitudinal CSR wakefield in Bend 6 when computed with and without the effects of the upstream bends 4 and 5, shown at distances of 5–40 cm into the bend.

CONCLUSIONS

A one-dimensional integrated Green function method using both a piecewise constant and a piecewise linear basis has been implemented to model longitudinal CSR wakefields using the models described in [4, 6]. This method does not require computing the derivative of the longitudinal charge density, and accurately includes the short-range behavior of the CSR interaction. The error is smaller than other methods which have a comparable rate of convergence. As an application of this method, we studied upstream transient wakefields within the spreader section of a proposed Next Generation Light Source.

ACKNOWLEDGEMENT

This work is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

REFERENCES

[1] John Corlett et al., "Next Generation Light Source R& D and Design Studies at LBNL", to appear in Proc. of IPAC 2012.

[2] R.D. Ryne et al, Proc. IPAC 2012, New Orleans, Lousiana, 1689 (2012).

[3] J. Murphy et al, Particle Accelerators, **57**, pp. 9–64 (1997).

[4] E.L. Saldin et al., Nucl. Instrum. Methods Phys. Res. A **398**, 373 (1997).

[5] G. Stupakov and P. Emma, Proc. EPAC 2002, Paris, France, 1479 (2002).

[6] D. Sagan et al., Phys. Rev. ST - Accel. Beams **12**, 040703 (2009).

[7] C. Mayes and G. Hoffstaetter, Phys. Rev. ST - Accel. Beams **12**, 024401 (2009).

[8] R.D. Ryne et al., `http://arxiv.org/abs/1202.2409` (2012).

[9] Ji Qiang et al., Nucl. Instrum. Methods Phys. Res. A **682**, 49–53 (2012).

THE CONVERGENCE AND ACCURACY OF THE MATRIX FORMALISM APPROXIMATION*

S. Andrianov†, SPbSU, St. Petersburg, Russia

Abstract

To the present time there has been developed a large number of different codes for the particles beam dynamics modeling. However, their precision, accuracy and reliability of the numerical results are not sufficiently guaranteed in the case of long-term evolution of particle beams in circular accelerators. Here we discuss convergence estimates of the matrix presentation for Lie series. We also consider some problems of the matrix formalism accuracy for constructing the evolution operator of the particle beam. In this article there is paid a special attention to problems of symplecticity and energy conservation for long time evolution of particles beams.

INTRODUCTION

The well known Lie methods for nonlinear dynamics allow us to constructive evaluate the corresponding maps and use them for accelerator physics [1]. But it should be note that evaluation procedures of corresponding Lie maps is enough time-consuming process. Besides, it is very difficult to evaluate corresponding procedures and estimate the required accuracy of the corresponding evaluations. This procedure is necessary for correct calculation for different problems of accelerator physics, particularly for long time beam evolution. Unfortunately the most popular methods have not enough practical instruments for accuracy estimates. The trend of accelerator physics leads us to necessity to have a tool to assess not only the accuracy of computational procedures, but also to preserve certain qualitative properties of the computational procedures (such as symplecticity, energy conservation and so on). Usually these problems are solved only in numerical mode up to some order of integration steps $\mathcal{O}(h^n)$. In the present paper we demonstrate some analytical estimates for the corresponding solutions using the matrix formalism for Lie approach [2, 3]. We also consider some problems of the matrix formalism accuracy for constructing the evolution operator of the particle beam.

THE ACCURATE EVALUATION OF TRUNCATED LIE MAPS

Matrix Series Presentations for Lie Maps

In some previous papers we presented the matrix formalism for Lie maps generated by ordinary differential equations, which can be written in the the following form

$$\frac{d\mathbf{X}}{dt} = \mathbf{F}(\mathbf{X}, t) = \sum_{k=1}^{\infty} \mathbb{P}_k(t)\mathbf{X}^{[k]}, \quad \mathbf{X}_0 = \mathbf{X}(t_0), \qquad (1)$$

where $\mathbf{F} = \{F_1, \ldots, F_{2n}\}^{\mathrm{T}}$ and $\mathbf{X}^{[k]}$ is a vector of all phase moments for the phase vector $\mathbf{X}$ ($\dim \mathbf{X} = 2n$), for example, $\mathbf{X} = \{x, p_x, y, p_y\}^{\mathrm{T}}$ and $\mathbb{P}^{1k}$ $\left(\text{with matrix size } \dim \mathbb{P}^{1k} = \binom{2n+k-1}{k}\right)$ are matrices containing partial derivatives

$$\left\{\mathbb{P}^{1k}(t)\right\}_{ij} = \frac{1}{k_1! \ldots k_{2n}!} \left.\frac{\partial^k F_i(x_j, t)}{\partial x_1^{k_1} \ldots \partial x_n^{k_{2n}}}\right|_{x_1=\ldots=x_{2n}=0}.$$

The eq. (1) generates the Lie operator [1] in according the following equality

$$\mathcal{L}_{\mathbf{F}} = \mathbf{F}^*(\mathbf{X}, t)\frac{\partial}{\partial \mathbf{X}} = \sum_{k=1}^{\infty} \left(\mathbf{X}^{[k]}\right)^{\mathrm{T}} \mathbb{P}_{1k}^{\mathrm{T}} \frac{\partial}{\partial \mathbf{X}} = \sum_{k=1}^{\infty} \mathcal{L}_{\mathbf{F}_k}.$$

So, for corresponding Lie map [1] we can write

$$\mathcal{M}\left(t|t_0; \mathcal{L}_{\mathbf{F}}\right) = \mathrm{T}\exp \int_{t_0}^{t} \mathcal{L}_{\mathbf{F}}(\mathbf{X}, \tau)d\tau =$$

$$= \mathrm{T}\exp \sum_{k=1}^{\infty} \int_{t_0}^{t} \mathcal{L}_{\mathbf{F}_k} d\tau = \mathcal{M}\left(t|t_0; \sum_{t_0}^{t} \mathcal{L}_{\mathbf{F}_k}\right) =$$

$$= \prod_{k=1}^{\infty} \mathcal{M}\left(t|t_0; \sum_{t_0}^{t} \mathcal{L}_{\mathbf{G}_k}\right), \qquad (2)$$

where the symbol "T" denotes the so called chronological ordering exponent Lie operator [3] (or the Dyson operator). Here k indicates the order of corresponding homogeneous polynomials. Evaluation of (2) can be realized in the frame of two following approaches:

1. The Magnuss representation for chronological exponent operators [3, 4].

2. The Zassenhaus formula for homogeneous polynomials $\mathbf{G}_k$, $k \geq 1$ calculation). Using these approaches we can write some symbolic formulas for Lie map $\mathcal{M}$ evaluation. For example, if we introduce the following notations [3] $\mathbb{P}_m^{k\,1} = \prod_{j=1}^{k} \mathbb{G}_m^{\oplus((j-1)(m-1)+1)}$ (here $\oplus$ denotes the Kronecker sum), then we have the following equality

$$\exp\left(\mathcal{L}_{\mathbf{G}_m}\right) \circ \mathbf{X} = \mathbf{X} + \sum_{k=1}^{\infty} \frac{\mathbb{P}_m^{k\,1}}{k!} \mathbf{X}^{[k(m-1)+1]}.$$

* Work supported by Federal Targeted Program "Scientific and Scientific-Pedagogical Personnel of the Innovative Russia in 2009-2013" (Governmental Contract no. p 793)

† sandrianov@yandex.ru

Similar formulae allow us to write the following formulae

$$\mathcal{M}_{\le 3}\circ\mathbf{X}=\mathbb{R}^{11}\left(\mathbf{X}+\sum_{m=2}^{3}\sum_{k=1}^{\infty}\frac{\mathbb{P}_m^{k1}}{k!}\mathbf{X}^{[k(m-1)+1]}+\right.$$
$$\left.+\sum_{l=1}^{\infty}\sum_{k=1}^{\infty}\frac{1}{k!l!}\mathbb{P}_2^{kl}\,\mathbb{P}_3^{l\,(k+1)}\,\mathbf{X}^{[2l+k+1]}\right),\quad(3)$$

where $\mathbb{R}^{11}$ is a fundamental matrix for linear approximation for motion equations. The formulae similar to (3) allow us to evaluate the necessary maps constructively and use them up to required order. Here we should mention additionally that the corresponding matrices $\mathbb{P}_m^{k1}$ in symbolic or numerical modes up to necessary order on phase vector $\mathbf{X}$.

The formulae similar to (3) allow us to evaluate up to necessary order corresponding matrix expansion for Lie maps. But for many beam physics problems it is very important to conserve some properties, for example the symplecticity property for hamiltonian systems. The symplecticity property imposes special conditions on the corresponding matrices, which have the form of linear algebraic equations [3]. It is necessary to emphasize that these restrictions of the corresponding matrices can not influence on the accuracy estimations.

The Problems of Accuracy Evaluations

The Lie Map. The above pointed methods of Lie maps evaluation will guarantee necessary convergence. As an example, we can give some estimates for convergence of Lie map evaluation in accordance with Magnus presentation of Lie map. According to this approach the chronological presentation (2) can written in the form of ordinary exponential operator [4]:

$$\mathcal{M}(t|t_0)=\exp\mathcal{W}(t|t_0;\mathcal{A}),\quad(4)$$

where $\mathcal{W}(t|t_0;\mathcal{A})$ – is a new vector field, generated by "old" vector field $\mathcal{A}=\mathcal{L}_{\mathbf{F}}$. Differentiation of eq. (4) gives us

$$\frac{d\mathcal{W}(t|t_0)}{dt}=\left[\mathcal{A}(t)\circ\frac{\mathcal{W}(t|t_0)}{1-\exp\left(-\mathcal{W}(t|t_0)\right)}\right].\quad(5)$$

Equation (5) can be written as a series:

$$\frac{d\mathcal{W}(t|t_0)}{dt}=\sum_{k\ge 0}\alpha_k\left[\mathcal{A}\circ\mathcal{W}^k\right]=\mathcal{A}+\frac{1}{2}\left\{\mathcal{A},\mathcal{W}\right\}+$$
$$+\frac{1}{12}\left[\mathcal{A}\circ\mathcal{W}^2\right]+\ldots,\quad(6)$$

where $\alpha_{2k+1}=0$, $k\ge 1$, $\alpha_0=1$, $\alpha_1=1/2$, $\alpha_{2k}=(-1)^{k-1}B_{2k}/(2k)!$, B_{2k} – Bernoulli numbers. We can rewrite eq. (6) in the following integral equation:

$$\mathcal{W}(t|t_0)=\sum_{k\ge 0}\alpha_k\int_{t_0}^{t}\left[\mathcal{A}(\tau)\circ\mathcal{W}^k(\tau|t_0)\right]d\tau.\quad(7)$$

Let be $\mathcal{W}_1(t|t_0)=\int_{t_0}^{t}\mathcal{A}(\tau)d\tau$, then the solution of (7) can be written using the the method of successive approximations:

$$\mathcal{W}_k(t|t_0)=\int_{t_0}^{t}\mathcal{A}(\tau)d\tau+\sum_{l\ge 2}\int_{t_0}^{t}\left[\mathcal{A}\circ\mathcal{W}_{k-1}^{l}(\tau|t_0)\right]d\tau,\quad(8)$$

where $k\ge 2$, $t\in[t_0,T_1]\subset[t_0,T]\subset R^1$.

We introduce the multiplicative norm, which coordinates with functional dependence elements on $t\in[t_0,T_2]$, $t_0\le T_2\le T_1$, that is $\|\mathcal{A}\circ\mathcal{B}\|_\Phi\le\|\mathcal{A}\|_\Phi\|\mathcal{B}\|_\Phi\ \ \forall\,\mathcal{A},\mathcal{B}\in\mathcal{L}\mathfrak{A}[\mathfrak{X}]$, where the index "$\Phi$" means a norm in the corresponding functional space. Let be, for example, $\mathcal{A}(t)$ – a summable function on the interval $[t_0,T_2]$, then we can define $A(t)=\int_{t_0}^{t}\|\mathcal{A}(\tau)\|d\tau$, $\forall\,t\in[t_0,T_2]$, where $\|\cdot\|$ – an arbitrary multiplicative norm in $\mathcal{L}\mathfrak{A}[\mathfrak{X}]$ and $A(t)$ – a scalar nonnegative continuous function $\forall\,t\in[t_0,T_2]$. According to the step-by-step method we should prove the sequence convergence. For these aim we use recurrent relation (8) and receive for $\mathcal{W}$:

$$\mathcal{W}(t|t_0)=\int_{t_0}^{t}\mathcal{A}(\tau)d\tau+\alpha_1\int_{t_0}^{t}\left\{\mathcal{A}(\tau),\int_{t_0}^{\tau}\mathcal{A}(\tau')\right\}d\tau+$$
$$+\alpha_1^2\int_{t_0}^{t}\left\{\mathcal{A}(\tau),\int_{t_0}^{\tau}\left\{\mathcal{A}(\tau'),\int_{t_0}^{\tau'}\mathcal{A}(\tau'')d\tau''\right\}d\tau'\right\}d\tau+\ldots$$

Let introduce the notation $\mathcal{W}=\sum_{k\ge 0}\mathcal{W}^k$, where $\mathcal{W}^k$ – a group of terms of series comprising k nested Lie brackets, then we obtain the following estimation

$$\|\mathcal{W}^k(t|t_0)\|\le A(t)\left(2A(t)\right)^k C_k,\quad k\ge 0,\quad(9)$$

where C_k satisfies to following recurrent relations: $C_{2l}=\alpha_{2l}+C_{2l-2}C_{2l-4}$, $C_{2l+1}=\alpha_{2l}+C_3C_{2l-1}$, $l\ge 2$, $C_0=1$, $C_1=1$, $C_2=\alpha_1^2+|\alpha_1|$, $C_3=\alpha_1^3+2|\alpha_1|$. The connection of coefficients α_{2k} with Bernoulli numbers leads us to the estimation

$$|\alpha_{2m}|\le\frac{2}{(2\pi)^{2m}}\sum_{k\ge 1}\frac{1}{2^{2k}}<\frac{4}{(2\pi)^{2m}}.$$

Introducing the following notation $M=\int_{t_0}^{T_2}A(\tau)d\tau$, one can estimate $\|\mathcal{W}^k\|_{L_1}\le 2^kM^{k+1}C_k$ under sufficiently great k. The majorant series with general term $2^kM^{k+1}C_k$ will be converge in according to limit test for convergence by D'Alember, if there is satisfied the inequality

$$\lim_{k\to\infty}\frac{2^{k+1}M^{k+2}C_{k+1}}{2^kM^{k+1}C_k}=q<1.$$

Let be (for definiteness) $k=2l$, then

$$q=\lim_{l\to\infty}\frac{2MC_{2l+1}}{C_{2l}}=\lim_{l\to\infty}\frac{2M(\alpha_{2l}+C_3C_{2l-1})}{\alpha_{2l}+C_{2l-2}C_{2l-4}}=2M.$$

So the majorant series converges under $M < 1/2$, and the series (7) is a absolutely convergent series. So the equation (5) has a continuous-time solution on the interval $[t_0, T_2]$. From here follows the theorem about absolutely convergence of the series (8).

The Ordinary Differential Equations. It is often the beam dynamics problems are described by the Cauchy problem

$$\frac{d\mathbf{X}}{dt} = \mathbf{G}(\mathbf{X}), \quad \mathbf{X}(t_0) = \mathbf{X}_0,$$

where $\mathbf{G} = \{G_1, \ldots, G_n\}^{\mathrm{T}}$ is an analytical function for any $\mathbf{X} \in \mathcal{X}_1 \subset R^{2n}$, $0 \in \mathcal{X}_1$ and $\mathbf{G}(0) = 0$. Then in a neighborhood of zero $\mathbf{X} = 0$ we have

$$\mathbf{G}(\mathbf{X}) = \sum_{k=0}^{\infty} \mathbb{A}_k \mathbf{X}^{[k]}, \tag{10}$$

From the convergence of series (10) follows

$$|(\mathbb{A}_k)^i| \le \frac{M_i}{(r)^k}, \quad (r)^k = r_1^{k_1} \ldots r_n^{k_n}, \tag{11}$$

where $|x_i| \le r_i$, $i = \overline{1,n}$ and r_i guarantees embedding $\mathcal{P}_r \subset \mathcal{X}_1 \subset R^n$. Here $(\mathbb{A}_k)^i$ – i-th row of matrix $\mathbb{A}_k$ and M_i are some positive constants. Using (11) one can write the following estimations:

$$|G_i(\mathbf{X})| = \left| \sum_{k=0}^{\infty} (\mathbb{A}_k)^i \mathbf{X}^{[k]} \right| \le M_i \sum_{k=0}^{\infty} \left(\frac{x}{r}\right)^k,$$

where $(x/r)^k = \prod_{i=1}^{n} (x_i/r_i)^{k_i}$. Then one can obtain the following estimations $\forall\, i = \overline{1,n}$:

$$\|G_i(\mathbf{X})\| \le \frac{M_G^m}{1-(x/r)^m}, r^m = \min_{i=\overline{1,n}} r_i,\ M_G^m = \max_{i=\overline{1,n}} M_i.$$

Let x^m satisfies inequalities $|x_i| \le x^m < r^m$, then $\|\mathbf{G}(\mathbf{X})\| \le M_G^m / (1 - x^m/r^m)^n = G^m(x^m)$, where $G^m(x^m)$ is a function of single variable x^m, defining the n-dimensional cube $\mathcal{Q}_n \subset \mathcal{P}_n^r \subset R^1$: $|x_i| \le x^m$. Above mentioned computations allow us to estimate $G_i(\mathbf{X})\partial/\partial x_i$. Let define a Lie operator

$$\mathcal{L}_{\mathbf{G}} = \sum_{i=1}^{n} G_i(\mathbf{X}) \frac{\partial}{\partial x_i} = \left(\mathbf{G}(\mathbf{X}), \frac{\partial}{\partial \mathbf{X}}\right) = \mathbf{G}^*(\mathbf{X}) \frac{\partial}{\partial \mathbf{X}},$$

and $\exp\left(t\mathcal{L}_{\mathbf{G}(\mathbf{X})}\right) \circ \mathbf{F}(\mathbf{X}) = \sum\limits_{k=0}^{\infty} (t^k/k!) \mathcal{L}^k_{\mathbf{G}(\mathbf{X})} \circ \mathbf{F}(\mathbf{X})$, then after some evaluations one can derive

$$\sum_{k=0}^{\infty} \frac{|t|^k}{k!} \left\| \mathcal{L}^k_{\mathbf{G}} \circ \mathbf{F}(\mathbf{X}) \right\| \le \sum_{k=0}^{\infty} \frac{|t|^k}{k!} \left(\mathcal{L}^m\right)^k \circ F^m(x^m),$$

where $\|\mathbf{X}\| \le x^m < r^m$. Finally we can obtain

$$\sum_{k=0}^{\infty} \frac{|t|^k}{k!} \left\| \mathcal{L}^k_{\mathbf{G}} \circ \mathbf{F}(\mathbf{X}) \right\| \le \frac{M_{\mathbf{G}}^m}{(1-x^m/r^m)^n} \times$$
$$\times \sum_{k=0}^{\infty} \alpha_k \left(|t| \frac{M_{\mathbf{F}}^m (n+1)}{r^m (1-x^m/r^m)^{(n+1)}} \right)^k, \tag{12}$$

where $\alpha_k \le \left(\left(1 - \frac{1}{n+1}\right)^{n+1}\right)^k$, and we obtain a convergent series, if there

$$|t| < \frac{r^m (1-x^m/r^m)^{n+1}}{n M_{\mathbf{F}}^m (n/n+1)^n} < \frac{r^m (1-x^m/r^m)^{n+1} e}{n M_{\mathbf{F}}^m}.$$

So, the series (12) converges absolutely, if there are the next inequalities

$$|t| < \frac{r^m (1-x^m/r^m)^{n+1}}{n M_{\mathbf{F}}^m} \left(1 + \frac{1}{n+1}\right)^n,$$
$$\|\mathbf{X}\| \le x^m < r^m.$$

The above mentioned formulae demonstrate the basic restrictions on the parameters of the problem. But on the practice it will be more practically to use the following formula for convergence estimation:

$$\|\overline{\mathbf{X}} - \mathbf{X}_N\| \le \sum_{k=N+1}^{\infty} \frac{k\, r^k L^{k+1} M}{(k-1)!} J_k(L, M), \tag{13}$$

where N is the truncation error for the given approximation order, r is the region in the phase space under study ($\|\mathbf{X}_0\| \le r$) and h – an integration step. In the estimation (13) we introduce the following definitions:

$$L = \sup_{t,\tau \in \mathcal{T}} \|\mathbb{R}^{11}(t,\tau)\|, \quad \sup_{t,\tau \in \mathcal{T}} \|\mathbb{R}^{jj}(t,\tau)\| \le jL^j,$$

$M = \int\limits_{\mathcal{T}} \varphi(t)dt$, where $\|\mathbb{P}^{ij}(t)\| \le \varphi(t)/(j-1)!$, and

$$J_i(L,M) = \begin{cases} \prod\limits_{k=3}^{i} \left\{ \frac{(k-1)L^{k-1}M}{(k-2)!} + 1 \right\}, i \ge 3, \\ 1, i = 2 \end{cases}$$

CONCLUSION

The above described estimations allow a researcher to evaluate truncated matrix expansion up to necessary order. If it is necessary, the corresponding matrices can be corrected for symplecticity property guarantee without loss of accuracy.

REFERENCES

[1] Dragt A.J. Lie Methods for Nonlinear Dynamics with Applications to Accelerator Physics. University of Maryland, College Park, www.physics.umd.edu/dsat/

[2] Andrianov S.N. A Matrix Representation of the Lie Algebraic Methods for Design of Nonlinear Beam Lines. in: AIP Conf. Proc. 1997. N 391. NY, P. 355–360.

[3] Andrianov S.N. Dynamical Modeling of Control Systems for Particle Beams. SPbSU, Saint Petersburg, 2004 (in Russian).

[4] Magnuss W. On the Exponential Solution of Differential Equations for a Linear Operator. Comm. Pure Appl. Math. 1954. Vol. 7. N 4. P. 649–679.

UNDULATOR RADIATION INSIDE A DIELECTRIC WAVEGUIDE

A.S. Kotanjyan, Department of Physics, Yerevan State University, Armenia
A.A. Saharian, Departamento de Fisica-CCEN, Universidade Federal da Paraiba, Brazil, and
Department of Physics, Yerevan State University, Armenia

Abstract

We investigate the radiation from a charge moving along a helix around a dielectric cylinder immersed in a homogeneous medium. The radiation intensity in the exterior medium at large distances from the cylinder has been considered previously and here we are mainly concerned with the radiation propagating inside the cylinder. Numerical examples are given for a dielectric cylinder in the vacuum. It is shown that the presence of the cylinder can lead to the considerable increase of the radiation intensity. The insertion of a dielectric waveguide provides an additional mechanism for tuning the characteristics of the undulator radiation by choosing the parameters of the waveguide. The radiated energy inside the cylinder is redistributed among the cylinder modes and, as a result, the corresponding spectrum differs significantly from the homogeneous medium or free-space results. This change is of special interest in the low-frequency range where the distribution of the radiation energy among small number of modes leads to the enhancement of the spectral density for the radiation intensity. The radiation emitted on the waveguide modes propagates inside the cylinder and the waveguide serves as a natural collector for the radiation.

INTRODUCTION

The motion of charged particle along a helical orbit is used in helical undulators for generating electromagnetic radiation in a narrow spectral interval at frequencies ranging from radio or millimeter waves to X-rays. The unique characteristics, such as high intensity and high collimation, have resulted in extensive applications of this radiation in a wide variety of experiments and in many disciplines (see, for instance, [1] and references given therein). These applications motivate the importance of investigations for various mechanisms of controlling the radiation parameters. From this point of view, it is of interest to consider the influence of a medium on the spectral and angular distributions of the radiation.

In [2, 3] we have investigated the radiation generated by a charge moving along a helical orbit around/inside a dielectric cylinder enclosed by a homogeneous medium. It has been shown that under certain conditions strong narrow peaks appear in the angular distribution of the radiation intensity in the exterior medium. At these peaks the radiated energy exceeds the corresponding quantity in the case of a homogeneous medium by several orders of magnitude. In these investigations we have considered the radiation at large distances from the cylinder. In the present paper we consider the radiation intensity inside a dielectric cylinder emitted by a charge moving along a helical trajectory around the cylinder (for the radiation from a charged particle rotating inside a dielectric cylinder see [4]).

RADIATION INSIDE A DIELECTRIC WAVEGUIDE

Consider a point charge q moving along the helical trajectory of radius ρ_0 outside a dielectric cylinder with radius ρ_1 and with permittivity ε_0. We will assume that this system is immersed in a homogeneous medium with dielectric permittivity ε_1 (magnetic permeability will be taken to be unit). We denote with $\omega_0 = 2\pi/T = v_\perp/\rho_0$ the angular velocity of the charge, $v_\parallel$ and $v_\perp$ are the particle velocities along the
axis of the cylinder and in the perpendicular plane, respectively.

Electromagnetic fields inside the cylinder can be presented in the form of the Fourier expansion

$$F_l(r,t) = 2\,\mathrm{Re}\sum_{m=0}^{\infty}{}' e^{im(\phi-\omega_0 t)}\int_{-\infty}^{\infty} dk_z e^{ik_z(z-v_\parallel t)} F_{ml}(k_z,\rho), \quad (1)$$

where $F = E$ and $F = H$ for electric and magnetic fields respectively. The radiation parts in the field are determined by the singular points of the integrand. The only poles of the Fourier components $F_{ml}(k_z,\rho)$ are the zeros of the function

$$U_m(k_z) = V_m\left(\varepsilon_0|\lambda_1|\rho_1\frac{J_m'}{J_m} + \varepsilon_1\lambda_0\rho_1\frac{K_m'}{K_m}\right) - m^2\frac{\lambda_0^2+|\lambda_1|^2}{\lambda_0^2|\lambda_1|^2}\left(\varepsilon_0|\lambda_1|^2+\varepsilon_1\lambda_0^2\right), \quad (2)$$

where $J_m = J_m(\lambda_0\rho_1)$ and $K_m = K_m(|\lambda_1|\rho_1)$ are the Bessel and Macdonald functions and the prime means the differentiation with respect to the argument of the function,

$\lambda_j^2 = \omega_m^2(k_z)\varepsilon_j/c^2 - k_z^2$, $\omega_m(k_z) = m\omega_0 + k_z v_\parallel$,

and

$$V_m = |\lambda_1|\rho_1\frac{J_m'}{J_m} + \lambda_0\rho_1\frac{K_m'}{K_m}.$$

The corresponding modes are exponentially damped in the region outside the cylinder. These modes are the eigenmodes of the dielectric cylinder and propagate inside the cylinder.

We denote by $\lambda_0\rho_1 = \lambda_{m,s}$, $s = 1,2,\ldots$, the solutions to equation $U_m = 0$ with respect to $\lambda_0\rho_1$ for the modes with $m \neq 0$. The corresponding modes $k_z = k_{m,s}^{(\pm)}$ are related to these solutions by the formula

$$k_{m,s}^{(\pm)} = \frac{m\omega_0\sqrt{\varepsilon_0}}{c(1-\beta_{0\|}^2)}\left[\beta_{0\|} \pm \sqrt{1+b_{m,s}^2(\beta_{0\|}^2-1)}\right],$$
$$b_{m,s} = \frac{c\lambda_{m,s}}{m\omega_0\rho_1\sqrt{\varepsilon_0}}, \quad (3)$$

where $\beta_{0\|} = v_\|\sqrt{\varepsilon_0}/c$. For $\beta_{0\|}<1$, the condition for $k_{m,s}^{(\pm)}$ to be real defines the maximum value for s, wich we will denote by s_m:

$$\lambda_{m,s_m} < m\omega_0\rho_1\sqrt{\varepsilon_0}(1-\beta_{0\|}^2)^{-1/2}/c < \lambda_{m,s_m+1}.$$

If the Cherenkov condition $\beta_{0\|}>1$ is satisfied, the upper limit s_m for s is determined by the dispersion law for the dielectric permittivity via the condition $\varepsilon_0(\omega_m) > c^2/v_\|^2$.

Under the condition $\beta_{0\|}<1$, for the radiation parts in the fields one finds

$$F_l^{(rad)}(\boldsymbol{r},t) = \sigma F_l^{(\sigma)}(\boldsymbol{r},t)$$
$$= \sigma 4\pi \,\mathrm{Re}\left[i\sum_{m=1}^{\infty} e^{im(\phi-\omega_0 t)} \sum_{s=1}^{s_m} \mathrm{Res}_{k_z=k_{m,s}^{(\sigma)}} e^{ik_z(z-v_\| t)} F_{ml}(k_z,\rho)\right], \quad (4)$$

where $\sigma = +(-)$ for $z - v_\| t > 0$ $(z - v_\| t < 0)$. This expression describes waves propagating along the positive direction of the axis z for $\sigma = +$ and for $\sigma = -$, $b_{m,s}<1$, and waves propagating along the negative direction to the axis z for $\sigma = -$, $1<b_{m,s}<1/\sqrt{1-\beta_{0\|}^2}$. Under the condition $\beta_{0\|}>1$ for the radiation fields one has $F_l^{(rad)}(\boldsymbol{r},t) = -\sum_{\sigma=\pm} F_l^{(\sigma)}(\boldsymbol{r},t)\theta(v_\| t - z)$, where $\theta(x)$ is the Heaviside unit step function. In this case the radiation field is behind the charge. The latter formula describes the waves propagating along the positive direction of the axis z for $\sigma = +$ and for $\sigma = -$, $b_{m,s}>1$, and waves propagating along the negative direction of the axis z for $\sigma = -$, $b_{m,s}<1$.

The radiation intensity inside the dielectric cylinder can be obtained by evaluating the energy flux through the cross-section of the dielectric cylinder:

$$I_\sigma^{(in)} = \sum_{m=1}^{\infty} I_{\sigma,m}^{(in)} = \frac{q^2c^2}{2\varepsilon_0}(\varepsilon_1-\varepsilon_0)^2$$
$$\times \sum_{m=1}^{\infty}\sum_{s=1}^{s_m} \frac{J_m^{-2}}{K_m^2}\frac{\lambda_{m,s}^2}{U_m'^2(k_{m,s}^{(\sigma)})}\frac{k_{m,s}^{(\sigma)}\rho_1^2}{\omega_m(k_{m,s}^{(\sigma)})}$$
$$\times \sum_{p=\pm1} G_{m,s}^{(p\sigma)}\left[\left(\frac{\omega_m^2(k_{m,s}^{(\sigma)})\varepsilon_0}{c^2}\right)G_{m,s}^{(p\sigma)} - \frac{\lambda_{m,s}^2}{\rho_1^2}G_{m,s}^{(-p\sigma)}\right], \quad (5)$$
$$\times\left[J_{m+p}'^2 + J_{m+p}^2\left(1-\frac{(m+p)^2}{\lambda_{m,s}^2}\right)\right]$$

where

$$G_{m,s}^{(p\sigma)} = \frac{v_\perp}{2c}(V_m + pmu)\sum_{l=\pm1}\frac{lK_{m+l}(\lambda_{m,s}^{(\sigma)}\rho_0/\rho_1)}{V_m - lmu}$$
$$+\frac{v_\|}{c}\frac{\lambda_{m,s}^{(\sigma)}V_m}{k_{m,s}^{(\sigma)2}\rho_1^2} \quad (6)$$
$$\times\left(\frac{k_{m,s}^{(\sigma)}\rho_1}{V_m - pmu} - \frac{\lambda_{m,s}}{k_{m,s}^{(\sigma)}\rho_1^2}\right)K_m(\lambda_{m,s}^{(\sigma)}\rho_0/\rho_1),$$

with $u = \lambda_{m,s}/\lambda_{m,s}^{(\sigma)} + \lambda_{m,s}^{(\sigma)}/\lambda_{m,s}$ and

$$\lambda_{m,s}^{(\sigma)} = \sqrt{(1-\varepsilon_1/\varepsilon_0)k_{m,s}^{(\sigma)2}\rho_1^2 - \varepsilon_1\lambda_{m,s}^2/\varepsilon_0}.$$

In the numerical examples below we consider the case $v_\| = 0$. In this case $I_+^{(in)} = I_-^{(in)}$. In Figures 1 and 2, by black points we display the number of the radiated quanta at a given harmonic m per period of the charge rotation, $N_m = 2TI_{+,m}^{(in)}/(\hbar m\omega_0)$. The graphs are plotted for $\varepsilon_0 = 3$, $\varepsilon_1 = 1$ and for the electron with the energy 2 MeV. For Figure 1 we have taken $\rho_1/\rho_0 = 0.95$ and for Figure 2 one has $\rho_1/\rho_0 = 0.99$. The red points correspond to the radiation in the vacuum in the absence of the cylinder. As it is seen from the graphs, for certain harmonics the radiation intensity inside the dielectric cylinder is essentially larger than the corresponding intensity for the synchrotron radiation in the vacuum. At these harmonics new modes of the dielectric cylinder appear. For example, in the case $\rho_1/\rho_0 = 0.95$ there are no modes for $m = 7$ and a single mode appears for $m = 8$. Similarly, for $\rho_1/\rho_0 = 0.99$, the number of eigenmodes increases by one for the harmonics $m = 7$ and $m = 26$.

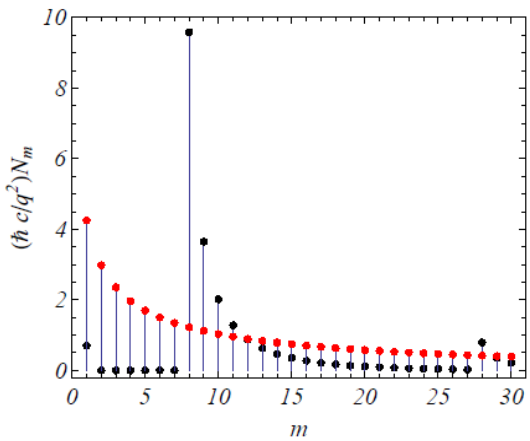


Figure 1: Number of the radiated quanta on a given harmonic m propagating inside the dielectric cylinder for $\rho_1/\rho_0 = 0.95$ (black points). The red points correspond to the radiation intensity in the absence of the cylinder.

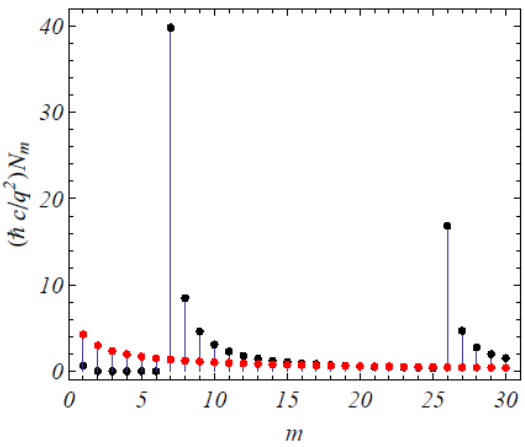


Figure 2: The same as in Figure 1 for $\rho_1 / \rho_0 = 0.99$.

CONCLUSION

In the present talk we have investigated the radiation from a charged particle moving along a helical trajectory around a dielectric cylinder immersed into homogeneous medium. The radiation intensity at large distances from the cylinder has been studied before in [3] and here we have considered the part of the radiation propagating inside the cylinder. We have specified the conditions for the existence of such radiation. Radiated energy inside the cylinder is redistributed among the cylinder modes which are solutions of Eq. (2). The spectral distribution of the radiation intensity differs significantly from that for the radiation in homogeneous medium or in free-space. Radiation emitted on the waveguide modes propagates inside the cylinder and the waveguide serves as a natural collector for the radiation. This eliminates the necessity for focusing to achieve a high-power spectral intensity. Note that, in addition to the part of the radiation propagating inside the cylinder, there are radiation fields in the exterior region localized near the surface of the dielectric cylinder. These fields are the tails of the eigenmodes for the dielectric cylinder and exponentially decrease with the distance from the cylinder. The corresponding energy flux through the plane perpendicular to the cylinder axis can be investigated in the way described above for the interior region. The details will be presented in the forthcoming paper.

REFERENCES

[1] A.Hofman, *The Physics of Sinchrotron Radiation*(Cambridge University Press, 2004).

[2] A. A. Saharian, A. S. Kotanjyan and M. L. Grigoryan, J. Phys. A 40 (2007) 1405; A. A. Saharian and A. S. Kotanjyan, J. Phys. A 40 (2007) 10641; S. R. Arzumanyan et al., Nucl. Instrum. Methods B 266 (2008) 3703.

[3] A. A. Saharian and A. S. Kotanjyan, J. Phys. A 42 (2009) 135402.

[4] A. A. Saharian and A. S. Kotanjyan, Int. J. Mod. Phys. B 26 (2012) 1250033.

STORAGE RING EDM SIMULATION: METHODS AND RESULTS

Yu. Senichev†, A. Lehrach, R. Maier , D. Zyuzin, Forschungszentrum Jülich, Germany
M. Berz, K. Makino, Michigan State University, USA,
S.Andrianov, A. Ivanov, Saint Petersburg State University, Russia
on behalf of the JEDI collaboration

Abstract

The idea of an electric dipole moment (EDM) search using the electrostatic storage ring with polarized beam is based on an accumulation of additional tiny spin rotations, about one billionth of a radian per second, which only occur in the presence of EDM. This method can be realized under conditions of long-time spin coherence ~1000 seconds. During this time, each particle performs about 10^9 turns in the ring moving on different trajectories. Under such conditions, the spin-rotation aberrations associated with various types of space- and time-dependent nonlinearities start playing a crucial role. Computer simulation is necessary to design such a ring, taking into account all the factors affecting the spin. We used COSY Infinity and an integrating program with symplectic Runge-Kutta methods in conjunction with analytical methods and T-BMT differential equation solving. We developed a new lattice based on the alternating spin rotation. As a result, we can achieve a spin coherence time (SCT) of ~5000 seconds. The difficulties of these studies are that aberration growth is observed on the scale of 10^9 turns and a few million particles. For this simulation, we used a supercomputer with a parallel computing process.

METHODS OF SIMULATION

At Forschungszentrum Jülich, two approaches are currently being considered for the EDM search in a storage ring: a method using a resonant RF spin flipper technique in the COSY ring [1] and frozen spin in the "magic" ring specially projected for the EDM search [2,3,4]. The resonance method has the character of preliminary studies of SCT and measuring techniques of spin decoherence arising for various reasons. In this paper, all results were obtained by the second method based on the "magic" ring conception.

The main difficulty in solving the problem of spin-orbital motion simulation together with the EDM signal lies in the fact that the signal expected directly from the EDM is extremely small. In particular, from sufficiently reliable estimations made in [2], it follows that due to EDM the spin in the "magic" ring rotates with the angular velocity of 10^{-9} radians per second. Taking into account the fact that the ring structure contains several hundred elements, and each particle performs about 10^6 revolutions per second, this means that EDM rotates the spin through an angle of approximately ~10^{-18} radians per element on average. Accordingly, the EDM signal is expected to change the value of the spin projections on the same tiny scale. Thus, in the EDM search we meet a problem that has not previously been encountered in accelerator physics: the arithmetic coprocessor has a mantissa length of 52 bits and can make a mistake in calculating spin projections after each element. This is a very serious limitation for using programs with the standard mantissa, and must be taken into account. Therefore using a powerful program we take a different approach, where the EDM signal is not implemented, and only the induced error signal is studied. This approach allows us to define the possible low level of the EDM signal and solve the problem from the opposite side.

In our studies, we use the following mathematical tools:

- COSY Infinity program [5], based on map generation using differential algebra and the subsequent calculation of the spin-orbital motion for an arbitrary particle;
- integrating program to study the effects that do not require a long numerical time;
- numerical integration of T-BMT differential equations for a spin in optics with smoothly approximated parameters of orbital motion;
- analytical approach.

Each of these methods is an integral part of our research.

COSY Infinity

COSY Infinity is known as a very powerful instrument for particle tracking in electromagnetic fields. COSY Infinity is a program for the simulation, analysis and design of particle optical systems, based on differential algebraic methods. Full spin-orbital tracking simulations of the entire experiment are absolutely crucial for systematically exploring the feasibility of the planned experiments. In the EDM search, it is the only program which allows the spin-orbit motion of millions of particles to be simulated over a real time scale experiment during n x1000 seconds. At present, we use the MPI (Message Passing Interface) version of the COSY Infinity program installed on a supercomputer with $3{\cdot}10^5$ processors. It is planned to use the COSY Infinity program and to include higher-order nonlinearities, normal form analysis, symplectic tracking and especially spin tracking upon the incorporation of RF-E and RF-B flippers into the program. In order to study subtle effects and simulate the particle and especially spin dynamics during accumulation and build-up of the EDM signal, custom-tailored fast trackers are needed capable of following up to 10–100 billions turns for samples of up to 10^4 -10^6 particles.

At the initial stage of EDM research, we use COSY Infinity to study the behaviour of the spin aberrations for a large number of particles and a long-time calculation.

†y.senichev@fz-juelich.de

This will allow us to answer the question of whether we can construct a facility in which the EDM signal can be seen against the background of various spin aberrations.

COSY Infinity seeks the solution in the form:

$$\vec{X} = M^1 \vec{X}_0 + M^2 \vec{X}_0^{[2]} + M^3 \vec{X}_0^{[3]} + \ldots + M^N \vec{X}_0^{[N]}, \quad (1)$$

where $\vec{X}^{[N]} = \underbrace{\vec{X} \otimes \ldots \otimes \vec{X}}_{Ntimes}$ – a Kronecker power of $\vec{X}$, $\vec{X}^k$ is a vector with C_{6+k-1}^k elements. Matrices M^k have the dimension $6 \times C_{6+k-1}^k$. Transfer maps M^k can be generated up to any order. And for spin motion we have initial spin coordinates $\vec{S}_0 = \{S_{x_0}, S_{y_0}, S_{z_0}\}$ under the condition $S_{x_0}^2 + S_{y_0}^2 + S_{z_0}^2 = 1$. After one revolution it is

$$\vec{S} = M_S \vec{S}_0, \quad (2)$$

where M_S is a spin rotation matrix.

Symplectic Runge-Kutta Integrating

The integrating Runge-Kutta program is intended to model the spin-orbital motion with fringe fields in elements and including the EDM signal directly in the simulation. The algorithm used in the program is not as fast as COSY Infinity by several orders of magnitude. Therefore, we use it mostly to investigate a short-time phenomenon that does not require long calculation periods.

In the program, the spin-orbital equations are written in the following form [6]:

$$\frac{d}{ds}\vec{Z} = F(s, \vec{Z}), \quad (3)$$

where $\vec{Z} = \{x, x', y, y', S_x, S_y, S_z\}$.

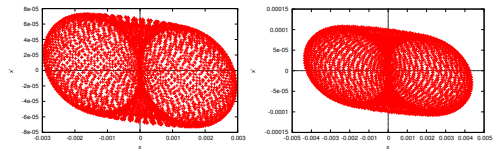
It allows us to use classical step-by-step integration methods to solve this system. As a basic method for the tracking program, a symplectic Runge-Kutta scheme was implemented [7]. According to this scheme, the solution of the spin-orbital equations can be presented in an iterative form:

$$\begin{aligned} \vec{Z}_{n+1} &= \vec{Z}_n + h\sum_{j=1}^{2} b_j \vec{F}(s + hc_j, \vec{X}^{(i)}) \\ \vec{Z}^{(i)} &= \vec{Z}_n + h\sum_{j=1}^{2} a_{ij} \vec{F}(s + hc_j, \vec{X}^{(i)}). \end{aligned} \quad (4)$$

Note that the symplectic scheme imposes the condition of a constant integration step. Moreover, this scheme requires implicit equations to be solved, and appropriate numerical methods can be used.

To compare the computation results of the two programs we used the lattice with cylindrical deflectors described in [4]. Comparing the results of tracking through a single element we found a coincidence with high accuracy in computational models of COSY Infinity and the integrating program.

Figure 1 shows simulation results of both programs for a rather complex phase motion in the radial plane with coupling to the longitudinal motion [6]. The spin behaviour also coincides in both programs.



(a) (b)

Figure 1: x–x′ motion in COSY Infinity (a) and the integrating program (b).

A different choice of the reference orbits and different symplectification methods does not permit absolutely the same numerical results to be obtained, but we can see from figure 1 that the dynamics and behaviour of particles are similar in both programs. At present, the integrating program for introducing the fringe fields and implementing the EDM signal is under development.

The last two methods, the numerical integration of T-BMT differential equations and the analytical formalism, will be discussed in detail later.

SPIN TUNE OF NON-MAGIC PARTICLE

Having the tools for spin-orbital tracking, we now proceed to the problem of formulating a lattice design that meets the requirements of the experiment. At the initial stage, the analytical formalism is the most powerful tool because it allows the issue to be seen as a whole. Moreover, the comparison of analytical estimates in the simplest cases with the program results provides extra computation control.

The spin oscillation equation therefore has the following form:

$$\begin{aligned} \frac{d\vec{S}}{dt} &= \vec{\omega}_G \times \vec{S} \\ \vec{\omega}_G &= -\frac{e}{m_0 \gamma c}\left(\frac{1}{\gamma^2 - 1} - G\right) \cdot (\vec{\beta} \times \vec{E}) \end{aligned} \quad (5)$$

As already mentioned, we consider the so-called “magic” [2], purely electrostatic ring for polarized proton, when for the reference “magic” particle $1/(\gamma_m^2 - 1) - G = 0$. Further spin projection indications will be made in the following line: z is orientated along the momentum, x and y are horizontal and vertical directions, respectively. Taking into account that the vertical and longitudinal electric field components are expected to be small and $\beta_x, \beta_y << \beta_z$ we can obtain an expression for the number of spin oscillations per one turn that is a spin tune:

$$\nu_s = \frac{e}{2\pi m_0 c^2} \cdot L_{orb} E_x \frac{1}{\gamma}\left(\frac{1}{\gamma^2 - 1} - G\right), \quad (6)$$

where L_{orb} is orbit length. For a particle of different energy from the “magic” value $\gamma \neq \gamma_m$ the factor

$G-1/(\gamma_m^2-1)\neq 0$ is not equal to zero, and the spin rotates with a tune dependent on particle energy. As we will show later, this leads to the spin tune aberrations. Expanding $G-1/(\gamma^2-1)$ in the Taylor series in the vicinity of $p=p_m$ we have:

$$\left(G-\frac{1}{\gamma^2-1}\right)_{p=p_m+\Delta p}=0\Big|_{p=p_m}+2G\cdot\frac{\Delta p}{p}-\frac{1+3\gamma^2}{\gamma^2}\cdot G\cdot\left(\frac{\Delta p}{p}\right)^2+\ldots \quad (7)$$

In a first approach, an incoherence of the spin tune can be estimated by the simple formula $\delta\nu_s=2\cdot\frac{eE_x\cdot R}{m_0c^2\gamma}\cdot G\frac{\Delta p}{p}$, where $E_x\cdot R$ is the rigidity of the ring depending on energy only and G is the anomalous magnetic moment. If to follow the definition of the spin coherence time (SCT) in [2] as the time during which the rms spread of the orientation spin of all particles in the bunch reaches one radian, then at the momentum spread $\Delta p/p=5\cdot10^{-5}$ ($\Delta W_{kin}/W_{kin}=10^{-4}$) the SCT is less than one millisecond. This disappointing fact perfectly coincides with the numerical simulation of COSY Infinity.

Following the previously proposed method [8], we then used the RF field to average the momentum deviation relative to the "magic" level. Under the RF field, the spin tune is modulated by longitudinal tune ν_z, which is two orders of magnitude higher than the spin tune ν_s, and therefore the spin oscillates with a very small amplitude $\Phi_{max}\sim(\nu_s/\nu_z)^2$ relative to a central position. However, taking into account $(\Delta p/p)^2$ in (7), the central position of the spin itself drifts very slowly. This drift term averaged over time gives the non-zero contribution:

$$\delta\nu_s=\frac{eE_x\cdot R}{m_0c^2\gamma}\cdot\frac{1+3\gamma^2}{\gamma^2}\cdot\frac{G}{2}\cdot\left(\frac{\Delta p}{p}\right)^2_{max} \quad (8)$$

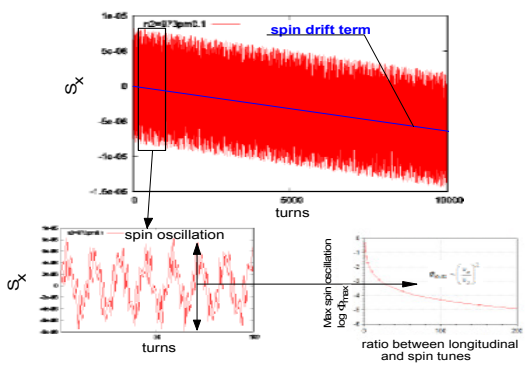


Figure 2: Oscillating and drift terms of spin behaviour.

Figure 2 shows the horizontal spin projection S_x behaviour when RF is turned on. Since the oscillating component is always within Φ_{max}, we will subsequently only be interested in the slow component drift. This depends on the square momentum spread and defines the spin tune incoherence: at the momentum spread ~$5\cdot10^{-5}$ SCT~ 180 sec. Thus the RF field increases the SCT by five orders. However, this is unfortunately only feasible for particles with zero initial deviation from the axis. Particles with non-zero deviation receive a new equilibrium orbit energy with the momentum shift $\delta p/p$ [9], which inevitably leads to a rapid increase of aberrations $\delta\nu_s=2\cdot\frac{eE_x\cdot R}{m_0c^2\gamma}\cdot G\frac{\delta p}{p}$. In an earlier paper [10] studying this phenomenon, we found a method of introducing additional oscillations on the momentum that gave the averaging of the equilibrium orbit itself. As a result, we can achieve a longer SCT time up to 500 sec at $<\Delta W_{kin}/W_{kin}>=10^{-4}$.

TIME-SPACE SPIN TUNE ABERRATION

The idea of the electric dipole moment search using the storage ring (SrEDM) with polarized beam is realized under conditions of long-time spin coherence of all particles. Following the requirements of the planned SrEDM experiment, the SCT should be more than 1000 seconds. During this time each particle performs about 10^9 turns in the storage ring moving on different trajectories through the optics elements. Under such conditions, the spin-rotation aberrations associated with various types of space and the time-dependent nonlinearities start to play a crucial role. Time-dependent aberration is a spin tune aberration due to the different time of flight of particles in the focusing-deflecting fields. The space-dependent spin aberrations are associated with differences in the focusing-deflecting fields on the trajectory of particles. In the previous section, we considered spin tune dependence on particle energy, which also introduces additional aberrations. Not all of these factors lead to unlimited growth aberrations. In some cases, the aberrations are periodic and remain within a small value that cannot affect the method of searching for EDM. For example, in periodic channels the aberrations oscillate with the betatron tune, while remaining within an acceptable range. However, in the general case it is unfortunately not so and aberrations increase. Assuming "magic" conditions, we define a variation of the spin tune through the finite differences up to the second order:

$$\delta\nu_s=\frac{e}{2\pi m_0c^2}\cdot\delta\left(\frac{1}{\gamma^2-1}-G\right)\cdot L_{orb}E_x\frac{1}{\gamma}\left[1+\frac{\delta L_{orb}}{L_{orb}}+\frac{\delta E_x}{E_x}+\gamma\cdot\delta\left(\frac{1}{\gamma}\right)\right] \quad (9)$$

Representing each of them through the Taylor series expansion in powers of the finite difference $\Delta p/p$ up to second order:

$$\begin{aligned}
&\delta\left(\frac{1}{\gamma^2-1}-G\right)=-2G\frac{\Delta p}{p}+\frac{1+3\gamma^2}{\gamma^2}G\left(\frac{\Delta p}{p}\right)^2+\ldots\\
&\frac{\delta L_{orb}}{L_{orb}}=\alpha_1\cdot\frac{\Delta p}{p}+\alpha_2\cdot\left(\frac{\Delta p}{p}\right)^2+\ldots\\
&\frac{\delta E_x}{E_x}=-k_1\frac{x}{R}+k_2\left(\frac{x}{R}\right)^2+\ldots\\
&\gamma\delta\left(\frac{1}{\gamma}\right)=-\frac{\gamma^2-1}{\gamma^3}\left(\frac{\Delta p}{p}\right)+\frac{(\gamma^2-1)^2}{2\gamma^5}\left(\frac{\Delta p}{p}\right)^2+\ldots
\end{aligned} \quad (10)$$

where α_1 and α_2 are the momentum compaction factor in the first and second approach, respectively; k_1 and k_2 are coefficients of the expansion of the field in the vicinity of the equilibrium orbit. As an example of the cylindrical deflector, the coefficients are $k_1 = 1$ and $k_2 = 1$. After averaging over time, for instance with RF on, the term $\Delta p/p$ makes a zero contribution to the spin tune. Substituting equations (10) to (9), and grouping the $\Delta p/p$ coefficients of powers up to the second order, we obtain:

$$\delta\nu_s = \frac{eL_{orb}E_xG}{2\pi m_0\gamma c^2}\cdot\left[F_2\left(\alpha_1,k_1,k_2,\frac{x}{R}\right)\cdot\left(\frac{\Delta p}{p}\right)^2 + 2F_1\left(k_1,k_2,\frac{x}{R}\right)\cdot\frac{\Delta p}{p}\right] \tag{11}$$

$$F_2\left(\alpha_1,k_1,k_2,\frac{x}{R}\right) = \frac{1+3\gamma^2}{\gamma^2}k_2\left(\frac{x}{R}\right)^2 - \frac{1+3\gamma^2}{\gamma^2}k_1\frac{x}{R} + \frac{5\gamma^2-1}{\gamma^2} - 2\alpha_1$$

$$F_1\left(k_1,k_2,\frac{x}{R}\right) = -k_1\frac{x}{R} + k_2\left(\frac{x}{R}\right)^2$$

Thus, the aberration of the spin is determined by a parabolic equation.

In our consideration, we did not include coefficients k_n with $n > 2$ and α_2 because we only consider the aberration growth up to the second order of $(\Delta p/p)^2$ and $(x/R)^2$. Figure 3 shows the two-dimensional parabolic dependence of spin tune aberration in a 3D representation, where one axis is a momentum spread in units of 10^{-4} and the other axis is a horizontal deviation in *mm*. The spin tune is normalized by a factor $N_F = \frac{eL_{orb}E_xG}{2\pi m_0\gamma c^2}$.

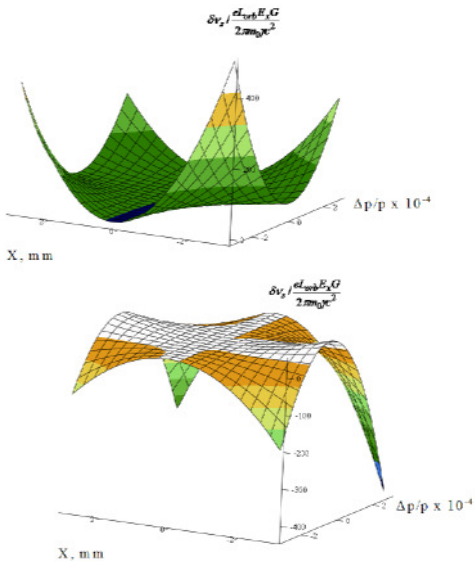


Figure 3: Spin tune aberration dependence on momentum spread and horizontal deviation at different k_1, k_2.

The coefficients k_1, k_2 depend on deflector shape and the momentum compaction factor is defined by the lattice as a whole.

Thus, these results show that it is impossible to exclude the growth aberrations of the tune spin for a non-monochromatic beam with non-zero emittance, that is at $\Delta p/p \neq 0$ and/or $x \neq 0$.

MINIMIZING OF ABERRATION

However, from the formula derived above we can perceive two methods of minimizing the spin aberrations. The first method is a choice of the lattice with compensation of the mutual influence of parameters k_1, k_2, α_1. In other words, we need to make a two-dimensional parabola maximally flat in the workspace of $(\Delta p/p)^2$ and $(x/R)^2$. To verify the analytical results we performed a full-scale simulation using the COSY Infinity program calculating the spin-orbital motion in the purely electrostatic lattice consisting of electrostatic deflectors and electrostatic quadrupoles only. Figure 4 shows the lattice in OptiM format [11].

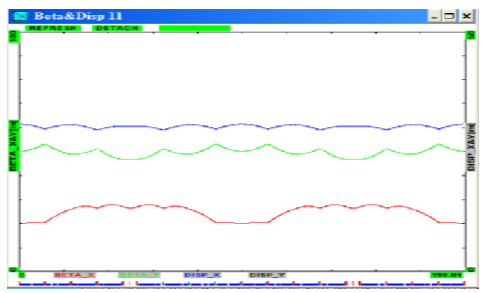


Figure 4: Twiss functions of electrostatic ring for ring and one cell.

The ring consists of two arcs, each arc has 4 FODO cells, and one cell has 4 electrostatic deflectors in each gap between quadrupoles F and D. As an example, the straight section is designed with one FODO cell. The horizontal and vertical tunes have values of 1.3 and 0.635. The electric field between the plates of the deflector is 17 MV/m.

The maximum flatness of surface (11) is reached by choosing the parameters of deflector k_1, k_2 and α_1 momentum compaction factor. The requirement for the momentum compaction factor is that it should be as large as possible. This obviously follows from the expressions (11) in the ring with a cylindrical or similar deflector geometry, when the electric field has the coefficients k_1 and k_2 close to unity.

Figure 5 shows the results of a numerical simulation with the optimum parameters of the deflector k_1=0.94 and k_2=0.96 in the whole range of operating parameters of the beam. The red curve is described by a parabola $\Delta\nu_s/NF = 0.012\cdot x^2$. At these parameters, the COSY Infinity result for the SCT is ~1000 seconds. Comparing this with purely cylindrical deflectors, we see that the

flatness in the workspace of the beam has improved by nearly factor ~20.

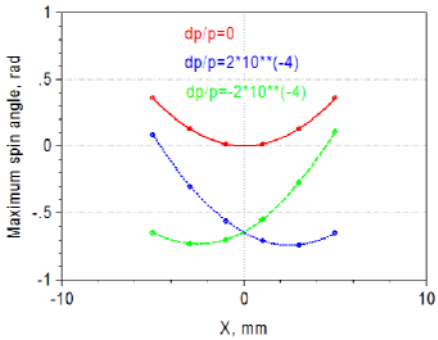


Figure 5: Maximum spin deflection angle after 10^9 turns versus x deviation at $\Delta p/p=0$ and $\pm 2\cdot 10^{-4}$.

ALTERNATING SPIN ABERRATION

The second method is to alternately change the deflector parameters and thereby alternate the sign of the spin aberration growth. In mathematical terms, this means minimizing all the factors F_0, F_1, F_2 by averaging them along an orbit, that is over time. For this purpose, we suggest the alternating spin aberration lattice which rotates spin, for instance, in one direction in even deflectors and in the other direction in odd deflectors. That is, the ring is equipped with two types of deflector having k_1 = const, and $k_2 = k_{av} \pm \Delta k$ changes from deflector to deflector. Figure 6 shows the results of the numerical simulation. We see that by choosing k_2=0.974±0.1 we can obtain practically zero aberration for particles with $\Delta p/p = 0$ and the function is described by a parabola $\Delta \nu_s/NF=0.004\cdot x^2$.

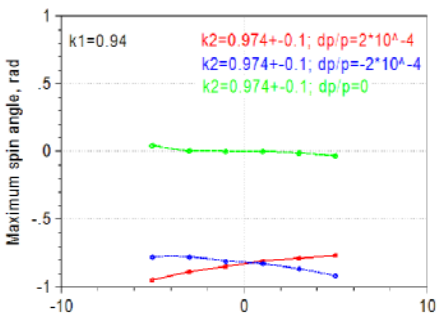


Figure 6: Maximum spin deflection angle after 10^9 turns versus x deviation in mm at $\Delta p/p=0$ and $\pm 2\cdot 10^{-4}$.

Comparing this again with the cylindrical deflector, we can see that the flatness in the workspace of the beam improved nearly ~100 times. However, the particle with non-zero momentum deviation has a finite value of the spin deflection with a parallel shift downward. It is impossible to remove this spread due to the final $\Delta p/p$ using the correct k_1 and k_2. As a result, the total spread of the spin deflection angle does not exceed ±0.5 radian after 10^9 turns, which corresponds to an SCT of about 5000 seconds.

The structure with the alternative geometry of the deflector allows us to tune the desired value of k_1, k_2. Raising the field strength between the plates in even deflectors and reducing it in the odd deflectors effectively adjusts the required coefficients k_1 and k_2. Another possibility is to create the required potential distribution due to potential changes in the stripline placed on the surface of the ceramic plates located at both ends of each deflector.

CONCLUSION

In the present work, we studied the behaviour of spin aberrations in the structure and developed techniques to minimize them. One of the most effective methods is the alternating spin aberration lattice. The analytical model allows us to find the general solution of aberration retention with an SCT of about 5000 seconds confirmed by COSY Infinity spin-orbital tracking.

ACKNOWLEDGMENTS

The authors would like to thank B. Lorentz, J. Pretz, F. Rathmann, H. Stroeher for cooperating in this work. The authors express also their deep appreciation to all IKP staff for fruitful discussions at our weekly scientific seminars.

REFERENCES

[1] A. Lehrach et al., Precursor experiments to search for permanent electric dipole moments of protons and deuterons at COSY, PSTP Proceedings, Dubna, 2011.
[2] F.J.M. Farley and Y.K. Semertzidis, Prog. Part Nucl. Phys. 52, 1 (2004).
[3] R. Talman, Ring lattice for the proton EDM measurement, 2011, Bad Honnef, Germany.
[4] Yu. Senichev, M. Berz, R. Maier, D. Zyuzin, Alternating spin aberration electrostatic lattice for EDM ring, IPAC proceedings, New Orleans, 2012.
[5] M. Berz, Computational Aspects of Design and Simulation: COSY Infinity NIM A298, p. 473 (1990).
[6] D. Zyuzin, et al., Comparison of different numerical modeling methods for beam dynamics in electrostatic rings, IPAC 2012 proceedings, New Orleans
[7] W. Oevel, M. Sofroniou, Symplectic Runge- Kutta Schemes II: Classification of Symmetric Methods, preprint, University of Paderborn, Germany, 1996.
[8] A.P. Lysenko et al., Part. Accel. 18, p.215 (1986).
[9] Yu. Senichev et al., The spin aberration of polarized beam in electrostatic rings, IPAC 2011, Spain
[10] Yu. Senichev and S.P. Møller, Beam dynamics in electrostatic rings, Proceedings of EPAC 2000, Vienna, Austria.
[11] V. Lebedev, OptiM manual, http://www-bdnew.fnal.gov/pbar/

IMPLEMENTING NEW BEAM LINE ELEMENTS INTO A MOMENT METHOD BEAM DYNAMICS CODE*

T. Roggen†, H. De Gersem, B. Masschaele
KU Leuven Kulak,
Wave Propagation and Signal Processing Research Group, Kortrijk, Belgium
W. Ackermann, S. Franke, T. Weiland
TU Darmstadt, Theorie Elektromagnetischer Felder (TEMF), Darmstadt, Germany

Abstract

Developing beam dynamics simulation tools using the moment method has advantages in terms of precision and efficiency when interests lie in average or rms dimensions of the beam, projected emittances or total energy. The moment method implemented in the V-Code solves the Vlasov equation by time integration, from an initial particle distribution represented by a discrete set of characteristic moments, accounting for all acting internal and external forces along the particle's path. The moment method delivers highly accurate beam dynamics results within a very small CPU time. This article proposes, illustrates and validates a new beam line element (BLE) for a radio frequency quadrupole (RFQ) for insertion in the V-Code. The focus will be on the RFQ cell structure, the electric field distribution and the insertion of the field distribution in the moment code.

INTRODUCTION

New particle accelerator projects rely heavily upon numerical simulations ranging from proof of concept to optimisation and fine tuning of the accelerator's individual components. Beam dynamics simulations are of vast importance to guarantee compatibility between all individual accelerator components in the optimisation cascade. Moment based algorithms have a major calculation time window advantage over full particle in cell (PIC) codes, while their accuracy is superior to macro-particle tracking algorithms. The V-Code is a sixth order moment Vlasov solver which takes into account space charge effects [1, 2]. An initial particle bunch is represented in a six dimensional phase space with longitudinal and transverse dynamics. For improved accuracy not only the average bunch coordinates and momenta, but also higher order correlation parameters (up to the sixth order) are accounted for in a phase distribution function $f(\tau, \vec{r}, \vec{p})$, $\tau = ct$ being the equivalent time, c the speed of light, t the time, $\vec{r}$ the space coordinates and $\vec{p}$ the normalized momentum ($\vec{p} = \vec{P}/mc$). V-Code allows an accelerator component's electromagnetic field distribution (exerting a force $\vec{F}$ on the particles) to be represented by a multipole expansion of the field data from finite element (FE) or finite difference time domain (FDTD) simulation results. Eventually, for $\gamma = E/mc^2$ the normalized energy, with energy E and mass m, all parameters of the Vlasov equation (Eq.1) are known, enabling to use time integration to find a solution.

$$\frac{\partial f}{\partial \tau} + \frac{\partial f}{\partial \vec{r}}\frac{\vec{p}}{\gamma} + \frac{\partial f}{\partial \vec{p}}\frac{\vec{F}}{mc^2} = 0 \qquad (1)$$

This contribution proposes, illustrates and validates a new BLE for a radio frequency quadrupole (RFQ) for insertion in the V-Code. The focus will be on the RFQ cell structure, the electric field distribution and the insertion of the field distribution in the moment code.

FIELD DATA FILES

The required BLE field component to reconstruct the electromagnetic fields in V-Code is calculated using 3D field information from the BLE model in CST Studio Suite 2012. BLE-specific data extraction and post-processing guarantee smooth field transitions at borders and inside a BLE. The field component is then stored in a *.dat field data file and read by V-Code to reconstruct a 3D field distribution using multipole expansion techniques.

RADIO FREQUENCY QUADRUPOLE

A radio frequency quadrupole (RFQ) is a low-velocity, high-current accelerator component that can accelerate hadrons from protons to uranium to energies of about $2q/A$ [MeV/n] (q being the ion charge and A the ion mass in amu). It allows for velocity independent electric focussing and adiabatic bunching, resulting in compact bunches and nearly 100 % capture and transmission efficiency. RFQs come in two flavours: four vane structures and four rod structures (Figure 1),with differences in required space, RF mode sensitivity and manufacturing techniques. The operating principle is analogue though.

The four electrodes in quadrupole configuration produce RF induced electric fields for focussing of the ion beam. With a longitudinal sinusoidal modulation of the electrodes, the vertical pair shifted out of phase by π, an accelerating longitudinal field component is introduced. Since the electric field between the rods finds its origin in the potential difference between the rods, and not in the RF electric field, a quasistatic approach is valid as long as the electrode gap is small compared to the RF wavelength [3]. The scalar potential function $U(r, \theta, z, t)$, as proposed in [4], is given in (2). r, θ, z are the polar coordinates, time t,

* This research is funded by grant "KUL 3E100118" "Electromagnetic Field Simulation for Future Particle Accelerators", project FP7-Euratom No. 269565 and the Belgian Nuclear Research Centre (SCK·CEN).

† toon.roggen@kuleuven-kulak.be

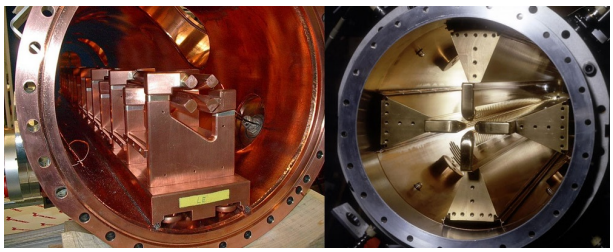


Figure 1: Left: four rod 217 MHz p^+ RFQ [5], right: four vane 101 MHz Pb^{27+} RFQ [6]

$\omega/2\pi$ the RF frequency, ϕ the initial phase shift, p, s and n indices with $n+s=2p+1$, I_{2s} the modified Bessel function, A_p and A_{ns} pole tip geometry dependent multipole coefficients in function of $m(z), a, k$ and V_0, with $m(z)$ the modulation parameter, minimal radius a, electrode potential difference V_0, $k=2\pi/2L$ with $2L=\beta_s\lambda$ the electrode modulation period, β_s the synchronous particle velocity and wave length λ (Figure 2).

$$U(r,\theta,z,t)=\sin(\omega t+\phi)\Big[\sum_{p=0}^{\infty} A_{0,2p+1} r^{2(2p+1)} \cos(2(2p+1)\theta) + \sum_{n=1}^{\infty}\sum_{s=0}^{\infty} A_{n,s} I_{2s}(knr)\cos(2s\theta)\cos(knz)\Big] \quad (2)$$

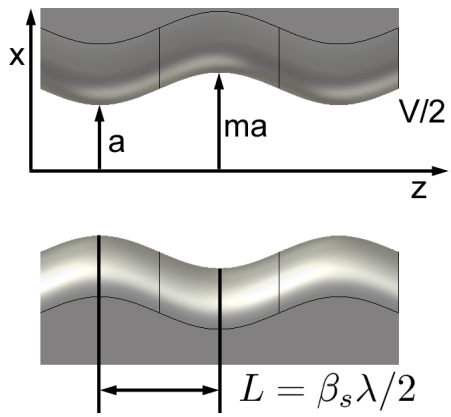


Figure 2: RFQ electrode tip geometry

TWO-TERM POTENTIAL FUNCTION

As a first approximation one may consider a hyperbolic rod geometry and only determine the quadrupole term $A_{0,1}$ of the first summation and the monopole term A_{10} of the second summation by solving the potential function on the rod tip for maximum and minimum vane modulation ($U=V_0/2$) at $z=0$. The electric fields are found by $\vec{E}=-\nabla U$.

$$A_{0,1}=\frac{V_0}{2a^2}\frac{I_0(ka)+I_0(kma)}{m^2I_0(ka)+I_0(kma)}=\frac{V_0}{2a^2}X \quad (3)$$

$$A_{1,0}=\frac{V_0}{2}\frac{m^2-1}{m^2I_0(ka)+I_0(kma)}=\frac{V_0}{2}B \quad (4)$$

$$E_r=\sin(\omega t+\phi)\frac{V_0}{2}\Big[X\frac{2r}{a^2}\cos(2\theta)+BkI_1(kr)\cos(kz)\Big] \quad (5)$$

$$E_\theta=\sin(\omega t+\phi)\frac{V_0}{2}\Big[-X\frac{r}{a^2}2\sin(2\theta)\Big] \quad (6)$$

$$E_z=\sin(\omega t+\phi)\frac{V_0}{2}\Big[-BkI_0(kr)\sin(kz)\Big] \quad (7)$$

IDENTIFICATION

A multipole expansion of $\vec{E}(r,\theta,z,t)$, replacing the Bessel function $I_n(kr)$ with its Taylor expansion, is integrated in V-Code to calculate a 3D field distribution for the volume occupied by the particle bunch. As a result of the quasistatic approach, time dependency of $\vec{E}$ can be understood as the product of $\vec{E}_{max}$ with $\sin(\omega t+\phi)$. V_0, X, B, a, and k being constant for each accelerating cell, each electric field component of the 3D field can be calculated providing:

- E_θ: $\frac{dE_\theta}{dr}$ at $\pi/4$
- E_z: on axis E_z
- E_r: using E_θ, E_z and $\nabla.\vec{E}=0$

All data is normalised.

HIGHER ORDER TERM POTENTIAL FUNCTION

Limitations of the two-term hyperbolic rod geometry are related to manufacturing, RF breakdown and RF power consumption in a region far away from the beam axis. Therefore higher order potential functions are mandatory to allow manufacturing of RFQs not suffering these limitations. Higher order terms can be defined identifying higher order coefficients according to additional collocation points. Rods being milled with a half circular transversal cross section, give rise to a four-term potential function with terms $A_{0,1}$, $A_{0,3}$, $A_{1,0}$ and $A_{1,2}$ to be defined using the collocation points in table 1.

Table 1: Four-term Potential Collocation Points

	U	z	r	θ
1	$V_0/2$	0	a	0
2	$-V_0/2$	0	ma	$pi/2$
3	$V_0/2$	0	$\sqrt{(a+\rho)^2+\rho^2}$	$\cos^{-1}(\frac{a+\rho}{r})$
4	$-V_0/2$	0	$\sqrt{(ma+\rho)^2+\rho^2}$	$\cos^{-1}(\frac{ma+\rho}{r})$

The resulting electric field using the four-term potential function is again calculated using $\vec{E}=-\nabla U$.

$$U(r,\theta,z,t)=\sin(\omega t+\phi)\Big[A_{0,1}r^2\cos(2\theta) + A_{0,3}r^6\cos(6\theta)+A_{1,0}I_0(kr)\cos(kz) + A_{1,2}I_4(kr)\cos(kz)\cos(4\theta)\Big] \quad (8)$$

$$E_r = \sin(\omega t + \phi)\Big[2A_{0,1} r \cos(2\theta) + 6A_{0,3} r^5 \cos(6\theta) + kA_{1,0} I_1(kr)\cos(kz) + kA_{1,2} I_5(kr)\cos(kz)\cos(4\theta)\Big] \quad (9)$$

$$E_\theta = \sin(\omega t + \phi)\Big[- 2A_{0,1} r \sin(2\theta) - 6A_{0,3} r^5 \sin(6\theta) - 4A_{1,2} I_4(kr)\cos(kz)\sin(4\theta)\Big] \quad (10)$$

$$E_z = \sin(\omega t + \phi)\Big[- kA_{1,0} I_0(kr)\sin(kz) - kA_{1,2} I_4(kr)\sin(kz)\cos(4\theta)\Big] \quad (11)$$

Implementation in V-Code of the four-term potential is foreseen for the near future.

3D FIELD MAPS OF RFQS

Typical RFQ lengths vary from 1 to several meters, which makes accurate 3D field map calculations of complete RFQs not recommendable. Additionally, a cell specific modulation parameter excludes the use of longitudinal symmetry conditions. As an alternative the minimum number of accelerating cells to be simulated to extract a stable 3D field in the middle cell n of interest was determined, independent of the cell length L. Modeling cells $n-6$ to $n+6$, one reaches a field accuracy of 10^{-4} %, equivalent to 1 V/m. RFQ beam dynamics are performed in a cell-per-cell cascade, enabling field parameter extraction from the numerical model using a denser mesh and convenient passing on of additional RFQ geometry parameters such as the aperture, modulation parameters and cell length.

BEAM DYNAMICS COMPARISON

To verify proper implementation of BLE and field file generation method, a beam dynamics comparison is set up between the V-Code and the CST particle tracker, and this for every BLE. Figure 3 shows good agreement for a Danfysik QSL040 quadrupole between V-Code and CST. A mesh of 20 million cells with symmetry settings in three dimensions was used to track 21000 computational particles representing individual electrons with an energy of 10 MeV (energy spread 0 %). The beam envelope is defined as 3σ, with σ the variance of the Gaussian bunch in the horizontal (x) or vertical (y) direction. The initial electron bunch has a transverse radius of 1 mm (3σ) and no transverse momentum. Both the horizontal and vertical beam envelopes of the bunch are plotted in function of the longitudinal propagation z of the particle in the BLE. The electron bunch enters the focussing quadrupole after a 166 mm drift tube. At 434 mm the bunch enters a second drift tube of 166 mm. The percental difference of the beam envelope size is less than 1 %, indicating good agreement between both particle trackers and confirming a proper implementation.

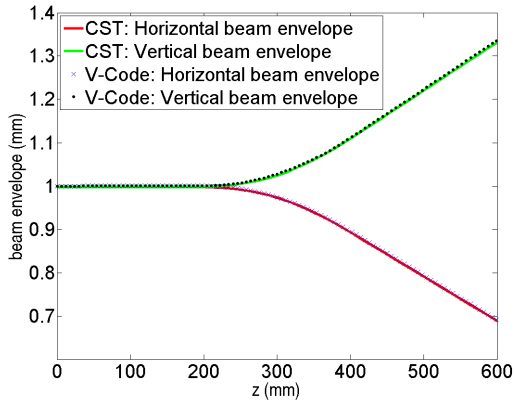


Figure 3: Beam envelope comparison between CST and V-Code for the Danfysik QSL040 quadrupole.

CONCLUSIONS

The RFQ as a new BLE of V-Code is implemented following a quasistatic approach. In a first phase the two-term potential function is incorporated, in a second phase the model will be replaced with the four-term potential function. The RFQ consists of a set of cell-by-cell components allowing accurate field parameters and convenient passing on of additional RFQ geometry parameters.

REFERENCES

[1] A. Novokhatski and T. Weiland, "Self-Consistent Model for the Beams in Accelerators", ICAP'98, Montery, September 1998, F-We21,

[2] S. Franke, W. Ackermann, T. Weiland, "A Fast and Universal Vlasov Solver for Beam Dynamics Simulations in 3D", ICAP'09, San Francisco, September 2009, TH1IODN01, pp. 208-211,

[3] T.P. Wangler, "Principles of RF Linear Accelerators", John Wiley & Sons Inc., NY, pp 225-257, ISBN: 978-0471168140

[4] I.M. Kapchinskii, V.A. Teplyakov, "Linear ion accelerator with spatially homogenious strong focussing", Prib Tekh Eksp, vol. 119, No. 2, pp. 19-22, March-April 1970

[5] ntg, 222b4e3b29, Available at: http://www.ntg.de/-typo3temp/pics/222b4e3b29.jpg [Accessed: August 14 2012]

[6] CERN, Das Innere des RFQ, Available at: http://www.lhc-facts.ch/img/rfq/rfq%20innenansicht.jpg [Accessed: August 14 2012].

MODELING OF COHERENT SYNCHROTRON RADIATION USING A DIRECT NUMERICAL SOLUTION OF MAXWELL'S EQUATIONS*

A. Novokhatski#, SLAC National Accelerator Laboratory,
Menlo Park, CA 94025, USA

Abstract

We present and discuss the properties of coherent electromagnetic fields of a very short, ultra-relativistic bunch, which travels in a rectangular vacuum chamber under the influence of a bending force of a magnet. The analysis is based on the results of a direct numerical solution of Maxwell's equations together with Newton's equations. We use a new dispersion-free time-domain algorithm which employs a more efficient use of finite element mesh techniques and hence produces self-consistent and stable solutions for very short bunches. We investigate the fine structure of the CSR fields. We also discuss coherent edge radiation. We present a clear picture of the field using the electric field lines constructed from the numerical solutions. This approach should be useful in the study of existing and future concepts of particle accelerators and ultrafast coherent light sources, where high peak currents and very short bunches are envisioned.

INTRODUCTION

The coherent synchrotron radiation (CSR) fields have a strong action on the beam dynamics of very short bunches, which are moving in the bends of all kinds of magnetic elements. They are responsible for additional energy loss and energy spread; micro bunching and beam emittance growth. These fields may bound the efficiency of damping rings, electron-positron colliders and ultrafast coherent light sources, where high peak currents and very short bunches are envisioned. This is relevant to most high-brightness beam applications. On the other hand, these fields together with transition radiation fields can be used for beam diagnostics or even as a powerful source of THz radiation.

A history of the study of CSR and a good collection of references can be found in [1]. Electromagnetic theory suggests several methods on how to numerically calculate CSR fields. The most popular method is to use Lienard-Wiechert potentials. Another approach is to numerically solve the approximate equations, which are a Schrodinger type equation. Some numerical algorithms and codes are described in [2]. We suggest that a direct solution of Maxwell's equations together with Newton's equations can describe the detailed structure of the CSR fields [3].

Modeling ultrafast phenomena requires a special algorithm for solving the electromagnetic equations. This algorithm must be free of frequency dispersion which means that all propagating waves must have their natural phase velocity, completely independent of the simulation parameters like mesh size or time step. We suggest an implicit algorithm which does not have stability issues and employs a more efficient use of finite element mesh techniques. This method can produce self-consistent stable solutions for very short bunches. We have already used this same approach long ago for wake field calculations. An implicit, dispersion-free time-domain algorithm was used in the computer code designed in 1976 for wake field dynamics studies at the Novosibirsk Electron-Positron Linear Collider VLEPP [4]. The algorithm details can be found in [5].

*Work supported by Department of Energy contract DE–AC03–76SF00515.
#novo@slac.stanford.edu

ELECTROMAGNETIC CSR SIMULATION

We may suggest that a direct solution of Maxwell's equations together with Newton's equations can describe the detailed structure of the CSR fields, the fields generated by an ultra-relativistic bunch of charged particles moving in a metal vacuum chamber inside a bending magnet. Electromagnetic components **E, B** must satisfy the equations

$$\frac{1}{c}\frac{\partial}{\partial t}\mathbf{E} = \nabla\times\mathbf{B} - \frac{4\pi}{c}\mathbf{j}_b \qquad \nabla\cdot\mathbf{E} = 4\pi\rho_b \qquad \mathbf{E}_{wall}\times\mathbf{n} = 0$$

$$\frac{1}{c}\frac{\partial}{\partial t}\mathbf{B} = -\nabla\times\mathbf{E} \qquad \nabla\cdot\mathbf{B} = 0 \qquad \mathbf{B}_{wall}\cdot\mathbf{n} = 0$$

A charge density and a charge current must satisfy a continuity equation

$$\rho_b = \sum_k \rho_k(\mathbf{x}) \qquad \boldsymbol{j}_b = \sum_k \rho_k(\mathbf{x})\mathbf{v}_k \qquad \nabla\cdot\mathbf{j}_b + \frac{1}{c}\frac{d}{dt}\rho_b = 0$$

A Newton force includes electromagnetic components and a bending magnetic field

$$\frac{d}{dt}\mathbf{p}_k = e\mathbf{E} + \frac{\mathbf{v}_k}{c}\times e\left(\mathbf{B} + \mathbf{B}_{bend}\right) \qquad \mathbf{p}_k = m\mathbf{v}_k \Big/ \sqrt{1-\frac{v_k^2}{c^2}}$$

There are a lot of finite-difference schemes, which numerically solve Maxwell's equations since the first one was published in 1966 [6]. Most of them are so-called ''explicit'' schemes, which means that the value of the field at the new time step is calculated only by the field values from the previous time step. Stability conditions for these schemes do not allow a time step to be greater than or equal to a space (mesh) step. This limitation brings an additional troublesome effect for wavelengths that are comparable to a mesh step. We state that this effect works like a frequency dispersion media, which is ''hidden'' in the finite-difference equation.

Dispersion of the Explicit Schemes

Let's check the explicit scheme for the two-dimensional case. When the field components satisfy the equation:

$$\frac{\partial^2\Phi}{c^2\partial t^2}=\frac{\partial^2\Phi}{\partial z^2}+\frac{\partial^2\Phi}{\partial x^2}$$

likewise the explicit scheme will be

$$\Phi^{n+1}-2\Phi^n+\Phi^{n-1}=\left(\frac{c\Delta t}{\Delta z}\right)^2\Delta_z^2\Phi^n+\left(\frac{c\Delta t}{\Delta x}\right)^2\Delta_x^2\Phi^n$$

The stability condition easily comes from the Fourier analyses $\Phi_k^n \sim e^{i\omega t+i\beta z+i\alpha x}$ and takes the form:

$$\sin^2\frac{\omega\Delta t}{2}=\left(\frac{c\Delta t}{\Delta z}\right)^2\times\sin^2\frac{\beta\Delta z}{2}+\left(\frac{c\Delta t}{\Delta x}\right)^2\times\sin^2\frac{\alpha\Delta x}{2}$$

For stability reasons we need the frequency ω to be real. This happens when the right part is less than one. To have stability for any longitudinal wave vector β and transverse wave vector α we need the following condition to be fulfilled

$$\left(\frac{c\Delta t}{\Delta z}\right)^2+\left(\frac{c\Delta t}{\Delta x}\right)^2\le 1 \quad \text{for } \Delta x=\Delta z \quad c\Delta t\le\frac{1}{\sqrt{2}}\Delta z$$

Now let's check the plane waves in free space. Without boundaries the plane waves must propagate at the speed of light. However the solution of the finite-difference equation shows that the propagation velocity depends upon the frequency.

$$V_{wave}=\frac{\omega}{\beta}=c\frac{2}{\beta c\Delta t}\times\arcsin\left(\frac{c\Delta t}{\Delta z}\sin\frac{\beta\Delta z}{2}\right)$$

This means that finite-difference equations contain something like a "hidden" dispersion media, which reveals itself at a wavelength comparable to the mesh size. A plot of propagation velocity as a function of frequency for different ratios of mesh steps to time steps is shown in Fig. 1.

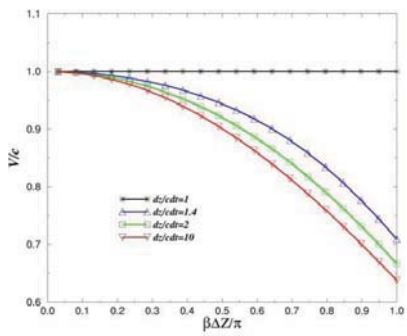


Figure 1: Propagation velocity of a plane wave in free space as a function of frequency for different ratios of mesh steps to time steps.

This numerical dispersion disappears only when the time step becomes equal to the mesh step, but in this case the scheme is unstable. The effect of the numerical dispersion may be very dangerous; it can greatly disturb the result. It may develop a strong diffusion of an initially smooth field distribution and reveal high frequency oscillations. Fig. 2 shows snap shots of a short wave bucket propagating in free space. This is the result of using the explicit scheme for the case when the length of a bucket is equal to two mesh sizes. We can see that a wave bucket has distortion, modulation and diffusion: everything that comes from a dispersion media. The numerical dispersion distorts the bucket shape. If we want to have a better result using the explicit scheme we need to decrease the mesh size least five times. In the wake field simulations this effect leads to an unphysical result like "self-acceleration" of a bunch head, which violates energy conservation. The explicit scheme ends up being not very good at calculating the wake fields of very short bunches.

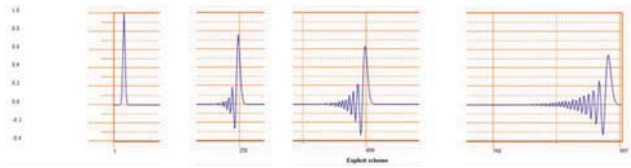


Figure 2: Snap shots of a propagating short wave bucket, calculated by an explicit scheme.

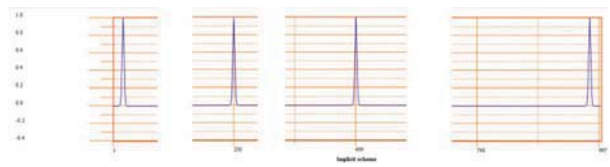


Figure 3: Snap shots of a propagating short wave bucket, calculated by an implicit scheme.

Implicit Scheme

As we stated before there will be no numerical dispersion if the time is equal to a mesh step. We can fulfil this condition by using the stable implicit scheme. In the implicit scheme for the calculation of the space derivatives we assume that the field at some time can be approximated by the average value of the field at a previous and a new time step. Of course the implicit algorithm is more complicated as it requires the solution of a system of equations. However, as we are interested in the waves that are propagating in the longitudinal direction we really only need a "partially" implicit scheme, which for our equation is the following:

$$\Phi^{n+1}-2\Phi^n+\Phi^{n-1}=\left(\frac{c\Delta t}{\Delta z}\right)^2\Delta_z^2\Phi^n+\left(\frac{c\Delta t}{\Delta x}\right)^2\Delta_x^2\frac{1}{2}\left(\Phi^{n+1}+\Phi^{n-1}\right)$$

The dispersion relation for this scheme is:

$$\sin^2\frac{\omega\Delta t}{2}=\frac{\left(\frac{c\Delta t}{\Delta z}\right)^2\sin^2\beta\frac{\Delta z}{2}+\left(\frac{c\Delta t}{\Delta x}\right)^2\sin^2\alpha\frac{\Delta x}{2}}{1+2\left(\frac{c\Delta t}{\Delta x}\right)^2\sin^2\alpha\frac{\Delta x}{2}}$$

This equation shows that this scheme is stable in the case of equal mesh and time step $c\Delta t=\Delta z$. Fig. 3 demonstrates the effectiveness of this scheme. Now a wave bucket keeps the same shape.

MAIN STRATEGY OF THE METHOD

The main strategy of our method is to use an implicit algorithm which does not have stability issues and employs a more efficient use of finite element mesh techniques. The scheme may have dispersion in the transverse

direction. However, electromagnetic fields, which interact with a beam, propagate in the vacuum chamber at small angles, so the effect of dispersion in the transverse direction is less important than dispersion in the longitudinal direction. To employ the implicit scheme we transform Maxwell's equations to second order equations.

We also use the Fourier series expansion in the vertical direction. This approach allows to make 3D simulations on the assumption that there is no vertical motion of the beam and the vertical size of a beam chamber is constant. We also assume that the conductivity of the wall of the beam chamber is infinite.

To decrease the amount of needed memory we use a traveling mesh. This is very important for simulations of real devices, like a bunch compressor where the distance between bends is tens of meters but the bunch length is of a micron size. Our mesh will move with the speed of light and we can definitely assume that the electromagnetic field in front of the bunch is zero; even if the bunch motion is not straight. Because a time delay or a longitudinal position delay due to the rotation in a bending magnet is very small, we really do not need more mesh space for the bunch. In our case a traveling mesh does not change the accuracy of the scheme or any conditions of stability.

To simulate the real shape of a non-monochromatic bunch moving under the action of the electromagnetic fields including vertical magnetic fields of bending magnets, we will use an ensemble of particles. We will track each particle and average the current (particle velocities) over the mesh. The charge density distribution will be integrated using the continuity equation for charge and current. This will help to smooth out errors of particle transitions from one cell to another. Later we will show a charge distribution of a bunch rotated by the vertical magnetic field. We also assume that initially a bunch has a Gaussian distribution in all directions and is travelling with a speed very close to the speed of light. In this case we can easily calculate the initial distribution of the electromagnetic field of a bunch equal to the field of a bunch traveling in infinite metallic beam chamber. More details of this method can be found in [7-8].

CSR FIELD DYNAMICS

Radiation in a Bend

Let us first try to understand how a bunch field remakes itself when a bunch is rotated in a magnetic field. We have calculated the electromagnetic field of a three dimensional Gaussian bunch that is initially moving along the vacuum chamber very close to the speed of light. At some point the bunch enters a vertical magnetic field of a bend. What happens after can be seen at Fig. 4, where snapshots of the electric field line distributions are shown at different time moments. In these plots the white boxes with the red arrows show a bunch contour and a bunch velocity direction. Before entering a bend the bunch has only a transverse field, which can be seen as a set of vertical lines. A new field that is generated in a bend is a set of ovals, which increase in size with a time. We can outline two time periods of the field formation. The first period is when a bunch is still inside the region of its initial transverse field. The first two plots in Fig. 4 are related to this first period. The second period starts when the bunch is delayed so much that it is out of the region of the initial transverse field. The bunch is delayed because the velocity vector rotates and the longitudinal component becomes smaller and smaller than the speed of light, however field lines that are not very far away "don't know" about this change and continue to propagate at the speed of light. The last plot in Fig. 4 shows this situation. We may consider these fields to be the fields of the edge radiation in a bend.

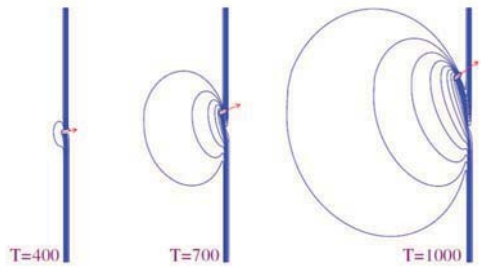


Figure 4: Snapshots of electric field lines of a bunch, which is moving in a magnetic field. White boxes show the bunch contour. Red arrows show the directions of the bunch velocity.

A more detailed picture of the field lines is shown in Fig. 5, where we also show the directions of the electric field lines by green arrows. If one examines this picture he can see that the upper field lines take the position of the lower lines and a part of the lower field lines take the position of the upper lines. However at the far ends the transverse field lines continue traveling in the same initial direction.

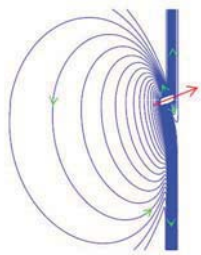


Figure 5: Detailed structure of the field pattern.

We can easily to explain such behavior if we present this field as a sum of two fields: $E = E_{db} + E_{in}$. The first part is the field of a dipole, which consists of two oppositely charged bunches. One bunch, which has a positive charge is the "real" one. This bunch is rotated in the magnetic field while the other bunch is a "virtual" one, which has an opposite charge and travels straight in the initial bunch direction. The second field $\mathbf{E}_{in}$ is the field of another "virtual", but positively charged bunch, which travels straight along the initial bunch direction. Naturally the virtual bunches together sum to zero. When we decompose the charges we decompose the fields and the

very complicated structure of the radiation field becomes simple. The decomposition of the field is shown in Fig. 6.

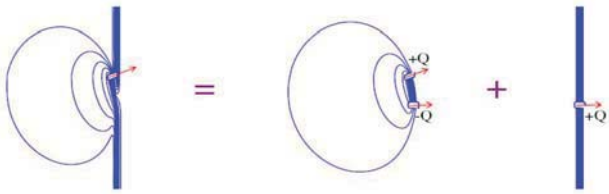


Figure 6: Decomposition of the field of a bunch moving in a magnetic field (left plot) into two fields: a field of a dipole (middle plot) and a field of a bunch moving straight in the initial direction (right plot). Red arrows show the directions of the bunch velocities.

The interaction of the bunch with the dipole field continues for a longer time. Fig. 7 shows the absolute value of the electric field on the horizontal plane in the vertical center of the vacuum chamber in consecutive time steps. The white oval shows the real bunch contour. When a dipole is created an electric field appears between a real bunch and a virtual bunch. This field increases in value and reaches a maximum when the bunches are completely separated and then decreases as the bunches move apart leaving the fields only around the bunches. The bunch acquires an energy loss while interacting with the electric dipole field.

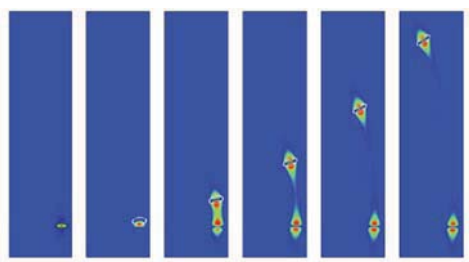


Figure 7: Absolute value of the electric field of a dipole.

Continuing the study of the radiation process we investigate the dense set of field lines in Fig. 5, or the fine structure of the field in front of a bunch. This region is common with classical synchrotron radiation. The characteristic wavelength of the synchrotron radiation or an equivalent value of the bunch length for this relativistic factor is

$$\sigma_{s.r.} \approx \frac{R}{\gamma^3}$$

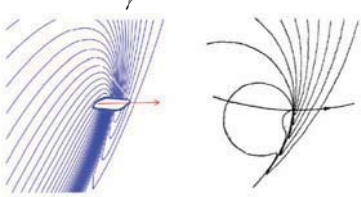


Figure 8: Fine structure of the field pattern in front of a bunch. The left plot shows field lines near a bunch. The right plot presents a picture from reference [9].

We chose reference [9], as it supplies a picture of the field lines of a particle moving in a circle with a relativistic factor of γ=6. The equivalent value of the bunch length for this relativistic factor is very close to our bunch length. Fig. 8 shows this finite structure together with a plot from [9]. We can state that the region before a bunch is very close for both cases.

Fields Acting Inside a Bunch

In order to study the fields acting on the particles inside the bunch we calculate the distribution of a collinear force $E_{||}$ and a transverse force $E_{\perp}$ as projections to the bunch velocity

$$\mathbf{F}_{||} = \mathbf{J}_b \cdot \mathbf{E} \quad \mathbf{F}_{\perp} = \left[\mathbf{J}_b \times \mathbf{E}\right]_x$$

We have found some very exciting fine structure from this force acting on the particles in the bunch. Fig. 9 shows a distribution of forces in the horizontal plane in the vertical center of the vacuum chamber at three time moments. The left three vertical plots in Fig. 9 show a bunch charge distribution. The starting plots at each set are at the bottom when the bunch just enters the magnetic field. The red arrows show the direction of the bunch velocity. The middle three vertical plots show a transverse force. Again, the red arrow shows the direction of the force. The transverse force is the well-known space-charge force, which probably is compensated by a magnetic force in the ultra-relativistic case. The right three vertical plots show the collinear force, which is responsible for an energy gain or an energy loss. The red color corresponds to acceleration and energy gain and the blue color corresponds to deceleration or energy loss. The red arrows are collinear or anti-collinear with the bunch velocity.

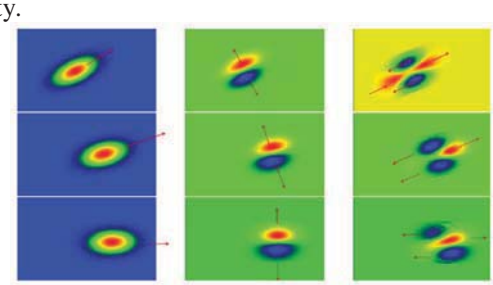


Figure 9: Bunch charge distributions, transverse forces and collinear forces on the horizontal plane in the vertical center of the vacuum chamber at three time moments.

We see here that the forces on the bunch are very complicated. The particles, which are in the center of the bunch, in front of the bunch and at the end are accelerated, whereas the particles at the boundaries are decelerating. This means that a bunch gets an additional energy spread in the transverse direction. The total effect is deceleration and the bunch loses energy. The asymmetry of the longitudinal fields can also be seen in Fig. 5, which shows the electric field line distributions. The bunch shape deformation due to the difference in the angular velocity along the radial position is usually small and can be seen only after some time; however the ultra-small beam emittance can be effected.

The integrated energy loss along the transverse direction as a function of the longitudinal coordinate is shown in Fig. 10 together with a bunch longitudinal

distribution. One can see that the head of the bunch and the tail are accelerated, when the rest of the bunch is decelerated. The shape of the energy loss distribution is compared with the analytical 1-D model [10] (green dashed line). We obtain a better agreement with the shape of the energy loss distribution for a larger bending radius and smaller bunch length. This comparison is shown at the right plot of Fig. 10. The transverse energy spread is smaller for a larger bending radius.

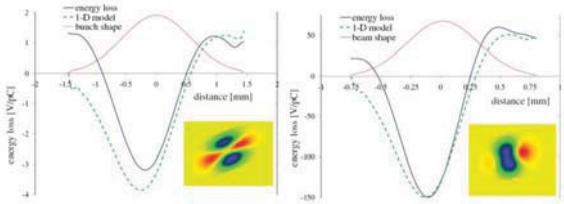


Figure10: Integrated energy loss along the transverse direction as a function of the longitudinal coordinate for two values of bending radius.

Coherent Edge Radiation

As we mentioned above, an ultra-relativistic bunch and CSR fields are moving together and interact for a long time. However one can see a field, which propagates straight ahead from the initial beam horizontal position. This field can be seen very well when the bunch gets a large horizontal displacement. Fig. 11 shows the distribution of the magnetic field on the horizontal plane in the vertical middle of the vacuum chamber. The large peak corresponds to the bunch field. A red arrow shows the initial bunch X-position and the direction of the bunch velocity. A blue arrow shows the direction of the bunch velocity at this time.

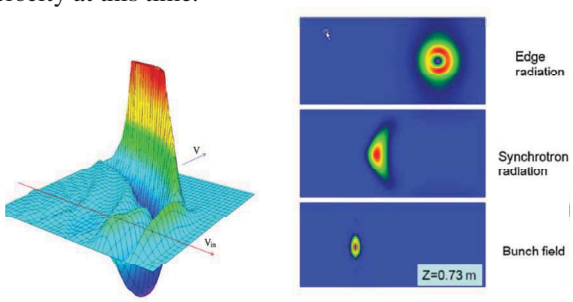


Figure11: Coherent edge radiation.

The right plot shows images of the coherent radiation in the form of transverse magnetic field distributions on the vertical planes of the vacuum chamber. At first we see an image of edge radiation, then the image of synchrotron radiation and finally a bunch field image. The calculated images of the coherent edge radiation look very similar to the images, which we have seen on the YAG screen after the dump magnets, which bend the beam down at LCLS.

Fields in the Beam Chamber of an Accelerator

As we mentioned before, the metal walls of a vacuum chamber of an accelerator change the distribution of the radiation fields, but the self-electromagnetic field of a bunch is also modified by the chamber geometry. This field is much stronger than the radiation fields. Fig. 12 shows the vertical electric field component inside the chamber when a bunch has left a magnet. The shape of the chamber follows the bunch trajectory. To see the radiation fields we need to magnify the amplitude 1000 times. When a bunch changes position in a chamber, its electromagnetic field also changes and the bunch must react back: loosing and then gaining the kinetic energy. In some cases this effect can be a much stronger radiation loss.

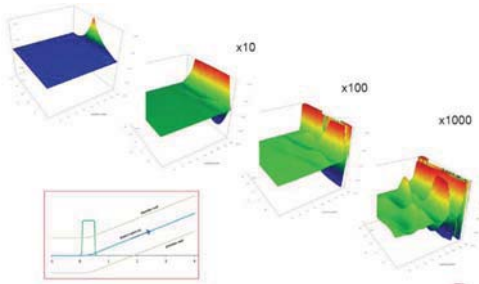


Figure 12: Magnified vertical component of the electric field of a bunch moving in a vacuum chamber after having been bent by a magnet.

CONCLUSIONS

We have found that there is much a more interesting and detailed structure to the CSR fields, which have not been described by any previous study. A very important result is discovering the structure of the complicated collinear forces. A bunch will get an additional energy spread in the transverse direction from the collinear force. This immediately leads to emittance growth and decoherence that could limit FEL lasing for very short bunches. We will continue to study this effect.

ACKNOWLEDGMENTS

The author would like to thank Mike Sullivan for help and valuable comments, Franz-Josef Decker, Paul J. Emma and Yunhai Cai for support and interest in this work and physicists of the Beam Physics Department for useful discussions.

REFERENCES

[1] J. B. Murphy, ICFA B. D. Newsletter, No.35 (2004).
[2] G. Bassi et al., Nuc. Instrum. Methods Phys. Res. A, 557, pp. 189-204 (2006).
[3] A. Novokhatski, M. Sullivan, IPAC'2010, Kyoto, May 2010.
[4] V. Balakin et al., Proc. of the All Union Accelerator Conference, Dubna, 1978, page 143. Translated to English: SLAC TRANS-188, November 1978.
[5] A. Novokhatski, SLAC-PUB-11251, May 2005.
[6] Yee KS. IEEE Transactions on Antennas and Propagation 1966; 14(3):302-307.
[7] A. Novokhatski, PAC'2011, New York, March 2011
[8] A. Novokhatski, Phys.Rev.STAB, 14 (2011) 060707.
[9] S. G. Arutyunyan, Sov. Phys. Usp. 29, 1053-1057 (1986)
[10] Ya. S. Derbenev et al., TESLA FEL-Report 1995-05 (1995)

RAPID INTEGRATION OVER HISTORY IN SELF-CONSISTENT 2D CSR MODELING*

Klaus Heinemann, David Bizzozero, James A. Ellison, Stephen R. Lau
Mathematics and Statistics, University of New Mexico, Albuquerque, NM, USA
Gabriele Bassi, Brookhaven National Laboratory, Upton, NY, USA

VLASOV-MAXWELL APPROACH TO BUNCH COMPRESSORS

This paper discusses three strategies in our work on the Vlasov-Maxwell (VM) system for a bunch compressor: the current paradigm, a modified paradigm, and a future paradigm. Each strategy only requires knowledge of the fields in the bunch. In the current paradigm we reduce the field calculation to a 2D integral over the 2D charge and current densities of the bunch and their time history. The current paradigm is implemented in our code VM3@A (**V**lasov-**M**axwell **M**onte-Carlo **M**ethod at **A**lbuquerque) using a time stepping algorithm. Here a major expense is in the integration over history at each time step. The modified paradigm relies on spatial Fourier transformations to reduce the 2D integral to 1D convolutions (one convolution per mode) over history, where we approximate the convolution kernel by a sum of exponentials. As a result the history dependence is effectively localized in time, and each time step then relies only on field information at the present and previous time steps. We present a toy model for the modified paradigm and results are encouraging. We then briefly discuss a future paradigm which involves a 3D Maxwell simulation, using a domain $\mathcal{D}$ which snuggly surrounds the bunch. Such a simulation will require radiation boundary conditions set on $\partial\mathcal{D}$.

The exact problem we deal with in all three strategies is the IVP for the coupled system consisting of the 6D Vlasov equation for the phase space density f and the Maxwell equations for the self field $\mathbf{E}, \mathbf{B}$ with boundary conditions on two perfectly conducting shielding plates. The initial self field is zero for the current paradigm and its modification. Details of the VM approach and the current paradigm are found in [1, 2, 3].

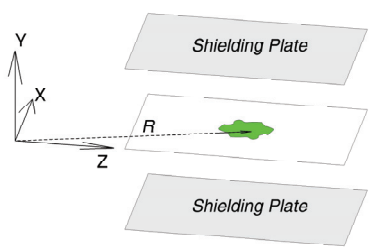


Figure 1: Geometry of sheet bunch model

* Work supported by US DOE grant DE-FG02-99ER41104

CURRENT PARADIGM

The current paradigm and its modification are based on a special case of the VM system, the *sheet bunch model*, where the bunch is confined to the mid-plane between the two shielding plates; whence

$$f(\mathbf{R}, Y, \mathbf{P}, P_Y; u) = \delta(Y)\delta(P_Y) f_{sheet}(\mathbf{R}, \mathbf{P}; u)$$
$$\rho_f(\mathbf{R}, Y; u) = \delta(Y)\rho_{sheet}(\mathbf{R}; u)$$
$$\mathbf{j}_f(\mathbf{R}, Y; u) = \delta(Y)\mathbf{j}_{sheet}(\mathbf{R}; u) \ ,$$

where $u = ct$, $\mathbf{R} = (Z, X), \mathbf{P} = (P_Z, P_X)$, and $\rho_f, \mathbf{j}_f$ are the charge and current densities associated with f. Our setup is shown in Fig. 1.

For the sheet bunch model $\mathbf{E}, \mathbf{B}$ are only needed in the part of the $Y = 0$ plane occupied by the bunch. Moreover, E_Y, B_X, B_Z vanish in the $Y = 0$ plane, and so we only need E_Z, E_X, B_Y. We define

$$\mathbf{F}(\mathbf{R}; u) := (E_Z(\mathbf{R}, 0; u), E_X(\mathbf{R}, 0; u), B_Y(\mathbf{R}, 0; u)) \ ,$$

and the exact solution in the $Y = 0$ plane is

$$\mathbf{F} = \mathbf{F}_0 + \sum_{m=1}^{\infty} \mathbf{F}_m \ , \quad \text{where} \tag{1}$$

$$\mathbf{F}_0(\mathbf{R}; u) = \frac{-1}{4\pi} \int_0^u \int_{-\pi}^{\pi} dv d\theta \mathbf{S}(\mathbf{R} + \mathbf{e}(\theta)(u - v), v) \tag{2}$$

is the nonshielding term and the $\mathbf{F}_m$ are image charge contributions needed to satisfy the boundary conditions. Here, $\mathbf{e}(\theta) = (\cos\theta, \sin\theta)$ and the source $\mathbf{S}$ is determined by $\rho_{sheet}, \mathbf{j}_{sheet}$. The nonshielding term alone is sufficient for some applications.

For the v-integration we use an adaptive Gauss-Kronrod integrator. The θ-integration is done with the trapezoidal rule. The domain of θ-integration is small for most v-values, as shown in Fig. 2 of [1]a.

The evolution of the source is governed by the Vlasov equation. We approximate this by a Monte Carlo particle method in accelerator coordinates [1]. At time u we know the particle phase space positions for $0 \le v \le u$. $\mathbf{F}(\mathbf{R}, u)$ is then computed as an integral over history and used to evolve the particles phase space positions to $u + \Delta$, where Δ is small enough so the fields are constant to good approximation. The new source at $u + \Delta$ is then determined using a density estimation procedure (e.g. kernel density estimation with a product of Epanechnikov kernels [3]) and the process is repeated.

The self field computation in (2) is the most time consuming part because of the v-integral over the bunch history. The flop count per time step is $O(N_X N_Z N_v N_\theta)$ where $N_X N_Z$ is the number of field grid points and $N_v N_\theta$ is the number of integration grid points. The whole v-integral over the history of the bunch must be done at each time step since $\mathbf{F}(R;u)$ cannot be used to calculate $\mathbf{F}(R;u+\Delta)$ and thus N_v increases with time. Therefore the process $\mathbf{F}(R;u)$ is not Markovian; the future depends on the past not only through the present.

MODIFIED PARADIGM

The main goal of this paper is to propose a modification of the current paradigm. This entails a method to calculate $\mathbf{F}_0(\mathbf{R};u)$ using its spatial Fourier transform $\widetilde{\mathbf{F}}_0(\mathbf{k};u)$ which we hope will speed up the computation. It can be shown that the spatial transforms $\widetilde{\mathbf{S}}(\mathbf{k};u)$ and $\widetilde{\mathbf{F}}_0(\mathbf{k};u)$ of $\mathbf{S}(\mathbf{R};u)$ and $\mathbf{F}_0(\mathbf{R};u)$ are related by

$$\widetilde{\mathbf{F}}_0(\mathbf{k};u) = -\frac{1}{2}\int_0^u dv \widetilde{\mathbf{S}}(\mathbf{k};v) J_0((u-v)|\mathbf{k}|)\ , \quad (3)$$

where J_0 is the zeroth order Bessel function of first kind and $|\mathbf{k}|$ is the Euclidean norm of $\mathbf{k}$. Eq. (3) is still history dependent, but this dependence now resides in J_0 rather than the source.

The key to our approach is the fact that, unlike J_0, an exponential kernel effectively removes the history dependence. To illustrate, consider

$$F(u) = \int_0^u dv g(v)\exp(\beta(u-v))\ , \quad (4)$$

for function g and complex constant β. Then for $\Delta > 0$

$$F(u) = \quad (5)$$
$$\exp(\beta\Delta)F(u-\Delta) + \int_{u-\Delta}^u dv g(v)\exp(\beta(u-v))\ .$$

Although the integral (4) remains history dependent, it solves the initial value problem $F' = \beta F + g(u)$ with $F(0)=0$, and thus is a Markov process.

The above suggests that the integral over the history of the bunch can be essentially localized through an approximation of J_0 by a sum of exponentials:

$$J_0(v) \approx \sum_{n=1}^{N_E} \alpha_n \exp(\beta_n v)\ . \quad (6)$$

Combination of this approximation with Eq. (3) yields

$$\widetilde{\mathbf{F}}_0(\mathbf{k};u) \approx \sum_{n=1}^{N_E} \alpha_n \widetilde{\mathbf{F}}_n(\mathbf{k};u)\ , \quad \text{where}$$
$$\widetilde{\mathbf{F}}_n(\mathbf{k};u) = -\frac{1}{2}\int_0^u dv \widetilde{\mathbf{S}}(\mathbf{k};v)\exp(\beta_n(u-v)|\mathbf{k}|)\ .$$

By analogy with (5) we obtain

$$\widetilde{\mathbf{F}}_n(\mathbf{k};u) = \exp(\beta_n\Delta|\mathbf{k}|)\widetilde{\mathbf{F}}_n(\mathbf{k};u-\Delta)$$
$$-\frac{1}{2}\int_{u-\Delta}^u dv \widetilde{\mathbf{S}}(\mathbf{k};v)\exp(\beta_n(u-v)|\mathbf{k}|)\ . \quad (7)$$

Formula (7) is the center piece of the modified paradigm, and it effectively removes the v-integration over bunch history. At the point in the code where we compute $\mathbf{F}(\mathbf{R},u)$, we know the particle phase space positions. In addition in the modified paradigm we will also know the $\widetilde{\mathbf{F}}_n(\mathbf{k},u-\Delta)$ and $\widetilde{\mathbf{S}}(\mathbf{k},u-\Delta)$. We first compute $\mathbf{S}(\mathbf{R},u)$ from the particle positions, $\widetilde{\mathbf{S}}(\mathbf{k},u)$ by FFT and then $\widetilde{\mathbf{F}}_n(\mathbf{k},u)$ from (7). Since $\widetilde{\mathbf{S}}(\mathbf{k},v)$ is slowly varying, the integral in (7) can be done after a linear interpolation of $\widetilde{\mathbf{S}}(\mathbf{k},v)$ using $\widetilde{\mathbf{S}}(\mathbf{k},\cdot)$ at $u-\Delta$ and u. From the $\widetilde{\mathbf{F}}_n(\mathbf{k},u)$ we find $\mathbf{F}(\mathbf{R},u)$ by an IFFT and then evolve the particles to $u+\Delta$. The FFT is key here, and a flop count per time step gives $O(N_X N_Z N_E \log(N_X N_Z))$ in contrast to $O(N_X N_Z N_v N_\theta)$. So $N_E \log(N_X N_Z)$ replaces the typically larger $N_v N_\theta$. In the case of a nonuniform grid an NFFT will be used [4].

Let us briefly comment on the construction of the approximation (6). Due to the asymptotic behavior of J_0, its sum-of-exponentials approximation (6) must break down for very late times. Therefore, we focus on finite-time approximation of J_0, and its Laplace convolution

$$(J_0 * g)(u) = \int_0^u dv J_0(u-v)g(v). \quad (8)$$

If A (here the exponential sum in (6)) approximates J_0, then we have the estimate [5]

$$\|A*g - J_0*g\|_{L_2(0,T)} \leq$$
$$e^{\eta T}\sup_{s\in\eta+\mathrm{i}\mathbb{R}}\left|\frac{\widehat{A}(s)-\widehat{J}_0(s)}{\widehat{J}_0(s)}\right| \|J_0*g\|_{L_2(0,T)}, \quad (9)$$

where $[0,T]$ is the approximation window in time. Since the Laplace transform $\widehat{J}_0(s) = (s^2+1)^{-1/2}$ of $J_0(v)$ is singular for $s = \pm\mathrm{i}$, the shift $\eta > 0$ above must be included. We have chosen $\eta = 10^{-6}$, and then focused on finding $\widehat{A}(s) = \sum_{n=1}^{N_E} \alpha_n/(s-\beta_n)$ to achieve

$$\sup_{s\in\eta+\mathrm{i}\mathbb{R}}\left|\frac{\widehat{A}(s)-\widehat{J}_0(s)}{\widehat{J}_0(s)}\right| < \varepsilon, \quad (10)$$

where ε is a prescribed tolerance.

Our algorithm for constructing the desired rational approximation $\widehat{A}(s)$ to $\widehat{J}_0(s)$ is essentially Xu and Jiang's (XJ) [6] adaptive version of the Alpert, Greengard, Hagstrom (AGH) compression algorithm [5] used in the context of radiation boundary conditions. We further modify the XJ algorithm to incorporate the parity conditions (see Fig. 2), obeyed by the real and imaginary parts of $\widehat{J}_0(s)$ for $s \in \eta + \mathrm{i}\mathbb{R}$, and we have also used quadruple precision arithmetic. Embedded within the XJ algorithm

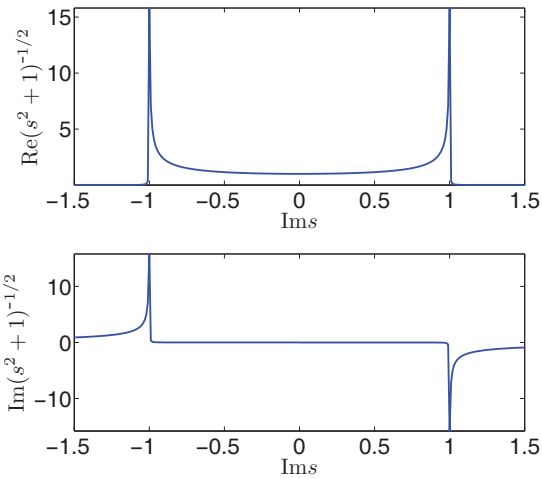


Figure 2: Profiles for $(s^2+1)^{-1/2}$. The real and imaginary profiles are shown along $s = \eta + \mathrm{i}(-\frac{3}{2}, \frac{3}{2})$. For the sake of the figure, here $\eta = 10^{-3}$ rather than $\eta = 10^{-6}$.

is the AGH algorithm for solving the following nonlinear least squares problem (here with $h(s) = \widehat{J}_0(s)$):

$$\min_{P,Q} \int_{\xi_1}^{\xi_2} \left| \frac{P(\eta+\mathrm{i}\xi)}{Q(\eta+\mathrm{i}\xi)} - h(\eta+\mathrm{i}\xi) \right|^2 \mathrm{d}\xi \,, \qquad (11)$$

that is minimization of the integral over the space of polynomials $P(s)$ and $Q(s)$ such that $\deg(Q) = \deg(P) + 1$. The XJ algorithm starts with an adaptive interval refinement of the inversion contour $\eta + \mathrm{i}\mathbb{R}$, one based on the smoothness of $h(\eta + \mathrm{i}\xi)$. The AGH algorithm is then applied recursively from the finest to the coarsest levels, with the goal of first isolating those pole locations β_n (and their corresponding strengths α_n) which lie closest to the contour (thereby somewhat allievating the ill-conditioning of the problem). Once a sum-of-poles approximation has been constructed via the XJ algorithm, we then verify (10). If it is not achieved, then we increment N_E and try again.

We studied the toy model

$$F(\mathbf{R}; u) = \int_0^u dv \int_{-\pi}^{\pi} d\theta G(\mathbf{R} + (u-v)\mathbf{e}(\theta); v) \,,$$

where "the source" is given by

$$G(\mathbf{R}; u) = \exp\Big(-\nu|\mathbf{R} - \mathbf{R}_c(u)|^2\Big) \,,$$
$$\mathbf{R}_c(u) = (a\cos(\omega u), b\sin(\omega u)) \,.$$

The source G is essentially a Gaussian moving on an ellipse. We take $\nu = 5, \omega = 2\pi, a = 1.2, b = 0.8$. In the spirit of the current paradigm and its modification we calculate $F(\mathbf{R}; n\Delta)$ on a $2D$ grid with $64 \cdot 48$ grid points in $Z \cdot X$. Figure 3 displays the expected quadratic growth in CPU time for the current paradigm where we use the quad2D integrator from Matlab. In contrast, in the modified

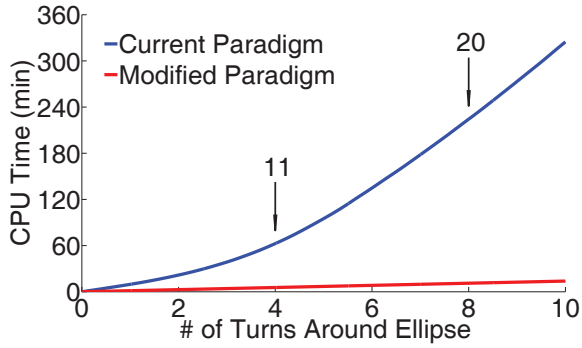


Figure 3: Comparison of CPU times between current paradigm and its modification. The numbers 11 and 20 are ratios of CPU times.

paradigm the CPU time grows linearly. Moreover the CPU times are consistent with flop counts and function evaluation counts. We used $64 \cdot 48$ grid points in $k_Z \cdot k_X$ with maximal $|\mathbf{k}|$-value around 26 so that the needed domain of J_0 is $[0, 260]$. We used $N_E = 56$ where the relative error (with respect to the envelope) of J_0 on the interval $[0, 10^6]$ is ≤ 0.01. The relative error only blows up beyond the interval $[0, 10^7]$.

In summary we have found in the toy model that the CPU time at large times u is considerably less for the modified paradigm and this is consistent with a function evaluation and flop count. For the bunch compressor problem there appears to be a considerable decrease in flop count for the modified paradigm, as discussed after (7). Thus the cpu times and function evaluation counts in the toy model and the flop counts in the current and modified paradigms for the bunch compressor give us some optimism that we can decrease the CPU time in our VM3@A code. Our next step is to revise the toy model to make it more realistic. If this looks good, we'll test the idea in our VM3@A code.

FUTURE PARADIGM

In the approaches discussed above we have replaced the VM system by an integro-differential system which involves the integral over history but only requires the evaluation of fields in the bunch. We now turn our attention to a direct time integration of the VM system of PDEs. This eliminates the integral over history but requires the fields outside the bunch. To deal with the latter we will consider radiation boundary conditions, e.g. as developed in [5]. For spatial discretization we will begin by investigating Discontinuous Galerkin methods [7] and their implementation in the code HEDGE (Hybrid Easy Discontinuous Galerkin Environment, see http://wiki.tiker.net/Hedge). This work will be part of a Ph.D. dissertation project by one of us (D.B.).

REFERENCES

[1] G. Bassi, J.A. Ellison, K. Heinemann, and R. Warnock, *Microbunching instability in a chicane: Two-dimensional mean field treatment*, Phys. Rev. ST Accel. Beams, **12**, 080704 (2009), and *Transformation of phase space densities under the coordinate changes of accelerator physics*, Phys. Rev. ST Accel. Beams, **13**, 104403 (2010).

[2] G. Bassi, J.A. Ellison, and K. Heinemann, *Self Field of Sheet Bunch: A Search for Improved Methods*. TH2IOPK01 in ICAP09 Proceedings, San Francisco, California. See JACoW.org website.

[3] K. Heinemann, *Two topics in particle accelerator beams*, Ph.D. Dissertation, Math&Stat, University of New Mexico, May, 2010.

[4] J.Keiner, S. Kunis, D. Potts, NFFT, Softwarepackage. http://www.tu-chemnitz.de/∼potts/nfft (20022005).

[5] B. Alpert, L. Greengard, and T. Hagstrom, SIAM J. Numer. Anal. 37, 1138-1164 (2000).
T. Hagstrom, T. Warburton, and D. Givoli, Journal of Computational and Applied Mathematics 234 (2010) 1988.

[6] K. Xu and S. Jiang, *A Bootstrap Method for Sum-of-Poles Approximations*, Journal of Scientific Computing, 2012, DOI: 10.1007/s10915-012-9620-9.

[7] J.S. Hesthaven and T. Warburton, *Nodal Discontinuous Galerkin Methods* (Springer, 2008).

BRINGING LARGE-SCALE ANALYTICS TO ACCELERATORS*

N. Malitsky#, BNL, Upton, NY 11973, USA

Abstract

The report presents a new approach for storing and processing both the accelerator control data and the experimental results. It is based on the analysis and consolidation of several modern technologies, such as the EPICS control infrastructure, the SciDB array-oriented data management and analytics platform, the HDF5 file format, and others.

INTRODUCTION

The efficient data management and processing systems are essential tools in the commissioning and operation of the accelerator facilities and large scientific experiments. For example, analysis of historical data is heavily involved in the troubleshooting process, detection and study of the composite behaviour patterns, comparison and design of the different operational scenarios, and many other operational tasks. The data acquisition system (DAQ) is responsible for collecting the detector measurements which are the primary results of the dedicated experiments.

Building these systems however represents a serious challenge for control developers that have to acquire and store heterogeneous data streams coming at different rates from tens of thousands of distributed devices. The conventional technologies, such as relational database management systems, were not designed for such requirements. As a result, many of the control teams had to build new proprietary tools from scratch or try to accommodate the existing technologies with some ad hoc extensions. For example, the Experimental Physics and Industrial Control System (EPICS [1]) collaboration maintains more than four archiver systems. Constrained by the associated technologies and available resources, each of these solutions cannot address all requirements and present some trade-off among different objectives: performance, reliability, extensibility, and others.

The scale, data rate, and complexity of the new light source facilities introduce the next challenge and demands for new approaches. Particularly, the BNL National Synchrotron Light Source II (NSLS II) and SLAC Linac Coherent Light Source II (LCLS II) projects shift the frontiers of control systems towards millions of control process variables and streaming rates of up to one million events per second. Similar requirements are introduced by other accelerator projects. Furthermore, recent progress in the development of light source detectors is leading to dramatic changes in the amount and complexity of available experimental data. These quantitative changes in requirements trigger two principal topics: demand for an efficient analytics-oriented database technology and consolidation of the control and experimental data management systems aiming to foster scientific discoveries.

Highly scalable data management and processing are two emerging topics in both industry and academics. Triggered by Google's web technologies, this domain represents an active factory of the new types of the analytics-oriented data storages, such as Casandra, CouchDB, Hbase, and MongoDB. Most of them are designed after Google's I/O stack: Google File System, Bigtable distributed storage system, and MapReduce processing framework.

Despite success in numerous projects, the web-oriented environment however cannot be directly applied to scientific applications. First, a Bigtable is a sparse, distributed multi-dimensional sorted map indexed by row key, column key, and a timestamp. It treats data as un-interpreted strings allowing flexible representations of structured and semi-structured formats. Unlike the web-oriented projects, scientific applications commonly relied on the multi-dimensional array-oriented data model. The map-based approach however is inefficient or incapable of addressing array-oriented queries. Second, the original MapReduce processing framework does not support complex iterative algorithms required by machine learning and scientific applications. This limitation prompted the new wave of hybrid data-intensive techniques merging or extending different data models and parallel paradigms: SciDB, GraphLab, SciHadoop, HadoopDB, MapReduce-MPI, and others.

To address the data challenges of the new modern light source facilities and the variety of technical solutions, we proposed an integrated configurable environment for connecting the different types of accelerator data with the modern large-scale analytics engines. The following sections subsequently overview the EPICS-based accelerator and beamline control system, the HDF5 file format [2] for storing experimental data of the modern light source facilities, two important representatives of the analytics approaches, the SciDB array-oriented parallel database [3] and the GraphLab distributed framework for processing complex algorithms on graphs [4], and finally the proposed large-scale integrated data management and analytics environment for accelerator and beamline experiments.

EPICS ACCLERATOR AND BEAMLINE CONTROL SYSTEM

EPICS is the Experimental Physics and Industrial Control System that has been started in 1989 at Los Alamos National Laboratory and in use today at a significant proportion of the particle physics laboratories

*Work supported by DOE contract DE-AC02-98CH10886
#malitsky@bnl.gov

in North America, Europe, and Asia. Its base infrastructure consists of two layers: distributed Input/Output Contollers (IOCs) and Operator Interface (OPI) applications connected by the Channel Access communication protocol. IOC provides a uniform interface to heterogeneous physical devices. According to the EPICS generic approach, each physical device is represented by a flat collection of process variables of predefined types. The existing EPICS-based projects usually deal with hundreds of thousands of such process variables. Monitoring and control of the distributed devices on the client level are provided by collection of the EPICS engineering tools and the physics applications. An archiver is one of the essential tools of this collection facilitating the analysis, operation, troubleshooting, and upgrade of the complex multi-system scenarios. Figure 1 illustrates the present EPICS two-tier environment.

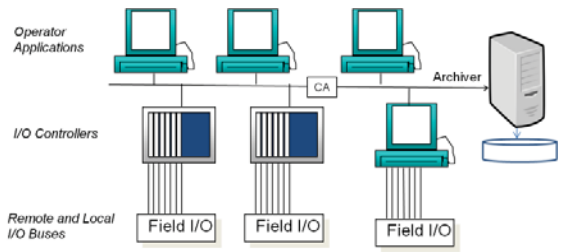


Figure 1: EPICS v3 two-tier environment.

The success of the EPICS infrastructure in the development and operation of numerous projects eventually generated new requests. One of these demanded features was the support of the three-tier model-based control architectures. Until recently, the middle layer servers have been implemented with other technologies like CORBA. In EPICS, this capability was hampered by a limited set of data types supported by the database and the corresponding communication protocol, particularly, by a lack of the structured data required by the high-level applications. To address this issue, the EPICS team started the development of the next version EPICS v4 adding the middle layer servers based on the novel concept of PV Data as show in Figure 2.

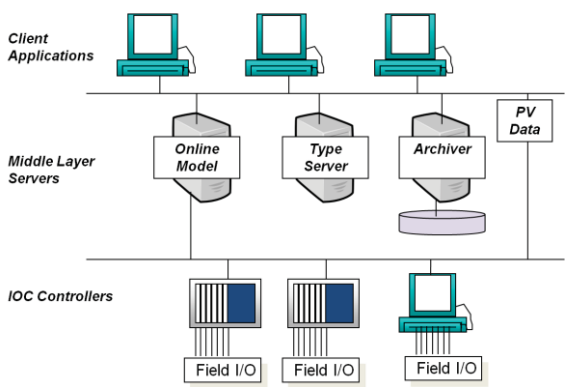


Figure 2: EPICS v4 three-tier high-level application environment.

PV Data is a generic self-described dynamic data container, allowing to generalize the present database records with the new hierarchical structures and to enhance all parts of the EPICS infrastructure: communication protocol, associated interfaces, and services. Moreover, it creates a basis for integrating data acquisition services of the beamline facilities.

HDF5 FILE FORMAT

HDF5 [2] is the latest version of the Hierarchical Data Format developed and distributed by the non-profit HDF group for storing scientific-oriented data of a wide range of application domains. The core of the HDF5 specification consists of a generic and compact data model based only on three primary concepts: dataset, data type, and group.

The HDF5 dataset is a multi-dimensional array of data elements with attributes and metadata including a description of the data elements, dimensions and all other information necessary for processing and storing data. Its structure and associated components are outlined in Figure 3.

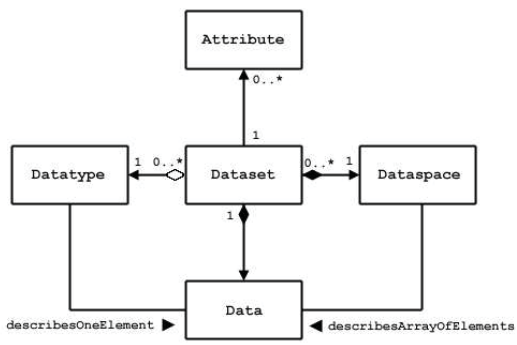


Figure 3: HDF5 dataset [2].

A data element of the HDF5 dataset can be any of several numerical or character types, small arrays, or even composite types similar to C structures. Its description and storage format is defined in the data type object. Data types are categorized into 11 classes. Each class is interpreted according to a set of rules and has a specific set of properties. The collection of data types can be extended by users. Moreover, the user-defined data type can be named, stored in a separate HDF5 file, and shared by other datasets located in the external HDF5 files. The HDF5 group serves for composing a collection of datasets and named types into the hierarchical structures.

The HDF5 software library is highly optimized and benchmarked on various environments ranging from laptops to massively parallel systems. For example, for boosting the data access performance, HDF5 added several important features like chunking, compression, and others. As a result, recently HDF5 is becoming a de facto standard for maintaining the complex experimental results produced in the modern light source facilities.

SCIDB ARRAY-ORIENTED DATABASE

SciDB [3] is a new open-source parallel database management system that brings the array-based analytics-oriented platform to large-scale scientific and commercial applications, such as astronomy, climate control, risk management and others. The multi-dimensional array is a fundamental data type used in various numerical computing environments (e.g. MATLAB) and is a natural model for describing both the time series of the control historical data and images in the beamline experiments. In recent years, there have been several approaches adding arrays into the relational databases. In contrast with the previous products, SciDB provides and implements a consistent array-oriented formalism including the conceptual model, a set of corresponding operators and a dedicated architecture.

In SciDB, an array can be created with the following command:

CREATE ARRAY Example <a1:integer, a2:float, a3:MyType> [Dim1=0:5 , Dim2=0:4]

The array definition contains three parts: name ("Example"), a cell structure, and a list of dimensions. Figure 4 illustrates what such an array might look like.

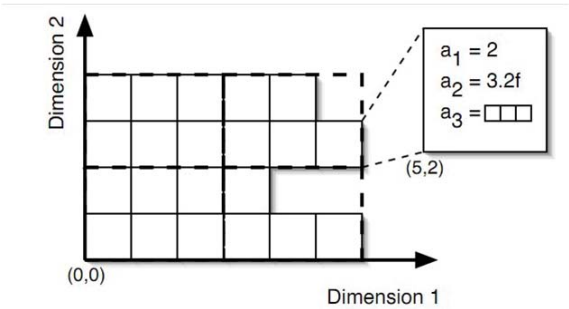


Figure 4: SciDB two-dimensional array [3].

A nominally rectilinear array can be sparse including empty cells and jagged edges. Every cell has the same data types for its values which can be scalars or composite structures defined according to the Postgres user-defined type (UDT) system.

For managing arrays, SciDB provides two flavours of the high-level languages: array query language (AQL) and array functional language (AFL). AQL is modelled after the SQL declarative language including two classes of queries, data definition language (DDL) and data manipulation language (DML). AFL is a proprietary language extending the AQL subset with the lower-level "physical" operators. The operators can generally be characterized based on where or not the operations manipulate an array in terms of its structure – the array's rank, and dimension indices – or by addressing the array's contents – the data values in attributes in cells. Some operators come in both structural and data-dependent forms.

The SciDB database represents the next logical step in the evolution of the modern column-oriented analytics platforms. In the traditional relational database management systems data values are stored and managed by rows. This organization is driven by the online transaction processing (OLTP) applications that generally create or modify one or a few records at a time. Conversely, online analytical processing (OLAP) applications often call upon the database to analyze selected attributes of vast number of rows or records. In such environments, the values for each single column are stored contiguously. The efficiency of the column-oriented approach has been demonstrated in the warehouse marketplace by products like Sybase IQ, Infobright, Vertica, and others.

The modern data warehouse systems are built after the OLAP cube model connecting data into multi-dimensional structures. In the context of the relational paradigm, the cube is usually created with the star or snowflake schemas. SciDB generalizes this approach by explicitly defining the array abstract data type (ADT) and map it into the common application programming interface and all parts of the multi-layer database architecture. Moreover, SciDB further extends the array model with the efficient chunk-based approach boosting the performance of associated data operations.

In recent years, there have been several approaches adding arrays into the relational databases, for example, using an array executor on top of blobs, or object-relational extensions. In order to properly compare and develop the most effective solution, the team of researches derived a benchmark called SS-DB. This benchmark was modelled after several scientific domains and includes a set of common queries and operations. The results clearly demonstrated the advantage of the SciDB explicit array-oriented architecture in comparison with the relation-based extensions.

GRAPHLAB COMPUTATIONAL FRAMEWORK

SciDB brings the array-oriented data model and associated use cases to the large-scale distributed processing environment based on the conventional database technology. Adherence to the existing architecture allowed to reuse and adapt the proven theoretical concepts and corresponding technical solutions, such as ACID, query processing and optimization framework, concurrency control and recovery techniques, high-level interface language and data description. At the same time, the rigid database methodology introduced the principle constraints reducing the scope of the potential applications and supporting algorithms. This limitation, which was consistent with a "one size does not fit all" principle, lead to the development of the alternative highly flexible data processing approach, MapReduce, based on only two functions for transforming and aggregating data.

Introduced by Google in 2004, the MapReduce programming model has been successfully adopted by many companies for the large-scale web applications and has triggered the development of a new class of so called NoSQL databases. The simple two-function model however was not able to address the full spectrum of scientific algorithms, particularly, requiring multiple asynchronous iterations and complex data structures. Recently, this topic became an active target of many research and commercial projects. One of most systematic approaches for solving the complex analytics tasks was suggested in the GraphLab project and implemented by the team of the computer scientists from the Carnegie Mellon University.

GraphLab is a distributed framework for developing a broad range of large-scale machine-learning applications, such as collaborative filtering, data-mining and computer vision. The framework consists of three main parts: data model, update function mechanism, and execution model.

In GraphLab, data is organized as a graph, the most universal and convenient structure of the scientific and machine learning models. In the distributed setting, the data is partitioned into multiple parts, where each sub-graph can be stored as a different file, representing a vertex in the global meta-graph structure (see Figure 5). This organization compliments the SciDB array-oriented data model and allows extending the hierarchical description of the HDF5 files to the larger scale.

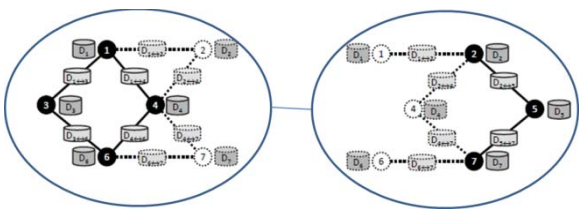


Figure 5: GraphLab meta-graph with two sub-graphs stored in the different files [3].

Similar to the SciDB approach, the algorithm-specific computation on the graph data is implemented via a mechanism of plug-in update functions. The GraphLab update function is a stateless procedure which is able to read and modify the scope of the input vertex including data associated with all adjacent vertices and edges. In addition to data processing, the function can return a set of tasks. Each task is a tuple consisting of an update function and the input vertex. All return tasks are placed in queue and eventually executed according to the scheduling order of the GraphLab execution engine.

This composite approach includes the update function mechanism and the execution queue provides the asynchronous iterative computation model for running parallel algorithms. To facilitate the development and computation of the complex algorithms, GraphLab enforces sequential consistency over update tasks with respect to the appropriate consistency model. The current version offers two alternative execution engines, Chromatic Engine and Locking Engine, which address these requirements using two different techniques, graph coloring and read-write locks.

INTEGRATED APPROACH

Figure 6 outlines the proposed environment designed to bring the modern distributed analytics engines to the analysis of the different types of accelerator large-scale data sets. Particularly, the approach addresses several actual tasks and issues:

- immediate upgrade of the existing EPICS archive system with the modern high-performance database technology;
- extension of the HDF5-based experimental file systems with the distributed quivery services.
- consolidation of the control and beamline data management systems;
- consistent transition to the new EPICS v4 three-tier infrastructure;
- enhancement of the proprietary retrieval engines with the modern analytics platforms for processing the complex array-oriented and data-mining algorithms.

The approach is designed after the three-layer I/O stack architecture consisting of the file system backend, middle-layer distributed data model and associated global services, and distributed computational framework. The data model represents the primary abstraction of the overall system and often determines the implementation of other components. In most data management systems, this architectural integrity is enhanced by the tighter integration of the different layers and encapsulation of the low-level interfaces. In our approach, we suggested to extend these systems with the driver framework enabling the interchange of the different file backends with the appropriate data processing engines.

In the context of the accelerator facilities, this approach resolves two conceptual tasks. First, it provides a consistent mechanism for in-situ integration of the multiple file formats used by the different teams and applications. Second, the driver-based mechanism facilitates the selection of the most optimal computational model (e,g, parallel database or MPI) for solving the particular category of the application tasks. The integrity of the overall system is still preserved by the common data model abstractions which determine requirements to the concrete backend and engine representations.

Many scientific data models can be described via three major components: multi-dimensional arrays, user-defined data types, and directed graphs. Among the various alternative file formats, HDF5 represents the most consistent and complete implementation of this data description. It has been successfully employed in various climate and space information systems. And our analysis and benchmark studies confirmed the advanced conceptual and technical features of this file format.

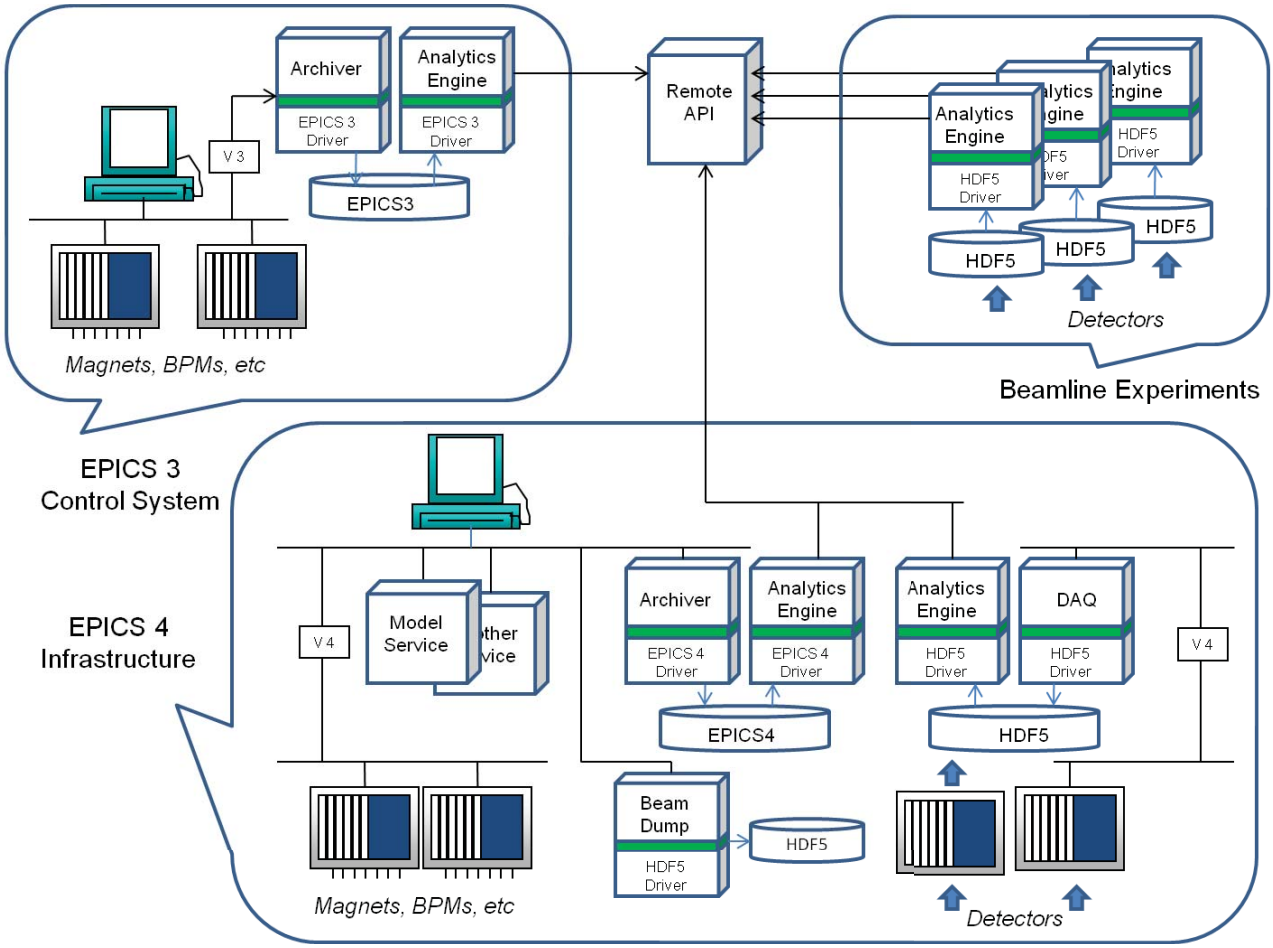


Figure 6: Large-scale integrated data management system.

With the exception of one project developed by Barrodale Computing Services, HDF5 was not generally considered as a backend of the traditional database management systems. The primary reason was associated with the mismatch between the relational and HDF5 scientific-oriented models. The new large-scale data processing tasks fundamentally changed the scope and requirements for the modern data management and analytics systems emphasizing the multi-dimensional arrays and graphs as the first class citizens. Following the incremental approach, the development of this integrated system has been started with more transparent and straightforward array-oriented applications including the large-scale correlation analysis of accelerator historical data based on the SciDB analytics engine and we expect to have a working prototype by the end of 2012 year.

REFERENCES

[1] L.Dalesio et al., "Experimental Physics and Industrial Control System Architecture," ICALEPCS'93. http://www.aps.anl.gov/epics

[2] M. Folk et al., "An Overview of the HDF5 Technology Suite and its Applications," AD'11. http://www.hdfgroup.org

[3] The SciDB Development Team, "Overview of SciDB," SIGMOD'10, http://www.scidb.org

[4] Y. Low et al., "Distributed GraphLab: "A Framework for Machine Learning and Data Mining in the Cloud," VLDB'12, http://graphlab.org

SIMULATION OF BASEBAND BTFs USING A PARTICLE-IN-CELL CODE

P. Görgen*, O. Boine-Frankenheim, TEMF, TU-Darmstadt, Darmstadt, Germany
W. Fischer, S. White, Brookhaven National Laboratory, Upton, NY, U.S.A.

Abstract

A simulation model for transverse bunched beam transfer functions (BTFs) at the base harmonic is presented. It is based on a code including different machine effects, most notably transverse space charge using a two-dimensional (2D) Poisson solver. A simplified model for the simulation of the strong-strong beam-beam effect was implemented using either 2D field data or analytic expressions under the assumption of Gaussian beams for the beam-beam interaction. The validity of the BTF model is verified based on the comparison of BTF and Schottky spectra features with analytic expectations from literature. The simulation model is then applied to the RHIC proton lattice. A linear transfer map is used between interaction points. BTFs including the beam-beam effect are simulated. Measurements are compared to simulation results at machine conditions.

MOTIVATION

Transverse BTFs of coasting beams were shown to give great diagnostic opportunities [1], for example enabling the direct measurement of the incoherent space charge tune shift. Our aim is to investigate the diagnostic opportunities of the transverse BTFs of bunched beams. We decided to start the investigation with the baseband beam transfer functions of bunched beams experiencing the beam-beam effect. The choice of baseband BTFs was made because both at Brookhaven National Laboratory (BNL) as well as GSI a high sensitivity BTF system based on the direct diode detection method developed at CERN [2] is in use. The choice of the beam-beam effect was made because it is localized and reduces overall simulation time.

THE BTF MODEL

In the measurement of BTFs, the beam is transversely excited using a kicker operating at a certain frequency. The response of the beam center of charge oscillation is measured using a pickup. The process is repeated over a range of frequencies and the relative complex response amplitude as a function of excitation frequency gives the beam transfer function. In our simulation model we started from an existing code [3] to implement a simplified baseband beam transfer function: The codebase provides particle tracking based on the 6D transfer maps computed by MADX [4] with the addition of different effects such as chromaticity, self-consistent space charge and others. The BTF simulation consists of a combination of two modifications:

1. In one cell of the tracking lattice, a transverse periodic excitation is added to the momenta of all particles. For the RHIC case where bunch lengths are far below 1/100 of the machine circumference we consider it safe to expect that all particles in one bunch get the same kick amplitude from an excitation signal at a frequency below the revolution frequency.
2. In the cell of the kick we also compute the offset of the center of charge $\langle x \rangle$ from the reference trajectory and store it together with the excitation amplitude and frequency.
3. The particle distribution is initialized and tracked for a few thousand turns to equilibrate possible matching errors due to the beam-beam effect. The equilibrated particle distribution is stored.
4. The simulation is run for a range of frequencies, simulating excitation for 3500 turns. The same number of samples is taken in the real-life BTF measurements. After each run, the equilibrated particle distribution is reloaded in order to save the time needed to simulate until the excited oscillation dies down.
5. In offline analysis, we compute the complex response amplitude and excitation amplitude exactly at the excitation frequency using the discrete time Fourier transform (DTFT). The division of the complex response amplitude by the excitation amplitude gives the BTF.

Testing of the BTF Model

In order to verify our implementation we compare our simulation with analytically accessible scenarios found for example in [5]. A popular example concerns a particle distribution with a Gaussian frequency spread. The analytic result [5] is given in units of the normalized frequency deviation u defined as a function of the mean particle betatron frequency $\overline{\omega}$, the driving frequency Ω and the frequency width of the distribution $\Delta\omega$ via the equation:

$$u = \frac{\overline{\omega} - \Omega}{\Delta\omega}. \tag{1}$$

For a Gaussian frequency distribution the analytic BTF $R(u) = f(u) + ig(u)$ is given by:

$$f(u) = \sqrt{\frac{2}{\pi}} e^{-u^2/2} \int_0^\infty \frac{\mathrm{d}y}{y} e^{-y^2/2} \sinh(uy) \tag{2}$$

$$g(u) = \sqrt{\frac{\pi}{2}} e^{-u^2/2} \tag{3}$$

We replicated the situation of a Gaussian frequency spread in the transverse direction for the PIC simulation of a longitudinally frozen bunch: We made use of Gaussian momentum distribution that acted via chromaticity as a Gaussian tune

* goergen@temf.tu-darmstadt.de

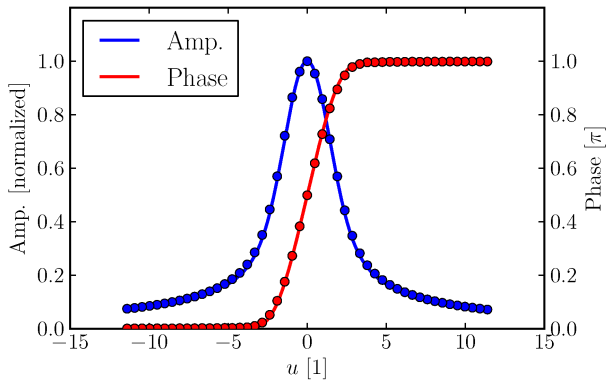


Figure 1: Comparison of amplitude and phase of the analytic prediction for the BTF of a beam with Gaussian tune spread (line) and the PIC BTF simulation results (points). The simulation agrees well with analytic predictions. u is the normalized frequency coordinate given in eq. (1).

spread in the transverse direction. The simulation results are shown in Fig. 1. We can see that our simulation replicates the analytic results well. We display the BTFs as is commonly done by splitting them into the absolute (amplitude) and angular (phase) part of the complex number.

For tune distributions we calculate the mean per-particle tunes for a subset of 10^4 particles using the FFT of the particle coordinates over 2^{11} turns and refine the peak frequencies using the bisection method to maximize the amplitude given by the DTFT.

MODEL FOR THE BEAM-BEAM EFFECT

In order to investigate BTF in RHIC a simplified model of the beam-beam effect was added to the code from [3]. In normal operation in RHIC, sets of six bunches (three per ring) couple with each other via the beam-beam interaction at two interaction points. For this a trivial parallelization scheme was implemented: Six copies of the original code are run, each utilizing the linear transfer map given by MADX [4] for the RHIC 2012 proton proton lattice [6] to translate a single bunch between interaction points. For simplicity we refer to a copy of the original code running for a single bunch simply as bunch. Communication between the bunches happens via the Message Passing Interface (MPI) [7]. At the interaction point two approximations to the beam-beam interaction were implemented and can be used for simulation:

Soft Gaussian approximation The beams are assumed to be of Gaussian transverse shape. Each bunch computes its center of charge position $\mu_{h,v}$ and rms width σ and exchanges them with its collision partner via MPI. The particles momentums are changed according to the analytic field amplitude for a Gaussian charge distribution [8] of given μ and σ.

Two-dimensional Fields Each bunch computes the particle density in the transverse plane on a discretized grid of 128^2 cells and by means of a two-dimensional Fast Fourier Transform (FFT) based Poisson solver the corresponding transverse field strength is derived. In the field calculation, open boundary conditions are assumed [9]. The field data is sent to the partner bunch via MPI and each bunch modifies particle momenta according to the field data.

In both of these the beam-beam interaction is approximated as a single kick on each particle, similar to what would happen if the bunches passed each other with all particles at the longitudinal position of the reference particle. We further assume that in the measurements we want to replicate, the excitation amplitude is sufficiently small to not influence the next BTF sample. This is justified from measurement where BTF can be repeated without significant change on the same beam.

The choice for the two-dimensionality of the interaction is made because of the computational resource limits: For computation of coherent BTFs a few thousand turns per scanned frequency need to be computed. At about 360 ms per turn of the scheme with 6 bunches, a BTF calculation takes quite some time. Simulation of a BTF including the beam-beam effect with 40 frequency samples for 100k particles and 3000 turns per excitation frequency takes approximately 11 hours on 6 CPUs. Usually we would like to have in the range of a hundred frequency samples to resolve subtleties of the spectra. In those cases we employ further trivial parallelization by running multiple simulations for different frequency ranges at once. An obvious extension of the code would include a sliced beam-beam interaction model which would allow for a more realistic simulation of the beam-beam interaction.

For the case with two interaction points, care has to be taken in the BTF model: In the real machine, the three bunches belonging to one beam pass the cells where excitation and recording of the response takes place at a distance of 1/3 of the ring circumference. To model this in simulation a phase shift in the excitation signal depending on the bunch number has to be introduced. Additionally it is necessary to preserve the order of the response data so that the later Fourier transform can easily be done over the combined position data of all three bunches, again corresponding to reality where each bunch contributes to the signal on the pickup.

Beam-beam Implementation Testing

In order to verify the correct implementation of the beam-beam effect we chose to simulate Schottky-type beam spectra. The simulation was run using the nominal RHIC lattice with identical tunes in the two rings, named Blue and Yellow of $Q_h = 28.695, Q_v = 29.685$. The beam-beam interaction was simulated but no BTF applied. We recorded the transverse position of the simulated bunch centres of charge every turn. Schottky spectra were derived using the FFT. In the resulting spectra, the positions of the coherent beam-beam modes were observed. The well-known expectations [10]

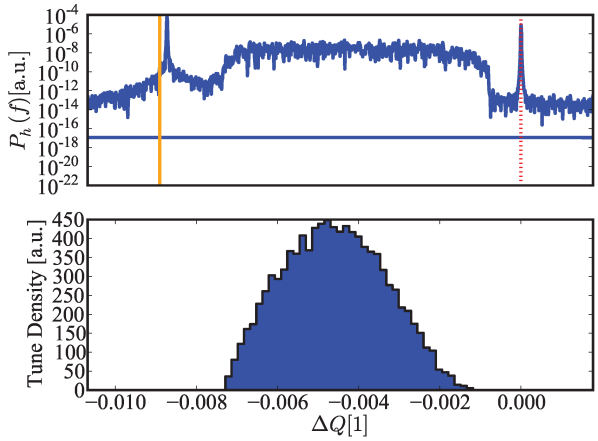


Figure 2: Example of a Schottky simulation of two bunches with one interaction point and identical tunes in both rings, the analytically expected position for the π-mode is shown in orange, the lattice tune (and expected position of the σ-mode) is indicated by a dashed line. The bottom plot shows the tune distribution. The simulation results correspond to the expectation. Similar results were obtained for the lattice with two interaction points when the beam-beam parameter was multiplied by the number of interactions per turn.

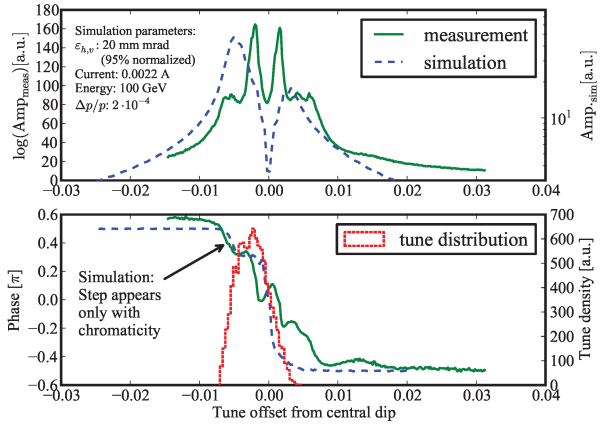


Figure 3: Comparison between simulated and measured BTF for the split tune conditions (tunes given in the text). The position of the central dip was set to zero to align measurement and simulation results.

for beams of identical tunes were met by simulation (the slight deviation is expected for beams with a beam-beam parameter above 0.01 [11]). As shown in fig. 2, the σ mode was found at the lattice tune, the π-mode was found at a the expected position for round beams of 1.215ξ with ξ the beam-beam parameter [10].

The beam-beam simulation using the Gaussian approximation gives a slightly smaller displacement of the π-mode with respect to the σ mode. This also is in accordance with expectation [11].

Comparison with Measurements

In March 2012 during the RHIC polarized proton-proton run there were test runs by S. White et al. [12] using a lattice different from the nominal one: The tunes in the blue ring ($Q_h = 28.691$, $Q_v = 29.689$) were chosen far from the tunes in the yellow ring ($Q_h = 28.735$, $Q_v = 29.725$). Measured BTF in the vertical plane showed an unusual peak under these conditions. We tried to replicate these settings using our simulation. The results are shown in fig. 3. While in the measurement the structure only shows up in one plane, in the simulation it is visible in every plane. However this is not too surprising: While in the real ring we have nonlinearities possibly damping coherent modes, our model only takes into account the beam-beam effect, chromaticity and the linear transfer matrices between the interaction points. Varying chromaticity we noticed that the step to the left of the phase jump in fig. 3 only appears after introduction of a finite chromaticity, the plot is for a chromaticity $\Delta Q/(\delta p/p)$ of 2. A rigid bunch model simulation [12] agrees in the prediction of coherent modes in all planes.

We consider the remaining differences to measurement are caused by the simplifications of our simulation, where we assume a linear lattice apart from the beam-beam effect.

OUTLOOK

With our BTF simulation tool we will investigate possible diagnostic opportunities of the BTF with respect to the incoherent tune distribution. This is desirable as diagnostics for the electron lens for head-on beam-beam compensation [13] to be installed at RHIC in the near future.

ACKNOWLEDGEMENTS

P. Görgen wishes to thank M. Blaskiewicz for many helpful suggestions during the implementation.

REFERENCES

[1] S. Paret et al., Phys. Rev. STAB. 13, 2010, pp. 022802.1ff.

[2] M. Gasior, R. Jones *CERN project report 853*

[3] O. Boine-Frankenheim, V. Kornilov, Proc. of ICAP, 2006, pp. 267ff.

[4] `http://madx.web.cern.ch/madx/`

[5] K.Y. Ng, *Physics of Intensity Dependent Beam Instabilities, Chapter 13*, World Scientific, Singapore, 2006

[6] S. Tepikian, private communication

[7] `http://mpi-forum.org`

[8] M. Bassetti, CERN-ISR-TH/80-06

[9] Kapin, V., Proc. of EPAC, 2002, pp. 1637ff

[10] K. Yokoya, K.Haruyo, Part. Accel. 27, 1990, pp. 181ff.

[11] K. Yokoya, Phys. Rev. STAB. 3, 2000, pp. 124401-1ff.

[12] S. White et al., Proc. of IPAC 2012, pp. 195ff

[13] W. Fischer et al., Proc. of PAC 2011, pp. 2223ff

DYNAMICS OF FERRITE CAVITIES AND THEIR EFFECT ON LONGITUDINAL DIPOLE OSCILLATIONS*

C. Spies†, U. Hartel, M. Glesner, TU Darmstadt, Darmstadt, Germany
H.-G. König, H. Klingbeil, GSI, Darmstadt, Germany

Abstract

In this paper, we consider ferrite cavities of the type that is currently used in the SIS18 at GSI and will be used in the future SIS100 which is being built in the frame of the FAIR project. We analyze the dynamics of the cavities in conjunction with their local control loops. An emphasis is put on the cavities' reaction to changes in the desired amplitude or resonant frequency. Using simulations, we show that the cavities' dynamics hardly influence longitudinal dipole oscillations, and conclude that a low-order model for the RF cavities is sufficient.

FERRITE CAVITIES

In a heavy-ion synchrotron, the frequency of the RF accelerating fields is typically relatively low ($< 10\,\mathrm{MHz}$) and highly variable [1]; for instance, in the planned SIS100 synchrotron at the GSI Helmholtz Centre for Heavy-Ion Research, $^{238}U^{28+}$ ions will be injected at about $0.56 \cdot c_0$ and accelerated to about $0.92 \cdot c_0$. Ferrite cavities for low RF frequencies are significantly smaller than, e. g. a $\frac{\lambda}{2}$ resonator cavity, and their resonance frequency can be tuned.

In a ferrite cavity, the metal beam pipe is interrupted by a short ceramic "gap" and surrounded by ferrite rings. Bias and excitation loops are wound around the ferrite rings (see Fig. 1). The RF current excites an RF magnetic field in the ferrite rings, which in turn induces an RF electric field along the beam pipe. This field enters the ceramic gap and accelerates the charged particles. The RF field also effects a reactive current in the metal beam pipe and casing.

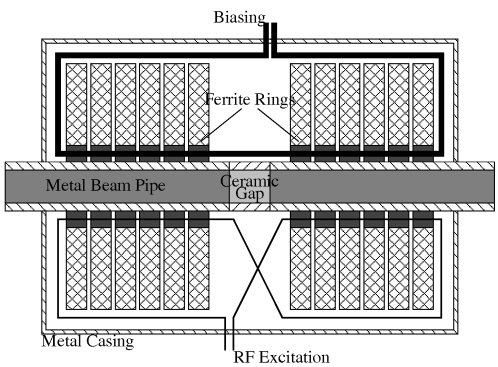


Figure 1: Schematic diagram of a ferrite cavity.

The cavity's behavior is equivalent to an RLC parallel oscillator (see Fig. 2) whose input is the RF excitation current [1]. The voltage across the RLC circuit is the voltage across the ceramic gap. The capacitance is primarily that of the ceramic gap and is constant. The inductance depends on the differential magnetic permeability of the ferrite material and can be tuned by the bias current. The resistance represents both Ohmic and magnetic losses and may be frequency and/or voltage dependent.

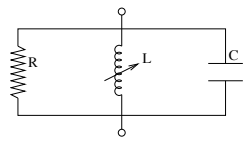


Figure 2: Equivalent circuit.

Resonance Frequency Control

From the equivalent circuit in Fig. 2, it is obvious that the accelerating voltage is maximized when the cavity is operated at its resonant frequency. In that case, the voltage is in phase with the excitation current. Therefore, each cavity is equipped with a local controller that measures the phase difference between excitation current and gap voltage. It outputs a low-level voltage that controls a solid-state amplifier which in turn generates the bias current.

The controller is an analog circuit based on operational amplifiers (OpAmps) and has a linear PI characteristic, i. e. its step response is the sum of that of a proportional (P) and that of an integral (I) controller. Additionally, there is a pilot control path that nonlinearly maps the setpoint frequency to an approximation of the required bias current.

Amplitude Control

Each cavity is also equipped with a local amplitude controller that measures the gap voltage (using a voltage divider) and compares it to the voltage setpoint. It outputs a low-level voltage that is used to modulate a low-level RF input signal, which is in turn fed into an amplifier chain that generates the excitation current.

The controller is an analog OpAmp circuit and has a linear PPT_1 characteristic, i. e. its step response is the sum of that of a P element and that of another P element with first-order lag (PT_1). Additionally, there is a constant-gain pilot control path.

Cavity Synchronization

Since the resonance frequency control described in section cannot detect phase shifts in the modulator-amplifier chain, another local controller measures the phase difference between the gap voltage and a reference signal and

* Work supported by the German Ministry of Education and Research, Grant No. 06DA9028I
† christopher.spies@mes.tu-darmstadt.de

adjusts the phase of the input signal in order to match the gap voltage's phase to that of the reference signal.

NONLINEARITIES

Many of the involved components exhibit nonlinear characteristics, some of which have been summarized in tbls. 1, 2 and 3.

Table 1: Frequency Dependency of Modulator / Amplifier Characteristics

Frequency	Amplifier Gain	Phase Shift
0.8 MHz	55 S	8.6°
2.0 MHz	31 S	54°
4.0 MHz	33 S	115°

Table 2: Frequency Dependency of Resonance Frequency Control Characteristics

Frequency	Amplifier Gain	Pilot Control Error
1.0 MHz	128 S	−5%
2.0 MHz	200 S	−22%
3.0 MHz	380 S	−23%
4.0 MHz	420 S	−20%

Table 3: Frequency Dependency of a Cavity's Resonance Frequency

Frequency	Bias Current	Inductance
1.0 MHz	20 A	33.4 μH
2.3 MHz	100 A	6.3 μH
2.9 MHz	150 A	4.0 μH
3.9 MHz	250 A	2.3 μH
4.8 MHz	400 A	1.5 μH

Experiments have shown the resistance in Fig. 2 to depend both on the operating frequency and on the voltage across the cavity. Figure 4 visualizes this dependency.

SYSTEM BEHAVIOR

Experiments conducted at GSI have shown that the amplitude control's plant (consisting of the amplitude modulator, the amplifier chain, the cavity and the voltage divider) exhibits a second-order frequency-dependent time lag of the order of some 10 ns and a dead-time of 3.9 μs. Only a part of the latter (about 1 μs) is caused by the cable delay to and from the cavity. The remainder must be attributed to an inherent dynamism of the plant.

The closed-loop resonance frequency control is capable of overshooting; its attenuation factor is 0.458, and its resonant frequency is 3.15 kHz in small-signal analyses.

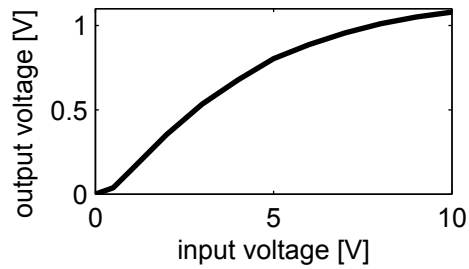


Figure 3: Modulator input-output characteristic.

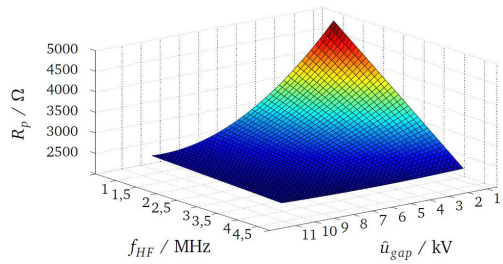


Figure 4: Cavity resistance.

Larger time constants are observed for larger steps in the frequency setpoint due to the limited slew rate of the bias current.

SIMULATION MODELS

We created two different simulation models. Each includes 14 cavities as well as their local control loops. The cavities were assigned random parameter deviations of up to 20%. In the first model, all the known nonlinearities as well as the dynamism of the cavity with respect to the amplitude have been included; Spline interpolations between the data points in sec. were used. In the second model, all input-output characteristics have been assumed to be linear, and all gains have been assumed to be constant. Their values have been chosen such that the open-loop gain becomes identical to the minimum open-loop gain of the nonlinear system. Therefore, the closed-loop time constant of the linear model equals the worst-case time constant of the nonlinear model.

We used preliminary ramp data for the future SIS100 synchrotron (under construction at GSI) as an input to the simulation. Figure 5 shows the RF amplitude and RF frequency setpoints during a $^{238}U^{28+}$ simulation cycle. We are primarily interested in the amplitude and phase of the effective RF voltage acting on the particles.

SIMULATION RESULTS

Figure 6 shows the actual RF amplitude obtained by simulating the first model along with the difference between both simulations. The maximum difference is about 1.5 kV. The relative difference is below 1% of the total RF amplitude except at the very end of the simulation, after flat-top has been reached.

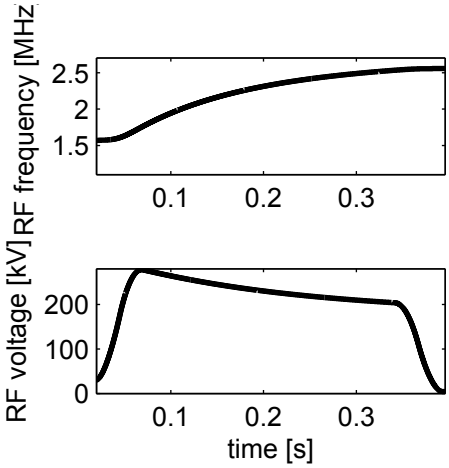


Figure 5: Frequency (top) and amplitude (bottom) ramps.

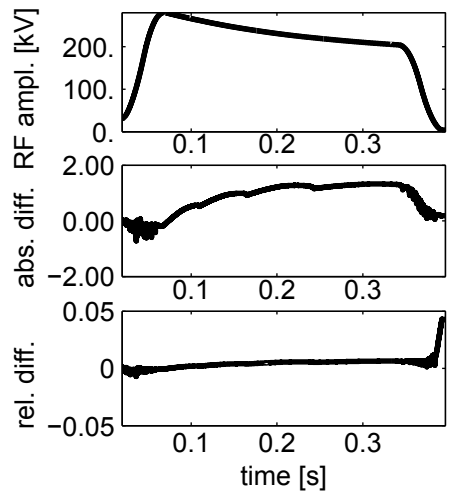


Figure 6: Actual RF amplitude (top), absolute difference (center) and relative difference (bottom).

Figure 7 shows the deviation of the RF phase from its setpoint in both simulations. The maximum deviation in both simulations is only about $0.5°$, which is below the accuracy of about $2°$ that the cavity synchronization can achieve in practice [2]. This difference is hardly noticeable.

These results suggest that the aforementioned nonlinearities have little impact on the RF amplitude and phase and consequently should have little to no impact on the beam. We therefore also simulated the influence of the RF voltage on the longitudinal motion of a single macro-particle. The resulting deviation of the beam phase from its setpoint in both simulations is shown in Fig. 8. There is no quantitative difference between both plots.

CONCLUSIONS

From the simulation results presented in this paper, we conclude that the second model (neglecting the nonlinearities) is a good fit. The local control loops mask the effect of the nonlinearies; the linear model would be a poor fit for the open-loop system. Neglecting the nonlinearities reduced the required simulation time by about 20%.

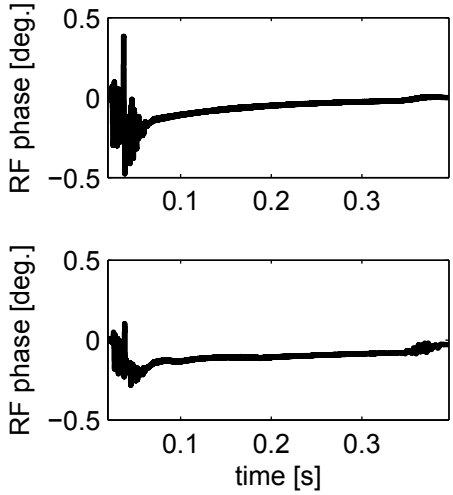


Figure 7: Actual RF phase.

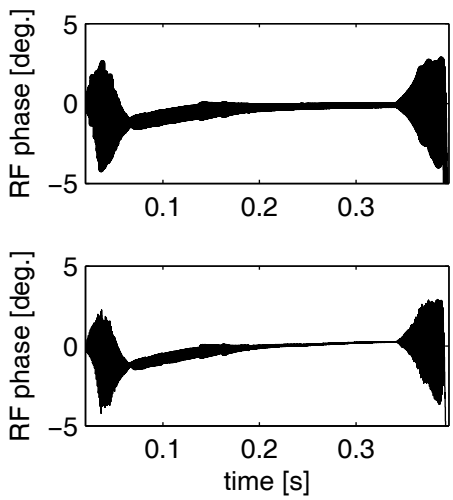


Figure 8: Beam phase.

ACKNOWLEDGMENTS

The authors would like to thank Martin Kumm for the system identification of the amplitude control plant and Jörg Mohr and Thomas Wöber for determining the input-output characteristics of the amplitude modulator.

REFERENCES

[1] H. Klingbeil, "Ferrite Cavities," *Proc. CERN Accelerator School*, 2010, pp. 299–317.

[2] H. Klingbeil, H. Damerau, M. Kumm, P. Moritz, G. Schreiber, B. Zipfel, "Commissioning of the SIS12/18 Cavity Synchronization System," GSI Scientific Report, 2005, p. 132.

DESIGN OF A COMPUTER BASED RESONATOR-SIMULATOR FOR TESTS OF RF CONTROL SYSTEMS*

T. Bahlo, C. Burandt, R. Eichhorn†, J. Enders, M. Konrad, P. Nonn
Institut für Kernphysik, TU Darmstadt, Germany

Abstract

In order to test RF control systems for accelerator cavities without being dependent on available prototypes, a resonator-simulator has been developed. The Simulator is based on a Xilinx VIRTEX-4 FPGA module and has been configured using MATLAB Simulink with a special Xilinx blockset. The underlying model for this configuration is a parallel RLC circuit that has been parameterised with common RF quantities like the resonance frequency, driving frequency, bandwidth and quality factor. This approach allows the behaviour of normal conducting cavities with quality factors up to 10^4 as well as superconducting cavities with quality factors up to 10^9 to be simulated. Besides, it can be operated in continuous-wave mode as well as in pulsed mode. We report on the mathematical model, its digital representation and on the benchmarking against realistic cavity behaviour.

INTRODUCTION

For experiments with intense antiproton beams, a dedicated proton LINAC is foreseen as an injector for the future Facility for Antiproton and Ion Research (FAIR). The proposed scheme with normal conducting crossed-bar H-mode cavities operated in a pulsed mode is planned to feed directly into the heavy-ion synchrotron SIS18 of the present GSI Helmholtz center for heavy-ion research [1]. The corresponding digital radio frequency control system for a test-bench setup is currently being developed at TU Darmstadt [2]. Due to the fact that there are no cavity prototypes available, a cavity simulator has been developed. This device is able to simulate normal conducting as well as superconducting cavities with common physical effects like beam loading, amplifier saturation and Lorentz-force detuning (LFD). In consequence of its ability to simulate superconducting cavities it also can be used to improve the digital control system [3] of the Darmstadt electron LINAC S-DALINAC [4] without being dependent on an available superconducting resonator or RF periphery.

THE RESONATOR MODEL

The simulator is based on a resonator model consisting of a parallel circuit of a resistor R, an impedance L and a capacity C that is driven by a generator current I_{gen} as shown in Fig. 1. The corresponding differential equation describing the voltage reads:

$$\ddot{U} + \frac{1}{RC}\dot{U} + \frac{1}{LC}U = \frac{1}{C}\dot{I}_{\mathrm{gen}} \qquad (1)$$

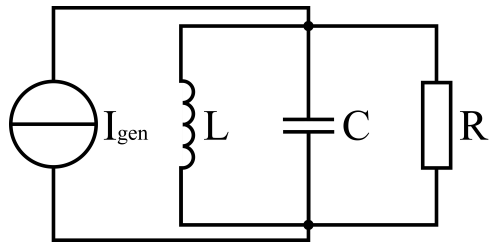


Figure 1: The equivalent circuit representing the cavity

It is convenient to rewrite this equation with more common RF parameters like the resonance frequency ω_r, the quality factor Q and the half width at half maximum $\Delta\omega_{1/2}$. With these the equation reads:

$$\ddot{U} + \frac{\omega_r}{Q_L}\dot{U} + \omega_r^2 U = \frac{\omega_r}{Q_L}\dot{U}_{\mathrm{gen}} \qquad (2)$$

To solve Eq. (2) one can use Eq. (3) as an ansatz which is motivated by the in-phase and quadrature procedure commonly used with RF control systems. The in-phase signal is denoted by V, while W represents the quadrature component.

$$U(t) = (V + iW) \cdot e^{i\omega t} \qquad (3)$$

Time discretisation and the introduction of several approximations (details described, e.g., in [5]) lead to:

$$V(t+\Delta t) = V(t) + \left[\omega_{1/2}\left(V_{\mathrm{gen}} - V(t)\right) - \Delta\omega W\right]\Delta t$$
$$W(t+\Delta t) = W(t) + \left[\omega_{1/2}\left(W_{\mathrm{gen}} - W(t)\right) + \Delta\omega V\right]\Delta t$$

HARDWARE

The simulator has been implemented on an Lyrtech VHS-ADAC board providing 8 DAC and 8 ADC channels with a resolution of 14 bit and a Xilinx VIRTEX-4 FPGA-module. The latter has been configured using the "Xilinx System Generator for DSP" [6] library for MATLAB/Simulink. This software package allows the designer to generate high-performance HDL code directly from the Simulink interface using pre-optimized algorithms for the Xilinx hardware. Using this tool greatly simplifies the configuration of the FPGA.

COMPARISON TO A REAL RESONATOR

A pulsed control system has to be able to tune the cavity's electrical field magnitude and phase to the wanted value as fast as possible. During this process it reacts on fast changes of these quantities. So for the simulator it is very crucial to reproduce the transient response of a real cavity with high precision.

* Work supported by the BMBF contract 06DA9024I
† now at: CLASSE, Cornell University, Ithaca, NY 14853, U.S.A.

On account of this the step-function response has been measured for the simulator and for a normal conducting pre-buncher with RF parameters shown in Table 1. The measurement was performed with resonant excitation as well as with non-resonant excitation by detuning the input signal by the amount of a full width at half maximum of the cavity. The results are shown in Figs. 2 and 3. For the shown dataset the resonantly excited curve holds an RMS error of $\Delta_{\mathrm{RMS}} = 6.9 \cdot 10^{-3}$ while the set with non-resonant excitation exhibits an error of $\Delta_{\mathrm{RMS}} = 4.02 \cdot 10^{-2}$. The latter error mainly results from the first μs.

Table 1: RF Parameters of the Prebuncher Used for the Comparison Measurement

Resonance frequency	(2994.1 ± 0.475) MHz
Bandwitdth (FWHM)	(475 ± 20) kHz
Loaded quality factor Q_L	6300 ± 265

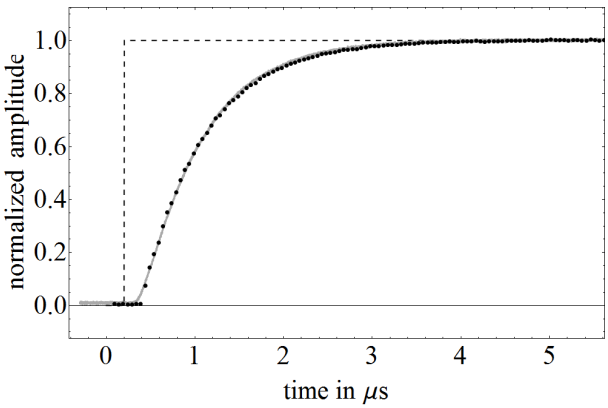


Figure 2: Response of the simulator (dotted) and a real cavity (gray) for resonant excitation by a step function (dashed)

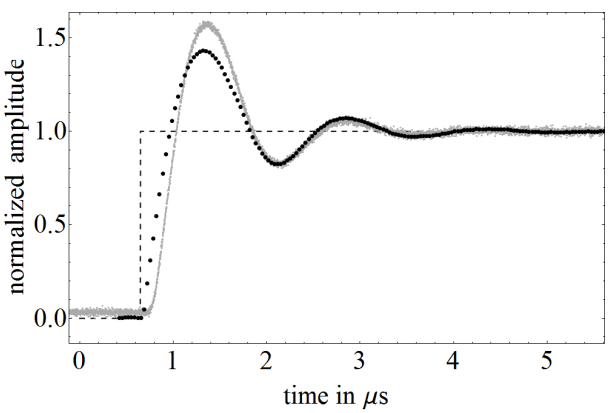


Figure 3: Response of the simulator (dotted) and a real cavity (gray) for non-resonant excitation following a step function (dashed). The used input signal was detuned by $\Delta f = \Delta f_{\mathrm{FWHM}}$ with Δf_{FWHM} denoting the full width at half maximum of the resonance centroid.

AMPLIFIER SATURATION

When RF amplifiers are operated near their maximum output level, most of them show a certain amout of saturation. This means the output characteristics is non-linear for high output levels. Hence, the control path for the control system also becomes non-linear. This has to be taken into account while developing new control systems. Therefore the simulator was implemented to reproduce this behaviour, too. It is possible to enter a four-point approximation of a known saturation curve into the simulator. The approximation is then used to transform the output of the simulator. The level of accuracy for the approximation, i.e., the number of points describing the output characteristics of the amplifier, is limited by the hardware of the simulator.

BEAM LOADING

A beam of charged particles traveling through a cavity interacts with the electro-magnetic field inside and converts field energy into kinetic energy. As the field amplitude inside the cavity decreases, the control system is in charge to increase the input power fast enough to prevent the amplitude from dropping significantly.

To be able to test this beam-loading effect with the simulator, an additional current source that has an opposite sign compared to the generator-current source, while sharing the same phase, is introduced. The length, period and delay of the virtual particle beam is determined by an internal counter. To achieve synchronisation between the external generator current and the internal beam counter, a trigger input port is used to reset the counter periodically. Figure 4 shows a result of the simulator inducing a beam-loading effect.

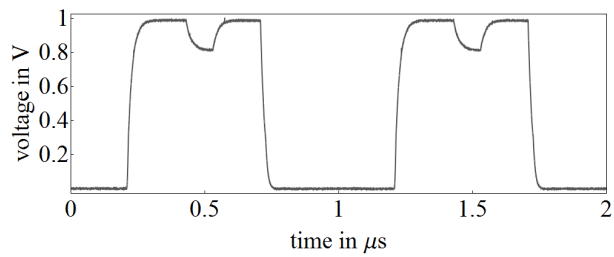


Figure 4: Output voltage of the simulator with beam loading

LORENTZ-FORCE DETUNING

For the simulation of superconducting cavities the transient response is less important, due to their high quality factor and slow step response. The small bandwidths, on the other hand, favour problems that affect the resonance frequency. For the mechanically rather unstable superconducting cavities any stress results in a relevant shift of the resonant frequency. Even the induced electromagnetic field deforms the geometry of the resonator and shifts the resonance frequency by a significant amount.

This Lorentz-force detuning is a current problem in accelerator physics [7]. The simulator uses the fact that LFD shifts the resonance frequency proportional to the square of the electric field amplitude in the cavity:

$$\Delta\omega_{\mathrm{LFD}} = -2\pi \cdot K \cdot E_{\mathrm{acc}}^2 \tag{4}$$

The factor K represents a constant of proportionality called "Lorentz-force-detuning constant". With the informations from Eq. (4) the simulator is able to reproduce this effect as shown in Fig. 5.

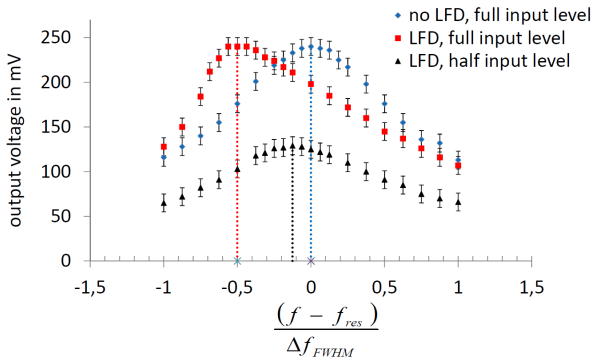


Figure 5: Measurement of the frequency shift due to LFD. The output voltage is shown as a function of the relative frequency, for the maximum amplitude input (squares) and half of the maximum amplitude (triangles). For comparison, the resonator resonance curve without LFD is indicated by diamond symbols.

GUI

The Xilinx System Generator for DSP provides the opportunity to communicate with the FPGA via Simulink. This allows the designer to create a graphical user interface, that simplifies the use of the simulator drastically. The current version of the GUI allows to enter the resonance frequency of the simulator, the parameters of the virtual beam like its length, period and magnitude and the Lorentz-force-detuning constant. Furthermore the GUI allows the user to read out some of the internal parameters of the simulator like the current resonance frequency, and the current frequency detuning.

CONCLUSION

The developed flexible cavity simulator is able to simulate normal conducting cavities as well as superconducting cavities. In both cases it is able to simulate relevant physical effects known from practical experience. The simulator can be used to easily test or improve control systems, without being dependent on RF or cryogenic equipment. In addition to that the strength of the simulated effects can be varied over a great range which would be difficult in reality. Furthermore the usage is easy and intuitive.

REFERENCES

[1] L. Groening, Technical Design Report FAIR Proton-Linac, Helmholtzzentrum für Schwerionenforschung GmbH, `http://www-win.gsi.de/fair-eoi/PDF/TDR_PDF/TDR_pLinac.pdf`

[2] M. Konrad, Development of a Digital Low-level RF Control System for the p-LINAC Test Stand at FAIR. Proc. IPAC, New Orleans, LA, 2012, p. 3461.

[3] M. Konrad et al., Digital base-band rf control system for the superconducting Darmstadt electron linear accelerator, Phys. Rev. ST Accel. Beams 15 (2012) 052802.

[4] A. Richter, Operational experience at the S-DALINAC, Proc. EPAC, Sitges, Spain, 1996, p. 110.

[5] T. Schilcher, Vector Sum Control of Pulsed Accelerating Fields in Lorentz Force Detuned Superconductiong Cavities, doctoral dissertation, Universität Hamburg, 1998.

[6] Xilinx System Generator for DSP, `http://www.xilinx.com/tools/sysgen.htm`

[7] V. Ayvazyan, Dynamic Lorentz Force Detuning Studies in TESLA Cavities, Proc. EPAC, 2004, Lucerne, Switzerland, 2004, p. 994.

SPACE CHARGE AND ELECTRON CLOUD SIMULATIONS

G. Franchetti*, GSI, Darmstadt, Germany
F. Zimmermann, CERN, Geneva, Switzerland

Abstract

Tracking of high intensity effects for few turns of a circular accelerator is at reach of present computational capabilities. The situation is very different when the prediction of beam behavior is extended to hundred of thousands of turns, where special approaches for the control of computer artifact are necessary sometimes to the expense of a complete physical modeling. The identification of the key physical ingredients helps to the development of computer algorithms capable of treating the long term tracking.

Here we present the latest status of long-term tracking simulations of high intensity bunches for SIS100, and also discuss a more realistic modeling of the incoherent effect of an electron cloud for the LHC.

INTRODUCTION

The motivation of these studies has started with the approval of the FAIR project at GSI. In the SIS100 synchrotron of the FAIR project [1] bunches of U^{28+} ions are stored for about one second and then accelerated: During this cycle beam loss cannot exceed 10% [2, 3, 4].

The simultaneous presence of space charge and the lattice induced nonlinear dynamics may create a diffusional regime leading to beam loss [5, 6]. The modeling of this unusual beam operation has been subject of studies in the past decade and relies on analytic models [7], which allow the suppression of a characteristic noise of PIC codes [8]. This approach neglects the coherent effects created by the Coulomb force as, for instance, the excitation of unstable modes of the beam [9]. On the other hand in the PIC self-consistent studies in Ref. [10] it is also shown that coherent nonlinear resonances are not of relevance for Gaussian beam distributions. Therefore in all studies for SIS100 it was adopted a modeling of "frozen" type and a quest to understand the dynamics was undertaken.

Over the years a systematic approach has been taken which had a two-fold purpose:

1) Understand the basics of the mechanism (lattice resonance affected by space charge created by a bunched beam) in order to more effectively improve SIS100 performances and to device possible cure for too high beam loss level [7]. Up to now the studies to understand the mechanisms of periodic resonance crossing have been confined to "frozen" algorithms.

2) Benchmarking the code prediction against experiments performed on a real machine. These studies started with an experimental campaign at the CERN-PS in the years 2002-2003 [11]. Few years later a new campaign was made at GSI in the SIS18 [6]. The results of these campaigns have shown that the simulation do not predict correctly the beam loss. The experiments always yield more beam loss of what is predicted with a computer code. The reason for this discrepancy is not fully understood as to the imperfection of the modeling implemented in computers also one has to account an imperfect knowledge of the real machine used for the benchmarking.

In the year 2004-2005 it was suspected that a similar diffusional mechanism could be created by the electron cloud during the electron pinch process driven by the passage of a proton bunch. Studies presented in Ref. [12] have shown that the effect of a resonance crossing is indeed at the base of a slow incoherent emittance growth. The issue of the modeling of the electron pinch is very central for the emittance growth prediction. The complexity of the dynamics of the electrons during the pinch process makes hard to use an analytic modeling. Attempts to investigate the emittance growth with analytic modeling of the electron pinch are reported in Ref. [5]. Differently from the space charge, the benchmarking with a beam experiment is here much more difficult because the knowledge of all electron cloud parameters is very difficult.

It is worth mentioning that the synergy between electron cloud and space charge studies allowed the development of a new theoretical framework which redefines the concept of fixed points for non adiabatic resonance crossing [13].

In these proceedings we report the state of the simulations for the SIS100 and we address recent development in the modeling of electron-cloud incoherent effects for the LHC.

SPACE CHARGE SIMULATION

As result of the studies on the periodic resonance crossing and all the considerations on nonlinear and high intensity effects the working point of the SIS100 was chosen as $Q_{x/y} = 18.84/18.73$. The studies presented in Ref. [5, 6] estimated the SIS100 beam loss, however without clear evidence that periodic resonance crossing is the issue. In recent studies we have found that this is the case.

Modeling of the Dynamics at Injection

Random errors, reference beam, and reference error seed. In SIS100 the nonlinearities are given by standard multipoles in sc dipoles [14, 15] now optimized with respect to those in Ref. [5, 6], and by the multipoles for sc quadrupoles [16]. Chromatic correction sextupoles are ignored. The systematic multipoles yield a short term dynamic aperture (10^3 turns) of 5.3σ for a reference beam of 8.75 mm-mrad rms emittance with the beam magnetic

* g.franchetti@gsi.de

rigidity at injection of 18 Tm. Magnet random errors (MRE) are introduced through a ±30% fluctuation for all computed multipoles of the sc dipoles [17]. Skew components, where missing, are introduced of the same rms strength as the corresponding normal. Also unavoidable residual closed orbit distortion (RCOD), after correction are included. We consider a reference vertical RCOD of 1 mm rms (1.6 mm horizontal), which contains 95% of the associated RCOD distribution. The feed down of magnets components for magnets displacement of $d_{x,rms} = d_{y,rms} = 0.32$ mm and MRE yields an average DA of $\simeq 4\sigma$ with a variance of $\simeq 0.2\sigma$, with a minimum at 3.4σ. We model the bunched beam with a Gaussian transverse distribution truncated at 2.5σ in amplitudes as result of a controlled beam shaping during transfer from SIS18 to SIS100. The reference emittances (2σ) are $\epsilon_{x/y} = 35/15$ mm-mrad (edge at $2.5\sigma <$ DA=3.4σ). For the next simulation we select a "reference error seed" (only MRE). The resonances excited by this seed are shown in Fig. 1 (left).

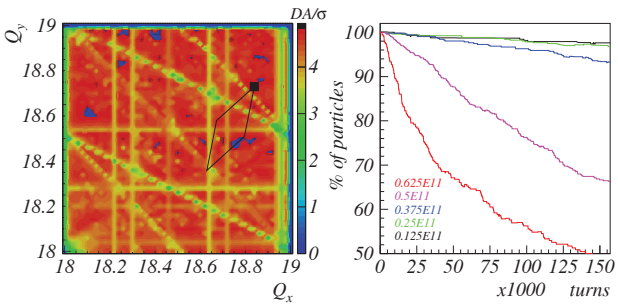


Figure 1: DA scan for the reference error seed and the expected tune-spread (left). First bunch survival evolution for several beam intensities (right).

Space charge induced beam loss Simulations with SC are made with MICROMAP including all previously discussed effects for the "reference error case". The SC is computed with a frozen model, which incorporates the local beam size defined by the beam optics [5, 6]. The space charge calculation are performed in the beam center of mass. For the total maximum nominal intensity of 5×10^{11} of U^{28+} in 8 bunches the SC peak tune-shifts are -0.21 / -0.37. In order to make sure that the space charge algorithm does not introduce artifacts we made a simulations in absence of lattice nonlinearities finding no beam loss in 1.57×10^5 turns. The beam survival at the end of the cycle (8 bunches) $N_T(t_{end})/N_T(inj)$ is obtained from the beam survival of the first bunch $N(t)/N_0$, with N_0 the number of particles in the first bunch, via the formula $N_T(t_{end})/N_T(inj) = 1/4 \sum_{i=1}^{4} N(t_{end} - t_i)/N_0$, with t_i injection time. In Fig. 1 (right) the first bunch survival is shown for the intensities: $0.625, 0.5, 0.375, 0.25, 0.125 \times 10^{11}$ ions. As shown by Fig. 1 (left), the SC dominated loss may be a result of the periodic crossing of: the second order resonance $2Q_y = 37$, the third order resonances $Q_x + 2Q_y = 56, 3Q_y = 56$, the fourth order resonances $2Q_x + 2Q_y = 75, 4Q_x = 75$. It should be noted here that the simulation model employed in this study lacks dynamical self-consistency. This is not expected to matter for losses at or below the few percent level. However, for larger losses, as for the cases $0.5, 0.625 \times 10^{11}$ ions, inclusion of full self-consistency (e.g. updating the SC force as a consequence of losses) could easily enhance or diminish the loss rate.

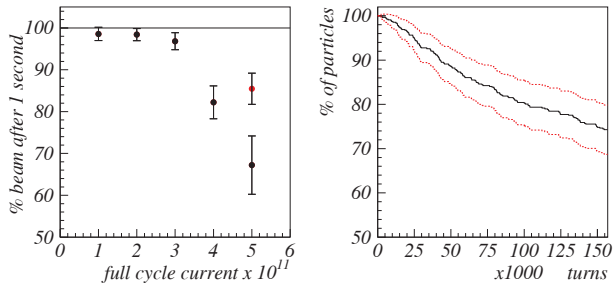


Figure 2: Left) Summary of the beam survival at the end of the cycle. Right) Beam survival of the first injected bunch with 0.625×10^{11} ions for dipoles with all multipoles reduced by a factor of 2.

Beam loss mitigation As in absence of lattice nonlinearities no beam loss is found, we first considered ideally improved dipoles. By reducing the nonlinear components of the dipoles by a factor 2 a simulation of the 0.625×10^{11} ions intensity, in Fig. 2 (right), shows a beam survival of $75\% \pm 2\%$ against the previous $\simeq 48\% \pm 2.7\%$ in Fig. 1(left) [error bars are described in Ref. [5, 6]]. In Fig. 2 (left) this is shown over the the full cycle by a red marker. We conclude that: 1) Better dipoles significantly improve the beam survival; 2) This finding does not yet prove that periodic resonance crossing is the underlying beam loss mechanism.

A more "realistic" approach, but still simplified, consists in removing only the 3rd order component in the dipoles. We find that, as expected, most of the 3rd order resonances vanish leaving the dynamic aperture unchanged [see Fig. 3 (left)]. A simulation of the first bunch for the intensity

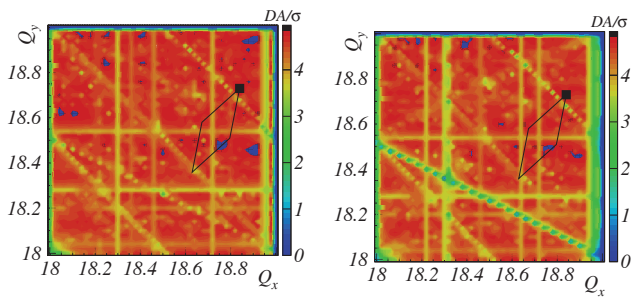


Figure 3: DA scan obtained by removing the 3rd order components in dipoles (left); Right) DA scan obtained by correcting $Q_x + 2Q_y = 56, 3Q_x = 56$.

0.625×10^{11} ions shows that the beam survival raise now to $97\% \pm 0.6\%$. This test proves that the third order resonances + space charge are responsible of the long term beam loss.

We then developed a resonance compensation scheme to reduce the strength of the 3rd order resonances $Q_x + 2Q_y =$

$56, 3Q_x = 56$, which cross the space charge tune-spread [Fig. 1(left)]. This approach was already suggested in Ref. [5, 6], but never implemented. We computed the driving term of the reference error seed, and those created by each of 12 dedicated corrector sextupoles. The compensation strategy is to cancel the total driving term at $Q_{x,c} = Q_{y,c} = 18.66$, the interception of the two resonances we intend to mitigate. The requirement is to reduce the total driving term at $(Q_{x,c}, Q_{y,c})$ leaving unaffected the dynamic aperture. After applying the correction scheme a new DA scan [see Fig. 3 (right)] confirmed the effectiveness of the resonance compensation: The resonances $Q_x + 2Q_y = 56, 3Q_x = 56$ have been compensated [compare with Fig. 1(left)]. We then repeated the simulation made in Fig. 1 (right) for the maximum intensity case and show the beam survival in Fig. 4 (right): We find that the beam survival rises to $97\% \pm 0.3\%$. Fig. 4 (left) shows the beam survival for the same beam but without resonance compensation [in blue the same curve of Fig. 1(right)].

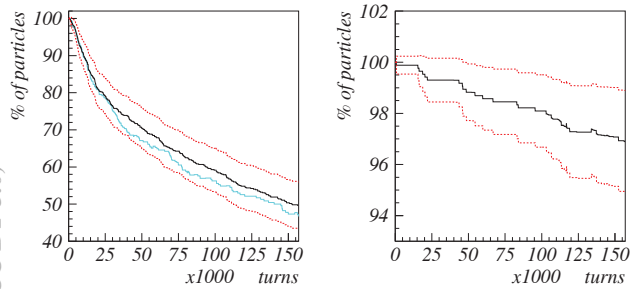


Figure 4: Survival of the first bunch beam for the case 0.625×10^{11} ions, without resonance compensation (left), and with resonance compensation (right).

Modeling of the Acceleration

After the last bunch is injected, the acceleration ramp of 4T/s starts [see Fig. 5]. During acceleration several processes happen simultaneously. We study here the acceleration without any beam loss mitigation scheme (resonance compensation). Our modeling rely on the following approximations/assumptions:

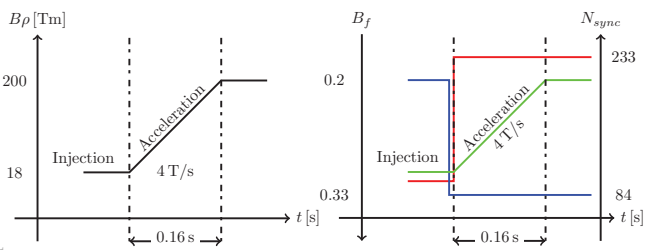


Figure 5: Schematic of the acceleration ramp (left); Change of bunching factor and of the synchrotron tune (right).

1) The SIS100 modeling is the same as the reference scenario, i.e. chromaticity, dispersion, RCOD, and MRE seed are included.

2) We assume at the beginning of the ramp the beam of the reference scenario. However, the longitudinal distribution is now rms matched to the acceleration bucket (change of bunching factor and synchrotron period, see Fig. 5).

3) The modeling of the acceleration takes into account of: a) The transverse beam emittance shrinking with $\beta\gamma$; b) The reduction of the space charge $\propto \gamma^{-2}$; c) The scaling of the synchrotron tune according to $(\beta^2\gamma)^{-1/2}$ in a linear bucket; d) The dynamic change of the dipole magnets multipole with $B\rho$ [14]; e) We also include the contribution of the eddy current, which we keep constant throughout the acceleration [18];

In order to assess possible beam loss during acceleration and to evaluate the effect of the fast ramping, simulations have been performed for the bucket used at the injection, and the bucket used during the ramp. We also computed the beam loss in presence or absence of the eddy current. The model with the bucket of the storage and no eddy current represents the case of an ideal adiabatic acceleration ramp. The algorithm used in the code is illustrated in Fig. 6

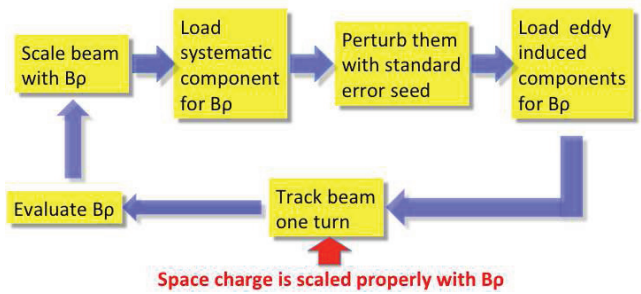


Figure 6: Computer algorithm used for modeling the acceleration dynamics with space charge.

Simulations show that for the adiabatic ramp beam loss is smaller than 1%, even adding the $4T/s$ eddy current. This is attributed to the fast damping of SC $\propto \gamma^{-2}$. Different is the case when the consistent bucket is used: The short bucket increases the space charge tune-spread $\simeq 60\%$ and 4% beam loss is found in the first 10^4 turns [Fig. 7(left)]. The more conservative case is obtained by the simultaneous presence of a small bucket and eddy current with an increase of beam loss to 5% [Fig. 7(right)]. These results indicate that beam loss for the reference beam should be expected in the level of $5 \pm 3\%$ in the first half of the ramp for the last injected bunch.

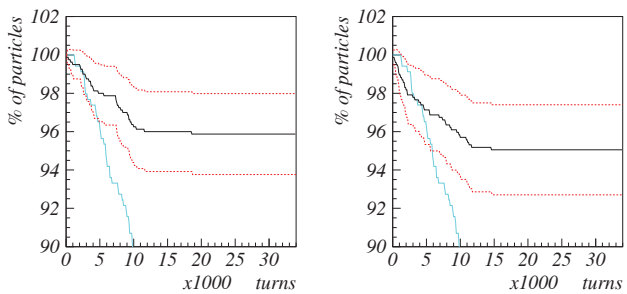


Figure 7: Beam loss during the $4T/s$ ramp without including eddy current (left). On the right picture the systematic eddy current is included. In blue the beam survival at injection plateau (for comparison).

Conclusion/Outlook

Our studies confirm that the working regime of SIS100 is subjected to a space charge induced periodic resonance crossing. For the selected "reference scenario" we proved that a proper compensation of the resonances across the tune-spread mitigates the damaging effect to 2.5% beam loss (5 % with safety margin). A preliminary study of the acceleration shows that beam loss of the order of 5% is found ($\sim$ 10% with a safety margin). The robustness of these results to other error seeds and an improved modeling of the beam dynamics during acceleration ramp will be subject of a future work.

ELECTRON CLOUD SIMULATIONS

We address here the modeling of the electron cloud incoherent effects which relies on the modeling of the pinch of the electrons. The details of this process are described in Ref. [5] where an ensemble of electrons are tracked and the structure of pinch is obtained. As a first approach, the modeling of the structure of electrons was made by using a simplified model as described in Ref. [5]. There the electron cloud structure has been modeled as a circular sheet with radius function of the position of the proton in the bunch reference frame. This model assumes that the average effect of the electron cloud on a proton of the bunch can be computed assuming the distribution of the electrons to be infinitely long. The resulting electric field has therefore no "z" component. Although the simplicity of the model, the resulting effect on the detuning experienced by a proton is rather complex as shown in Ref. [19]. This model is clean and allows the construction of very clear frequency maps, as shown in Ref. [5]. However, in this model the pinch process is assumed to be the same in each interaction point, which is not the case as the motion of the electrons is strongly influenced by the type of magnetic field created by the accelerator elements present at the location of the electron cloud. This effect was studied for pinch in drift and in a dipole in Ref. [5], but more in general the electron pinch structure becomes more complex for higher order magnetic fields. For example in Ref. [20] a study of the electron pinch in a quadrupolar magnetic field is presented. In particular the symmetry of the electron structure is very sensitive to the transverse position of the proton bunch with respect to the magnetic center of the element in which the electron cloud is located. This dependence is shown in Fig. 8. It is considered a LHC bunch with transverse rms size σ_r of 0.88 mm, an rms bunch length σ_z of 11.4 cm, with a bunch population $N_p = 1.15 \times 10^{11}$ protons, and a beam energy of 450 GeV. The initial electron distribution is uniform in a circle of radius $R = 10\sigma_r$, and it is always considered centered in the vacuum chamber, the number of macro-electrons is $N = 5 \times 10^5$. The left column of pictures show the pinch of electrons when the proton bunch is centered. The right column pictures show pinch process when the beam is shifted horizontally of 3 times the beam transverse radius. From Fig. 8 it is evident that the structure of the electrons cannot be modeled by simple analytic models. A relevant issue for the phenomenon of periodic resonance crossing is to establish what is the maximum tune-shift created by the pinch of the electrons. The larger is the tune-shift the more pronounced is the effect on the proton beam. In a study presented at the ECLOUD'12 workshop it is shown that in a good approximation the maximum tune-shift has to be expected along the closed orbit of the proton bunch [21]. An attempt of characterization of the detuning from a pinch process requires the inclusion in the proton dynamics of the full electron dynamics. To this purpose the following procedure was adopted:

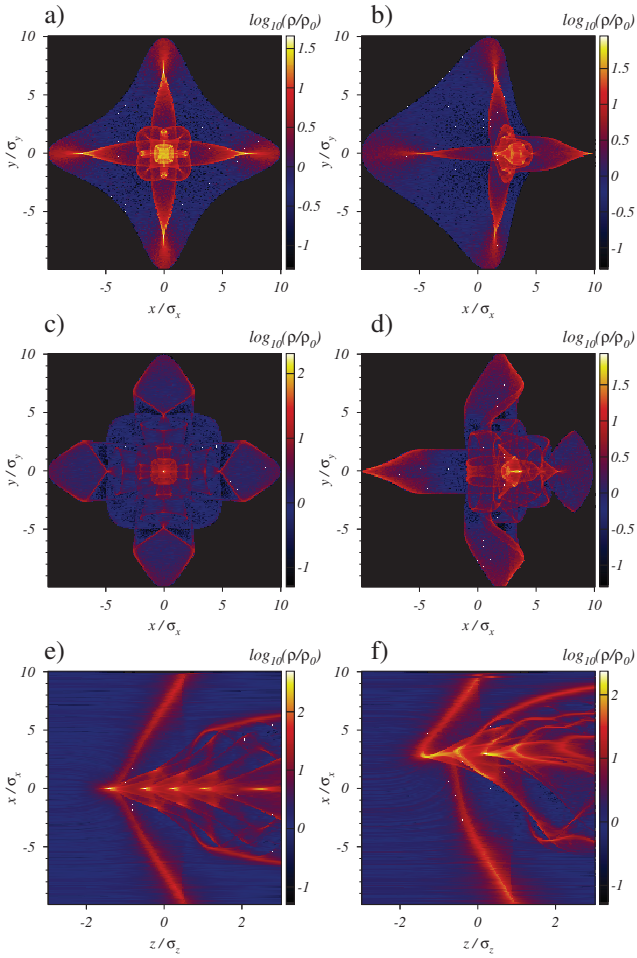


Figure 8: Electron density enhancement for the proton beam on axes (left column), and the same simulation with the beam shifted of $\Delta x_b = 3\sigma_r$ (right column). ab) the $x-y$ plane at $z = 0$; cd) the $x-y$ plane at $z = 1$; ef) the $z-x$ plane at $y = 0$.

1) We compute the "normalized transverse force" E_x, E_y created by electrons at each location of the bunch. The force is modeled as $\propto 1/r$ with a cut off [21]. By normalized it is meant that the electric field E_x, E_y is computed for a reference charge density assigned to each macro electron. As consequence the realistic force is obtained by re-scaling this force of a proper factor F, proportional to the electron cloud density.

2) The field E_x, E_y is stored as function of x, y, z on a

$200 \times 200 \times 200$ grid that include the bunch itself. The grid extends to $[-10\sigma_r, 10\sigma_r]$ in both transverse axis, while on the longitudinal axes it extends in the range $[-3\sigma_z, 3\sigma_z]$.

3) The actual force on a proton when it passes through the electron cloud at the longitudinal position z (in the bunch reference frame), is obtained via tri-linear interpolation from the grid data.

4) The previous procedure is applied for a pinch of electrons in a drift, dipole, and a quadrupole. We define in this way 3 new elements "EC kick" which are consistently applied in the neighbor of each element of the circular accelerator structure.

At the moment this procedure remains inconsistent as it does not take into account of the differences in optics at different locations where electron pinch will take place. Clearly these optics differences are responsible of deforming the transverse section of the proton beam which consequently will produce a "deformed" electron pinch. Hence it becomes necessary, but it is left to future studies, to establish if there is a scaling property of the structure of electron cloud with β_x, β_y at the location in which the pinch takes place. In Fig. 9 we show the detuning along the longitudinal axis obtained applying this method for a reference example Ref. [21]. Here the electron cloud density is artificially enhanced for the purpose of testing the algorithm ($n_e \sim 10^{16}\mathrm{m}^3$). The completion of the tests is left for future studies.

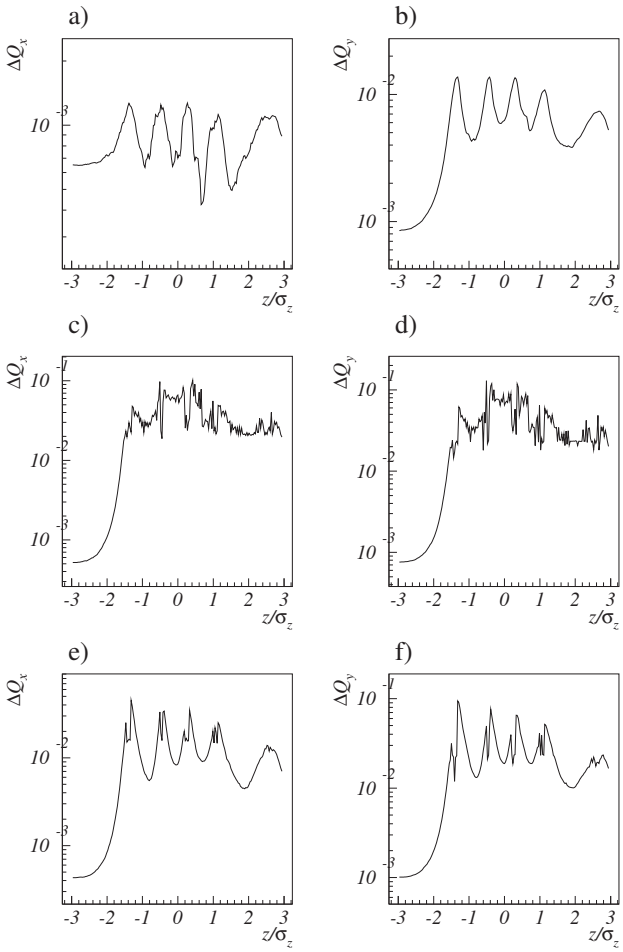


Figure 9: Detuning along the bunch for test particles at transverse amplitude of $0.1\sigma_r$. The right column is the horizontal tune, the left column is the vertical tune. ab) one EC kick in a dipole; cd) one EC kick in a drift; ef) one EC kick in a quadrupole.

ACKNOWLEDGMENTS

The authors thank O. Boine-Frankenheim, H. Eickhoff, E. Fischer, I. Hofmann, V. Kapin, A. Mireau, A. Parfenova, P. Spiller, P. Schnizer, S. Sorge for many fruitful discussions and contributions to these studies.

The author G.F. thanks the European Commission under the FP7 Research Infrastructures project EuCARD, grant agreement no. 227579, for the financial support to participate in this conference.

REFERENCES

[1] P. Spiller *et al.*, Proc. of EPAC 2008, p. 298, MPOC100. G. Franchetti *et al.*, Proc. of EPAC 2006 p. 1882, THPCH005. SIS100 Technical report, `http://www.gsi.de/fair/reports/btr.html`.

[2] E. Mustafin *et al.*, Proc. of EPAC 2004, p. 1408, TUPLT112. E. Mustafin *et al.*, Proc. of PAC 2005, p. 3943, FPAE075

[3] H. Kollmus *et al.*, Proc. of EPAC 2006, p. 1426, TUPCH174; A.W. Molvik *et al.*, Phys. Rev. Lett. **98** 054801 (2006).

[4] C. Omet, Proc. of EPAC2008 p. 295, MOPC099.

[5] G. Franchetti, I. Hofmann, W. Fischer, F. Zimmermann, PRSTAB **12**, 124401(2009);

[6] G. Franchetti *et al.*, PRSTAB **13**, 114203 (2010).

[7] G. Franchetti and I. Hofmann, Nucl. Instr. and Meth. A 561, (2006), 195-202.

[8] J. Struckmeier, Phys. Rev. ST Accel. Beams 3, 034202 (2000).

[9] I. Hofmann *et al.*, Phys. Rev. ST Accel. Beams 6, 024202 (2003).

[10] I. Hofmann, G. Franchetti, J. Qiang, R. Ryne, Proc. 29th ICFA Advanced Beam Dynamics Workshop on Beam Halo Dynamics, Diagnostics, and Collimation HALO 03 (AIP, New York, 2003), 693, 65.

[11] G. Franchetti, I. Hofmann, M. Giovannozzi, M. Martini, E. Metral, Phys. Rev. ST Accel. Beams 6, 124201 (2003).

[12] E. Benedetto, G. Franchetti, F. Zimmermann, Phys. Rev. Lett. 97, 034801 (2006).

[13] G. Franchetti, F. Zimmermann, submitted to PRL.

[14] P. Akishin *et al.*, GSI Note, 3rd June 2010.

[15] V. Kapin, G.Franchetti, ACC-note-2010-004 (2010).

[16] A. Kovalenko, private communication.

[17] P. Spiller, private communication.

[18] R. Kurnishov *et al.*, GSI Note, May 17, 2008.

[19] G. Franchetti and F. Zimmermann, Proc. of Beam 07 , Oct. 1-6, 2007, CERN, Switzerland.

[20] G. Franchetti, F. Zimmermann Proc. of IPAC2011, S. Sebastian, Spain. MOPS001. p. 586.

[21] G. Franchetti, F. Zimmermann Proc. ECLOUD'12 to be published.

SIMULATING THE WIRE COMPENSATION OF LHC LONG-RANGE BEAM-BEAM EFFECTS

T.L. Rijoff, F. Zimmermann, CERN, Geneva, Switzerland

Abstract

The performance of the Large Hadron Collider (LHC) and its minimum crossing angle are limited by long-range beam-beam collisions. Wire compensators can mitigate part of the long-range effects. We perform simulations to explore the efficiency of the compensation at possible wire locations by examining the tune footprint and the dynamic aperture. Starting from the weak-strong simulation code BBTrack we developed a new Lyapunov calculation tool, which seems to better diagnose regular or chaotic particle behavior. We also developed faster ways to execute the simulation and the post-processing. These modifications have allowed us to study different wire positions (longitudinal and transverse), varying wire currents, several wire shapes, and a range of beam-beam crossing angles, in view of a prototype wire installation in the LHC foreseen for 2014/15. Our simulations demonstrate that the wire can provide a good compensation, including for reduced crossing angle. Among the benefits of an LHC wire compensator are a better overlap of colliding bunches, as well as the possibility of smaller β^* or higher beam current.

INTRODUCTION

To reach high energy and high luminosity it is important to compensate the negative effects due to the electromagnetic interactions between the two beams before and after the collision points, the so called beam-beam effects [1]. This possible limitation of the collider performance can be partially mitigated with a DC wire compensator [2]. A very good compensation should be obtained when the wire has a distance from the beam equal to the average long range distance (for nominal crossing angle this means 9.5 σ), and with a wire current that depends on the number of long range interactions n_{LR} according to the formula $I_{\mathrm{opt}} = n_{\mathrm{LR}} c e N_b / L_w$ (c speed of light and e elementary charge). If we consider 32 LR interactions (n_{LR}) in total at one Interaction Point (IP), with 1.15×10^{11} particles per opposite bunch (N_b), and a wire length (L_w) of 1 m, we obtain I_{opt} = 176.8 A.

From the longitudinal point of view, the best compensation is obtained when the $\beta_x + \beta_y$ at the wire and when the betatron phase advance between the LR collision points and the wire is as small as possible. For the nominal LHC optics an optimal location has been found at 104.9 m from the IP [3], in the MAD optics this location carries the label "BBC" (Beam Beam Compensator).

For technical reasons we need to explore solutions different from the one indicated above: so (1) we analysed what happens moving the wire into the shadow of the collimator (for nominal crossing angle this means to 11 σ), (2) we checked what happens if we use the same electric current as at 9.5 σ (I=176.8 A) as when scaling the current quadratically (I = 237 A) (3) we tried different longitudinal wire locations: (4) we tested a modified optics [4].

We analyzed the following scenarios: Head On (HO): 2 head-on collisions at IPs1 and 5, Head On Long Range (HOLR): 2 HO collisions plus 16 LR collisions at each side of the IP1 and IP5, Beam Beam Compensator (BBC): HOLR plus a wire at 105 m after IP1 and IP5, Tertiary Collimator Target (TCT): HOLR plus a wire at 147 m before IP1 and IP5, TCT opt β: HOLR plus a wire at 150 m after IP1 and 147 m before IP5, TCT opt β: 2 HOLR plus a wire at 147 m before IP1 and 150 m after IP5 and Quadrupole 5 (Q5): HOLR plus a wire at 199 m after IP1 and IP5

We find that BBC offers the best compensation, but the simulated performance is promising also for TCT if we use the modified optics, which could be studied experimentally in the LHC from about 2015 onward, and for TCT opt β with nominal LHC optics.

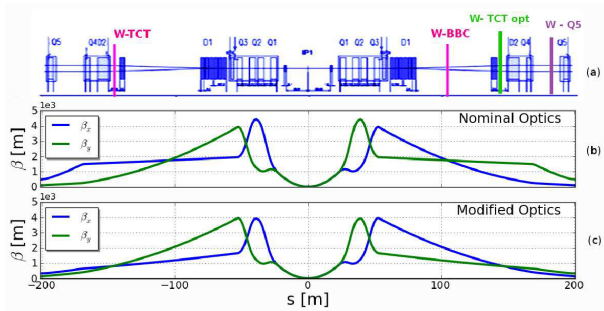


Figure 1: IP1: possible wire positions (row 1), β functions for LHC nominal optics (row 2) and modified optics (row 3)

For the most interesting cases we varied the crossing angle in the range [12,6.3] σ, and found that suitably placed LHC wire compensators should allow for a reduction of the crossing angle by the equivalent of at least 1-2 σ while maintaining the same stable region in phase space as for the larger crossing angle without compensator.

SIMULATION TOOLS

We used the weak-strong code BBTrack [5] to track the particles, and we developed new postprocessing tools to analyse tune footprints and particle stability [6].

Stability Analysis

We tracked each particle, together with a twin particle launched with a small transverse offset of 10^{-8} m , for at least 300,000 turns. To determine the stability of a particle trajectory, on each turn (j) a Lyapunov indicator, $\lambda[j]$, is computed from the time evolution of the normalized distance d in phase space between the two twin particles.

Specifically, a particle is considered as unstable if $\lambda[j]$ exceeds a certain threshold value (taken to be equal to 3 in our tests). The original formula for $\lambda[j]$

$$\lambda[j]\ (\text{old}) = \frac{d_r[j] - d_r[0]}{2d_r[j/2]} \tag{1}$$

flags some stable particles as unstable, for example the first particle plotted in Fig. 2. We introduced the improved formula

$$\lambda[j](\text{new}) = \frac{\langle d_r[\frac{j}{2} : j]\rangle - \langle d_r[0 : \frac{j}{2}]\rangle}{\langle d_r[j/4 : 3j/4]\rangle}\ , \tag{2}$$

with $\langle d_r[m1 : m2]\rangle$ denoting the average value of d between turns $m1$ and $m2$. Using the new formula the top case in Fig. 2 is correctly identified as stable, the bottom case as unstable.

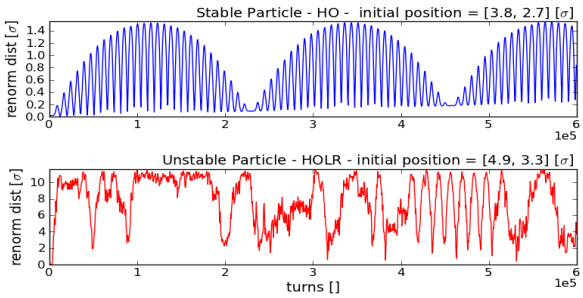


Figure 2: Normalized distance as a function of turn number for a stable (top) and an unstable trajectory (bottom) .

Figure 3 illustrates, for an example, that the stable region in amplitude space does not change when we further increase the number of turns. In the figure, the horizontal (vertical) axis refers to the horizontal (vertical) start amplitude in units of σ_x (σ_y). The color code indicates the number of turns after which an instability has been detected.

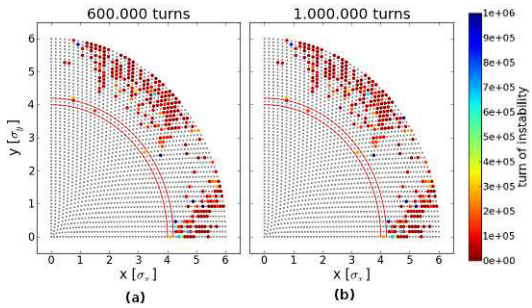


Figure 3: Particle stability for HOLR with a reduced crossing angle corresponding to a LR separation of 6.3σ, when tracking over 6×10^5 (left) or 10^6 turns (right).

Tune Footprint Analysis

We considered a particle distribution in [0,6.5] σ with more particles in [3,6.5] σ, since the long range collisions affect particles with larger amplitude. For each particle we recorded its position at IP1 for the first 50.000 turns and we calculated the tune taking the maximum of the real fast fourier transform (rfft) of positions as function of the turn number . We plotted the fractional part of the tune checking if we touched any resonance lines with order ≤ 9.

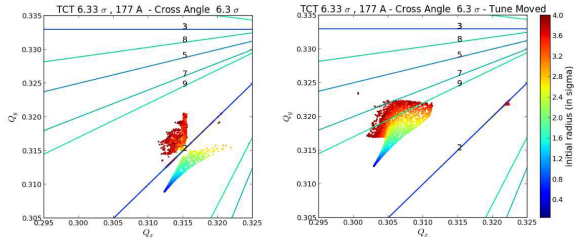


Figure 4: Tune footprint for TCT wire at 6.3σ with current 177 A if we suppose a reduced crossing angle corresponding to a LR separation of 6.3σ, without modification (left), and moving back the central tune (right)

In some cases, the central tune is too close to, or in the worst cases crosses, one of the resonance lines, causing instabilities. In these cases we moved the central tune back with a bbtrack feauture and the formula

$$\Delta Q_u = \pm\frac{r_0 I_w L_w \beta_u}{2\pi\gamma e c d_w^2} \tag{3}$$

where d_w is the wire distance and $-$ refers to the plane of crossing.

RESULTS

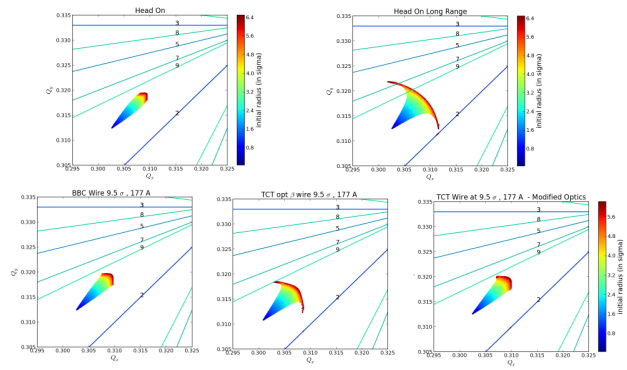


Figure 5: Tune footprint for HO (top left), HOLR (top right) wire compensator with current 177 A and transversal position 9.5 σ located at BBC (bottom left), TCT opt β (bottom center), TCT with modified optics (bottom right)

From the tune analysis we obtain the best results with a current of 177 A and a transversal position of 9.5σ. As visible in the Fig. 5 (bottom left), a wire located at BBC gives a tune footprint almost identical to the one obtained without the long range effects. Also a wire at the TCT location with modified position (bottom center) or optics (bottom right) give satisfactory results.

On the other hand, in the stability analysis (Fig. 6) the best results are obtained with a current of 237 A and a transversal position of 11σ. For example using a wire at BBC position we pass from 5.7% of unstable particles of HOLR tests (top right) to 0.5% of unstable particles if the wire is located at 11 σ and has a current of 237 A. For a transverse position of 9.5 σ and a current of 177 A we obtain 2.2% of unstable particles, and similar results for the wire at TCT with opt β or with modified optics (in both these cases we have 2.4% of unstable particles).

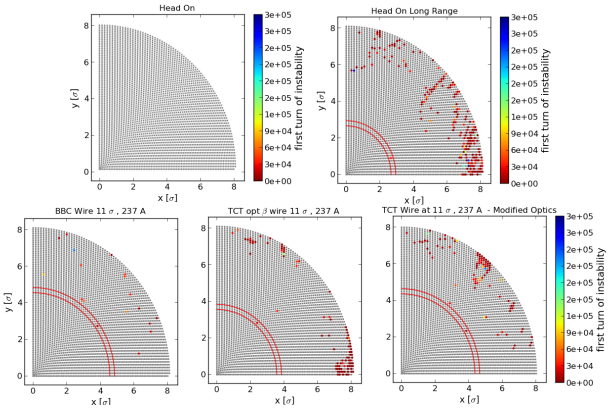


Figure 6: Particle stability for HO (top left), HOLR (top right) wire compensator with current 237 A and transversal position 11 σ located at BBC (bottom left), TCT opt β (bottom center), TCT with modified optics (bottom right)

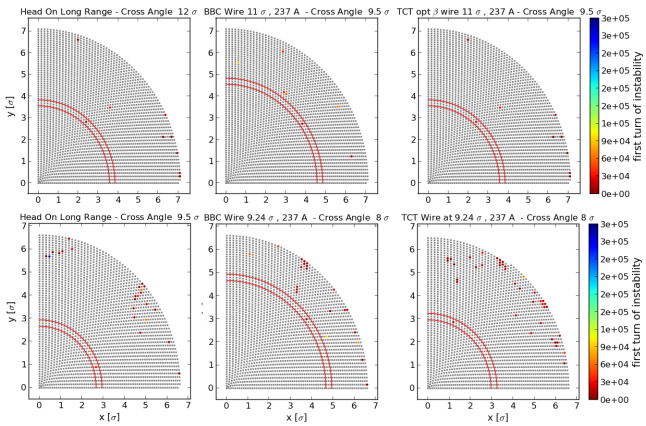


Figure 7: Particle stability: top left: HOLR crossing angle of 12 σ; crossing angle of 9.5 σ, wire at 11 σ, with a current of 237 A at BBC position (top center) and at TCT opt β position (top right). bottom left: HOLR crossing angle of 9.5 σ; crossing angle of 8 σ a wire at 9.24 σ, with a current of 237 A at BBC position with nominal optics (bottom center) and at TCT position with modified optics (bottom right)

Comparing the stability plots for different crossing angles (Fig.: 7) we notice that suitable wire compensators allow us to reduce the crossing angle by at least 1-2 σ maintaining the same stable region in phase space. For example, the stable region for a nominal crossing angle (9.5 σ) with wire compensators at BBC or TCT opt β is comparable to the stable region if the two beams have an average separation of 12 σ (first row). In a similar manner second row of Fig. 7 shows that with a wire compensator at the BBC with nominal optics or at the TCT location with modified optics we can reduce the crossing angle to 8 σ and obtain the same stable region as for the HOLR with a crossing angle of 9.5 σ.

If the beam beam average distance is reduced to 6.3 σ the long range beam beam effects become more dangerous and for several particles the tune crosses a resonance line of order 2 as illustrated in Fig. 8 (top). The bottom pictures line show how wire compensators can correct the long range effect also at this crossing angle.

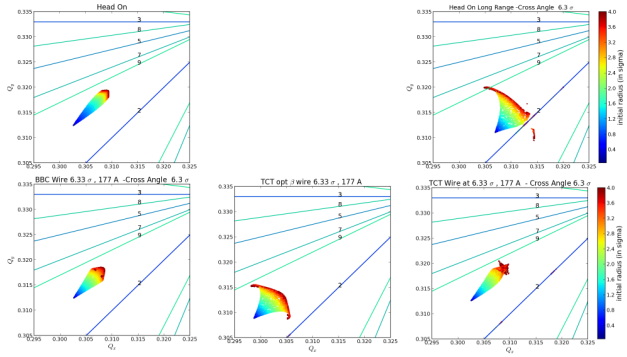


Figure 8: Tune footprint crossing angle 6.3 σ: top: HOLR. bottom: Wire at 6.33 σ with a current of 177 A. bottom left: at BBC, bottom center: at TCT opt β, bottom right: at TCT with modified optics.

CONCLUSIONS

We analyzed the possible compensation of long range beam beam effect in the LHC through an electric wire. We based our work on the bbtrack code [5]. We noticed that in some cases the stability analysis flags some stable particles as unstable. Therefore we developed a new way to calculate the Lyapunov coefficient. We verified that this automatic method is consistent with the identification based on direct data inspection and the results are stable when we increase the number of turns. With some Python scripts to speed up the input file generation, tracking execution and postprocessing analysis we explored the behaviour of the wire compensator for various longitudinal and transversal positions, with different currents, two different optics and of crossing angle. We saw that a wire at BBC location provides the best compensation and allows reducing the crossing angle by 1-2 σ, maintaining the same stable region in phase space. Compensation effects are also promising for the TCT location and the modified optics, as well as for a modified TCT location (TCT opt β) with the nominal LHC optics.

REFERENCES

[1] W. Chou, D. Ritson, Proc. PAC97, Vancouver, p. 113; Y. Papaphilippou, F. Zimmermann, PRST-AB 2, 104001 (1999); M. Albert et al, CERN-ATS-Note-2011-058 MD

[2] J.-P. Koutchouk, PAC Proc. PAC2001, Chicago, p. 1681

[3] U.Dorda "Compensation of long-range beam-beam interaction at the CERN LHC" (May, 2008)

[4] S. Fartoukh private comunications (2012)

[5] U. Dorda, BBTrack *http://ab-abp-bbtrack.web.cern.ch*

[6] T. Rijoff, "Testing long range beam-beam compensation for the LHC luminosity upgrade" (July, 2012)

ELECTRON CLOUD SIMULATIONS WITH PyECLOUD

G. Iadarola*, Università di Napoli Federico II , Napoli, Italy and CERN, Geneva, Switzerland
G. Rumolo, CERN, Geneva, Switzerland

Abstract

PyECLOUD is a newly developed code for the simulation of the electron cloud (EC) build-up in particle accelerators. Almost entirely written in Python, it is mostly based on the physical models already used in the ECLOUD code but, thanks to the implementation of new optimized algorithms, it exhibits a significantly improved performance in accuracy, speed, reliability and flexibility. PyECLOUD simulations have been already broadly employed for benchmarking the EC observations in the Large Hadron Collider (LHC). Thanks to the new feature of running EC simulations with bunch-by-bunch length and intensity data from machine measurements, the scrubbing process of the LHC beam pipes could be reconstructed from heat load measurements in the cryogenic dipoles. In addition, PyECLOUD simulations also provide the estimation of the bunch-by-bunch energy loss, which can be compared with the measurements of the stable phase shift.

INTRODUCTION

The Electron Cloud (EC) has been recognized as a possible limitation to the performances of the Large Hadron Collider (LHC). [1] In 2011, a one week scrubbing run was enough to lower the Secondary Electron Yield (SEY) of the LHC beam screens to values which allow an almost EC free operation with 50ns bunch spacing.

On the other hand Machine Development (MD) sessions with 25ns beams showed that a severe EC is still developing with this bunch spacing, the main observables being the heat load on the cryo-magnets, the dynamic pressure rise as well as detrimental effects on the beam, namely fast EC induced instabilities, which could be avoided with high chromaticity settings, and slower incoherent effects as particle losses, emittance growth and energy losses, especially on the last bunches of the injected trains.[2]

Analysis and predictions on the EC formation rely most entirely on numerical simulations. CERN has a long experience in the EC build-up simulation, mostly carried out with the ECLOUD code, developed and maintained at CERN since 1997. [3]

Unfortunately this code, written in FORTRAN 77, resulted not very suitable to be adapted to the length and the irregular structure of the beams employed in the MD sessions.

Therefore we have decided to write a fully reorganized code, in a newer and more powerful language, considering that the initial effort would be compensated by a significantly increased efficiency in future developments and debugging.

The new code has been called PyECLOUD, since it is almost entirely written in Python and is largely based on the physical models of the ECLOUD code. On the other hand, several features have been modified, and in some cases completely redesigned, with respect to ECLOUD, with substantial improvements in terms of reliability, accuracy, speed and usage flexibility. [4]

PyECLOUD

As ECLOUD, PyECLOUD is a 2D macroparticle (MP) code, where the electrons are grouped in MPs in order to reduce the computational burden.

The dynamics of the MP system is simulated following the stages sketched in Fig. 1.

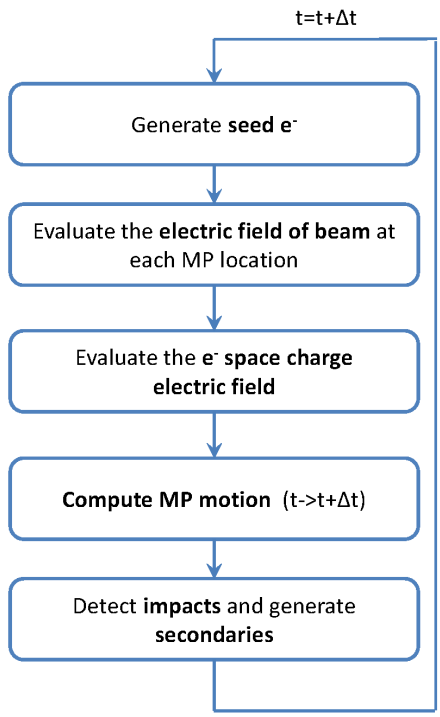


Figure 1: Flowchart representing PyECLOUD main loop.

At each time step Δt, *seed* electrons, due to residual gas ionization and/or to synchrotron radiation induced photoemission from the walls, are generated with the same time evolution of the beam and with transverse position and momentum determined by theoretical or empirical models.

Then the electric field acting on each MP is evaluated: the field of the beam is precomputed on a suitable rectangular grid, loaded from file and obtained at each MP location by a linear (4 points) interpolation; the *space charge*

* giovanni.iadarola@cern.ch

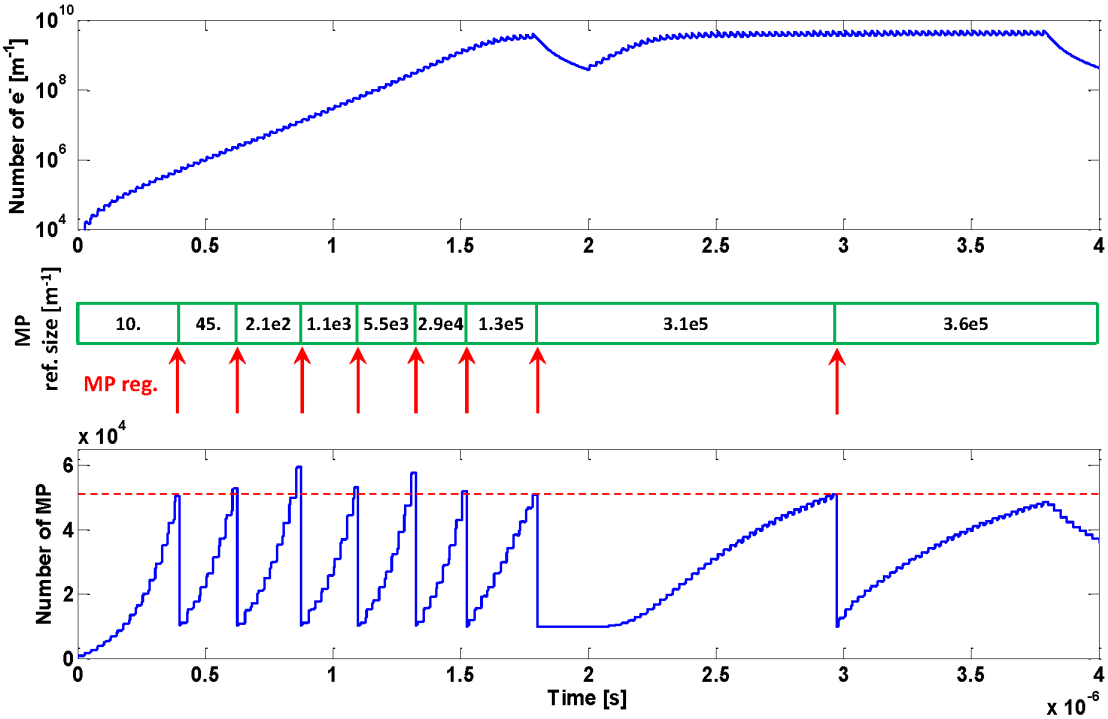


Figure 2: Top: evolution of the number of electrons in the beam pipe for a 25ns LHC beam in the SPS (2 trains of 72 bunches); middle: evolution of the reference MP size; bottom: evolution of the number of MPs, the regeneration threshold is highlighted in red.

contribution of the electron system itself is calculated by a classical Particle in Cell (PIC) algorithm, where the finite difference method is employed to solve the electrostatics equation with perfectly conducting boundary conditions.

Once the total electric field at each MP location is known, MP positions and momenta are updated by integrating the dynamics equation; at this stage the presence of an externally applied dipolar magnetic field can also be taken into account.

At each time step, a certain number of MPs can hit the wall. In these cases a proper model of the secondary emission process is employed to generate charge, energy and angle of the emitted electrons. According to the size of the emitted charge, a rescaling of the impinging MP can be performed or new MPs can be emitted.

MP Size Management

One of the peculiarities of the EC build-up process is the fact that, due to the multipacting effect, the electron number can spread several orders of magnitude (see Fig. 2 - top). As a consequence, it is impossible to choose a MP size which is suitable for the entire simulation, allowing both a satisfactory description of the phenomena and a computationally affordable number of MPs. The MP size management in PyECLOUD has been significantly improved with respect to ECLOUD and will be briefly described in this subsection.

MP sizes are not rigidly imposed but are determined by "decisions" taken during the execution. For this purpose a reference MP size N_{ref}, adaptively changed during the simulation, is employed to control the number of electrons per MP. In particular:

- The size of MPs generated by *seed* mechanisms is exactly N_{ref};
- When a MP hits the wall, it is simply rescaled according to the SEY if the emitted charge is less than $1.5N_{ref}$, otherwise "true" secondary MPs are generated so that the resulting MP size is as close as possible to N_{ref};
- Once per bunch passage, a cleaning procedure is performed, which deletes the MPs with charge smaller than $10^{-4}N_{ref}$.

N_{ref} is changed whenever the total number of MPs becomes larger than a certain threshold (typical value $\sim 10^5$), which means that the computational burden is becoming too high. When this happens, a *regeneration* of the set of MPs is performed, by the following procedure (see Fig. 2):

- Each MP is assigned to a cell of a uniform grid in the 5-D space (x, y, v_x, v_y, v_z) obtaining an approximation of the phase space distribution of the electron gas;

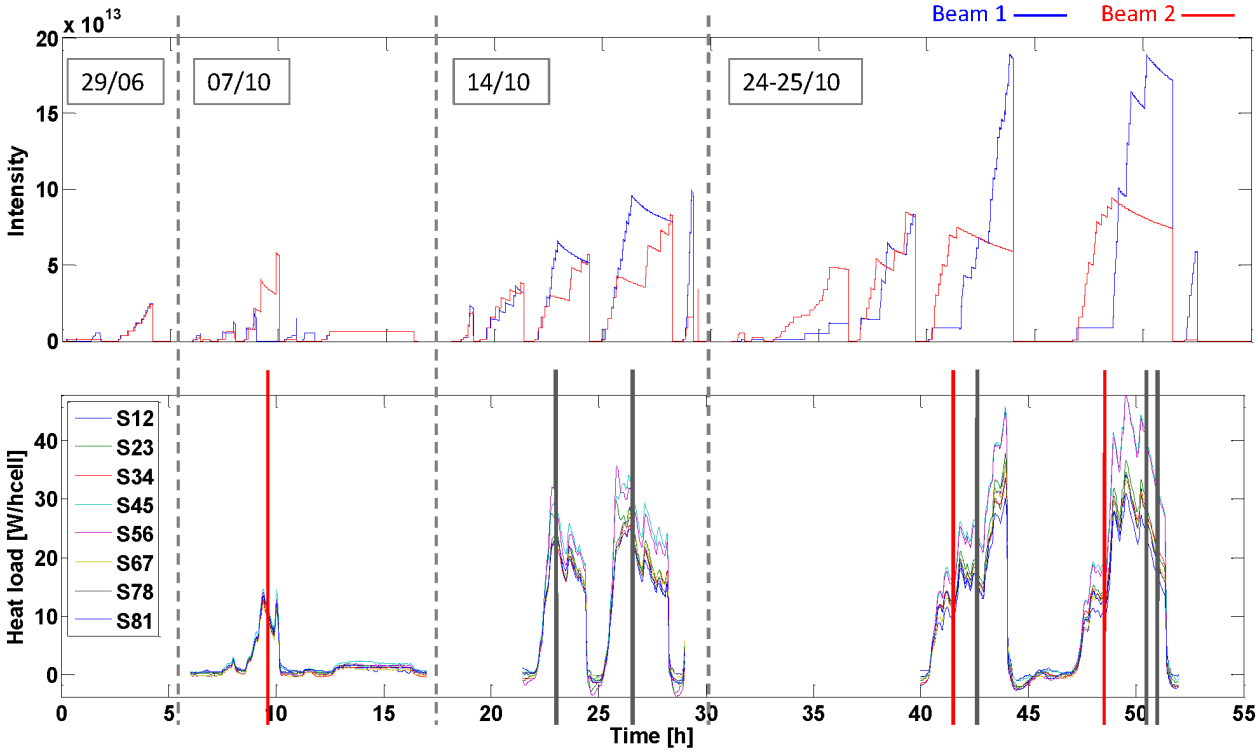


Figure 3: Beam intensity and heat load in the eight arcs of the LHC during the MD sessions with 25ns beams. The vertical bars represent the measurement points used to compare heat load with electron cloud simulations. The red vertical bars correspond to the measurement points in which only (or mainly) beam 2 was in the machine.

- The new N_{ref} is chosen in order to get a target number of MPs (typically 5-10 times smaller than the regeneration threshold), which still allows an accurate simulation but with a more reasonable computational effort;
- A new set of MPs, having the new reference size, is generated according to the computed distribution.

Several numerical test have shown that the errors on the total charge and the total energy which are introduced by this procedure, are never larger than 1-2%.

SIMULATIONS FOR THE LHC ARCS

PyECLOUD simulations have been extensively used to benchmark and analyze EC observations in the LHC. [4, 5, 6] In particular they have been employed for the estimation of the SEY of the copper coated beam screen of the cryogenic arcs.

The cryogenic system, provides data on the total power dissipated (in W/half-cell) on the beam screens of both beams 1 and 2. The heat load evolution during the MD sessions with 25ns in 2011 are shown in Fig. 3.

We know that the EC in the beam chamber at a certain time, and consequently also its effects, strongly depends on the beam structure at the same time. Therefore, in order to compare the heat load data with the simulation results, we have taken several cuts in time (marked as vertical lines in Fig. 3) and carried out EC simulations using the correct beam structures (measured bunch lengths and bunch intensities) at those times for both beam 1 and beam 2. This is crucial for a reliable estimation of the SEY, since, due to EC effects on the beam, both bunch intensities and bunch lengths are different along the train and strongly changing with time.

PyECLOUD simulations were run scanning the maximum of the SEY function ($\delta_{\max}$), so that the curves of the simulated heat loads $\Delta W^{(i)}_{\mathrm{b1-sim}}(\delta_{\max})$ and $\Delta W^{(i)}_{\mathrm{b2-sim}}(\delta_{\max})$, could be produced for each i-th measurement point. The electron reflectivity at zero energy was fixed to the value of 0.7 for all simulated cases. The solution $\delta^{*(i)}_{\max}$ for the i-th heat load measurement considered is then found from the equation:

$$\Delta W^{(i)}_{\mathrm{b1-sim}}(\delta_{\max}) + \Delta W^{(i)}_{\mathrm{b2-sim}}(\delta_{\max}) = \Delta W^{(i)}_{\mathrm{meas}} \quad (1)$$

This procedure allows to draw the evolution of $\delta_{\max}$ of the beam screens reported in Fig. 4. From this curve we can see that $\delta_{\max}$ was about 2.28 at the beginning of the scrubbing run, then it had already decreased to 2.1 by the time the first 25ns beams were first injected into the LHC, and to 1.52 thanks to the scrubbing tests with 25ns beams. Comparing these values with the EC threshold evaluated from simulations (red line in Fig. 4) it is possible to identify the

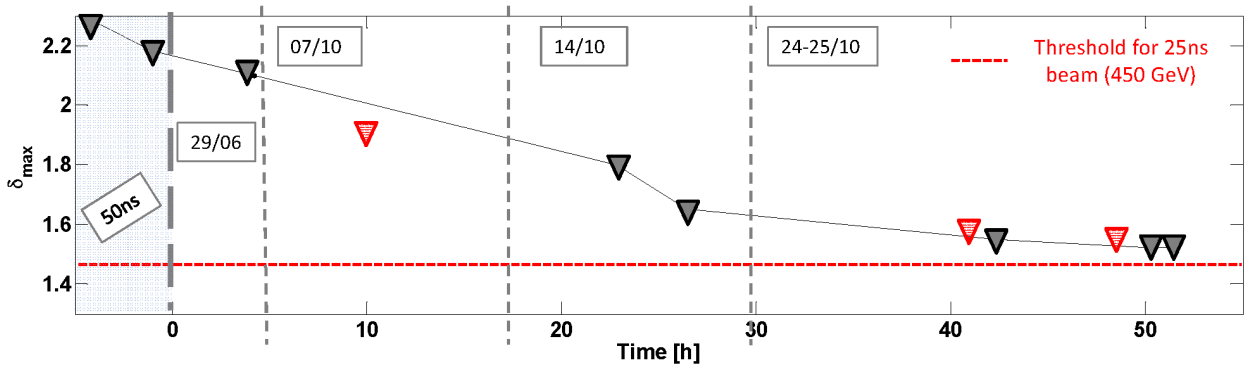


Figure 4: Estimated evolution of $\delta_{\max}$ on the inner surface of the beam screen in the dipole chambers. The first two points correspond to the measurements done with the 50ns beams during and after the scrubbing run, while the other points corresponding to the vertical bars of Fig. 3.

further amount of scrubbing that is necessary to mitigate the EC effects at the injection energy.

The three points marked with red vertical bars in Fig. 3 and red triangles in Fig. 4 correspond to situations in which the total heat load could be only (or mainly) attributed to beam 2, allowing a separated estimation of $\delta_{\max}$ for the beam screen of beam 2. They show that, while it seems plausible that at the beginning of the 25ns MDs the beam screen of beam 2 was more quickly scrubbed than that of beam 1, the conditioning status of the two screens has become later equalized, the last two points for beam 2 being hardly distinguished from the values obtained from the total heat load.

PyECLOUD also provides the calculation of the bunch-by-bunch energy loss per turn. This is based on a simple balance on the energy of the electron cloud. The difference between the total energy of the electrons (electrostatic and kinetic) before and after the bunch passage plus the energy lost in electron-wall collisions during the bunch passage represents the net energy transfered from the bunch to the electrons — and therefore, lost by the bunch. The bunch-by-bunch energy loss per turn thus calculated can be directly compared with the one estimated from stable phase shift measurements from the RF system [7].

The data acquired at the time of the last measurement from Fig. 3 have been plotted in Fig. 5 together with the simulated energy loss, as resulting from the PyECLOUD simulation. The best fit for the measured data has been found with $\delta_{\max} = 1.5$ (very close to the value estimated from the heat load) and a 10% uncaptured beam present in the gaps between trains and also in the abort gap.

The absolute values, as well as both the intra-batch and batch-to-batch trend, seem to be very well caught by the simulation. Furthermore, when zooming on single batches (Fig. 6), we can see that the simulations can successfully reproduce the measurements down to a surprisingly high level of detail.

The EC distribution calculated by PyECLOUD simulations (with $\delta_{\max}$ estimated from heat load measurements) has been also employed for beam dynamics simulations with the HEADTAIL code in order to benchmark fast instabilities observed on 25ns beams in the LHC with low chromaticity settings. Details on the simulation method and results can be found in [6]. The good agreement obtained on the onset of the instability along the batch gives a further indication that our model and simulation code provide a good description of the EC formation in the LHC.

CONCLUSIONS

PyECLOUD is a new EC build-up simulation code, which has been developed for reliable and efficient benchmarking and analysis of the EC observations in LHC. The consistence of simulation results with heat load, stable phase-shift and instability observations in the machine provide strong indications on the reliability of the employed models and numerical solutions.

ACKNOWLEDGMENTS

The authors would like to express their gratefulness to G. Arduini, H. Bartosik, C. Bhat, V. Baglin, R. De Maria, O. Dominguez, M. Driss Mensi, J. Esteban-Muller, K. Li, H. Maury Cuna, G. Miano, E. Métral, H. Neupert, G. Papotti, E. Shaposhnikova, M. Taborelli, L. Tavian, C. Y. Vallgren, and F. Zimmermann for the support they provided in the code development, the simulation work and the machine data analysis as well as for valuable comments and discussions.

REFERENCES

[1] O. Bruning *et al.*, "LHC Design Report", CERN, 2004

[2] G. Rumolo *et al.*, "LHC experince with different bunch spacings in 2011 (25, 50 and 75ns)", Proceedings of the LHC Performance Workshop - Chamonix 2012, 6-10 February 2012, Chamonix, France

[3] G. Rumolo *et al.*, "Simulation of the electron-cloud build up and its consequences on heat load, beam stability, and diagnostics", Phys. Rev. ST Accel. Beams 4 (2001), 1

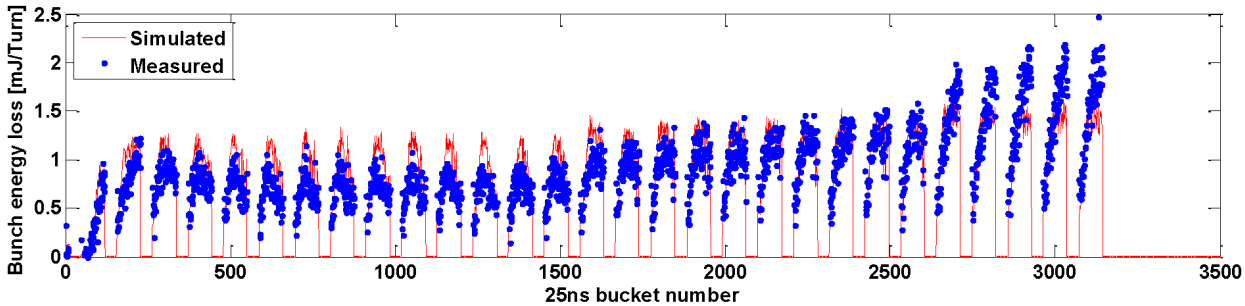


Figure 5: Bunch by bunch energy loss (for beam 1) at the time of the last measurement in Fig. 3.

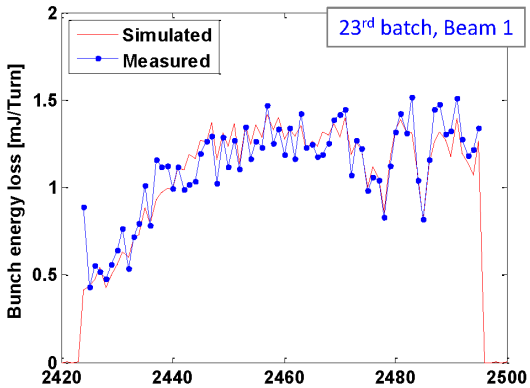


Figure 6: Close up on a selected batch in Fig. 5.

[4] G. Iadarola and G. Rumolo, "PyECLOUD and build-up simulations at CERN", Proceedings of the ECLOUD12 workshop, 5-9 June 2012, La Biodola, Isola d'Elba, Italy

[5] H. Maury Cuna *et al.*, "Simulation of electron-cloud heat load for the cold arcs of the Large Hadron Collider", Proceedings of the 3rd International Particle Accelerator Conference 2012, 20 - 25 May 2012, New Orleans, USA

[6] H. Bartosik *et al.*, "Benchmarking HEADTAIL with e-cloud observations with LHC 25ns beam", Proceedings of the ECLOUD12 workshop, 5-9 June 2012, La Biodola, Isola d'Elba, Italy

[7] J. Esteban Muller *et al.*, "Synchronous Phase Shift at LHC", Proceedings of the ECLOUD12 workshop, 5-9 June 2012, La Biodola, Isola d'Elba, Italy

SIMULATIONS FOR ION CLEARING IN AN ERL *

G. Pöplau†, A. Markoviḱ, U. van Rienen, University of Rostock, Rostock, Germany
A. Meseck, HZB, Berlin, Germany

Abstract

Light sources of the next generation such as ERLs (Energy Recovering Linacs) require very high quality bunches in order to achieve the research objectives they are designed for. It is essential to avoid sources of instabilities such as ionized residual gas. In order to enable detailed numerical studies the tracking code MOEVE PIC tracking developed by the research group at the University of Rostock was expanded to simulations of the behavior of an ion cloud in the environment of electromagnetic fields. In this paper we take further steps to study possible designs of clearing electrodes - a common countermeasure for the clearing of ions. We present simulations with different compositions of the residual gas. The numerical results taking into account the design of the ERL B*ERL*inPro show how the clearing time depends on the percentage of heavy ions in the ion cloud.

INTRODUCTION

Energy Recovery Linacs (ERLs) being the most promising candidates for next-generation light sources put very high demands on preservation of beam brightness and reduction of beam losses. Thus, it is mandatory to avoid the impact of ionized residual gas considered as a source for instabilities in accelerators [1].

Recently, we have presented simulations for the clearing of ionized residual gas with electrodes performed with an upgraded version of software package MOEVE PIC Tracking. It has been currently further developed to model the interaction of the ions with the electron beam in presence of external electromagnetic potentials such as the field of clearing electrodes [2, 3]. The tracking code allows for studies on clearing times for electrodes with different voltage as well as detailed studies of the behavior of the ions in the environment of the electrodes.

In this paper, we especially consider the influence of different compositions of the residual gas on the clearing times. For our numerical studies, we have chosen two different compositions of ion species in the residual gas due to [1] and [4], respectively. The numerical results indicate that the essential influence on the clearing time for the special composition is caused by the varying percentage of ions much heavier than H_2^+. Furthermore, the parameters for the bunch are taken from the design of B*ERL*inPro due to [5].

* Work supported by BMBF under contract number 05K10HRC
† gisela.poeplau@uni-rostock.de

MOEVE PIC TRACKING FOR THE SIMULATION OF ION CLOUDS

The new version of the software package MOEVE PIC Tracking presented in [3] allows for the computation of the interaction of the ion cloud with the bunch and includes the field generated by the clearing electrodes. All simulations take into account the 3D space charge fields of both bunch and ion cloud. Hence, this approach allows for very detailed numerical studies of the dynamics of ion clouds beyond established theoretical results [6, 7]. For a more detailed description of the upgrade of MOEVE PIC Tracking we refer to [3] and references therein.

Further, the clearing times for different voltages of the electrodes are studied in [3]. The simulations with parameters of B*ERL*inPro confirmed the consideration of the peak current of the bunch rather than the average current in order to assign the voltage of the electrodes. We restricted the ion species in the cloud to H_2^+ ions.

In this paper we present simulations for ion clouds with different compositions of ion species. For the first test case (mixture 1) the residual gas consists of H_2^+ ions (98%), CH_4^+ (1%) and CO^+ ions (1%) due to the studies for the Cornell X-ray ERL [1]. Mixture 2 was taken with H_2^+ ions (48%), CH_4^+ (16%), CO_2^+ (18%), CO^+ (14%) and CO_2^+ (17%) ions due to measurements at SPEAR3 [4]. The main difference of these two compositions lies in the percentage of the ions much heavier than H_2^+: for mixture 1 it is 2% and in mixture 2 it is 52%. The data of these residual gas compositions are summarized in Table 2. Beam pipe and electrodes are modelled as in [3], i. e. we set a circular beam pipe with a radius of 2 cm and button-like electrodes with a diameter of 16 mm. The electrodes are located on opposite sides of the beam pipe and the voltage of each is set to the same value of -2700 V due to the observations in [3].

The ion clouds are modelled with a total number of 1 million ions with the mixtures specified in Table 2. The ions are distributed over the whole pipe's cross-section

Table 1: Main Parameters of B*ERL*inPro

maximum beam energy	50 MeV
maximum beam current I	100 mA
nominal beam charge Q	77 pC
maximum repetition rate	1.3 GHz
normalized emittance	10^{-6} m
bunch length σ_t	2 ps

Table 2: Mixtures of Ionized Residual Gas Used for the Simulations

	ion species	percentage	mass number
mixture 1	H_2^+	98 %	2
	CH_4^+	1 %	16
	CO^+	1 %	28
mixture 2	H_2^+	48 %	2
	CH_4^+	5 %	16
	H_2O^+	16 %	18
	CO^+	14 %	28
	CO_2^+	17 %	44

along a length of 1 cm. The bunch was generated by the program *generator* of ASTRA [8] with 100,000 macro electrons uniformly distributed in a cylinder due to the parameters given in Table 1.

RESULTS

The first question we considered was how many ions we have to take into account for the simulations. Assuming that all molecules of the residual gas are ionized a pipe with a radius of 2 cm is filled with 31 million ions per cm at a vacuum pressure of 10^{-10} mbar. Simulation results for these parameters were presented in [3]. Since in practice we expect rather a vacuum pressure of 10^{-8} mbar related simulations would require 3,100 million ions per cm which is impossible to run on a normal PC. Hence, the question is what could be concluded from a simulation with e. g. 1 million ions for the dynamics of 31 million ions and more.

As a test case we compared the results from simulations with ion clouds - consisting of H_2^+ only - of 1 million, 31 million (vacuum pressure of 10^{-10} mbar) and 150 million ions (vacuum pressure of ca. $5 \cdot 10^{-10}$ mbar), respectively. Figure 1 represents the clearing time for these ion clouds scaled up to 150 million in order to make the results comparable. It turns out that independently of the number of ions most of the ions are cleared after 1 µs. Furthermore the dynamics of the ions at different densities is quite similar as shown in Figure 2. As a consequence we performed the simulations with mixed ions only with a total number of 1 million ions.

Figure 3 shows the clearing time for ionized residual gas with compositions due to Table 2. As expected the clearing time increases for increasing percentage of heavier ions in the ionized residual gas. Figure 4 represents the transversal dynamics of the different ion clouds. Furthermore, Figure 5 shows the evolution of the ion cloud with mixture 1 within the first µs, where only H_2^+ (dark blue) and CH_4^+ (light blue) are plotted. The originally uniformly distributed ion species are separated due to their different mass and consequently different velocity.

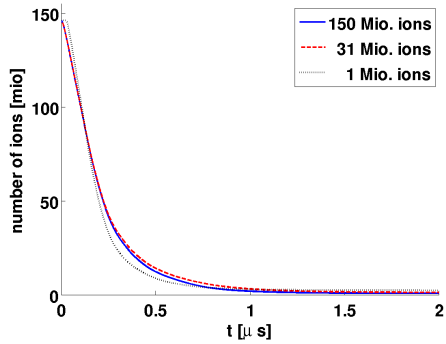


Figure 1: Clearing times of ion clouds with H_2^+ ions cleared by the field of electrodes with a voltage of -2700 V.

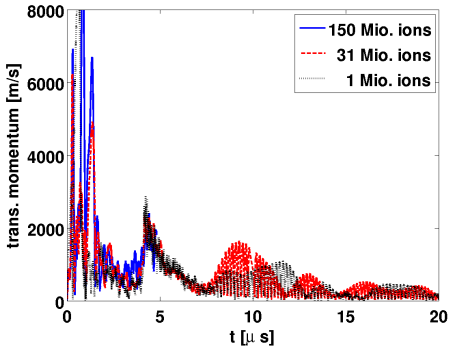


Figure 2: Transversal momentum of ion clouds with H_2^+ ions in the field of electrodes with a voltage of -2700 V.

CONCLUSION

In this paper we presented a simulation study on the behavior of an ion cloud in the potentials of clearing electrodes and electron beam. The simulations are performed for different compositions of ions in the residual gas. The software package MOEVE PIC Tracking allows for a detailed analysis of the dynamics of an ion cloud which

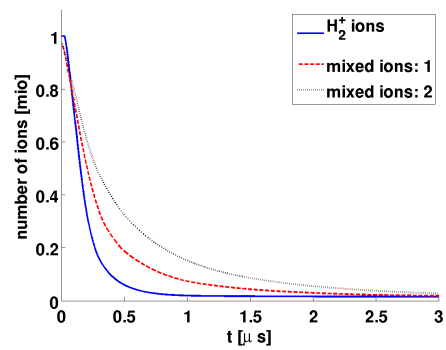


Figure 3: Clearing times of clouds with different compositions of ions compared to H_2^+ ions cleared by the field of electrodes with a voltage of -2700 V.

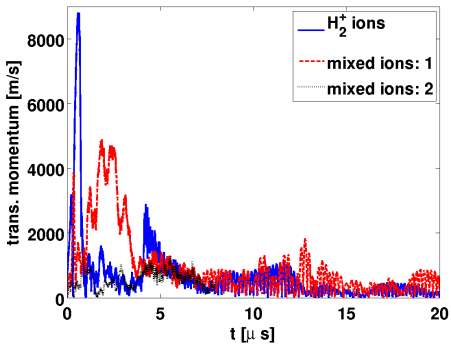


Figure 4: Transversal momentum of ion clouds with different compositions of ions compared to H_2^+ ions in the field of electrodes with a voltage of -2700 V.

undergoes the potentials of the electron beam and clearing electrodes simultaneously. The deep understanding of this dynamics is important for further design studies for B*ERL*inPro. The simulation results indicate that the percentage of the ions with much heavier mass than H_2^+ ions have a significant impact on the clearing time. The higher the percentage of the heavy ions the longer the clearing time.

REFERENCES

[1] G.H. Hoffstaetter and M. Liepe. Ion clearing in an ERL. *Nuclear Instruments and Methods in Physics Research Section A*, 557(1), 205– 212, 2006.

[2] G. Pöplau, U. van Rienen, S.B. van der Geer, and M.J. de Loos. Multigrid algorithms for the fast calculation of space-charge effects in accelerator design. *IEEE Transactions on Magnetics*, 40(2):714–717, 2004.

[3] G. Pöplau, A. Meseck, and U. van Rienen. Simulation of the behavior of ionized residual gas in the field of electrodes. In *Proceedings of IPAC 2012, New Orleans, USA*, 283–285, 2012.

[4] L. Wang, Y. Cai, T. O. Raubenheimer, and H. Fukuma. Suppression of beam-ion instability in electron rings with multi-bunch train beam fillings. *Phys. Rev. ST Accel. Beams*, 14:084401, Aug 2011.

[5] A. Knobloch and et al. Status of the B*ERL*inPro Energy Recovery Linac Project. In *Proceedings of IPAC 2012, New Orleans, USA*, 601–603, 2012.

[6] Y. Baconnier. Neutralization of accelerator beams by ionization of the residual gas. CERN-PS-84-24-PSR-REV-2, CERN, 1985.

[7] A. Poncet. Ions and neutralization. In M. Dienes, M. Month, and S. Turner, editors, *Frontiers of Particle Beams: Intensity Limitations*, volume 400 of *Lecture Notes in Physics*, Springer Berlin / Heidelberg, 488–508, 1992.

[8] K. Flöttmann. *ASTRA*. DESY, Hamburg, `www.desy.de/~mpyflo`, 2000.

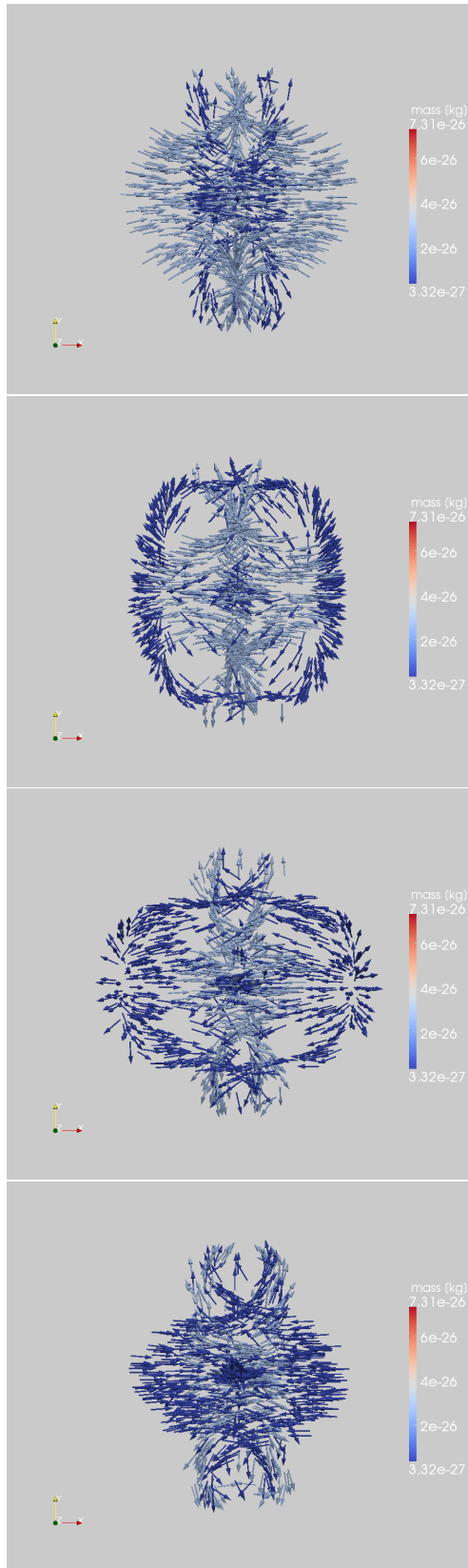


Figure 5: Evolution of an ion cloud (dark blue: H_2^+, light blue: CH_4^+) in the field of electrodes with a voltage of -2700 V after 0.23 μs, 0.38 μs, 0.54 μs and 0.69 μs (from top).

NUMERICAL STUDIES ON THE INFLUENCE OF FILL PATTERNS ON ION CLOUDS*

A. Meseck, HZB, Berlin, Germany
G. Pöplau†, U. van Rienen, University of Rostock, Rostock, Germany

Abstract

Energy Recovery Linacs (ERLs) are the most promising candidates for next-generation light sources now under active development. An optimal performance of these machines requires the preservation of the high beam brightness generated in the injector. For this, the impact of the ionized residual gas on the beam has to be avoided as it causes instabilities and emittance growth. Obviously, the vacuum chamber has to be cleared out of ions but as the potential of the electron beam attracts the ions, it is not enough to install vacuum pumps.

One measure for ion clearing is the generation of gaps in the bunch train. In this paper, we present numerical studies of the behavior of an ion cloud that interacts with bunch trains taking into account the effects of the clearing gaps. In the studies different longitudinal distributions for the particles in the bunch and different "fill patterns" i.e. different spacings of the clearing gaps are investigated. The simulations are performed with the software package MOEVE PIC Tracking developed at Rostock University where recently new features for the simulation of ion effects have been implemented. The presented numerical investigations take into account the parameters of the ERL B*ERL*inPro with the objective to deduce appropriate measures for the design and operation of B*ERL*inPro.

INTRODUCTION

The expected extraordinary performances of the next generation light sources, such as ERLs, require intense hight brightness electron beams. Therefore, the ion-caused instabilities and emittance growth degrading the beam quality are serious threats to these sources and have to be avoided.

In an electron accelerator, several effects like collision with the electron beam, synchrotron radiation and field emission can lead to ionization of the rest gas in the beam pipe. The negative potential of the electron beam can trap the positive ions which in turn leads to an increase of the beam halo, to emittance blowup and to transverse and longitudinal instabilities by interacting resonantly with the beam. There are several measures of avoiding the ion-trapping such as utilizing clearing electrodes or using short or long clearing-gaps. A long gap between two trains of high-repetition rate bunches allows for ions to drift out of the beam potential and thus to reach the wall of the vacuum chamber, while a short clearing-gap causes large oscillations of ions around the beam which in turn clears the vicinity of the bunches from the ions.

Table 1: Main Parameters of B*ERL*inPro

maximum beam energy	50 MeV
maximum beam current I	100 mA
nominal beam charge Q	77 pC
maximum repetition rate	1.3 GHz
normalized emittance	10^{-6} m
bunch length σ_t	2 ps
vacuum pressure	10^{-10} mbar

However, the applicability of the long-gap approach in ERLs suffers from two effects. The first effect is the transient effect of the RF systems which requires very long ion-clearing-gaps [1]. The second effect is the fast beam-ion instability. This is because a long gap does not preclude ions from accumulating in a single bunch train and interacting resonantly with the beam [1, 2]. Therefore, short clearing-gaps seem to be favorable, in particular as experimental and numerical studies indicate that multibunch trains with short gaps are very effective in suppressing the ion trapping process [3]. Please note that in an ERL, also for a short clearing-gap the transient RF effects have to be avoided by ensuring that the clearing-gap in the accelerating and in the decelerating beam coincide.

Although short bunch gaps in the ERL-beam allow for suppression of the ion accumulation in the vicinity of the beam, they are not totally welcome because they change the cw characteristics of the radiation delivered by the ERL. Therefore, to keep the time characteristic as cw-like as possible, frequent but extremely short bunch gaps are favorable. In the presented paper we study whether a bunch gap as short as one accelerating bucket occurring very frequently can excite oscillations large enough to clear the vicinity of the beam. For this, we use the software package MOEVE PIC Tracking [4, 5] to investigate numerically the behavior of an ion cloud that interacts with bunch trains with short gaps betwixt. For the simulations the parameter settings planned for B*ERL*inPro [6] have been used. The parameters that are relevant for the simulations in this paper are given in the Table 1.

SIMULATION TOOL MOEVE

For the presented study, the behavior of an ion cloud that interacts with bunch trains and clearing-gaps has been sim-

* Work supported by BMBF under contract number 05K10HRC
† gisela.poeplau@uni-rostock.de

ulated with a further development of the software package MOEVE PIC Tracking [4, 5]. It tracks particles, where the space charge fields are taken into account in each time step of the tracking procedure. For more details we refer to [4, 7] and citations therein.

The electron beam parameters for the simulations are chosen according to the nominal B*ERL*inPro parameters published in [6]. The electron bunch is modelled with 100,000 macro particles using Gaussian and uniform distributions of cylindrical shapes. For the generation of these distributions the program *generator* of the tracking code ASTRA was applied [8].

In this study we restrict the model of the ion cloud to H_2^+ ions with a relative atomic mass of $A = 2$. Since the residual gas consists mainly of H_2^+ these ions are the most relevant species for investigations of ion effects in B*ERL*inPro. The density of the ion cloud was taken with $2.4 \cdot 10^6$ cm^{-3} according to a vacuum pressure of 10^{-10} mbar. Hence, the ion cloud was modelled with 85,000 ions per cm.

SINGLE-BUCKET CLEARING-GAPS

Assuming that the ions are generated with near thermal velocities, it is a simple conclusion that they are not free to drift in the vacuum chamber. They experience the space charge potential of the electron beam as long as the beam is not fully neutralized. Of course there are also additional external magnetic fields such as dipole and quadrupole fields which we omit in the presented study. Their effects on the ions are subject to further studies in future. The beam potential – to be specific the electric field component of the electron beam – attracts the ions towards the centre of the beam thus causing transverse oscillations of the ions with an amplitude given by their transverse position at "birth" [9].

In contrary to a long clearing-gap which allows for the ions to fully escape the beam potential and reach the chamber wall, in a short clearing-gap the ions have not enough time to escape the beam. However, the short gap gives the beam potential a time structure. This leads to a time dependent attractive force with many frequency components which in turn excites large ion oscillations clearing the vicinity of the beam [1, 3].

However, also very short clearing-gaps change the cw characteristics of the electron beam. In order to keep the time characteristic as cw-like as possible, frequent but extremely short bunch gaps are favorable. This rises the question whether a bunch gap as short as one bunch occurring very frequently can excite oscillations large enough to clear the vicinity of the beam.

Figure 1 represents simulation results for three different repetition rates of the single-bucket-gaps in the B*ERL*inPro beam. It shows the transverse oscillation amplitude of the ion cloud around the beam as a function of time for each gap repetition rate. In the top of Figure 1 the single-bucket-gap occurs every 1000^{th} bunch. For better comparison, the case without clearing-gaps is also shown. The effect of the

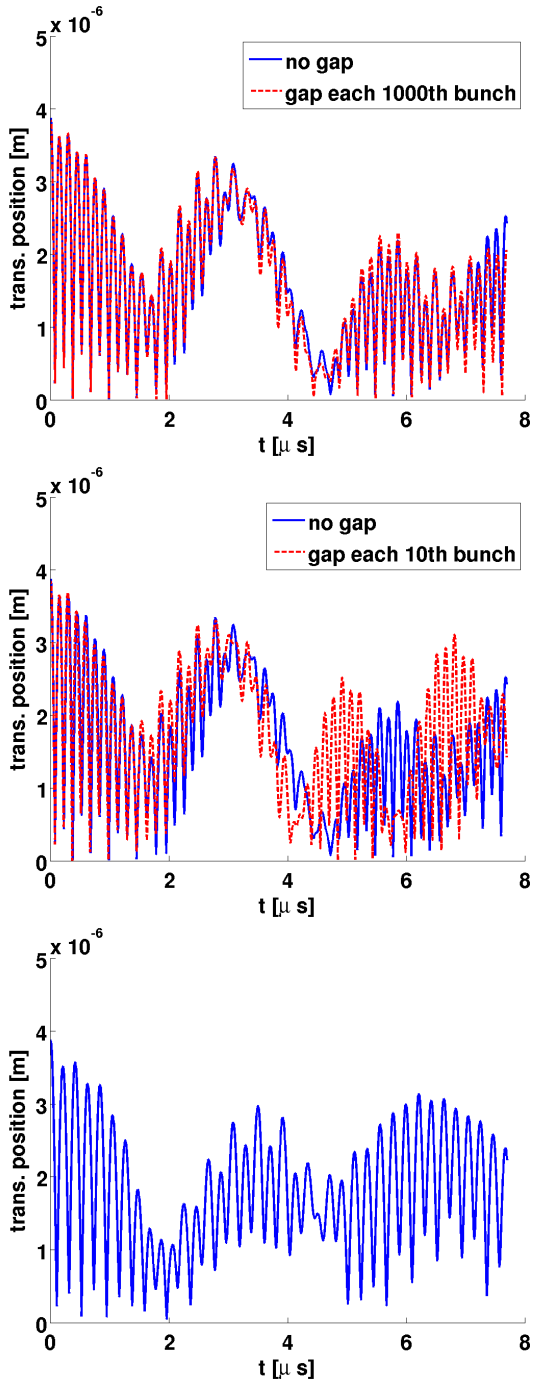


Figure 1: Transverse oscillations of an ion cloud around the beam as a function of time for three different repetition rates of the single-bucket-gap. In the top figure the single-bucket-gap occurs every 1000^{th} bunch, in the middle every 10^{th} bunch and in the bottom figure every second bunch. The longitudinal electron distribution is Gaussian and the beam parameters are given in Table 1.

clearing-gap seems to be very small in this case. The second plot of Figure 1 shows the oscillation amplitude when the single-bucket-gap occurs every 10^{th} bunch. There is an increase in the oscillation amplitude nicely visible after some few microseconds and also a change in the oscillation frequency. However the increase in the amplitude is not strong enough to clear significantly the vicinity of the beam. A clearing-gap occurring every second bunch, shown in the bottom of Figure 1, does not seem to increase the oscillation amplitude either. Please note that the average current of the beam has been kept constant in these simulations by increasing the bunch charges accordingly. However, this correction has not been applied to the last case, where the clearing-gap occurs every second bunch. This is because an increase of a factor of two to the bunch charge would change bunch dynamics in B*ERL*inPro significantly and can not be assumed without further investigation on bunch dynamics. The bunches used for these simulations have Gaussian longitudinal profiles. Their other parameters are listed in Table 1.

Figure 2 represents simulation results for two different repetition rates of the single-bucket-gap. The bunches used for these simulations have uniform longitudinal profiles. In the top of Figure 2 the single-bucket-gap occurs every 1000^{th} bunch. Similar to the Gaussian case, also for a uniform longitudinal distribution, the effect of the clearing-gap seems to be rather small. The oscillation amplitude for single-bucket-gap occurring every 10^{th} bunch, shown in the bottom of Figure 2, is not strong enough to clear significantly the vicinity of the beam.

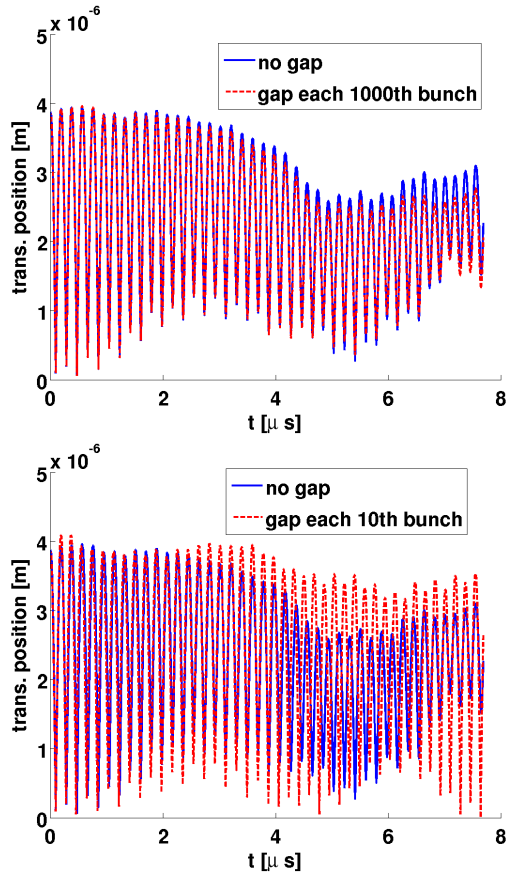


Figure 2: Transverse oscillation of an ion cloud around the beam as a function of time for uniform longitudinal electron distributions. In the top figure the single-bucket-gap occurs every 1000^{th} bunch, in the bottom every 10^{th} bunch. The beam parameters are given in Table 1.

CONCLUSION

We have presented numerical studies investigating the effect of a single-bucket clearing-gap on ion clouds interacting with a high repetition rate multibunch beam. In the studies different longitudinal distributions for the particles in the bunch and different spacings of the clearing-gap have been investigated. The simulations were performed with the software package MOEVE PIC Tracking. The parameters of the ERL B*ERL*inPro have been used for the studies. Although the single-bucket-gap clearly leads to an increase of the ion oscillation amplitudes around the beam, the excited amplitudes are not large enough to clear the beam vicinity of ions. We conclude that either the clearing-gap has to be longer or other clearing measures such as utilizing clearing electrodes have to be envisaged for the operation of B*ERL*inPro.

REFERENCES

[1] G. H. Hoffstaetter and M. Liepe. Ion clearing in an ERL. *Nuclear Instruments and Methods in Physics Research Section A*, 557(1), 205– 212, 2006.

[2] F. Zimmermann, T. O. Raubenheimer and G. Stupakov. A fast beam-ion instability. *Report SLAC-PUB-95-6792*, 1995.

[3] L. Wang, Y. Cai and T. O. Raubenheimer, G. Stupakov. Suppression of beam-ion instability in electron rings with multi-bunch train beam fillings. *Phys. Rev. ST Accel. Beams* **14**, *084401,(2011)*.

[4] G. Pöplau, U. van Rienen, S.B. van der Geer, and M.J. de Loos. Multigrid algorithms for the fast calculation of space-charge effects in accelerator design. *IEEE Transactions on Magnetics*, 40(2):714–717, 2004.

[5] A. Markoviḱ, G. Pöplau, and U. van Rienen. 3D PIC computation of a transversal tune shift caused by an electron cloud in a positron storage ring. In *Proceedings of IPAC 2010, Kyoto, Japan*, 1928–1930, 2010.

[6] A. Knobloch and et al. Status of the B*ERL*inPro Energy Recovery Linac Project. In *Proceedings of IPAC 2012, New Orleans, USA*, 601–603, 2012.

[7] G. Pöplau, A. Meseck, and U. van Rienen. Simulation of the behavior of ionized residual gas in the field of electrodes. In *Proceedings of IPAC 2012, New Orleans, USA*, pages 283–285, 2012.

[8] K. Flöttmann. *ASTRA*. DESY, Hamburg, `www.desy.de/~mpyflo`, 2000.

[9] A. Poncet. Ion Trapping and Clearing. in *Cern Accelerator School*, Third Advanced Accelerator Physics Course CERN-90-04, CERN, 1990.

PARTICLE TRACKING IN ELECTROSTATIC FIELDS WITH ENERGY CONSERVATION

A. Ivanov*, Saint-Petersburg State University, Russia

Abstract

The key idea of the research is to consider spin dynamics in electrostatic fields. Due to the fact, that spin rotation frequency explicitly depends on velocity of the particle and its kinetic energy is changed in electrostatic fields it is important to use some technique that provides both conservation energy and simplicity condition. An appropriate mathematical model is described and the results of numerical calculation are shown. In conclusion, fringe fields influence is examined and compared with case of ideal fields.

INTRODUCTION

In the article particle dynamics is considered in 8-dimensional space. A state of dynamic system is described as $(x,\ x',\ y,\ y',\ S_x,\ S_y,\ S_s,\ t)$ vector, where x, x' and y, y' are transverse and vertical displacement and velocity respectively; $S_x,\ S_y,\ S_s$ are components of spin vector in curvilinear coordinate system (see Fig. 1); t is time variable. Note, that a state vector depends on arc length s, which is chosen as an independent variable.

The article consist of three parts. Firstly, mathematical models of the particle and spin dynamics are discussed. In the second part the numerical step-by-step integration approach is presented. And the last part is numerical experiment of fringe fields modeling, where the energy conservation is especially important condition.

MODEL DESCRIPTION

This section is devoted to the mathematical models of particle motion and spin dynamics. Both trajectory and spin equations are presented in generalized form along the design orbit. In case of straight orbit equations are similar to description in Cartesian coordinates. Equations along the arc of a circle are presented without derivation.

Particle dynamics in electrostatic fields is described by the Newton-Lorenz equation

$$\frac{dp}{dt} = qE, \tag{1}$$

where p is the momentum, q is the charged of the particle, E is the electric field.

For spin description BMT equation is used [1]

$$\begin{aligned} &\frac{dS}{dt} = \omega \times S, \gamma = \frac{1}{\sqrt{1 - v^2/c^2}}, \\ &\omega = \frac{Q}{m_0^2 c^2}\frac{1}{\gamma}\left(G + \frac{1}{1+\gamma}\right) p \times E, \end{aligned} \tag{2}$$

* 05x.andrey@gmail.com

where $G = (g-2)/2$, g is the anomalous spin factor, γ is the Lorentz factor.

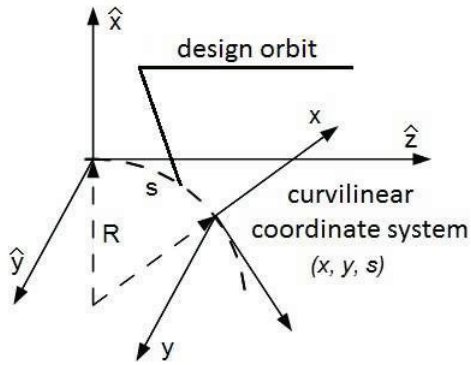


Figure 1: Curvilinear coordinate system.

Trajectory Equations

Derivation of the trajectory equations that describe the orbital motion uses generalized coordinates. The design orbit is chosen in accordance to symmetry of field distribution. For example, in quadrupole lenses it is a straight line, in cylindrical or spherical deflectors it is arc of a circle. Along the arc length the equations are following

$$\begin{aligned} &x'' + (1 - \frac{v^2}{c^2})^{1/2}\frac{HG}{v}x' - (1 + \frac{x}{R})\frac{1}{R} = \\ &\qquad = \frac{QH}{m_0 v}(1 - \frac{v^2}{c^2})^{1/2} HE_x/v, \\ &y'' + (1 - \frac{v^2}{c^2})^{1/2}\frac{HG}{v}y' = \frac{QH}{m_0 v}(1 - \frac{v^2}{c^2})^{1/2} HE_y/v, \end{aligned} \tag{3}$$

where H, G is functions of variable x, x', y, y', R, R is a radius of curvature of the design orbit.

Spin Dynamics

The BMT equation in case of arc design orbit is presented as

$$\begin{aligned} S'_x &= S_s/R + \frac{Q}{m_0 c^2}\left(G + \frac{1}{1+\gamma}\right)\big((h_s E_x - x' E_s)S_s - \\ &\quad - (x' E_y - y' E_x)S_y\big), \\ S'_y &= \frac{Q}{m_0 c^2}\left(G + \frac{1}{1+\gamma}\right)\big((x' E_y - y' E_x)S_x - \\ &\quad - (y' E_s - h_s E_y)S_s\big), \end{aligned} \tag{4}$$

$$S'_s = -S_x/R + \frac{Q}{m_0c^2}\left(G + \frac{1}{1+\gamma}\right)\left((y'E_s - h_sE_y)S_y - \right.$$
$$\left. - (h_sE_x - x'E_s)S_x\right).$$

Kinetic Energy Conservation

Note that right-hand sides of the Equations (4) and (3) contain velocity magnitude v. For the energy conservation satisfying we can calculate it as

$$v = c\sqrt{1 - \left((1 - v_0^2/c^2)^{-1/2} - \frac{Q}{m_0c^2}\Phi\right)^{-2}}, \quad (5)$$

where v_0 is the velocity out of electric field, $\Phi = \Phi(x, y, s)$ is the potential.

NUMERICAL INTEGRATION

In ideal fields the Equation (5) provides the energy conservation itself. But due to the errors in numerical methods and artificial fields it is not enough. The next mathematical model allows to control the velocity and correct the particle motion.

The Equations (4) and (3) can be written as

$$\frac{d}{ds}X = F(s, X),$$
$$\frac{d}{ds}v_0 = 0, \quad (6)$$

where $X = (x, x', y, y', S_x, S_y, S_s)$. Derivative of velocity is equal to zero, that indicates to the energy conservation.

This allows as to use classical step-by-step integration methods to solve this system. Article [2] provide both symplectic Runge-Kutta integration schemes, and the algorithm for it derivation up to the 12 order. For the current research a symplectic 2-stage Runge-Kutta scheme of 4 order was implemented as a basic approach. Moreover the set of methods was also implemented (symplectic Euler scheme, symplectic average point, etc.)

Table 1: 2-stage 4-order Implicit Runge-Kutta Scheme

$b_1 + \tilde{c_1}$	$b_1/2$	$b_1/2 + \tilde{c_1}$
$b_1 - \tilde{c_1}$	$b_1/2 - \tilde{c_1}$	$b_1/2$

$$b_1 = 1/2, 2b_1\tilde{c_1}^2 = 1/12$$

According to this scheme (Table 1), the solution of the Equations (6) can be presented in iterative form

$$\mathbf{X}_{n+1} = \mathbf{X}_n + h\sum_{j=1}^{2} b_j\mathbf{F}(s + hc_j, \mathbf{X}^{(i)}),$$
$$\mathbf{X}^{(i)} = \mathbf{X}_n + h\sum_{j=1}^{2} a_{ij}\mathbf{F}(s + hc_j, \mathbf{X}^{(i)}).$$

This integration method provide a symplectic solution by choosing the corresponding coefficients a_{ij}, b_j, c_j. Note that this symplectic scheme imposes the condition of constant integration step h. The simple software environment for designing and modeling was developed (see Fig. 2) and all these parameters can be set.

Moreover this scheme requires to solve of implicit equations and appropriate numerical methods can be used. In this research standard Newton method was used. So the precision of the numerical approach depends on step value and error tolerance for solution of implicit equations.

The scheme does not satisfy to energy conservation at all. For this purpose velocity derivative was added to the model. Though the velocity magnitude can be calculate directly by the x, x', y, y', s, t values, but the equation for it derivation let us introduce an addition control parameter.

For example, this control important to estimate influence of fringe fields. It is difficult task to introduce fringe fields that satisfy to Laplas equation. So the simple model for fringe fields is chosen, where such correction is necessary.

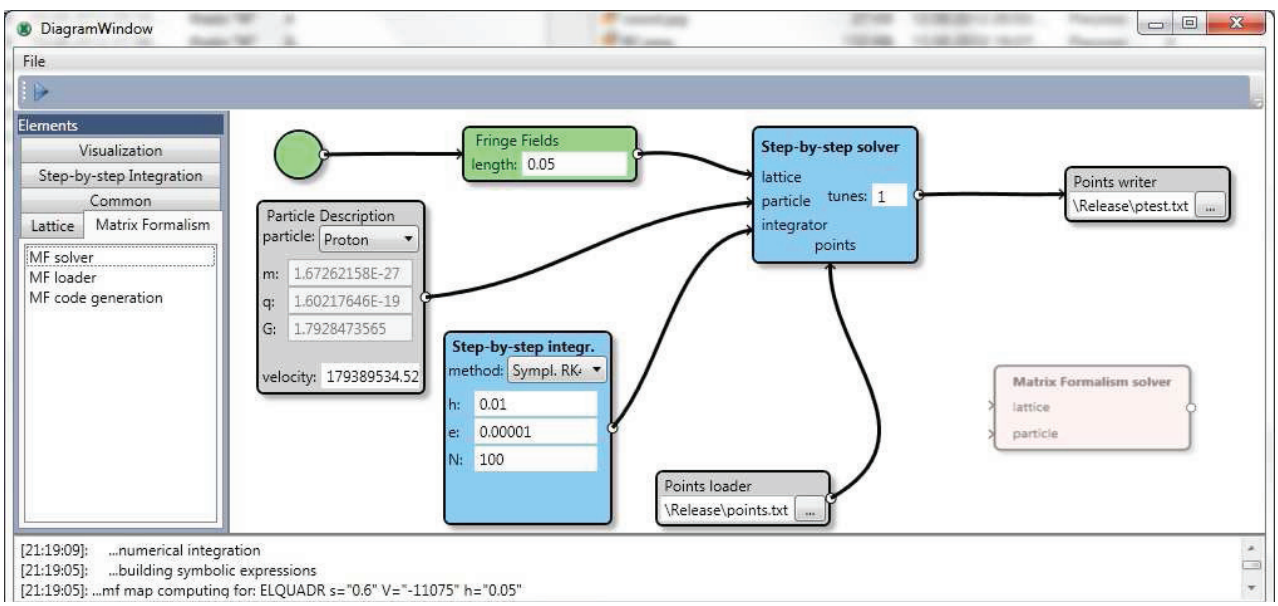


Figure 2: GUI for particle tracking.

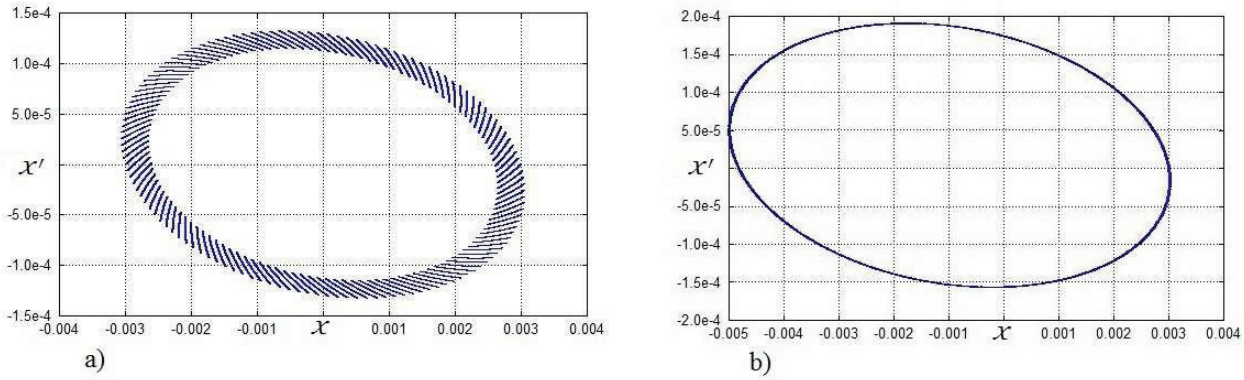


Figure 3: Fringe fields influence. a) Tracking without energy conservation. b) Energy conservation correction.

SIMULATION

The approach was used to examining particle dynamics in electrostatic storage ring. The lattice consist of hyperbolic quadrupoles, cylindrical deflectors and drifts.

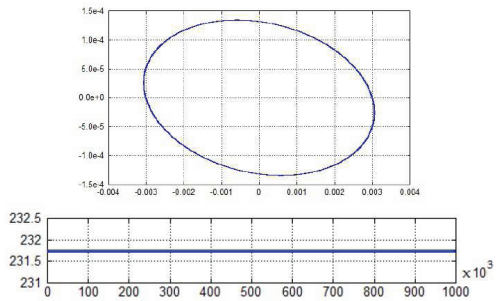


Figure 4: Symplectic phase plane and energy conservation.

Figure (4) shows the result of simulation of a particle. Ellips in $x - x'$ phase plane indicates to the symplectic mapping. The kinetic energy is also conserved.

Fringe Fields Influence

To introduce fringe fields simple model was used, when the potential in fringe fields change as

$$\tilde{\Phi}(x, y, s) = k(s) \cdot \Phi(x, y, s), \qquad (7)$$

where $\Phi(x, y, s)$ is the potential in ideal fields, $k(s)$ is a function of changes of the potential along the design orbit. For example, $k(s)$ can described by linear function or Enge function.

Although the electric fields which described by the Equation (7) do not satisfy to the physic laws, it can be use for modeling. Simulation in Cartesian coordinates indicates to the excessive acceleration of particles (see Fig. 3, a). This allows us to measure it and introduce to the Equation (6) a correction $v_0' = const$.In such way nonphysical acceleration is compensated by artificial chang of the mathematical model.

In Fig. (3, b) phase plane with fringe fields influence is shown. Center of the ellipse is shifted relative to the origin and symplectic condition is satisfied.

CONCLUSION

The approach described above is devoted to the high precession step-by-step integration. On the other hand there are exist mapping algorithms for beam dynamic simulation. Such methods allows to build map corresponded to the dynamic system. In the paper [3] matrix formalism for solving of ODE is presented. The same mathematical model is used for beam dynamics description.

Step-by-step integration allows us to obtain the correction of velocity derivation in fringe fields. This correction can be introduce to the mapping approach. So tracking algorithms aare used for precision estimation and examining of the particle dynamics, and mapping approaches are used for modeling of long-term evolution of beam.

ACKNOWLEDGMENTS

The research is a part of JEDI[1] collaboration. The author would like to thank Yu. Senichev for advices and remarks, D. Zyuzin for comparative calculations on COSY Infinity and special thanks for my scientific supervisor S. Andrianov.

REFERENCES

[1] V.V. Balandin, N.I. Golubeva, "Hamiltonian methods for the study of polarized proton beam dynamics in accelerators and storage rings," `arxiv.org/pdf/physics/9903032`

[2] W. Oevel, M. Sofroniou, "Symplectic Runge-Kutta schemes II: classification of symmetric methods," `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.46.5060`

[3] A. Ivanov, S. Andrianov, "Matrix formalism for long-term evolution of charged particle and spin dynamics in electrostatic fields," WEACC3, these proceedings.

[1] Juelich Electric Dipole Moment Invenstigations, Spokes-persons: A. Lehrach, J. Pretz, and F. Rathmann

TRAVELING POLES ELIMINATION SCHEME AND CALCULATIONS OF EXTERNAL QUALITY FACTORS OF HOMS IN SC CAVITIES*

T. Galek†, T. Flisgen and U. van Rienen, Rostock University, Rostock, Germany
B. Riemann, Technical University Dortmund, Dortmund, Germany
A. Neumann, Helmholtz-Zentrum Berlin, Berlin, Germany

Abstract

The main scope of this work is the automation of the extraction procedure of the external quality factors Q_{ext} of Higher Order Modes (HOMs) in Superconducting (SC) radio frequency (RF) cavities. The HOMs are generated by charged particle beams traveling through a SC cavity at the speed of light ($\beta \approx 1$). The HOMs decay very slowly, depending on localization inside the structure and cell-to-cell coupling, and may influence succeeding charged particle bunches. Thus it is important, at the SC cavity design optimization stage, to calculate the Q_{ext} of HOMs. Traveling Poles Elimination (TPE) scheme was used to automatically extract Q_{ext} from the transmission spectra and careful eigenmode analysis of the SC cavity was performed to confirm TPE results. The eigenmode analysis also delivers important information about band structure, cell-to-cell coupling and allows rapid identification of modes that could interact with the charged particle bunches.

INTRODUCTION

The SC RF cavity presented in this article is a 1.3 GHz 7-cell Cornell design modified TESLA cavity with JLab HOM waveguide couplers as shown in figure 1. The discussed SC RF cavity will be used in the Berlin Energy Recovery Linac Project (B*ERL*inPro), which is currently under development for a CW LINAC technology and expertise required to drive next-generation Energy Recovery Linacs (ERLs) [1]. The main priority on the current stage of the cavity design requires strongly damped HOMs in order to obtain high performance of the linac.

Using a modern simulation software one can efficiently calculate all the necessary quantities during the optimization steps. Simulations used to obtain results presented in this article can be divided into two main categories: eigenmode simulations and frequency domain simulations. The eigenmode simulations give us important information about all the modes that can exist in the model structure in the given frequency range. Important quantities can be calculated as a post processing step, e.g. R/Q which is a measure of a mode interaction with the charged particle beam, E_{peak}/E_{acc} and H_{peak}/E_{acc} which are relevant to suppression of field emission and thermal break down. The frequency domain simulations are used to obtain S-parameter spectra from which Q_{ext} factors of HOMs can be extracted. For this purpose we present an automated procedure that is using vector fitting with rational functions to express the S-parameter transmission spectra with a set of poles. The Traveling Poles Elimination (TPE) scheme is a simple iterative procedure which main purpose is to detect static poles and calculate external quality factors. All the simulations were performed using CST Microwave Studio 2012 (CST MWS) [2].

* Work funded by EU FP7 Research Infrastructure Grant No. 227579 and by German Federal Ministry of Research and Education, Project: 05K10HRC.

† tomasz.galek@uni-rostock.de

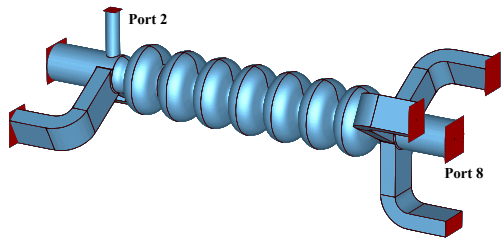


Figure 1: 7-cell TESLA cavity with coaxial input and HOM waveguide couplers

POLE FITTING

Rational Fitting of S-Parameter Spectra

For the extraction of the external quality factors Q_{ext} from S-parameter spectrum the fast implementation of the Vector Fitting (VF) algorithm was used [3]. The vector fitting is an iterative procedure of pole relocation by solving a linear least squares problem until the convergence criterion is met. The VF employs a method to ensure stable poles by flipping unstable poles into the left half complex plane. To achieve a faster convergence the algorithm uses, during the pole identification step, a relaxed non-triviality constraint and utilizes the matrix structure [4, 5, 6].

The S-parameter spectra are assumed to follow the complex rational function approximation

$$S(f) = \sum_{k=1}^{N} \frac{a_k}{2\pi i f - p_k} + R_k, \qquad (1)$$

where $i^2 = -1$ is the imaginary unit, a_k the residues, p_k complex conjugate pairs of poles and R_k a frequency-independent residual summarizing all other contributions. The complex pole $p_k = \alpha_k + i\omega_k$ contains a resonance frequency $\omega_k = 2\pi f_k$ and an attenuation constant α_k. The quality factor Q_k for a given pole can be obtained using

$$Q_k = -\frac{\Im(p_k)}{2\Re(p_k)} = -\frac{\omega_k}{2\alpha_k}. \quad (2)$$

A purely imaginary pole ($\alpha = 0$) would correspond to an infinite quality factor, thus any sharp peak appearing in the S-parameter spectrum will have very high Q_{ext}. This applies to passbands with very narrow bandwidth as well.

Traveling Poles Elimination Scheme

The Traveling Poles Elimination (TPE) scheme is a simple iterative procedure using the VF algorithm, with a goal to detect static poles among a set of unstable poles. The TPE procedure can be summarized in few steps:

- The first TPE iteration uses a set of starting poles. The number and location of the starting poles depends on number of peaks in the S-parameter spectrum. A simple peak finding procedure automatically finds peaks and assigns starting complex conjugate pair of poles to each peak.
- In subsequent TPE iterations the set of fitted poles from the previous TPE iteration is expanded by a number of additional pairs of poles added randomly. The number of freshly added poles should not exceed 10% of the total number of poles from previous TPE iteration.
- When all the TPE iterations are done, the poles are sorted and close neighbors are detected. The set of poles from the last TPE iteration is taken as a target set. For each pole in the target set, the sorting procedure is searching for close neighbors in pole sets from the previous TPE iterations. The closest neighbors are found within frequency TOL_f and quality factor TOL_Q tolerances defined by the user.

Validation of the TPE Procedure

The TPE procedure was validated on a simpler model of the 7-cell cavity. All the geometrical parameters were kept the same only the test structure consists of two end-cells, and the same port setup, as used for the S-parameter simulations. The frequency domain simulations were performed using fast resonant frequency solver in CST MWS, using hexahedral mesh with 15 steps per wavelength, resulting in total of 1.5 million mesh cells, and frequency range 1.2 - 3.3 GHz. The spectrum used for validation is a transmission spectrum S8(3)2(1), notation used is similar to that of CST MWS, from coaxial port 2 using TEM mode 1 to the beam pipe port 8 using TM_{01} monopole mode 3. Additionally to cross check the results, eigenmode simulations were performed for the 2-cell test cavity using JDM eigenmode solver in CST MWS, which allows to calculate quality factors for all the modes.

Figure 2 shows the S8(3)2(1) transmission spectrum, Q_{ext} factors extracted using TPE procedure and Q_{ext} factors calculated using eigenmode solver. There are many eigenmode Q_{ext} factors not matched with static poles, these modes are mostly waveguide modes, starting to appear around 1.578 GHz (cutoff frequency of the first waveguide mode) and have low quality factors $Q_{ext} < 10^3$. The waveguide modes are irrelevant in S-parameter spectrum because these modes arise only in eigenmode simulations due to boundary conditions enforced by the eigenmode solver. The other poles found by the TPE procedure are in good agreement with the Q_{ext} factors from eigenmode calculations.

EIGENMODE ANALYSIS

The eigenmode analysis was performed in a similar way to the one presented by R. Wanzenberg for the 9-cell TESLA cavity [7]. Two separate eigenmode simulations for just a single cell, with periodic boundary conditions (PBC) at x-min and x-max limits, were computed. While using the PBC one can control the phase shift from one cell to the other, in this case it is 0° and 180° at the PBC. The eigenmode simulations were performed using a tetrahedral mesh with curved elements (2nd order). The frequency range was set to 1.2 - 3.3 GHz, and 26 modes per simulation were calculated.

What one can learn from such an approach is the resonance frequency of the fundamental 0-mode f_0 (0° phase shift at PBC) and the π-mode f_π (180° phase shift at PBC). The same rules apply to higher order modes. The passband width of a given mode is given by f_0 and f_π, and all the resonance frequencies of the modes within one band follow a cosine-like dispersion curve, and the number of modes in the band depends on the number of the cells in the cavity. To calculate the cell-to-cell coupling factor k_{cc} one needs only f_0 and f_π [7, 8]

$$k_{cc} = 2 \cdot \frac{f_\pi - f_0}{f_\pi + f_0}. \quad (3)$$

The k_{cc} factor can be either positive, i.e., $f_0 < f_\pi$, or negative, i.e., $f_0 > f_\pi$. In addition the k_{cc} factor specifies the passband width, smaller k_{cc} gives narrower passbands. For small ($|k_{cc}| \leq 0.01$) values there is a danger that if the given mode is excited by the beam, e.g., somewhere in the middle of the cavity, it will propagate out and decay very slowly. Thus the k_{cc} factor gives us a preliminary knowledge of which modes can be dangerous or trapped. In table 1 the results of the eigenmode analysis are gathered, including the f_0 and f_π of all the modes, mode type, cell-to-cell coupling k_{cc} factor, R/Q and Q_{ext} extracted from the 7-cell cavity S-parameter transmission spectra using TPE procedure.

The eigenmode analysis results are in good agreement with the 9-cell TESLA cavity [7] results. There are of course some differences in the band frequencies, due to the fact that it is a bit different geometry. There is one inconsistency though, an additional TE monopole mode (TE M1 in table 1, mode number 9 for 0° phase shift and mode number 13 for 180° phase shift) that is not in the report by R. Wanzenberg [7]. Additionally current eigenmode results contain three sextupole bands not included in [7]. The

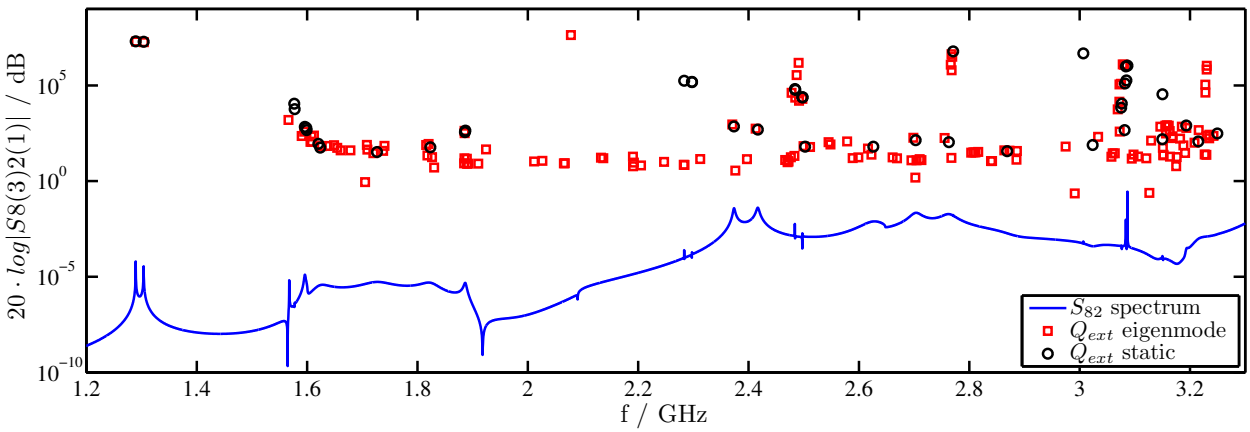


Figure 2: The S-parameter spectrum for the 2-cell test structure (blue solid line). The external quality factors were obtained during eigenmode simulations of the test structure (red squares) and using TPE procedure (black circles).

Table 1: Overview of the Eigenmode Analysis Results

		Phase advance 0°					Phase advance 180°				
Mode	**f/GHz**	**R/Q**	**Q_{ext}**	**R/Q Q_{ext}**	**Mode**	**f/GHz**	**R/Q**	**Q_{ext}**	**R/Q Q_{ext}**	**Type**	**k_{cc}**
1	1.273	1.44E+02	1.97E+07	2.83E+09	1	1.3	111.411	1.75E+07	1.95E+09	TM M1	0.021
2, 3	1.593	3.80E-09	1.29E+04	4.91E-05	4, 5	1.814	4.85E-08	5.82E+01	2.82E-06	TE D1	0.13
4, 5	1.888	2.56E-08	4.34E+02	1.11E-05	2, 3	1.782	7.97E-08	3.60E+01	2.87E-06	TM D1	-0.058
6, 7	2.278	4.17E-08	1.75E+05	7.31E-03	6, 7	2.304	1.91E-07	1.50E+05	2.87E-02	TE Q1	0.011
8	2.438	26.1306	4.92E+02	1.29E+04	8	2.354	23.7545	7.14E+02	1.70E+04	TM M2	-0.035
9	2.484	3.23E-08	1.81E+07	5.84E-01	13	2.49	1.04E-08	1.03E+07	1.08E-01	TE M1	0.0024
10, 11	2.492	3.09E-07	4.79E+04	1.48E-02	11, 12	2.467	8.53E-07	6.50E+04	5.54E-02	TM Q1	-0.01
12, 13	2.552	1.21E-07	7.80E+01	9.43E-06	9, 10	2.444	7.70E-08	7.20E+01	5.54E-06	TE D2	-0.043
14	2.675	0.416747	1.38E+02	5.76E+01	18	2.789	83.8406	1.00E+02	8.38E+03	TM M3	0.042
15, 16	2.765	1.21E-08	3.31E+07	3.99E-01	14, 15	2.768	2.81E-07	1.52E+07	4.26E+00	TE S1	0.0011
17, 18	3.012	1.38E-07	7.46E+01	1.03E-05	16, 17	2.785	8.37E-06	1.08E+02	9.02E-04	TE D3	-0.078
19, 20	3.075	3.65E-08	1.13E+04	4.11E-04	21, 22	3.071	2.02E-08	7.50E+03	1.52E-04	TM S1	-0.0013
21, 22	3.11	5.14E-07	9.80E+01	5.03E-05	19, 20	3.065	1.01E-06	2.86E+01	2.88E-05	TE D4	-0.015
23, 24	3.165	5.99E-07	2.84E+02	1.70E-04	25, 26	3.278	1.71E-08	2.84E+02	4.84E-06	TE Q2	0.035
25, 26	3.227	9.75E-11	2.83E+02	2.76E-08	23, 24	3.228	6.10E-09	2.83E+02	1.73E-06	TE S2	0.0003

R/Q values have been calculated on the beam axis for all the modes.

CONCLUSIONS

In this work we have presented a very simple numerical method to automatically extract external quality factors and resonance frequencies from the transmission S-parameter spectra. The traveling pole elimination scheme was validated on a simplified model of the cavity with only two cells. Test structure geometric dimensions were kept the same as for the full 7-cell cavity. Afterward the VF and TPE procedures were used to study real SC RF 7-cell TESLA cavity. The Q_{ext} factors were extracted for all the modes in 1.2 - 3.3 GHz frequency range. These results in combination with results obtained from eigenomde simulations give detailed insight into HOMs behavior in 1.2 - 3.3 GHz frequency range. The optimization of the 7-cell cavity for the B*ERL*inPro main linac requires investigations of HOMs in higher frequency ranges, thus the same methodology will be applied in the future.

REFERENCES

[1] A. Neumann et al., "Status of the HOM Calculations for the B*ERL*inPro Main Linac Cavity", ID: 1462 - FRAAC3 (this conference).

[2] CST AG, http://www.cst.com

[3] http://www.energy.sintef.no/Produkt/VECTFIT/

[4] B. Gustavsen and A. Semlyen, IEEE Trans. Power Delivery, 14(3):10521061, July 1999.

[5] B. Gustavsen, IEEE Trans. Power Delivery, 21(6):15871592, June 2006.

[6] D. Deschrijver et al., IEEE Microwave and Wireless Components Letters, 18(6):383285, June 2008.

[7] Rainer Wanzenberg, "Monopole, dipole and quadrupole passbands of the TESLA 9-cell cavity", In TESLA reports 2001-33, 2001.

[8] S. Belomestnykh and V, Shemelin, "High-β Cavity Design - A Tutorial", In 12th International Workshop on RF Superconductivity (SRF2005), number SRF060424-03, 2005.

COMPARISONS OF DIFFERENT ELECTROMAGNETIC SOLVERS FOR ACCELERATOR SIMULATIONS*

J. Xu†, X. Zhufu, R. Zhao, Institute of Software, Beijing, 100190, China
C. Li, X. Qi, L. Yang, Institute of Modern Physics, Lanzhou, 730000, China
M. Min, Argonne National Laboratory, Argonne, IL 60439, USA

Abstract

Electromagnetic simulations are fundamental for accelerator modeling. In this paper two high-order numerical methods will be studied. These include continuous Galerkin (CG) method with vector bases, and discontinuous Galerkin (DG) method with nodal bases. Both methods apply domain decomposition method for the parallelization. Due to the difference in the numerical methods, these methods have different performance in speed and accuracy. DG method on unstructured grid has the advantages of easy parallelization, good scalability, and strong capability to handle complex geometries. Benchmarks of these methods will be shown on simple geometries in detail first. Then they will be applied for simulation in accelerator devices, and the results will be compared and discussed.

INTRODUCTION

Time dependent electromagnetic simulations are bases for many accelerator simulations. It is an important area in computational electromagnetics. Many numerical methods have been developed till now, such as finite difference (FDM), finite volume (FVM) and finite element methods (FEM). FEM has the advantage of handling complex geometries, therefore many efforts have been made with FEM. For FEM methods, there are two types of methods have been proved successful, they are CG with vector base and DG with Nodal base. In this paper we will study the performance of these two methods.

The paper is organized in the following way: the numerical method is explained and algorithms are compared in section 2, validation is shown in section 3, then benchmarks results are shown in section 4, and a comparison on wakefield simulations is given in section 5. At last, the conclusion is drawn in section 6.

NUMERICAL METHOD

Maxwell's Equation

In 3D domain Ω, time dependent Maxwell's equations can be written as:

$$\frac{\partial \mathbf{B}}{\partial \mathbf{t}} = -\nabla \times \mathbf{E}, \qquad \frac{\partial \mathbf{D}}{\partial \mathbf{t}} = \nabla \times \mathbf{H} + \mathbf{J} \tag{1}$$

$$\nabla \cdot \mathbf{D} = \rho, \qquad \nabla \cdot \mathbf{B} = \mathbf{0}, \quad \mathbf{x} \in \mathbf{\Omega}, \tag{2}$$

$$\hat{\mathbf{n}} \times \mathbf{E} = \mathbf{0}, \qquad \hat{\mathbf{n}} \cdot \mathbf{H} = \mathbf{0} \quad \mathbf{x} \in \partial\mathbf{\Omega}, \tag{3}$$

where, the electric field $\mathbf{E}$, electric flux density $\mathbf{D}$, as well as the magnetic field $\mathbf{H}$ and the magnetic flux density $\mathbf{B}$ are related through the constitutive relations $\mathbf{D} = \epsilon\mathbf{E}$, $\mathbf{B} = \mu\mathbf{H}$.

CG Formulation

Maxwell's Equation (1) can be written in following inhomogeneous wave equation

$$\nabla \times \frac{1}{\mu_r} \nabla \times \mathbf{E} + \frac{\epsilon_\mathbf{r}}{\mathbf{c_0^2}} \frac{\partial^\mathbf{2} \mathbf{E}}{\partial \mathbf{t^2}} = -\mu_\mathbf{0} \frac{\partial \mathbf{J}}{\partial \mathbf{t}} \tag{4}$$

which times edge bases and apply electric (PEC) boundary condition to arrive at

$$\mathbf{M} \frac{\mathbf{1}}{\mathbf{c_0^2}} \frac{\mathbf{d^2 e}}{\mathbf{dt^2}} + \mathbf{Se} = -\vec{\mathbf{f}} \tag{5}$$

where $\mathbf{M}$ and $\mathbf{S}$ are mass and stiffness matrices in the following

$$\mathbf{M_{ij}} = \int_\Omega \epsilon_r \vec{W}_i \cdot \vec{W}_j d\Omega \tag{6}$$

$$\mathbf{S_{ij}} = \int_\Omega \frac{1}{\mu_r} \nabla \times \vec{W}_i \cdot \nabla \times \vec{W}_j d\Omega \tag{7}$$

$$f_i = \int_\Omega \mu_0 \vec{W}_i \cdot \frac{\partial \vec{J}}{\partial t} d\Omega \tag{8}$$

Based on the Newmark-Beta formulation [1, 2], Equation (5) can be solved to obtain e^{n+1} with

$$\begin{aligned} e^{n+1} = {} & (M + \beta(c_0\delta t)^2 S)^{-1} \cdot \\ & \{(2M - (1-2\beta)(c_0\delta t)^2 S)e^n \\ & -(M + \beta(c_0\delta t)^2 S)e^{n-1} \\ & (c_0\delta t)^2(\beta f^{n+1} + (1-2\beta)f^n + \beta f^{n-1})\} \end{aligned} \tag{9}$$

More detailed information can be found in [1, 2, 3]. The CG method is based on Nédélec edge bases, which is explained in detail in [4], and we omit it due to the constraint of page limit.

DG Formulation

The discrete form of DG is:

$$\frac{\mathbf{dE_N}}{\mathbf{dt}} = \mathbf{M^{-1}S} \times \mathbf{H_N} + \mathbf{S^E}$$

* Work supported by Funding from Chinese Acdemy of Sciences
† xu_jin@iscas.ac.cn

$$+ \quad \mathbf{M^{-1}F}(\hat{\mathbf{n}} \times \frac{\mathbf{Z}^+[\mathbf{H_N}] - \hat{\mathbf{n}} \times [\mathbf{E_N}]}{\mathbf{Z}^+ + \mathbf{Z}^-}) \Big|_{\partial \mathbf{D}} \quad (10)$$

$$\frac{\mathbf{dH_N}}{\mathbf{dt}} = \mathbf{M^{-1}S} \times \mathbf{E_N} + \mathbf{S^H}$$

$$+ \quad \mathbf{M^{-1}F}(\hat{\mathbf{n}} \times \frac{\hat{\mathbf{n}} \times [\mathbf{H_N}] + \mathbf{Y}^+[\mathbf{E_N}]}{\mathbf{Z}^+ + \mathbf{Z}^-}) \Big|_{\partial \mathbf{D}} \quad (11)$$

where $\hat{\mathbf{n}}$ is the normal vector on boundary, ϕ is the test function in Ω. $[\mathbf{E_N}] = \mathbf{E_N^+} - \mathbf{E_N^-}$ and $[\mathbf{H_N}] = \mathbf{H_N^+} - \mathbf{H_N^-}$ for the solutions $\mathbf{E_N^-}$ and $\mathbf{H_N^-}$ in the local domain and $\mathbf{E_N^+}$ and $\mathbf{H_N^+}$ in neighboring elements. M, S are mass and stifness matrixes respectively.

More details on DG method and nodal base Finite Element method can be found in [5, 6, 7, 8, 9], which also be skipped due to the page limit.

We adopt low-storage five-stage fourth-order explicit Runge-Kutta (LSERK) scheme has been used [5].

Methods Comparison

- As the order of the vector base is the order of nodal base minus 2, they have different accuracy for the interpolation, derivative and curl operators.
- Total degree of freedom when using CG with vector base is much less than the total degree of freedom when using DG nodal base.
- Both algorithms solve the Maxwell's equation in time domain, but since the CG method use nodal base, it has no extra cost for transforming the fields to the spectral space.
- CG method needs to inverse global matrixes, while DG method only needs to inverse local matrixes. This makes the DG method to have better scalability and faster speed.
- DG method need to exchange face information for each element, this usually needs more memory and the communication is more than CG method.
- DG method uses 4th order Runge-Kutta integration scheme, while CG method uses 2nd order time integration scheme.

VALIDATION

Analytical Solution

In order to further verify our results in 2D case, we conduct similar tests in 3D. The analytic solution with periodic boundary condition in $[-\pi, \pi]^3$ are:

$$\begin{aligned}
E_x &= 0.0 \\
E_y &= \cos(x) \cdot \sin(y) \cdot \sin(z) \cdot \cos(\sqrt{3}t) \\
E_z &= \cos(x) \cdot \cos(y) \cdot \cos(z) \cdot \cos(\sqrt{3}t) \\
H_x &= 2.0 \cdot \cos(x) \cdot \sin(y) \cdot \cos(z) \cdot \sin(\sqrt{3}t)/\sqrt{3} \\
H_y &= -\sin(x) \cdot \cos(y) \cdot \cos(z) \cdot \sin(\sqrt{3}t)/\sqrt{3} \\
H_z &= \sin(x) \cdot \sin(y) \cdot \sin(z) \cdot \sin(\sqrt{3}t)/\sqrt{3}
\end{aligned} \quad (12)$$

Table 1: Comparison of DOFs (E=203)

Polynomial Order	3	4	5
CG vector base	165	1038	3231
DG nodal base	4902	10184	18310

The errors verses time have been plotted in Fig. 1. As can be seen that the errors with both method decrease as polynomial order increases. Since the order of the vector base (PM) equals the order of nodal base (P) minus 2, for the same P, the error of vector base is larger than the nodal base. Another difference is the errors for the DG method oscillate around a constant level, while the errors for the CG method increase linearly.

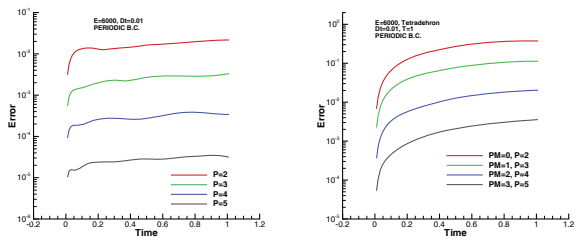


Figure 1: Errors versus polynomial orders: Nodal base (left); vector base (right)

BENCHMARKS

In order to compare performance in detail, we study them in following different perspectives.

Degree of Freedom

First we compare degrees of freedom in Table 1, as can be seen that DOF of vector is much less than the DOF of nodal base. This makes the size of global mass and stiffness matrices smaller when using vector base than using nodal base. This usually leads to better condition number for the matrix and can achieve fast speed. As DG method been used instead of CG, inverse of global matrix has been avoided which eliminate this shortcoming.

Iteration Convergence Speed

In order to see the differences of speed with different vector base orders, the comparison has been given in Table 2. From the table, the increase of time is nonlinear. This make it difficult to use for higher order vector base.

Speed Comparison

Next, we compare their speeds with 6000 tetrahedra and running for T=1. Currently CG solver has not been fully optimized. Table 3 shows that for P=2 (PM=0), CG with

Table 2: Iteration Steps for Inverse Operation with Vector Base (E=6000)

Nodal Base Order (P)	2	3	4
Vector Base Order (PM=P-2)	2	3	4
Iteration steps	36	111	276

Table 3: Comparison of Speeds (E=6000)

Nodal Base Order	2	3	4	5
DG Nodal Base	6.35	14.1	24.3	46.1
Vector Base Order	0	1	2	3
CG Vector Base	2.14	32.86	333.95	2561.78

vector is faster. When P larger than 2, DG with nodal base is faster. As the inverse takes more and more time, the speed for CG with vector base increase nonlinearly and this is a big challenge for using CG method.

Comparison of Wakefield Simulations

We have used these two solvers for a wakefield simulation. A Gaussan beam has been simulated in the device shown in Fig. 3 . The longitudinal wake potential has been calculated at r=1 for different beam bunch sizes, $\sigma_z = 0.25, 0.5, 0.75, 1.0$ and $\sigma_r = 0.1$. A 2D poisson equation has been solved to get the initial electric field and current has been activated to simulate the charged beam.

Left plot in Fig. 2 shows the wakepotential comparison. Solid line is the result using DG nodal base, and dash line is the result for CG vector base. They are consistent, this means both solvers produce correct results and they both can be used for the wake field simulations. The right one shows the beam distribution function for different $\sigma_z = 0.25, 0.5, 0.75, 1.0$.

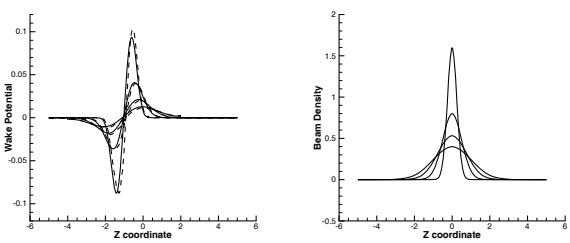


Figure 2: Comparison of wakepotential(left: Nodal base-solid, vector base-dash); beam distribution for (σ_z=0.25,0.5,0.75,1.0)

Figure 3 shows the electric field contours from the simulation. Totally 561883 tetrahedra elements have been used, and P=2.

CONCLUSION

In this paper, we have compared performance of two types time dependant EM solvers. They both can be used for accelerator simulations. Comparisons have been performed in terms of DOFs, iteration steps, accuracy, and speed. At last, results for a wakefield simulation has been compared. CG method with vector base has smaller DOFs than CG with nodal base, but due to the nature of solving global matrix, CG method with vector base has the challenge when DOF becomes large. On the other side, the memory cost of DG is much larger than CG method, which is a challenge for DG method.

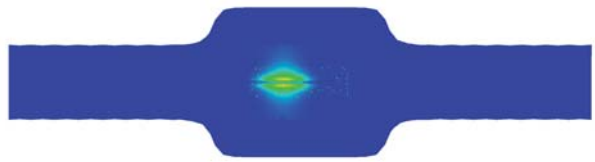


Figure 3: Contour of electric field

ACKNOWLEDGMENT

Thanks supporting groups from Institute of Software, Institute of Modern Physics and Argonne National Laboratory.

REFERENCES

[1] N. M. Newmark, A method of computation for structural dynamics, Journal of Eng. Mech. Div., ASCE, vol. 85, pp. 67-94, July 1959.

[2] O.C. Zienkiewicz, A new look at the Newmark, Houboult and other time stepping formulas. A weighted residual approach., Earthquake Engineering and Structural Dynamics, vol. 5, pp. 413-418, 1977.

[3] A. Candel, A. Kabel, L. Lee, Z. Li, C. Limborg, and C. Ng, Parallel 3D Finite Element Particle-In-Cell Code for High-Fidelity RF Gun Simulations, Proceedings of LINAC08, Victoria, BC, Canada, 2008.

[4] J. Jin, The Finite Element Method in Electromagnetics, second edition John Wiley & Sons, Inc., New York, (2002).

[5] J.S. Hesthaven and T. Warburton, Nodal Discontinuous Galerkin Methods: Algorithms, Analysis and Applications, Springer, New York, (2008).

[6] M. Min, P.F. Fisher, and T. Chae, Spectral Element Discontinuous Galerkin Simulations for Wake Potential Calculations: NEKCEM, Proceedings of PAC07, Albuquerque, New Mexico, USA, 2007.

[7] M. Dubiner, Spectral Methods on triangles and other domains, *J. Sci. Comp.*, **6**, 345, (1991).

[8] T. Koornwinder, Two-variable analogues of the classical orthogonal polynomials, in *Theory and Applications of Special Functions*, edited by A.R.A. (Academic Press, San Diego, 1975).

[9] W.H. Reed and T.R. Hill, Triangular mesh methods for the beutron transport equation, Tech. Report LA-UR-73-479, Los Alamos Scientific Laboratory, Los Alamos, NM, 1973.

STOCHASTIC RESPONSE SURFACE METHOD FOR STUDYING MICROPHONING AND LORENTZ DETUNING OF ACCELERATOR CAVITIES*

J. Deryckere, T. Roggen†, B. Masschaele, H. De Gersem, KU Leuven - Kulak, Kortrijk, Belgium

Abstract

The dependence of the resonant frequency of an RF cavity on its geometry is represented by a stochastic response surface model, which is constructed on the basis of a few eigenmode solutions extended with sensitivity information. The response surface model is used for statistic analysis and for calculating the effect of Lorentz detuning.

INTRODUCTION

High energy cavities are used within a very small frequency range. Any mechanical deformation, albeit small, may lead to an unacceptable shift of the frequency of the applied eigenmode [1]. To achieve a design that is robust against microphoning and Lorentz detuning [2–5], the simulation tool should deal with small changes in geometry in a consistent way and should be able to calculate the resonance frequencies with a relative accuracy of 10^{-5}. The simple and straightforward procedure which changes and remeshes the geometry and repeats the eigenmode solving, turns out to be inefficient because the introduced discretisation errors will mask the small changes in eigenfrequency, unless an prohibitively fine mesh is used [6]. This paper introduces two techniques to overcome this problem: (a) the eigenmode solver also delivers the sensitivities of the eigenfrequency and thereby increases the amount of information obtained for a single set of geometric parameters; (b) the eigenfrequency is modelled by a stochastic response surface method which allows reliable interpolation and uses the concept of uncertainty to deal with errors introduced by remeshing.

CAVITY EIGENMODE SOLVERS

The eigenmodes of the cavity are calculated by solving one of

$$\nabla \times \left(\frac{1}{\mu} \nabla \times \vec{E} \right) = \omega^2 \varepsilon \vec{E} \quad (1)$$

$$\nabla \times \left(\frac{1}{\varepsilon} \nabla \times \vec{H} \right) = \omega^2 \mu \vec{H} \quad (2)$$

with $\vec{E}$ and $\vec{H}$ the electric and magnetic field strengths, ω the angular frequency, ε the permittivity and μ the permeability [7, 8]. Only the eigenmodes with the lowest eigenfrequencies are relevant. When discretised by the finite-element (FE) method or the finite-integration technique (FIT), (1) and (2) become

$$\mathbf{K}_{\frac{1}{\mu}} \widehat{\mathbf{e}} = \omega^2 \mathbf{M}_{\varepsilon} \widehat{\mathbf{e}} \quad (3)$$

$$\mathbf{K}_{\frac{1}{\varepsilon}} \widehat{\mathbf{h}} = \omega^2 \mathbf{M}_{\mu} \widehat{\mathbf{h}} \quad (4)$$

and will further be expressed generically by $\mathbf{K}_\alpha \mathbf{u} = \omega^2 \mathbf{M}_\beta \mathbf{u}$ where $(\alpha, \beta, \mathbf{u})$ either stands for $(\frac{1}{\mu}, \varepsilon, \widehat{\mathbf{e}})$ or $(\frac{1}{\varepsilon}, \mu, \widehat{\mathbf{h}})$ and $\widehat{\mathbf{e}}_j$ and $\widehat{\mathbf{h}}_j$ are the degrees of freedom for the electric and magnetic field strengths respectively. The matrix coefficients are

$$\mathbf{K}_{\alpha,i,j} = \int_V \alpha \left(\nabla \times \vec{w}_i \right) \cdot \left(\nabla \times \vec{w}_j \right) \mathrm{d}V \quad (5)$$

$$\mathbf{M}_{\beta,i,j} = \int_V \beta \vec{w}_i \cdot \vec{w}_j \, \mathrm{d}V \quad (6)$$

with $\vec{w}_i(x, y, z)$ FE or FIT shape functions and V the computational domain.

In many situations, the cavity has a cylindrical symmetry and only modes with $\vec{E} = (E_r, 0, E_z)$ and $\vec{H} = (0, H_\theta, 0)$ are relevant. Then, a substantial saving of computation cost is achieved by only triangulating the rz-cross-section of the cavity and discretising (2) by the shape functions

$$\vec{w}_j = \frac{N_j(r,z)}{2\pi r} \vec{e}_\theta \quad (7)$$

$$N_j^{(k)}(r,z) = \frac{a_j^{(k)} + b_j^{(k)} r^2 + c_j^{(k)} z}{2S^{(k)}} \quad (8)$$

where $\vec{e}_\theta$ is the peripheral unit vector, $N_j(r,z)$ are nodal shape functions associated with mesh node j and (8) expresses $N_j(r,z)$ in element k as a function of the coefficients $a_j^{(k)}$, $b_j^{(k)}$ and $c_j^{(k)}$. $N_j(r,z)^{(k)}$ is features a quadratic dependence on r such that a homogeneous electric field can be represented exactly on the mesh [9].

SENSITIVITIES OF THE RESONANCE FREQUENCY

As will become clear below, highly accurate sensitivities of the cavity eigenmode with respect to geometric parameters are of paramount importance for studying microphoning and Lorentz detuning and for a stochastic analysis or optimisation of the design. The sensitivities of the eigenfrequency ω_p of eigenmode $(\omega_p, \mathbf{u}_p)$ with respect to the geometric parameters ζ_q are obtained directly from the eigenvalue solver by [10–12]

$$\frac{\mathrm{d}\omega_p}{\mathrm{d}\zeta_q} = \frac{1}{2\omega_p} \frac{\mathbf{u}_p^H \left(\frac{\mathrm{d}\mathbf{K}_\alpha}{\mathrm{d}\zeta_q} - \omega_p^2 \frac{\mathrm{d}\mathbf{M}_\beta}{\mathrm{d}\zeta_q} \right) \mathbf{u}_p}{\mathbf{u}_p^H \mathbf{M}_\beta \mathbf{u}_p} \quad (9)$$

* Work supported by the grant KUL_3E100118, grant KUL_3E080005, project FP7-Euratom No. 269565 and the Belgian Nuclear Research Centre (SCK·CEN).

† toon.roggen@kuleuven-kulak.be

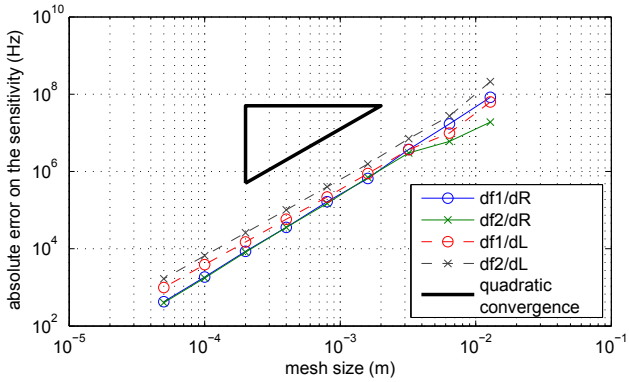


Figure 1: Convergence of the absolute discretisation error for the sensitivities of the first two eigenfrequencies (f_1, f_2) of a pill-box resonator with respect to changes in the radius R and length L of a pill-box resonator.

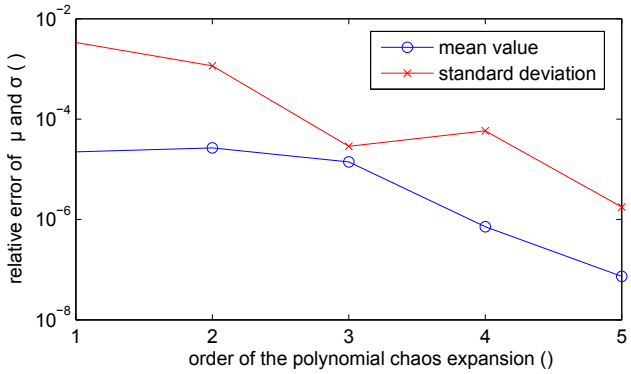


Figure 2: Convergence of the approximation error for the mean value μ and the standard deviation σ with respect to the order of the polynomial chaos expansion.

The calculation of $\frac{\mathrm{d}\mathbf{K}_\alpha}{\mathrm{d}\zeta_q}$ and $\frac{\mathrm{d}\mathbf{K}_\beta}{\mathrm{d}\zeta_q}$ involves the derivatives $\frac{\mathrm{d}\mathbf{w}_j}{\zeta_q}$ of the shape functions with respect to the geometric parameters [13]. The convergence of the discretisation error for the sensitivity calculated by (9) is of the same order as the one for the eigenvalues themselves. An exemplary validation for a 2D eigenmode solver based on (7) and applied to a pill-box resonator is shown in Fig. 1.

STOCHASTIC RESPONSE SURFACE METHOD

The study of microphoning or Lorentz detuning of a resonating cavity and the optimisation of the design require a large number of eigenmode solutions. A direct calculation may become too time consuming. Instead, we solve the eigenmodes for a restricted number of geometries, use these to construct a surrogate model, which is then exploited for calculating the properties for slightly modified geometries. Moreover, we prefer a *stochastic* response surface method (RSM) instead of a standard RSM, in order to account for both model uncertainties and simulation inaccuracies [14, 15]. The uncertain geometric parameters are based on Gaussian variables and the eigenfrequencies are approximated by a series expansion of multi-dimensional Hermite polynomials. Their coefficients are determined by regression.

As a rule of thumb, the data of $2N$ eigenmode solutions are sufficient to obtain a robust estimate of N coefficients. The sensitivities of the resonant frequency with respect to M input parameters can be exploited to diminish the number of required collocation points to $2N/(M+1)$. This is particularly beneficial because the additional calculation of M sensitivities is substantially less time consuming than the eigenmode calculation for a single geometry.

The insertion of the stochastic RSM between the field solver and an outer calculation procedure is illustrated for a stochastic analysis of a cylindrical cavity with radius $R = 30$ mm and length $L = 100$ mm. The standard deviations are $\sigma_R = 1.5$ mm and $\sigma_L = 5$ mm respectively. A reference solution for the eigenfrequency $f = 3.849$ GHz and its standard deviation $\sigma_f = 194.1$ MHz is obtained by 1000 2D FE eigenmode solves. The stochastic RSM preserves the high accuracy of the FE solver. The relative numerical error falls below 10^{-5} for a polynomial chaos expansion of fifth order (Fig. 2). A stochastic RSM of fifth order accuracy set up with 42 FE eigenmode solves achieves the same accuracy as the reference solution. When, additionally, the FE solver provides the sensitivities of the eigenfrequency to R and L, a stochastic RSM for only using 14 sets of geometric parameters is sufficient.

The stochastic RSM allows to organise an analysis with changing geometric parameters as long as their range is covered by the RSM. This is commonly the case for statistical analyses and for studying microphoning and Lorentz detuning.

EXAMPLE: LORENTZ DETUNING

The accelerating eigenmode of the TESLA cavity [16] is computed for a single cell by a 2D eigenmode solver based on (2) discretised by the shape functions (7) and (8). The Lorentz force density or radiation pressure acting on the inner cavity wall is

$$f_{\mathrm{Lor}} = \frac{1}{2}\mu \left(\sum_j \widehat{\mathbf{h}}_j \frac{N_j(r,z)}{2\pi r} \right)^2 \qquad (10)$$

The mechanical deformation is calculated by a 1D axisymmetric shell solver on the boundary of the 2D mesh, yielding the radial displacements $w(z)$, axial displacements $u(z)$ and bending angles of the mesh nodes (Fig. 3) [17,18]. Assuming elastic behaviour, the deformation changes linear with the radiation pressure. However, the radiation pressure is quadratically dependent on the magnitude of the electric field. Also the dependence of the eigenfrequency on the magnitude of the deformation is nonlinear. It is assumed that the pattern of the deformation is

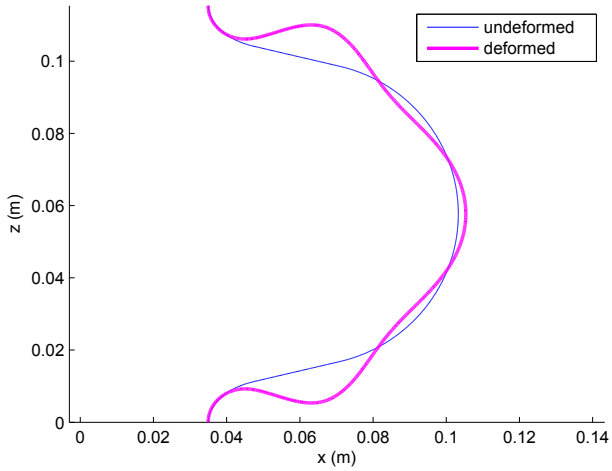


Figure 3: Displacement wall TESLA cavity due to radiation pressure with exaggeration factor 100.

only marginally influenced by the loading such that the deformed cavity wall can be parametrised by

$$\begin{cases} R(z) & \leftarrow & R(z) + \alpha w(z) \\ z & \leftarrow & z + \alpha u(z) \end{cases} \tag{11}$$

where $(R(z), z)$ is the design geometry, $(w(z), u(z))$ is the deformation pattern and α represents the size of the deformation. α is sized such that $\alpha = 1$ corresponds to the deformation when an energy of 1 J is stored in a single cell.

The sensitivity of the eigenfrequency to the deformation pattern is determined without changing the topology of the FE mesh. For numerical reasons, the deformation is spread out over all mesh nodes.

$$\begin{cases} r_j & \leftarrow & r_j + \alpha u(z_j)\frac{r_j}{R(z_j)} \\ z_j & \leftarrow & z_j + \alpha w(z_j)\frac{r_j}{R(z_j)} \end{cases} \tag{12}$$

The dependence of the FE shape functions on α causes $\frac{\mathrm{d}\vec{w}_j}{\mathrm{d}\alpha} \neq 0$ and hence $\frac{\mathrm{d}\mathbf{K}_{\frac{1}{\varepsilon}}}{\mathrm{d}\alpha} \neq 0$ and $\frac{\mathrm{d}\mathbf{K}_{\mu}}{\mathrm{d}\alpha} \neq 0$. The sensitivity of the eigenfrequency on the magnitude of the deformation is calculated from (9) and found to be $\frac{\mathrm{d}f}{\mathrm{d}\alpha} = 75.615$ Hz.

CONCLUSIONS

Significant shifts of the eigenfrequency of a cavity because of small geometric deformations are adequately calculated by an eigenmode solve delivering sensitivity information combined with a stochastic response surface model.

REFERENCES

[1] P. Pierini, D. Barni, A. Bosotti, G. Ciovati, and C. Pagani, "Cavity design tools and applications to the TRASCO project," in *Proceedings of the 1999 Workshop on RF Superconductivity*, Santa Fe, New Mexico, USA, 1999, pp. 380–383.

[2] Z. Conway, J. Fuerst, M. Kelly, K. Shepard, G. Davis, and J. Delayen, "Electro-mechanical properties of spoke-loaded superconducting cavities," in *SRF 2007*, Beijing, China, 2007.

[3] H. Gassot, "Mechanical stability of the RF superconductive cavities," in *EPAC*, Paris, France, 2002, pp. 2235–2237.

[4] V. Akçelik, K. Koa, L. Lee, Z. Lia, C. Nga, and L. Xiaoa, "Shape determination for deformed electromagnetic cavities," *J. Comput. Phys.*, vol. 227, no. 3, pp. 1722–1738, Jan. 2008.

[5] E. Zaplatin, T. Grimm, W. Hartung, M. Johnson, M. Meidlinger, and J. Popielarski, "Structural analysis for a half-reentrant superconducting cavity," in *EPAC*, Edinburgh, Scotland, 2006, pp. 424–426.

[6] W. Ackermann, G. Benderskaya, and T. Weiland, "State of the art in the simulation of electromagnetic fields based on large scale finite element eigenanalysis," *ICS Newsletter*, vol. 17, no. 2, pp. 3–12, 2010.

[7] J. Billen and L. Young, "Poisson Superfish," Los Alamos, Tech. Rep. LA-UR-96-1834, 1996.

[8] U. van Rienen and T. Weiland, "Triangular discretization method for the evaluation of RF-fields in cylindrically symmetric cavities," *IEEE Trans. Magn.*, vol. 21, pp. 2317–2320, 1985.

[9] F. Henrotte, H. Hedia, N. Bamps, A. Genon, A. Nicolet, and W. Legros, "A new method for axisymmetrical linear and nonlinear problems," *IEEE Trans. Magn.*, vol. 29, no. 2, pp. 1352–1355, Mar. 1993.

[10] D. Meidlinger, "A general perturbation theory for cavity mode field patterns," in *SRF 2009*, Berlin, Germany, 2009.

[11] N. Burschäpers, S. Fiege, R. Schuhmann, and A. Walther, "Sensitivity analysis of waveguide eigenvalue problems," *Adv. Radio Sci.*, vol. 9, pp. 85–89, July 2011.

[12] K. Brackebusch, H.-W. Glock, and U. van Rienen, "Calculation of high frequency fields in resonant cavities based on perturbation theory," in *IPAC*, San Sebastian, Spain, Sept. 2011.

[13] T. N. Nguyen and J.-L. Coulomb, "High order FE derivatives versus geometric parameters. implementation on an existing code," *IEEE Trans. Magn.*, vol. 35, no. 3, pp. 1502–1505, May 1999.

[14] S. Isukapalli, A. Roy, and P. Georgopoulos, "Stochastic response surface methods (SRSMs) for uncertainty propagation: application to environmental and biological systems," *Risk Analysis*, vol. 18, no. 3, pp. 351–363, 1998.

[15] D. Gorissen, I. Couckuyt, P. Demeester, T. Dhaene, and K. Crombecq, "A surrogate modeling and adaptive sampling toolbox for computer based design," *Journal of Machine Learning Research*, vol. 11, pp. 2051–2055, 2010.

[16] B. Aune and et al., "Superconducting tesla cavities," *Phys. Rev. ST Accel. Beams*, vol. 3, p. 092001, Sep 2000. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevSTAB.3.092001

[17] K. Surana, "Geometrically nonlinear formulation for the axisymmetric shell elements," *Int. J. Numer. Meth. Engng*, vol. 18, p. 477502, 1982.

[18] C. Polat and Y. Calayir, "Nonlinear static and dynamic analysis of shells of revolution," *Mechanics Research Communications*, vol. 37, pp. 205–209, 2010.

REALISTIC 3-DIMENSIONAL EIGENMODAL ANALYSIS OF ELECTROMAGNETIC CAVITIES USING SURFACE IMPEDANCE BOUNDARY CONDITIONS

Hua Guo*, Peter Arbenz†, Computer Science Department, ETH Zurich, Zurich, Switzerland
Benedikt Oswald‡, Paul Scherrer Institute, Villigen, Switzerland

Abstract

The new X-ray Free Electron Laser (SwissFEL) at the Paul Scherrer Institute (PSI) employs, among many other radio frequency elements, a transverse deflecting cavity for beam diagnostics. Since the fabrication process is expensive, an accurate 3-D eigenmodal analysis is indispensable. The software package Femaxx has been developed for solving large scale eigenvalue problems on distributed memory parallel computers. Usually, it is sufficient to assume that the tangential electric field vanishes on the cavity wall (PEC boundary conditions). Of course, in reality, the cavity wall is conductive such that the tangential electrical field on the wall is nonzero. In order to more realistically model the electric field we impose surface impedance boundary conditions (SIBC) arising from the skin effect model. The resulting nonlinear eigenvalue problem is solved with a nonlinear Jacobi–Davidson method. We demonstrate the performance of the method. First, we investigate the fundamental mode of a pillbox cavity. We study resonance, skin depth and quality factor as a function of the cavity wall conductivity. Second, we analyze the transverse deflecting cavity of the SwissFEL to assess the capability of the method for technologically relevant problems.

FORMULATION OF THE PROBLEM

We wish to calculate the resonant frequencies and the corresponding field distribution in a dielectric electromagnetic cavity. The cavity wall Γ is assumed to be of arbitrary shape; there is no aperture or hole in Γ. The surface conductivity σ_s of Γ is large but finite. The interior Ω of the cavity is assumed to be source-free, and is characterized by $(\mu_0\mu_r, \varepsilon_0\varepsilon_r)$. μ_0 and ε_0 are the magnetic permeability and electric permittivity in free space. μ_r and ε_r are relative magnetic permeability and relative electric permittivity, respectively. At microwave frequencies, μ_r and ε_r can be assumed to be non-dispersive.

In the time-harmonic regime, after eliminating the electric field $\mathbf{E}(\mathbf{x})$, the magnetic field $\mathbf{H}(\mathbf{x})$ satisfies

$$\begin{aligned} \nabla \times (\varepsilon_r^{-1} \nabla \times \mathbf{H}(\mathbf{x})) - k_0^2\, \mu_r\, \mathbf{H}(\mathbf{x}) = \mathbf{0}, \quad \mathbf{x} \in \Omega, \\ \nabla \cdot (\mu_r \mathbf{H}(\mathbf{x})) = 0, \quad \mathbf{x} \in \Omega. \end{aligned} \tag{1}$$

Here, $k_0 = \tilde{\omega}\sqrt{\mu_0\varepsilon_0}$ is the *complex* wave number in free space, $\tilde{\omega} = \omega + i\alpha$ is the *complex* angular frequency with ω the angular frequency and α the exponential decay rate.

*hguo@inf.ethz.ch
†arbenz@inf.ethz.ch
‡benedikt.oswald@psi.ch

We use the surface impedance boundary condition (SIBC) on Γ [1]

$$\mathbf{n} \times (\mathbf{n} \times \mathbf{E}(\mathbf{x})) = Z_s\, \mathbf{n} \times \mathbf{H}(\mathbf{x}), \qquad \mathbf{x} \in \Gamma. \tag{2}$$

Here, Z_s is the *complex* surface impedance and $\mathbf{n}$ the surface normal vector pointing outwards.

We employ Z_s based on the theoretical skin effect model [2]

$$Z_s = \frac{1+i}{\sigma_s \delta}, \tag{3}$$

where σ_s is the surface conductivity, and δ is the skin depth. The real part of Z_s is the surface resistivity, i.e.,

$$R_s = \mathrm{Re}(Z_s) = \frac{1}{\sigma_s \delta}. \tag{4}$$

The skin depth δ is [2]

$$\delta = \sqrt{\frac{2}{\omega \mu_0 \mu_r \sigma_s}}. \tag{5}$$

δ depends on the angular frequency ω. Note that the skin effect model is appropriate only if σ_s is large enough such that (according to [2]): (1) the conduction current is given by Ohm's law and the net charge density is zero; (2) the displacement current is negligible in comparison with the current, i.e., $\omega\varepsilon_r\varepsilon_0 \ll \sigma_s$. With the above two assumptions, we consider the conductor is good, and the loss of the cavity is small. In other words, the decay rate $\alpha \ll \omega$, and thus $\omega \approx \tilde{\omega} = k_0 c$, implying that

$$\delta \approx \sqrt{\frac{2}{k_0 c \mu_0 \mu_r \sigma_s}}. \tag{6}$$

The finite element method (FEM) is a suitable method for arbitrary geometrical scales. In order to apply the FEM we use the weak form of Eq (1), see [3],

$$\begin{aligned} &\textit{Find } k_0 \in \mathbb{C} \textit{ and } \mathbf{H} \in V,\ \mathbf{H} \neq \mathbf{0}, \textit{ such that for all} \\ &\mathbf{f} \in V \textit{ and all } q \in W \\ &\int_\Omega \left[\frac{1}{\varepsilon_r} \nabla \times \mathbf{H} \cdot \nabla \times \mathbf{f} - k_0^2 \mu_r \mathbf{H} \cdot \mathbf{f} \right] d\mathbf{x} \\ &\qquad + i k_0 \frac{1}{Z_0} \int_\Gamma (\mathbf{n} \times \mathbf{E}) \cdot \mathbf{f}\, ds = 0, \\ &\int_\Omega \mu_r\, \mathbf{H} \cdot \nabla q\, d\mathbf{x} = 0. \end{aligned} \tag{7}$$

Here, V denotes the functions in $H(\mathbf{curl}; \Omega)$ that satisfy the SIBC boundary conditions and $W = H_0^1(\Omega)$ [1]. $Z_0 = \sqrt{\mu_0/\varepsilon_0}$ is the characteristic impedance of free space.

Plugging in the SIBC (2), and using (3) and (6), we get

$$\begin{aligned} & ik_0 \frac{1}{Z_0} \int_\Gamma (\mathbf{n} \times \mathbf{E}) \cdot \mathbf{f}\, ds \\ & = ik_0 \frac{Z_s}{Z_0} \int_\Gamma (\mathbf{n} \times \mathbf{H}) \cdot (\mathbf{n} \times \mathbf{f})\, ds \\ & = (i-1) k_0^{\frac{3}{2}} \sqrt{\frac{c\mu_0\mu_r}{2\sigma_s {Z_0}^2}} \int_\Gamma (\mathbf{n} \times \mathbf{H}) \cdot (\mathbf{n} \times \mathbf{f})\, ds. \end{aligned}$$

With this, the weak form (7) becomes

$$\begin{aligned} & \textit{Find } k_0 \in \mathbb{C} \textit{ and } \mathbf{H} \in V,\ \mathbf{H} \neq \mathbf{0}, \textit{ such that for all} \\ & \mathbf{f} \in V \textit{ and all } q \in W \\ & \int_\Omega \left[\frac{1}{\varepsilon_r} \nabla \times \mathbf{H} \cdot \nabla \times \mathbf{f} - k_0^2 \mu_r \mathbf{H} \cdot \mathbf{f} \right] d\mathbf{x} \\ & \quad + (i-1) k_0^{\frac{3}{2}} \sqrt{\frac{c\mu_0\mu_r}{2\sigma_s {Z_0}^2}} \int_\Gamma (\mathbf{n} \times \mathbf{H}) \cdot (\mathbf{n} \times \mathbf{f})\, ds = 0, \\ & \int_\Omega \mu_r\, \mathbf{H} \cdot \nabla q\, d\mathbf{x} = 0. \end{aligned} \tag{8}$$

We discretize problem (8) with the finite element Ritz-Galerkin method [1] employing appropriate finite element subspaces of V and W. To that end we triangulate Ω by tetrahedra. The magnetic vector functions in V are then approximated by Nédélec edge elements, while the scalar functions in W are approximated by Lagrange nodal finite elements [1]. This approach avoids the generation of spurious eigensolutions, and imposing the boundary conditions is straightforward [1].

Let the vector functions $\mathbf{N}_i$, $1 \le i \le n$, be the Nédélec basis functions, while the scalar functions N_ℓ, $1 \le \ell \le m$, denote the Lagrange basis functions. Eventually, we obtain a constrained complex nonlinear eigenvalue problem

$$T(\lambda)\mathbf{x} = A\mathbf{x} + \lambda^{\frac{3}{2}} R\mathbf{x} - \lambda^2 M\mathbf{x} = \mathbf{0}, \tag{9a}$$

$$C^T \mathbf{x} = \mathbf{0}. \tag{9b}$$

Here $\lambda (= k_0)$ is the eigenvalue and $\mathbf{x}$ is the eigenvector. The matrices A, R, M, and C in (9) have the entries

$$\begin{aligned} a_{ij} &= \int_\Omega \varepsilon_r^{-1} (\nabla \times \mathbf{N}_i) \cdot (\nabla \times \mathbf{N}_j)\, d\mathbf{x}, \\ m_{ij} &= \int_\Omega \mu_r \mathbf{N}_i \cdot \mathbf{N}_j\, d\mathbf{x}, \\ r_{ij} &= (i-1) \int_\Gamma \sqrt{\frac{c\mu_0\mu_r}{2\sigma_s {Z_0}^2}} (\mathbf{n} \times \mathbf{N}_i) \cdot (\mathbf{n} \times \mathbf{N}_j)\, d\mathbf{x} \\ c_{i\ell} &= \int_\Omega \mu_r \mathbf{N}_i(\mathbf{x}) \cdot \nabla N_\ell(\mathbf{x})\, d\mathbf{x}, \\ & \qquad 1 \le i, j \le n,\ \ 1 \le \ell \le m. \end{aligned}$$

We solve the nonlinear eigenproblem (9a) with the nonlinear Jacobi–Davidson (NLJD) method. The algorithm is taken from Betcke and Voss [4,5]. To impose the divergence-free condition (9b), we construct an appropriate projector to assert that each vector in the NLJD search space is in the null space of C^T, see [6,7].

After having solved (9) the angular frequency ω and the decay rate α are derived from λ. The magnetic field $\mathbf{H}$ is obtained from the calculated eigenvector, the electric field $\mathbf{E}$ by differentiation.

The system's total stored energy U in Ω, and the average power loss P_s in the surface conductor Γ, are computed as [2]

$$\begin{aligned} U &= \frac{\mu_0\mu_r}{2} \int_\Omega |\mathbf{H}|^2\, d\mathbf{x}, \\ P_s &= \frac{R_s}{2} \int_\Gamma |H_t|^2\, d\mathbf{x}. \end{aligned} \tag{10}$$

Here, H_t is the tangential component of the magnetic field on Γ [2]. The systems quality factor can then be defined by

$$Q = \frac{\omega U}{P_s}. \tag{11}$$

NUMERICAL EXPERIMENTS

All simulations have been performed on the Cray XT6 at the Swiss National Supercomputing Centre (CSCS) [8].

Pillbox Cavity

Before simulating the transverse deflecting cavity, we validated the correctness and reliability of our method by means of the elementary pillbox cavity. Let radius and length of the pillbox be $r = 0.05$ m and $h = 0.1$ m, respectively. The mesh contains 306'337 tetrahedra. By using quadratic Nédélec elements, the finite element discretization counts 1'986'080 degrees of freedom (dof). We compute the fundamental TM010 mode with varying surface conductivity σ_s. σ_s is large enough that the two assumptions for a good conductor are satisfied. The results we obtained are listed in Table 1.

Table 1: Numerical analysis for TM010 mode of the pillbox cavity. $f = \frac{\omega}{2\pi}$ is the resonance frequency; δ is the skin depth (5); Q is the quality factor (11).

σ_s (S/m)	f (GHz)	δ (mm)	Q
10^2	2.238100	1.1	30
10^3	2.277208	0.33	99
10^4	2.289538	0.11	316
10^5	2.293434	0.033	1001
$5.8 \cdot 10^7$	2.295160	0.0014	24160
10^{10}	2.295229	0.00011	317255

If σ_s decreases, the field can penetrate into the surface conductor with larger skin depth δ. We understand an increasing δ is equivalent to increasing the volume of the cavity Ω. Therefore, the resonant wavelength increases and the frequency f decreases. Our numerical results clearly show this trend in δ and f.

We also calculated the case of perfect electric conductor (PEC) boundary conditions. Then, $\sigma_s = \infty$

and $\mathbf{n} \cdot \mathbf{E} = 0$ on Γ. The computed resonance is $f_{\text{pec}} = 2.295234\,\text{GHz}$ [9]. f_{pec} is slightly larger than $2.295229\,\text{GHz}$, where $\sigma_s = 10^{10}\,\text{S/m}$. We observe that if σ_s is $5.8 \cdot 10^7\,\text{S/m}$ (the conductivity of copper), a further increase of σ_s generates almost negligible shift in f, i.e., less than 1 MHz. Therefore, copper is a material well suited for the cavity wall.

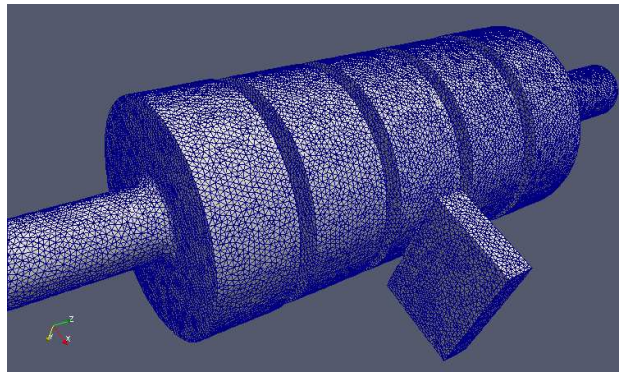
The quality factor decreases significantly as σ_s decreases. If σ_s reduces by a factor of 100, then δ_s, R_s and thus P_s increase by a factor of about 10. Therefore, the quality factor reduces by approximately a factor 10.

Transverse Deflecting Cavity

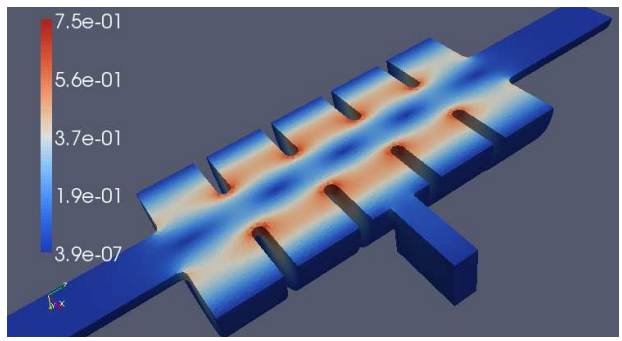
We simulated the 5-cells transverse deflecting cavity. Each cell is a pillbox, and the middle cell is coupled to a rectangular feeding waveguide, see Fig. 1(a). The detailed geometrical parameters of this cavity are given in [10].

We computed the operating TM110 mode with the surface of the conductor being copper, i.e., $\sigma_s = 5.8 \cdot 10^7\,\text{S/m}$. In order to maintain accuracy, we use a mesh containing 872'261 tetrahedra, see Fig. 1(a). By employing quadratic Nédélec elements, the finite element discretization counts 5'726'536 degrees of freedom. 512 cores are used on the Cray XT6.

We obtained the frequency $f = 2.995488\,\text{GHz}$ and the quality factor $Q = 15504$. They match well with the design parameters [10]. The skin depth is $\delta \approx 1.2\,\mu\text{m}$, much smaller than the radius of the cells (58.28 mm). If we replace the SIBC with PEC boundary conditions, the resonance increases slightly from $f = 2.995488\,\text{GHz}$ to $f_{\text{pec}} = 2.995930\,\text{GHz}$. The electrical field distribution $|\mathbf{E}|$ is plotted in Fig. 1(b).



(a)



(b)

Figure 1: (a) The geometry and the associated tetrahedral mesh of the 5-cells transverse deflecting cavity. (b) The electric field distribution of the TM110 mode (in half of the domain).

ACKNOWLEDGMENT

It is our pleasure to thank Antonio Falone and Gian-Luca Orlandi (PSI) for sharing thoughts on the transverse deflecting cavity. The work of the first author (H. Guo) was supported in part by grant no. 200021-117978 of the Swiss National Science Foundation.

REFERENCES

[1] J. Jin. *The Finite Element Method in Electromagnetics*. John Wiley, New York, NY, 2nd edition, 2002.

[2] S. Ramo, J. R. Whinnery, and T. Van Duzer. *Fields and Waves in Communication Electronics*. John Wiley, New York, NY, 2nd edition, 1994.

[3] J. Wang and N. Ida. Eigenvalue analysis in electromagnetic cavities using divergence free finite elements. *IEEE Trans. Magn.*, 27(5):3978–3981, 1991.

[4] T. Betcke and H. Voss. A Jacobi–Davidson-type projection method for nonlinear eigenvalue problems. *Future Gener. Comput. Syst.*, 20(3):363–372, 2004.

[5] H. Voss. Iterative projection methods for large-scale nonlinear eigenvalue problems. *Comput. Techn. Rev.*, 1:187–214, 2010.

[6] P. Arbenz, M. Bečka, R. Geus, U. Hetmaniuk, and T. Mengotti. On a parallel multilevel preconditioned Maxwell eigensolver. *Parallel Comput.*, 32(2):157–165, 2006.

[7] R. Geus. *The Jacobi–Davidson algorithm for solving large sparse symmetric eigenvalue problems*. PhD thesis no. 14734, ETH Zurich, 2002.

[8] Homepage of the Swiss National Supercomputing Centre (CSCS): `http://www.cscs.ch/`.

[9] H. Guo, A. Adelmann, A. Falone, C. Kraus, B. Oswald, and P. Arbenz. Computation of electromagnetic modes in the transverse deflecting cavity. In *Proceedings of the First International Particle Accelerator Conference*, pages 1847–1849, May 2010.

[10] M. Pedrozzi. SwissFEL injector conceptual design report. Technical report, Paul Scherrer Institute, 2010.

MODEL-BASED ANALYSIS OF DIGITAL SIGNAL PROCESSING BLOCKS IN A BEAM PHASE CONTROL SYSTEM*

C. Spies†, H. Klingbeil, M. Glesner, TU Darmstadt, Darmstadt, Germany

Abstract

A beam phase control system comprises digital phase detectors and band pass filters to detect coherent longitudinal dipole and higher order bunch oscillations. These digital signal processing functions can be implemented in several ways, e. g. in software or on a programmable logic device. In this paper, we consider different possible implementations and compare them in terms of their real-time performance and their system resource consumption. For the phase detectors, a software implementation is compared against different hardware implementations. For the band pass filters, different possible architectures are considered.

BEAM PHASE CONTROL

Longitudinal oscillations in synchrotrons may occur due to an initial energy mismatch or may be excited by discontinuities in the input ramps, by wake fields, etc. These oscillations can be characterized by two mode numbers m and n [1]. In order to eliminate undesired dipole oscillations ($m = 1$, $n = 0$), a beam phase control (BPC) system has been proposed for the SIS100 synchrotron under construction at GSI [1].

The inputs to the BPC system are the signal of a beam position monitor, which is proportional to the beam current, and a low-level signal proportional to the RF voltage across the gap of a reference cavity. An analog preprocessing stage [2] that includes low-pass filters reduce the wideband beam signal to a narrow-band signal at the harmonic frequency. Two RF phase detectors are used to determine the phase of each of those signals with respect to a common, arbitrary reference signal. The difference between both phases is the phase of the bunch center of gravity with respect to the RF voltage. A frequency-variable comb filter [3] is then used to extract the component at the characteristic synchrotron frequency from the phase difference; this component corresponds to longitudinal dipole oscillations. In the future, the magnitude of the beam signal will also be detected and a filter tuned to twice the synchrotron frequency will be used to identify quadrupole oscillations [4].

A system-level overview is shown in Fig. 1, and a block diagram of the digital signal processing blocks involved is shown in Fig. 2. In the following, we will focus on the phase detector and band-pass filter blocks.

* Work supported by the German Ministry of Education and Research, Grant No. 06DA9028I

† christopher.spies@mes.tu-darmstadt.de

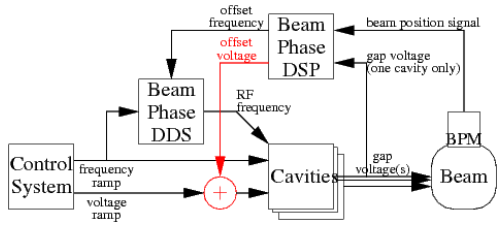


Figure 1: System level block diagram.

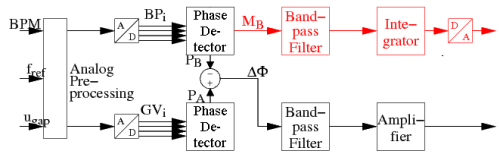


Figure 2: Signal processing block diagram.

PHASE DETECTORS

The analog preprocessing mentioned previously mixes the (variable-frequency) input signals to a fixed intermediate frequency (IF) and makes sure that the phase detector receives exactly 4 samples per IF period [2]. Assuming the phase φ of the input signals with respect to the reference is constant (or changes very little) during one IF period, 4 subsequent samples $s_{1,2,3,4}$ correspond to $+\hat{x}\sin\varphi$, $+\hat{x}\cos\varphi$, $-\hat{x}\sin\varphi$, $-\hat{x}\cos\varphi$, where $\hat{x}$ is the amplitude of the input signal. An inphase/quadrature (IQ) phase detector can be used to compute

$$i = \frac{s_1 - s_3}{2}, \quad (1)$$

$$q = \frac{s_2 - s_4}{2}, \quad (2)$$

$$\varphi = \arctan\frac{q}{i}. \quad (3)$$

The arctan function is not available on many simple digital signal processors (DSPs). Different possible approximations exist. In order to fairly compare them to each other, we subjected each implementation to the same inputs. The desired result is shown in Fig. 3 and corresponds to the expected beam phase in the SIS100 synchrotron during an $^{238}U^{28+}$ acceleration cycle. The individual approximations investigated are described in the remainder of this section. Table 1 compares them in terms of their maximum absolute error, mean absolute error and standard deviation vs. the expected result (computed using MATLAB), and Table 2 compares the hardware implementations in terms of their resource consumption.

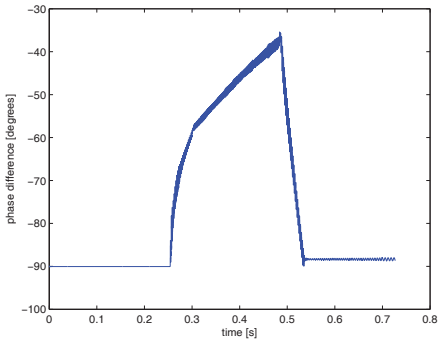


Figure 3: Phase detector test input.

Table 1: Error of Different Phase Detectors

Implementation	Max.	Mean	Std. Dev.
DSP software	0.280°	0.094°	0.081°
1024 × 12 bit LUT	0.0654°	0.0302°	0.0150°
16-stage CORDIC	0.0274°	0.0051°	0.0038°

Software Implementation

In early versions of the phase detector [2], the following approximation formula was used:

$$\arctan x \approx \frac{x}{1 + 0.28 \cdot x^2} \tag{4}$$

for $|x| \leq 1$. The DSP (a TI TMS320C6713 [2] at 225 MHz) requires 230 cycles (about 1 μs) to compute the phase difference between both inputs. A maximum sampling rate of 3.91 $\frac{\text{MS}}{\text{s}}$ could be sustained.

Hardware Implementation

Running the filter on the same DSP as the phase detector would not have been possible due to its limited computing power. It was therefore decided to offload the phase detector into a field-programmable gate array (FPGA). All hardware implementations use 16 bit integer arithmetic.

Look-Up-Table The initial implementation [5] used a look-up table (LUT). The quotient $\frac{q}{i}$ is computed using a pipelined divider and used as an index into a LUT of precomputed values of arctan. This architecture can sustain a maximum sampling rate of about 150 $\frac{\text{MS}}{\text{s}}$ [5].

Coordinate Rotation Digital Computer The LUT-based implementation was replaced by the well-known CORDIC algorithm [6] in order to improve accuracy and reduce resource consumption [7]. This architecture can sustain a maximum sampling rate of about 169 $\frac{\text{MS}}{\text{s}}$. The CORDIC algorithm has the added advantage of simultaneously computing the magnitude of each input signal. This enables the detection of longitudinal quadrupole oscillations of the beam [4].

Table 2: FPGA Footprint of Different Phase Detectors

Implementation	Slices	Flipflops	LUTs
1024 × 12 bit LUT	878	910	1,026
16 bit, 16-stage CORDIC	478	788	792

BAND PASS FILTER

The computed phase difference is fed into a band-pass filter in order to detect oscillations at the synchrotron frequency (dipole oscillations) or integer harmonics thereof (quadrupole, sextupole, ... oscillations). The synchrotron frequency can vary from a few Hz to a few kHz, i. e. it is highly variable but much smaller than the sampling frequency. This necessitates a very large sample memory (16,384 samples) and it was decided to use a sparse coefficient vector (at most 64 nonzero taps) [3, 8]. The output y is computed from the current input x_0 and past inputs $x_1, \ldots, x_{N-1}$ as

$$y_k = a \cdot y_{k-1} + g \cdot \sum_{j=1}^{64} b_j \cdot x_{d_j}, \tag{5}$$

where a is a feedback factor selecting between FIR ($a = 0$) and IIR ($a \neq 0$) operation, g is the filter gain, b_j is the j-th nonzero coefficient and d_j is the position of the j-th nonzero coefficient in the coefficient vector.

Typically, only 3 nonzero coefficients are used ($b_0 = 1$, $b_1 = -2$, $b_2 = 1$). They are equally spaced ($d_0 = 0$, $d_1 = d$, $d_2 = 2 \cdot d$) and the tap distance is

$$d = (2 \cdot f_C \cdot T_S)^{-1}, \tag{6}$$

where f_C is the desired center frequency and T_S is the sampling interval.

An iterative implementation of the filter consumes 3,874 FPGA slices and can sustain a data rate of 1.56 $\frac{\text{MS}}{\text{s}}$ [8]. Pipelined implementations can sustain higher data rates (up to 100 $\frac{\text{MS}}{\text{s}}$) at the expense of FPGA resources. Note that the phase detectors consume 4 successive input samples to compute one phase value, so the input data rate of the filter is only $\frac{1}{4}$ of the phase detector's input data rate.

Table 3 (taken from [8]) compares different possible filter implementations in terms of their FPGA footprint.

CONCLUSIONS AND FUTURE WORK

A frequency offset is computed from the filter output and added to the RF frequency in order to dampen any longitudinal oscillations. This frequency offset is sent from the controller to the synthesizer generating the RF frequency over an optical direct link [9]. The maximum data rate of this link is 40 $\frac{\text{Mbit}}{\text{s}}$ and the packet length is 55 bit. Each packet contains one frequency value. The maximum data rate at the filter output is therefore limited to 728 $\frac{\text{kS}}{\text{s}}$; otherwise, the optical link would be overloaded. This, in turn,

Table 3: FPGA Footprint of Different Filter Architectures [8]

Implementation	Slices	RAM Blocks	Multipliers
64-stage pipeline	4,826	16	56
32-stage pipeline	4,770	16	32
16-stage pipeline	4,760	16	16
8-stage pipeline	4,769	16	8
4-stage pipeline	4,528	16	4
iterative	3,874	16	1

limits the maximum data rate of the phase detectors to $2.91\,\frac{\mathrm{MS}}{\mathrm{s}}$. Any of the investigated implementations is capable of this data rate; however, the used DSP cannot run the phase detectors and the filter at the same time. Therefore, the limiting factor is the communication link, and contrary to previous publications [8], an improvement of the throughput of the signal processing hardware is uncalled for.

The actual data processing takes little time (well below $2\,\mu$s with any combination of implementations) compared to the total delay including data transmission (about $10\,\mu$s [1]), so the computational latency is also not an issue.

All investigated implementations are sufficiently accurate; the maximum error is well below the accuracy achievable in practice (about 1° due to noise). Again, contrary to previous publications [10, 11], an improvement of the accuracy is uncalled for.

We conclude that future work should concentrate on minimizing the hardware resource consumption in order to be able to offload more functionality onto the existing device.

ACKNOWLEDGMENTS

The authors would like to thank Surapong Pongyupinpanich for his help in preparing this paper.

REFERENCES

[1] H. Klingbeil, B. Zipfel, M. Kumm, P. Moritz, "A Digital Beam-Phase Control System for Heavy-Ion Synchrotrons," IEEE Trans. Nucl. Sci. 54(6), 2007, pp. 2604–2610.

[2] H. Klingbeil, "A Fast DSP-Based Phase-Detector for Closed-Loop RF Control in Synchrotrons," IEEE Trans. Instrum. Meas. 54(3), 2005, pp. 1209–1213.

[3] P. Surapong, C. Spies, M. Glesner, H. Klingbeil, "Design of frequency-variable digital filters for beam phase control," GSI Scientific Report, 2011, p. 340.

[4] D. Lens et al., "Damping of Longitudinal Modes in Heavy-Ion Synchrotrons by RF-Feedback," *Proc. Intl. Conf. Control Applications*, 2010, pp. 1737–1742.

[5] A. Guntoro et al., "Implementation of Realtime and High-speed Phase Detector on FPGA," in K. Bertels, J. M. P. Cardoso, S. Vassiliadis (Eds.), "*Reconfigurable Computing*", Springer, 2006, pp. 1–11.

[6] P. K. Meher et al., "50 Years of CORDIC — Algorithms, Architectures, and Applications," IEEE Trans. Circ. Syst. 56(9), 2009, pp. 1893–1907.

[7] A. Guntoro, M. Glesner, "Resolving longitudinal amplitude and phase information of two continuous data streams for high-speed and real-time processing," Advances in Radio Science 7, 2009, pp. 133–137.

[8] P. Surapong, M. Glesner, H. Klingbeil, "Implementation of Realtime Pipeline-Folding 64-Tap Filters on FPGA," *Proc. Conf. Ph. D. Research in Microelectronics and Electronics*, 2010, pp. 1–4.

[9] H. Klingbeil, U. Laier, K.-P. Ningel, S. Schäfer, C. Thielmann, B. Zipfel, "New digital low-level rf system for heavy-ion synchrotrons," Phys. Rev. ST AB 14(10), 2011, pp. 102802/1–9.

[10] P. Surapong, M. Glesner, "Pipelined Floating-Point Architecture for a Phase and Magnitude Detector based on CORDIC," *Proc. Intl. Conf. Field Programmable Logic and Applications*, 2011, pp. 382–384.

[11] F. A. Samman, P. Surapong, C. Spies, M. Glesner, "Floating-point-based hardware accelerator of a beam phase-magnitude detector and filter for a beam phase control system in a heavy-ion synchrotron application," *Proc. Intl. Conf. Accelerators and Large Experimental Physics Control Systems*, 2011, pp. 683–686.

DESIGN OF SRF CAVITIES WITH CELL PROFILES BASED ON BEZIER SPLINES*

Bernard Riemann†, Thomas Weis, TU Dortmund University, Dortmund, Germany
Axel Neumann, Helmholtz-Zentrum Berlin, Berlin, Germany

Abstract

Elliptical cavities have been a standard in SRF linac technology for 30 years. In this work, we present a novel approach [1] using Bezier spline profile curves. By using different degrees of spline curves, the number of free parameters can be varied to suit a given problem (endcell tuning, basecell figures of merit), thus leading to a high flexibility of the spline approach. As a realistic example, a cubic spline SRF multicell cavity geometry is calculated and the figures of merit are optimized for the operational mode. We also present an outline for HOM endcell optimization that can be realized using available 2D solvers.

INTRODUCTION

In modern ring-based and linac facilities, high acceleration gradients and duty cycles are achieved using superconducting radio frequency (SRF) cavities. To protect these cavities from quenching and RF breakdown, a sufficiently small ratio of peak surface electric field and accelerator gradient $E_{\mathrm{surf}}/E_{\mathrm{acc}}$ is required. Therefore the shapes must not contain edges, and thus a smooth profile curve for the figure of rotation is needed.

As had been done in normal-conducting cavities, edges of cylindrical/iris-loaded cavities were rounded off in a first step. To further enhance performance, the rounding circles were substituted by ellipses [2] (Fig. 1), leading to modern SRF "re-entrant", "low-loss" and "TESLA-shape" [3] cavities, in which resonant multipacting of secondary electrons is highly suppressed in contrast to cylindrical shapes. While elliptical shapes have shown superior performance, this parameterization is used by historical contingency, as to the authors' knowledge, no other parameterizations of similar smoothness (modulated sinusoid curves, splines) have been researched for ultra-relativistic particles.

Spline Geometry Parameters

The spline cavity geometry is such an alternative shape with additional desirable properties like continuous curvature, and a number of free parameters that can be chosen lower than for elliptical shapes (but may be arbitrarily increased).

* this work is supported by BMBF contract 05K10PEA
† bernard.riemann@tu-dortmund.de

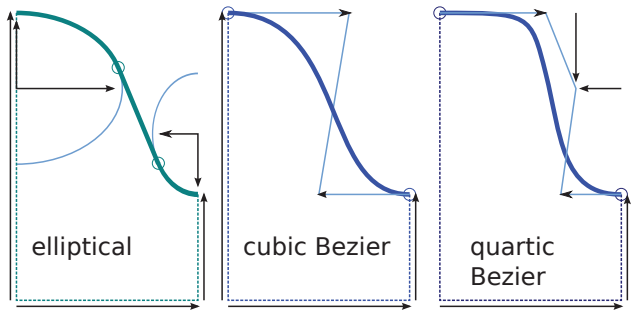


Figure 1: Sketch and cavity parameters (arrows) for different cavity shape parameterizations.

A Bezier spline is a parameterized curve [4]

$$\vec{s}(t) = \vec{a}_0 + \vec{a}_1 t + \vec{a}_2 t^2 + ... = \sum_{n=0}^{N} \vec{s}_n b_{n,N}(t) \quad (1)$$

with Bernstein polynomials $b_{n,N}$ and $t \in [0,1]$. The $N+1$ points $\vec{s}_n$ define the so-called control polygon and contain the free parameters of the spline. Let us consider the simplest possible case of a cubic ($N = 3$) Bezier spline as shown in Fig. 1. The 8 geometry parameters are reduced by a) zero slope of the curve at iris and equator radius of the cavity, and b) fixed relation of the coordinate system to the equator plane. 5 cavity parameters remain, compared to 7 for an elliptical cavity.

Applying two further elementary RF problem constraints c) operation frequency and d) transit-time factor (cell length is $\beta\lambda/2$), only three optimization parameters remain for the specific application. Instead of specifying these constraints, the low dimensionality of the remaining problem allows to scan the remaining parameter space, thus creating a "map" for the cubic Bezier cavity that can be used to speed up complex design optimization tasks.

IMPLEMENTATION

For the single cell computations with periodic boundary conditions, the 2d code SUPERFISH [5] was employed. Since splines are not implemented in SUPERFISH, a small wrapper for MATLAB was developed. The spline geometry was discretized by calculating a 200 point (halfcell) polygon from its parameters (Fig. 2), which was then used as input in SUPERFISH, also specifying a special localized mesh that matches the polygon discretization accuracy.

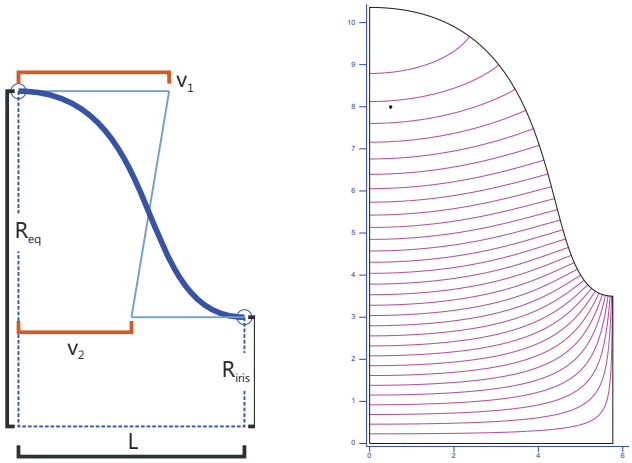


Figure 2: Parameterization for cubic Bezier shape (left) and exemplary result for electric field lines ($v_1 = 0.9, v_2 = 0.65, R_{\mathrm{iris}} = 35\,\mathrm{mm}, R_{\mathrm{eq}} = 103.60\,\mathrm{mm}$, right).

Frequency Tuning and Parameter Scan

Since the spline parameterization is not known to SUPERFISH, the R_{eq} tuning procedure had to be reimplemented in MATLAB. The procedure almost exactly imitates the SUPERFISH procedure [5] by reading out $D(k^2)$ (from OUTFIS.TXT) after each SUPERFISH run. After fitting a polynomial $D(R_{\mathrm{eq}})$ to the data of previous iteration values, the outer equator radius is tuned to the frequency (π-mode) $\nu_\pi = 1\,3\,\mathrm{GHz}$.

For a fixed iris radius, the $(v_1, v_2)/L$ parameter space (Fig. 2) was scanned in an arbitrary range $v_i/L \in [0.05, 0.95]$ that was partly chosen due to meshing considerations (no "re-entrant"-type shapes in this region). The sequence proceeds in a spiral pattern, starting from the central point $(0.5, 0.5)$ in 0.05 steps, so that the changes of the tuned equator radii and their slopes between the steps were small and fast convergence (1-3 iterations) could be achieved. Knowing the matching equator radii for the π-mode boundary condition, a second scan for the 0-mode was performed to obtain the corresponding eigenmode frequency ν_0 needed for the intercell coupling constant (see Results). The procedure for one fixed iris radius took approx. 3.5 hours on a standard PC.

RESULTS

Single Cavity Map

The results of the parameter scan for four different figures of merit [6] with two different iris radii are shown in Figs. 3 and 5. These figures are the shunt impedance per quality factor R/Q (wall-loss independent shunt impedance parameter), the surface electric field to effective acceleration gradient factor $E_{\mathrm{surf}}/E_{\mathrm{acc}}$ up to 5.0, the Geometry factor G (inversely related to cryogenic losses) and the intercell coupling constant $\kappa = 2(\nu_\pi - \nu_0)(\nu_\pi + \nu_0)$.

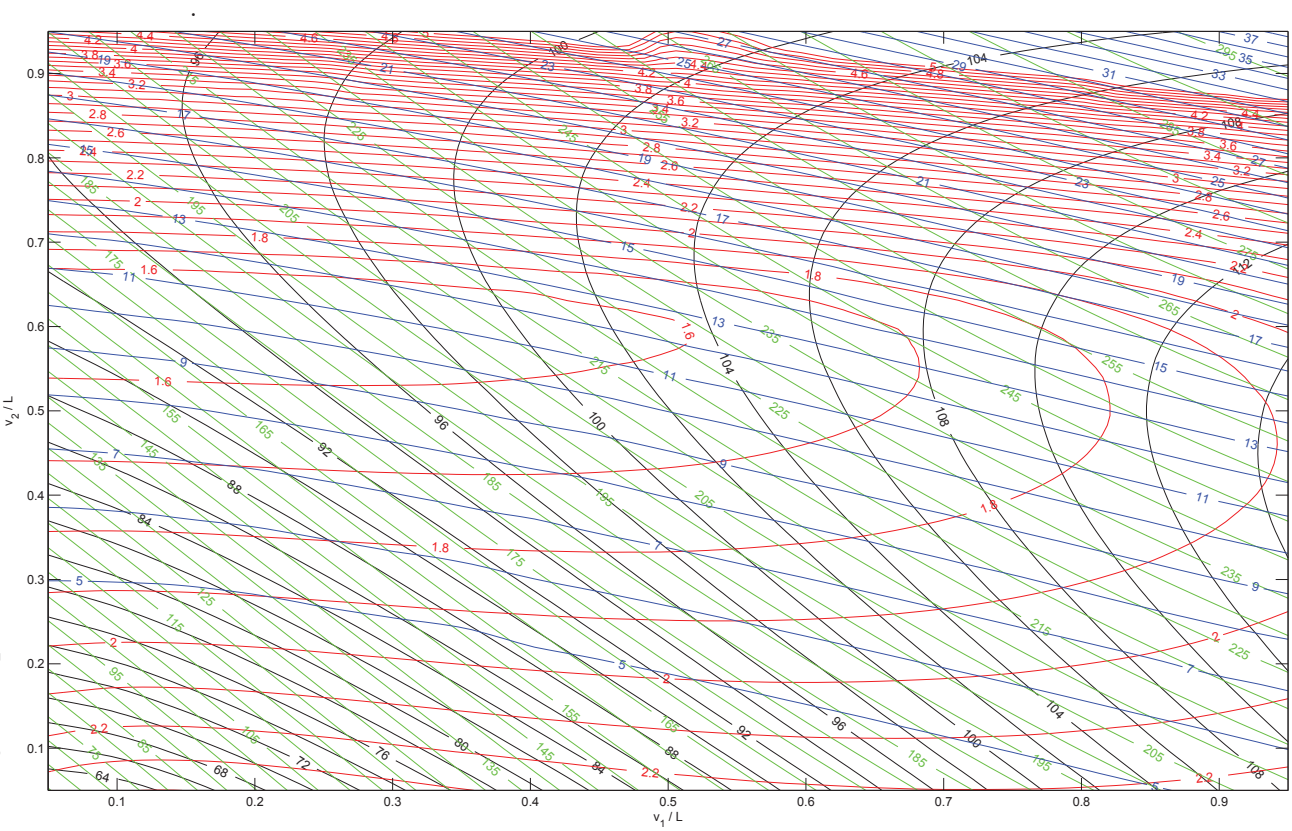


Figure 3: Cubic Bezier map for a 1.3 GHz cavity with iris radius 35 mm. The map shows the cubic interpolated isolines of R/Q [Ω] in black, $E_{\mathrm{surf}}/E_{\mathrm{acc}}$ in red, the Geometry factor [Ω] in green, and the intercell coupling constant $[10^{-3}]$ in blue.

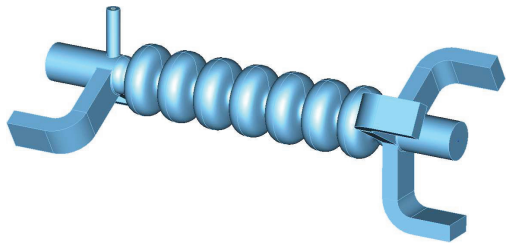


Figure 4: Rendering of the HOM-damped seven-cell Spline cavity.

It can be seen by comparison e.g. with [3] that the cubic spline shape geometry performance is in roughly the same range as the standard elliptical cavity geometries, although as stated, the cubic shape has two less free parameters.

By inspecting the maps, the trade-off between different figures of merit in the design process becomes visible (see e.g. red and black lines). The differences in $E_{\mathrm{surf}}/E_{\mathrm{acc}}$ (e.g. the 1.6 plateau) and intercell coupling strength for different radii are also of interest.

Multicell Structure

Within the scope of BERLinPro main linac research [7, 8, 9], a heavily HOM-damped 7 cell 1.3 GHz cubic Bezier spline design with an iris radius of 36mm was computed and optimized for the operational passband. Due to the fact that the BERLinPro linac uses waveguide absorbers for HOM damping (no rotational symmetry), CST MWS [10] was used to calculate the structure's eigenmodes.

For the cavity with radius-tuned endcells to reach field homogenity, $R/Q = 768\,\Omega$ could be achieved. The frequency difference of the operational mode to the nearest mode of the passband is 1.3 MHz. With these values, the Spline multicell structure has almost equal performance for the operational mode as the elliptical BERLinPro cavity [8].

OUTLOOK

It was shown that cubic Bezier spline cavity geometries yield roughly comparable figures of merit to elliptical cavities and can be used for rapid cavity prototyping using the map approach.

The initial parameters can then be used as starting points for optimization of higher order Bezier splines, due to the possibility of increasing the order of a given spline without changing its actual shape (degree elevation). This property may allow for further optimization of central cavity cells. By iterative degree elevation and a subspace optimization scheme, it may be possible to approximate an idealized free-form cavity for a given goal function.

Furthermore, degree elevation is of special interest for HOM tuning of endcells, where a large number of eigenmode fields need to be optimized to allow sufficient HOM damping, e.g. for ERLs. While asymmetric endcells are often used in elliptical cavities (each HOM mode couples out only on one side), the flexible number of spline parameters could circumvent this problem, allowing stronger damping.

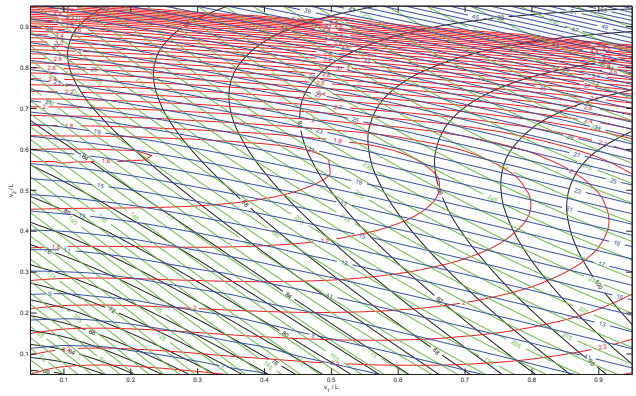


Figure 5: Cubic Bezier map for a 1.3 GHz cavity with iris radius 40 mm. The figures may be scaled for design of 1.5 GHz cavities with $\approx$ 35 mm radius. Isoline legend is the same as for Fig. 3.

REFERENCES

[1] B. Riemann et al., "SRF multicell cavity design using cubic and higher order spline cavity profiles", T 80.9, Verhandlungen DPG Göttingen 2012

[2] P. Kneisel, R. Vincon and J. Halbritter. "First results on elliptically shaped cavities", *Nuclear Instrum. Methods*, 1981, 188, 669 - 670

[3] Aune et al., "Superconducting TESLA cavities", *Phys. Rev. ST Accel. Beams* Vol. 3 pp. 092001 (2000)

[4] Paul De Casteljau. "Courbes à pôles", National Industrial Property Institute (France), 1959

[5] K. Halbach and R.F. Holsinger. "SUPERFISH - a computer program for evaluation of RF cavities with cylindrical symmetry", *Part. Acc.*, 7:213–222, 1976.

[6] T. Wangler, *RF Linear Accelerators*, 2nd ed., Wiley-VCH, 2007

[7] A. Jankowiak et al., "BERLinPro - a compact demonstrator ERL for high current and low emittance beams", Proc. LINAC 10, Tsukuba, Japan, 2010.

[8] A. Neumann et al., "Status of the HOM calculations for the BERLinPro main linac cavity", FRAAC3, Proc. ICAP'12, Rostock, Germany, 2012.

[9] T. Galek, "Traveling Poles Elimination Scheme and Calculations of External Quality Factors of HOMs in SC Cavities", WEP07, Proc. ICAP'12, Rostock, Germany, 2012.

[10] CST Microwave Studio (CST Studio Suite), CST AG, 2012.

TOOLS FOR ANALYSIS AND IMPROVEMENT OF LINAC OPTICS DESIGN FOR HIGH BRIGHTNESS ELECTRON BEAMS

S. Di Mitri#, M. Cornacchia, C. Scafuri, Sincrotrone Trieste S.C.p.A., Basovizza, 34149, Italy

Abstract

The optics design of single pass high brightness electron linacs usually aims at the preservation of the transverse emittance. Collective effects mainly impose constraints to the optics design such as at the low-beta interaction points in colliders and magnetic compressors in free electron lasers (FELs). Other constraints are from the trajectory correction scheme, performance of diagnostics, collimation systems and physical space limitations. Strong focusing is typically prescribed for all the aforementioned cases, although it may hamper the main goal of emittance preservation through the excitation of optical aberrations. Strong focusing also potentially leads, through focusing errors, to large beam optics mismatch. Based on these sometimes conflicting requirements, we have developed tools for the analysis and improvement of electron linac optics. They are based on the Elegant code [1] and allow the user to identify: i) local sources of phase space distortions and emittance dilution, ii) lattice areas particularly sensitive to focusing errors, iii) poor trajectory steering. The analysis does not require massive particle tracking since it deals with the single particle motion in the normalized phase space.

ELEGANT ON-LINE

In order to use Elegant as an on-line machine model [2], we have developed a set of utilities for interfacing the simulator with the accelerator. We have exploited the Elegant capability to read and write a set of element parameters from a file in SDDS format via the load_parameters and write_parameters commands. The SDDS file is filled in with the actual parameters of the running accelerator by means of a dedicated utility which examines the SDDS files and maps the requested parameters to the appropriate control system variables by means of database tables. The requested variables are then acquired from the control system and scaled if necessary, e.g. converting beam postion monitors (BPM) readings from millimeter to meter). The reverse path is followed for setting parameters: Elegant writes the new values to the SDDS files, another dedicated utility reads the new parameters from the SDDS file, maps them to control variables and set the values via standard control system calls. Since Elegant works with normalized machine physics quantities, we have developed a set of specialized control system servers, called Tango devices [3], which perform the conversion from engineering quantities (e.g. current) to machine physics quantities (e.g. quadrupole strength) by means of calibration tables. Such tables are directly handled by the Tango server.

#simone.dimitri@elettra.trieste.it

OPTICS MATCHING

The goal of optics matching is to impose the design values of the Twiss functions to the electron beam. This is typically done with at least four quadrupole magnets. For beam energies lower than ~100 MeV, the beam optics cannot be predicted with sufficient accuracy in Elegant since the particles move in the space-charge dominated regime. For this reason, it is very important to measure the beam optics at the *end* of the injector, where the electron spatial distribution is frozen to any practical purpose. The matching loop is illustrated in Fig. 1.

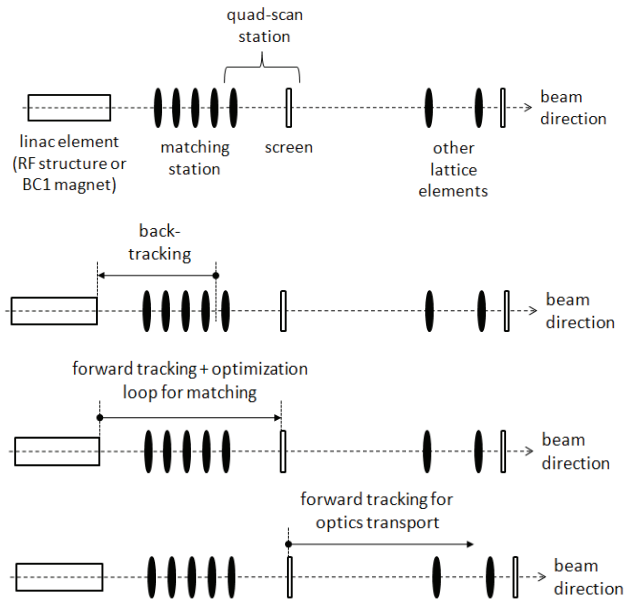


Figure 1: Illustration of the optics matching loop [2]. From top to bottom: i) the beam Twiss parameters are measured with the quadrupole scan technique [4] at the entrance of the last quadrupole magnet of the matching station; ii) the present machine configuration is read by Elegant and the measured Twiss parameters are back-tracked to a point upstream of the matching station; iii) starting from the present machine configuration, Elegant starts optimizing the quadrupole strengths to match the beam Twiss parameters to the design values; iv) once the matching has been performed, the beam is transported through the downstream lattice.

The matching loop has been coded in MATLAB [5] and a Graphical User Interface (GUI) is available as a standard control room application. A theoretical betatron mismatch parameter is defined as follows [6]:

$$\xi = \frac{1}{2}\left(\bar{\beta}\gamma - 2\bar{\alpha}\alpha + \bar{\gamma}\beta\right) \quad (1)$$

where $\overline{\beta},\overline{\alpha},\overline{\gamma}$ are the design Twiss parameters, for each transverse plane. β,α,γ are the Twiss parameters computed by Elegant at the end of the optimization process. By definition, $\xi \geq 1$; the closer this value is to 1, the closer the Elegant solution is to the design optics. A graphical output is displayed that shows the betatron functions as they are computed in the back-tracking mode and in the forward tracking mode. All intermediate data and results of the matching procedure are exchanged via SDDS format files and can be plotted with standard SDDS based tools.

OPTICS SENSITIVITY

The capability of the perturbed magnetic focusing to generate optics mismatch is here investigated. The focusing error may be due to a large energy, a mean energy mismatch or a gradient calibration error. We initially refer to [7] and assume that: i) the radiofrequency focusing can be neglected compared to the magnetic focusing; ii) the beam optics mismatch at the end of the line is dominated by the focusing errors along the lattice, *not* by the optics mismatch at the injection point. Unlike [7], however, we assume an identical relative focusing error $k\delta$ for all the quadrupoles, so that the final mismatch parameter is computed in each plane as follows:

$$m_\delta^f \cong 1+\frac{1}{2}\left[\left(\sum_{i=1}^{N} k_i\beta_i L_i\delta_i \cos(2\Delta\mu_i)\right)^2+\left(\sum_{i=1}^{N} k_i\beta_i L_i\delta_i \sin(2\Delta\mu_i)\right)^2\right] \quad (2)$$

The sum in Eq.2 is over N quadrupoles, k is the nominal quadrupole strength, L is the quadrupole magnetic length, δ is the fractional strength error and $\Delta\mu$ is the betatron phase advance. We now consider Eq.2 for one local source of mismatch at the time (one quadrupole), while all the other magnets have the nominal strength ($\delta = 0$). We thus obtain the optics mismatch induced by each individual quadrupole magnet:

$$m_\delta^q - 1 \cong \frac{1}{2}(k\beta L\delta)_q^2 \equiv \varsigma_q \quad (3)$$

In the case of perturbations that may lead to (small) emittance dilution, it can be shown through the beam matrix formalism that Eq.3 also describes the relative emittance growth induced by the error kick. We define ς_q as the optics sensitivity to focusing error. If we assume that the emittance dilution or mismatch parameter at the end of the line $m_\delta^f = 1 + C$, is the result of many identical, uncorrelated and small perturbations $\varsigma_q = \varsigma << 1$, the maximum sensitivity allowed to each quadrupole (tolerance) is of the order of $m_\delta^q - 1 = \varsigma_q \approx C/\sqrt{N}$. For typical numbers $C = 5\%$ and $N = 100$, the lattice regions characterized by $\varsigma_q \geq 0.5\%$ (see Fig. 2) should be reviewed either by imposing tighter tolerances on the quadrupole setting or by adopting a weaker focusing.

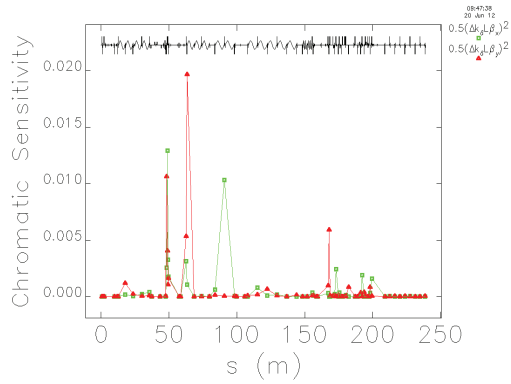


Figure 2: Optics sensitivity to the energy spread in the FERMI FEL [8].

CHROMATICITY

The chromatic properties of the design optics are evaluated by varying the beam mean energy by δ (*i.e.*, +/-1.0%) at the injection point, in small steps. The Twiss function, *f(s)*, is computed at each step along the entire lattice (see Fig. 3). A linear fit provides the coefficient $df/d\delta$, at any location along the lattice. The amount of local mismatch can be estimated as $df(s)/d\delta$ times δ, this being the energy deviation relative to the initial energy.

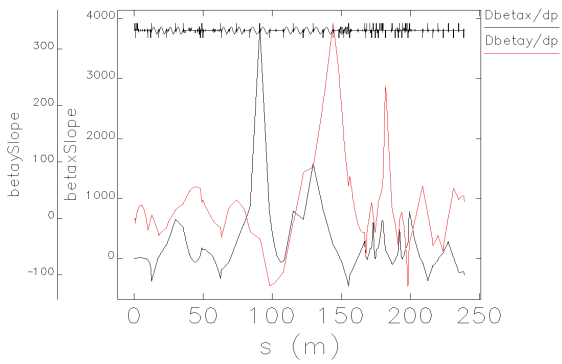


Figure 3: $d\beta/d\delta$ for the FERMI baseline optics.

COURANT-SNYDER INVARIANT

To evaluate the impact of chromatic and geometric aberrations on the distortion of the transverse particle distribution, we mapped the Courant-Snyder (C-S) invariant of the bunch centroid along the lattice. To do this, transport matrices up to the second order in the particle coordinates are used in Elegant. To move the centroid on a nonzero amplitude trajectory we excite a betatron oscillation through an initial angular kick (typically as large as 1 mrad or so) and compute the invariant for the design optics (see Figure 4). We experienced that tracking only 10^3 particles or so is enough to obtain reliable indications of the phase space distortion.

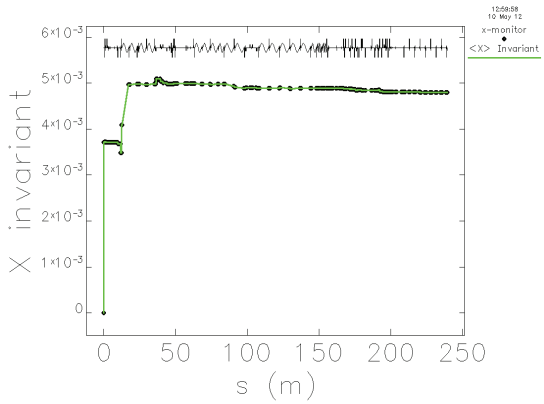


Figure 4: The horizontal C-S invariant of the bunch centroid along the FERMI FEL.

H-FUNCTION

The design H-function, $H = \beta\eta'^2 + 2\alpha\eta\eta' + \gamma\eta^2$ plays a role for the collective effects. On the one hand, it should be made as small as possible to minimize the emittance growth induced by the emission of coherent synchrotron radiation (CSR) [9]:

$$\gamma\varepsilon = \gamma\varepsilon_0\sqrt{1+\frac{H}{\varepsilon_0}\sigma_\delta^2} \qquad (4)$$

A computational approach to the dependence of the CSR instability on the Twiss functions can be found in [10]. However, Eq.4 is considerably cheaper in terms of CPU time and the *H*-function can be used as a constraint in the Elegant optimization loop. On the other hand, a large *H* function determines a large path length difference of particles with different energies:

$$\sqrt{\langle \Delta l^2 \rangle} = \sqrt{H\varepsilon}\,, \qquad (5)$$

where the integral is computed over a generic dispersive path and the final RMS path length is averaged over the beam particle ensemble (see Fig.5).

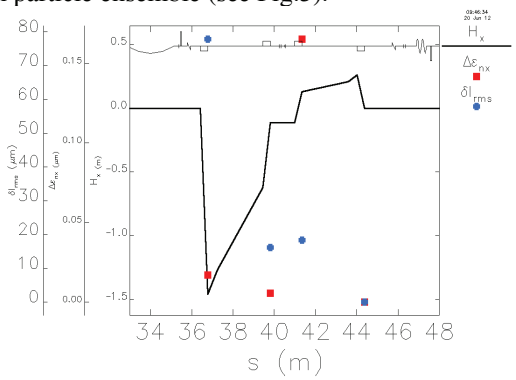


Figure 5: The design *H*-function in the first magnetic compressor of FERMI, together with the emittance growth (Eq. 4) and the RMS phase mixing (Eq. 5) for each dipole magnet.

The spread of the path length translates into the particles longitudinal phase mixing that washes out energy and density modulations such as those due to the microbunching instability [11]. Because of the conflicting requirements on the design of *H*, an optics design should include from the very beginning a proper tuning of the Twiss functions in the dispersive regions traversed by a short, high charge beam or characterized by a high gain of the microbunching instability. This tuning would allow the manipulation of the *H* function and eventually reach a compromise between the aforementioned effects.

FLOQUET SPACE

The amount and the phase advance separation of steering magnets and BPMs determine the beam trajectory sampling and control. At least 4 BPMs per betatron period ensure an accurate reconstruction of the beam trajectory. The trajectory control is then very efficient if the steering magnets are close to local maxima of the betatron function and if each consecutive steerer and BPM are separated by π/2 phase advance. To analyze the efficiency of the beam steering we plot in Fig.6 the bunch centroid position normalized to the local √β versus the betatron phase advance.

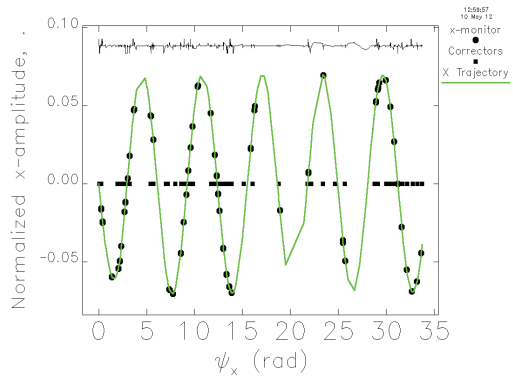


Figure 6: The horizontal bunch centroid trajectory in the Floquet coordinates along the FERMI FEL. BPMs (on the curve) and steering magnets (on the central straight line) are superimposed to the trajectory.

REFERENCES

[1] M. Borland, APS LS-287 (2000).
[2] S. Di Mitri et al., Phys. Rev. ST – Accel. Beams **15**, 012802 (2012).
[3] A. Gotz et al., MP705, ICALEPCS 2003, Gyeongiu, Korea, (2003). http://www.tango-controls.org/
[4] M. G. Minty and F. Zimmermann, SLAC-R-621 (2003).
[5] http://www.mathworks.com/products/matlab/
[6] M. Sands, SLAC-AP-85 (1991).
[7] V. Balandin and N. Golubeva, in Proc. of the 23rd Part. Accel. Conf., Vancouver, Canada, 2009, TH6PFP031.
[8] C. Bocchetta et al., FERMI@Elettra FEL Conceptual Design Report No. ST/F-TN-07/12, 2007
[9] H. Braun et al., Phys. Rev. Lett. **84**, 658 (2000).
[10] F. Stulle et al., Phys. Rev. ST Accel. Beams **10**, 031001 (2007).
[11] E. Saldin et al., Nucl. Instrum. Methods Phys. Res., Sect. A **490**, 1 (2002).

ANALYTICAL PRESENTATION OF SPACE CHARGE FORCES*

S. Andrianov†, SPbSU, St. Petersburg, Russia

Abstract

This paper presents an analytical description of space charge forces generated by charged particle beams. Suggested approach is based on some set of models for particle distribution function. All necessary calculations have analytical and closed form for different models for beam density distributions. These model distributions can be used for approximation of real beam distributions. The corresponding solutions are included in general scheme of beam dynamics presentation based on matrix formalism for Lie algebraic tools. Computer software is based on corresponding symbolic codes and some parallel technologies. In particular, as computational tools we consider GPU graphic card NVIDIA. As an example, we consider a problem of beam dynamics modelling for microprobe focusing systems.

INTRODUCTION

Most of space charge modelling methods are based on numerical methods and corresponding programming technologies. In particular, matrix processors became very popular in the last time for numerical evaluations (i. e. using the Particle-in-Cell method). But for long time beam evolution problems these methods demonstrate rather low effectiveness. The nature of the corresponding beam dynamics needs effective algorithms for space charge field reconstruction during the beam evolution along the reference orbit (with coordinate s). For this purpose in this paper we consider the following two methods: a beam with elliptical form in transverse four-dimensional phase space (*unbunched beam model*) and a beam presented as six-dimensional ellipses (*bunched beam model*). For the both models we use Ferrer's integrals method [1]. This method allows to derive corresponding formulae for beam generated field in symbolic forms. Similar presentation is compatible with presentation of the corresponding beam dynamics using Lie algebraic tools formalism [2]. In this paper we demonstrate evaluation method of electrical field generated by different distribution functions of beam particles. The analytical presentation of corresponding solutions is matched with matrix formalism for Lie algebraic tools [3]. Furthermore this approach can be realized for solution of Maxwell–Vlasov equations [4].

* Work supported by Federal Targeted Program "Scientific and Scientific-Pedagogical Personnel of the Innovative Russia in 2009-2013" (Governmental Contract no. p 793)

† sandrianov@yandex.ru

DISTRIBUTION FUNCTION IN PHASE SPACE

In this section we demonstrate basic idea of our approach using the first models – the unbunched beam model in four-dimensional space with following coordinates $\{x, p_x, y, p_y\}$. The corresponding distribution function $f(x, p_x, y, p_y, s)$ can be presented as a function of $\varkappa^2 = \mathbf{X}^{\mathrm{T}}\mathbb{A}\mathbf{X}$, where $\mathbf{X} = (x, y, p_x, p_y)^{\mathrm{T}}$ and $\mathbb{A}$ is a symmetric nonsingular matrix. Let us consider some popular functions to present space charge distribution.

1) *Uniform distribution*:

$$\varphi(\varkappa^2) = \frac{2\sqrt{\det \mathbb{A}}}{\pi^2}\Theta(1-\varkappa^2),\ \Theta(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

2) *Vladimirov–Kapchinsky* (*microcanonical*) *distribution*:

$$\varphi(\varkappa^2) = \frac{\sqrt{\det \mathbb{A}}}{\pi^2}\delta(1-\varkappa^2).$$

3) *Normal* (*Gauss*) *distribution*:

$$\varphi(\varkappa^2) = \frac{\sqrt{\det \mathbb{A}}}{4\pi^2}\exp\left(-\frac{\varkappa^2}{2}\right).$$

Of course, three indicated types of distributions don't exhaust all variety of admissible distributions family. But these distributions are interesting from point of view of configuration distributions of charge density in the configuration space (in the next we omit the variable s for reduction).

$$\rho(x, y) = \int\limits_{R^2} f(x, p_x, y, p_y)\, dp_x\, dp_y.$$

Indeed, practically, it is very difficult to measure the function $f(\mathbf{X})$, but the distribution $\rho(x, y)$ can be obtained with measurement accuracy. More over, only function $\rho(x, y)$ defines the electrical potential (or electrical field) which is necessary for beam dynamics evaluation. Let present the matrix $\mathbb{A}$ in a block matrix form

$$\mathbb{A} = \begin{pmatrix} \mathbb{A}^{11} & \mathbb{A}^{12} \\ \mathbb{A}^{21} & \mathbb{A}^{22}, \end{pmatrix},\quad \mathbb{A}^{kk} = \left(\mathbb{A}^{kk}\right)^{\mathrm{T}},\ \mathbb{A}^{21} = \left(\mathbb{A}^{12}\right)^{\mathrm{T}},$$

then we can write

$$\varkappa^2 = \mathbf{X}^{\mathrm{T}}\mathbb{A}\mathbf{X} = \mathbf{X}_1^{\mathrm{T}}\mathbb{A}^{11}\mathbf{X}_1 + \mathbf{X}_1^{\mathrm{T}}\mathbb{A}^{12}\mathbf{X}_2 + \\ + \mathbf{X}_2^{\mathrm{T}}\mathbb{A}^{21}\mathbf{X}_1 + \mathbf{X}_2^{\mathrm{T}}\mathbb{A}^{22}\mathbf{X}_2,$$

where $\mathbf{X}_1 = (x, y)^{\mathrm{T}}$, $\mathbf{X}_2 = (p_x, p_y)^{\mathrm{T}}$. After some evaluations one can obtain the following equalities.

1) *Uniform distribution*:

$$\rho(x, y) = \frac{2qN_0}{\pi}\sqrt{\frac{\det \mathbb{A}}{\det \mathbb{A}^{22}}}(1-\varkappa_r^2)\Theta(1-\varkappa_r^2),$$

where $\varkappa_r^2 = \mathbf{X}_1^{\mathrm{T}}\mathbb{C}\mathbf{X}_1$, $\mathbb{C} = \mathbb{A}^{11} - \mathbb{A}^{21}\left(\mathbb{A}^{22}\right)^{-1}\mathbb{A}^{12}$.
2) KV-*distribution*:

$$\rho(x,y) = \frac{qN_0}{\pi}\sqrt{\frac{\det\mathbb{A}}{\det\mathbb{A}^{22}}}\delta(1-\varkappa_r^2).$$

3) *Gauss-distribution*:

$$\rho(x,y) = \frac{qN_0}{2\pi}\sqrt{\frac{\det\mathbb{A}}{\det\mathbb{A}^{22}}}\exp\left(-\frac{\varkappa_r^2}{2}\right).$$

Where q – a particle charge, N_0 – a number of particle in transverse section.

For our long beam model we have $I(s,t) = v_0\int_{R^2}\rho(x,y,s,t)dxdy = \mathrm{const}$, where v_0 – a longitudinal particle velocity. For definition of the matrices $\mathbb{A}^{ik}$ and $\mathbb{C}$ we impose some natural conditions: 1) $v_0\int_{R^2}\rho_i(x,y)\,dx\,dy = I = \mathrm{const}$, $\forall\, i = \overline{1,3}$; 2) $\rho_i(0,0) = \rho_0$. As reference values we choose corresponding values for KV-distribution (the case of an "ideal beam")

$$\rho_0 = \frac{qN_0}{\pi}\sqrt{\frac{\det\mathbb{A}}{\mathbb{A}^{22}}},\quad I = v_0qN_0\sqrt{\frac{\det\mathbb{A}}{\det\mathbb{A}^{22}}\det\mathbb{C}}\ .$$

Excepting N_0 one evaluates $\rho_0 = I\sqrt{\det\mathbb{C}}/\pi v_0$.

The matrix $\mathbb{C}_i$ describes the ellipse for the i-th case, we can suppose $\mathbb{C}_i = \alpha_i^2\mathbb{C}$, that is we consider as an ellipse. Then 2), 3) can be rewritten in the form

$$\rho_1(x,y) = \rho_0(1-\alpha_1^2\varkappa^2)\Theta(1-\alpha_1^2\varkappa^2),$$
$$\rho_2(x,y) = \rho_0\Theta(1-\alpha_2^2\varkappa^2),\quad \rho_3(x,y) = \rho_0\exp(-\alpha_3^2\varkappa^2),$$

where $\varkappa^2 = \mathbf{X}_1^{\mathrm{T}}\mathbb{C}\mathbf{X}_1$, $\mathbb{C}$ – a matrix describing an ellipse in configuration space for the ideal beam. The second condition for definition of our free parameters we get from the equality $\rho_1 = \rho_2$. After some simple evaluations we get $\alpha_1 = 2/3$ and $\alpha_2 = 1$. It means that function ρ_1 leads to the same value as homogeneous distribution ρ_2, if it be defined for ellipse $X_1^{\mathrm{T}}\mathbb{C}X_1 \le 9/4$. Analogical evaluations can be performed for any distributions. We should note that function $\rho_3(x,y)$ has not a compact support. For similar distributions we should define the probability for ingressing to the interior of ellipse $\mathbf{X}_1^{\mathrm{T}}\mathbb{C}\mathbf{X}_1 \le \lambda^2$. Here λ^2 defines a "beam boundary" with some probability. For example, for probability $P = 0.9$ we obtain $\alpha_3 = 9/\left(5\sqrt{\pi}\,\mathrm{erf}(1)\right)$ and $\lambda = 1/\alpha_3$.

Besides the mentioned distributions we point the following two distribution: quadratic – $\rho_4(x,y) = \rho_0(1-\alpha_4^4\varkappa^4)\Theta(1-\alpha_4^2\varkappa^4)$ and cosine-like – $\rho_5(x,y) = \rho_0\cos^2(\alpha_5^2\varkappa^2)\Theta(1-\varkappa^4/2)$. Like in previous cases we can evaluate α_4 and α_5 (this coefficient evaluates via known Fresnel function). We should note that described types of modelling functions (both in phase and in configuration spaces) demonstrate the suggested approach, which can be used practically for any distribution function. In particular, one can use instead of the function $\rho(x,y)$ a polynomial of quadratic form variable $\varkappa^2$ with some polynomial order agreed with known nonlinear order for control electromagnetical fields. As an additional step we can suggest the method of Padé approximation, which makes possible to select corresponding coefficients with desired accuracy.

SELF-FIELD OF SPACE CHARGE

For many practical problems of beam physics we can suggest that phase space distribution function $f(\mathbf{X})$ and configuration distribution function ρ are functions of transverse coordinates only. In other words $f = f(\mathbf{X})$ and $\rho = \rho(x,y)$ for some small interval of variable s. In this case Poisson equation for electrical potential Ψ (inside the beam ellipsoid), can be written in form

$$\frac{\partial^2\Psi}{\partial x^2} + \frac{\partial^2\Psi}{\partial y^2} = -\frac{1}{\varepsilon_0}\rho(x,y). \qquad (1)$$

According to the first section we assume $\rho(x,y) = \rho_0\varphi(\varkappa^2)$, where φ – certain function (e. g. one of pointed five functions) of argument $\varkappa^2 = \mathbf{X}_1^{\mathrm{T}}\mathbb{C}\mathbf{X}_1$, $\mathbf{X}_1 = (x,y)^{\mathrm{T}}$. For small enough interval Δs the resulting essential variation of ellipse axes can be negligible. So, in coordinates (ξ,η) the ellipse will be canonical: $\varkappa^2 = \xi^2/a^2 + \eta^2/b^2$. We need to use elliptic coordinates $\xi = h\cosh\alpha\cos\beta$, $\eta = h\sinh\alpha\sin\beta$ to solve (1). Poisson equation transforms to the following form

$$\frac{\xi^2}{\alpha^2} + \frac{\eta^2}{\beta^2} = -\frac{1}{\varepsilon_0}\rho h^2\left(\cosh^2\alpha - \cos^2\beta\right),$$

with solution in coordinates ξ, η:

$$\Psi(\xi,\eta) = -\frac{1}{4\varepsilon_0}\rho\left(\xi^2+\eta^2 - \frac{a-b}{a+b}\left(\xi^2-\eta^2\right)\right) + \Psi_0,$$

or finally in coordinates x, y:

$$\Psi(x,y) = -\frac{1}{4\varepsilon_0}\rho\left(x^2+y^2 - \right.$$
$$\left. - \frac{a-b}{a+b}\left(x^2\cos 2\gamma - y^2\cos 2\gamma + 2xy\sin 2\gamma\right)\right) + \Psi_0,$$

where $\gamma = \arctan\left(2c_{12}/(c_{22}-c_{11})\right)$ – the rotation angle for coordinates transformation with rotation matrix $\mathbb{T}$.

Let consider the known approach based on Ferrer's integral [1]. According to this approach we obtain for inner potential of elliptic cylinder

$$\Psi^{(i)} = -\frac{\pi ab}{eps_0}\int_0^\infty \Phi\left(\varkappa^2(u)\right)\frac{du}{\Delta(u)},$$

where $\Phi(\varkappa^2) = \int_0^{\varkappa^2}\rho(t)dt$, $\Delta^2(u) = (a^2+u)(b^2+u)$, $\varkappa^2(u) = \xi^2/(a^2+u) + \eta^2/(b^2+u)$. One can obtain electrical field generated by space charge

$$E_\xi^{(i)}(\xi,\eta) = \frac{2\pi ab}{\varepsilon_0}\int_0^\infty \frac{\xi}{a^2+u}\rho\left(\varkappa^2(u)\right)\frac{du}{\Delta(u)},$$

$$E_\eta^{(i)}(\xi,\eta) = \frac{2\pi ab}{\varepsilon_0}\int_0^\infty \frac{\eta}{b^2+u}\rho\left(\varkappa^2(u)\right)\frac{du}{\Delta(u)}.$$

On the next step we evaluate components of the electrical field E_ξ, E_η for our charge distribution functions. Using inverse transformation $\mathbb{T}^{-1}$ we can obtain E_x, E_y (electrical field components in initial coordinate system). After some evaluations one can determine (using Euler substitution $t = \Delta(u)/(a^2+u)$, $u = b^2 - a^2t^2/(t^2-1)$, $t(0) = b/a$, $t(\infty) = 1$):

$$E_\xi = \frac{4\pi\rho_0}{\varepsilon_0}\frac{ab}{a(a+b)}\xi, \quad E_\eta = \frac{4\pi\rho_0}{\varepsilon_0}\frac{ab}{b(a+b)}\eta.$$

We will denote it as E^0_ξ and E^0_η. For the linear distribution in $\varkappa^2$ we obtain for the components of electrical field correspondingly

$$E_{\xi,\eta} = E^0_{\xi,\eta} + \Delta E_{\xi,\eta}, \tag{2}$$

$$\Delta E_\xi = -\frac{4\pi\rho_0}{\varepsilon_0}\frac{a_1b_1}{a_1^2+b_1^2}\xi\left(\frac{2a_1+b_1}{3a_1^3}\xi^2 + \frac{1}{a_1b_1}\eta^2\right),$$
$$\Delta E_\eta = -\frac{4\pi\rho_0}{\varepsilon_0}\frac{a_1b_1}{a_1^2+b_1^2}\eta\left(\frac{1}{a_1b_1}\xi^2 + \frac{2b_1+a_1}{3b_1^3}\eta^2\right).$$

The described approach can be used for all of our distributions. We obtain formulae similar to (2) with corresponding expressions for $\Delta E_{\xi,\eta}$.

After rotation for Gauss distribution $\rho_3(x,y) = \rho_0\exp\left(-\varkappa^2\right)$ we have $\varkappa^2 = \xi^2/a^2 + \eta^2/b^2 \le 1$. Using the Euler substitution and integration on ξ, η we can obtain

$$E_{\xi,\eta} = E^0_{\xi,\eta} + \Delta E_{\xi,\eta},$$

$$\Delta E_\xi = -\frac{4\pi\rho_0}{\varepsilon_0}ab\xi\left[\exp\left(\frac{\xi^2-\eta^2}{b^2-a^2}\right) - 1\right]\times$$
$$\times\sum_{k=1}^{\infty}\left[\chi^1_k\left(\frac{\eta}{\xi}\right) - \left(\frac{b}{a}\right)^{2k+1}\chi^1_k\left(\frac{\eta a^2}{\xi b^2}\right)\right],$$
$$\Delta E_\eta = -\frac{4\pi\rho_0}{\varepsilon_0}ab\xi\left[\exp\left(\frac{\eta^2-\eta^2}{b^2-a^2}\right) - 1\right]\times$$
$$\times\sum_{k=1}^{\infty}\left[\chi^2_k\left(\frac{\eta}{\xi}\right) - \left(\frac{b}{a}\right)^{2k-1}\chi^2_k\left(\frac{\eta a^2}{\xi b^2}\right)\right],$$

where $\chi^1_k(x) = \frac{\xi^{2k}}{k!(b^2-a^2)^{k+1}}\sum\limits_{j=0}^{k}\frac{(-1)^{k-j}C^j_k}{2k+1-4j}x^{2j}$, $\chi^2_k(x) = \frac{\xi^{2k}}{k!(b^2-a^2)^{k+1}}\sum\limits_{j=0}^{k}\frac{(-1)^{k-j}C^j_k}{2k-1-4j}x^{2j}$. For distributions $\rho_4(x,y) = \rho_0(1-4\varkappa^4/9)\Theta(1-4\varkappa^4/9)$ and $\rho_5(x,y) = \rho\cos^2(\varkappa^2\frac{\pi}{4})\Theta(2-\varkappa^2)$ rotation and scaling lead to $\rho_5(\xi,\eta) = \cos^2(\pi\varkappa_5^2/2)\Theta(1-\varkappa_5^2)$, $\varkappa_5^2 = \xi^2/a_5^2+\eta_5^2/b_5^2 \le 1$, $a_5 = \sqrt{2}a$, $b_5 = \sqrt{2}b$ we also obtain the corresponding series.

We should note that any space charge distribution $\rho(x,y)$ with polynomial of n-order according $\varkappa^2$ leads us to polynomial $\Delta E_{\xi,\eta}$ order $2n+1$.

The above exemplified formulae can be essentially simplified for a circular beam ($a = b = R$). For example, for *Cosine-like distribution* we have:

$$\Delta E_\xi = \frac{2\pi\rho_0}{\varepsilon_0}\xi\left(\frac{2R^2}{\pi(\xi^2+\eta^2)}\sin\frac{\pi(\xi^2+\eta^2)}{2R^2} - 1\right),$$
$$\Delta E_\eta = \frac{2\pi\rho_0}{\varepsilon_0}\eta\left(\frac{2R^2}{\pi(\xi^2+\eta^2)}\sin\frac{\pi(\xi^2+\eta^2)}{2R^2} - 1\right).$$

After returning to the initial coordinate system we can obtain components $E_{x,y}$.

CONCLUSION – PARALLEL CALCULATIONS FOR BEAM DYNAMICS

In the previous sections we have considered methods which allow to evaluate a power series expansion for electrical field. The form of these series allows us to use matrix formalism for beam dynamics both in external (control) and self-field generated by beam [5]. The presentation of these fields in the same form allows us to use the method predictor-corrector for beam dynamics evaluation [6] for nonlinear fields also. The matrix presentation practically of all objects allows us to parallelize necessary operations. More over the necessary processors load it is not so essential as compared with parallelization of "usual" approach based on PIC or other similar methods. The computer simulation of some beam transport lines demonstrated necessary effectiveness of the described approach. The computation process divides into several levels of parallelization (parallelization tree) that allows us to constraint effective programs using GPU graphic card TESLA S2050, OpenCL and CUDA technologies [7].

REFERENCES

[1] Ferrers N.M. On the Potentials of Ellipsoids, Ellipsoidal Shells, Elliptic Laminae, and Elliptic Rings of Variable Densities. Quart. J. Pure and Appl. Math. 1877. Vol. 14. N 53. P. 1–22.

[2] Dragt A.J. Lie Methods for Nonlinear Dynamics with Applications to Accelerator Physics. University of Maryland, College Park, www.physics.umd.edu/dsat/

[3] Andrianov S.N. A Matrix Representation of the Lie Algebraic Methods for Design of Nonlinear Beam Lines. in: AIP Conf. Proc. 1997. N 391. NY, P. 355–360.

[4] Andrianov S.N., Dvoeglazov A.I. Matrix Representation for the Vlasov-Maxwell Equation Solution. Proc. of EPAC'2000, Vienna, Austria. 2000. P. 1459–1461.

[5] Andrianov S.N. Dynamical Modeling of Control Systems for Particle Beams. SPbSU, Saint Petersburg, 2004 (in Russian).

[6] Andrianov S.N. High-Order Optics with Space-Charge: Analytical Approach. Proc. of EPAC 1998, Stockholm, Sweden. P. 1091–1093.

[7] Andrianov S.N., Tereshonkov Yu.V. Parallel and Distributed Modeling of Beam Dynamics. Proc. of the 8-th Intern. Conf. High-performance Computation on Cluster Systems. Kazan, Russia. 2008. P. 61-68 (in Russian).

DYNAMICS OF ENERGY LOSS OF A BUNCH INTERSECTING A BOUNDARY BETWEEN VACUUM AND DIELECTRIC IN A WAVEGUIDE*

T. Yu. Alekhina# and A. V. Tyukhtin, St. Petersburg State University, St. Petersburg, Russia

Abstract

We analyze radiation of a small bunch crossing a boundary between two dielectrics in a cylindrical waveguide. The main attention is paid to investigation of dynamics of a charge energy loss and the effect of the boundary on the electromagnetic field (EMF). Algorithms of computations for the field and the energy loss are founded upon certain transformations of integration path. We consider two instances in detail: the bunch is flying from dielectric into vacuum and from vacuum into dielectric. In both cases we compare the energy losses by transition radiation (TR) and by Cherenkov one (CR). Our investigation shows, for example, that energy loss is negative at certain segments of the bunch trajectory.

INTRODUCTION

One of the problems being important for the wakefield acceleration technique and for new methods of generation of microwave radiation consists in analysis of effect of the boundary on the wave field when a bunch flies into a dielectric structure or from one. It should be noticed that energetic characteristics of TR at a single boundary in a waveguide and in the case of a dielectric plate were investigated in papers [1,2]. However, the most attention was paid to study the energetic spectrums of generated modes. Dynamics of an energy loss as well as an EMF structure has not been analyzed.

Our research is based on original approach used previously for the case of the vacuum-plasma boundary [3]. But Cherenkov radiation is not generated in such situation; therefore it varies radically from the case under consideration. The electromagnetic field structure of the point charge was partially investigated in our works [4,5]. Now we analyze dynamics of the energy losses which allows of better understanding physical phenomena in this situation.

We consider a point charge q moving in a metal circular waveguide of radius a along its axis (z-axis) and intersecting the boundary $(z=0)$ between two homogeneous isotropic non-dispersive dielectrics with permittivity ε_1 $(z<0)$ and ε_2 $(z>0)$ at the moment $t=0$. The charge passes uniformly with a velocity $\vec{V}=c\beta\vec{e}_z$ (c is a light speed in vacuum).

The analytical solution of the problem is traditionally found for the spectral harmonics of the vector potentials as an expansion into a series of eigenfunctions of the transversal operator [3,6]. Expressions for components of the EMF can be easily derived from the formulae for the general case of the boundary between two arbitrary homogeneous isotropic media [3].

We investigate the dynamics of the charge energy loss per unit length of the charge motion:

$$W=-qE_z\big|_{\substack{z\to\beta ct\\ r\to 0}}. \qquad (1)$$

As it follows from general expressions for EMF the charge energy losses in both media have two summands: $W_{1,2}=W^q_{1,2}+W^b_{1,2}$. The first one $\left(W^q_{1,2}\right)$ is connected with so-called by V. L. Ginzburg [7] "forced" field that is EMF of the charge in a regular waveguide. It contains CR if $\beta>\beta_{C1,2}$, $\beta_{C1,2}=\varepsilon_{1,2}^{-1/2}$. The second summand $\left(W^b_{1,2}\right)$ is connected with the "free" field that is determined by the influence of the boundary and includes TR. Each summand is decomposition in an infinite series of normal modes:

$$W^{q,b}_{1,2}=\sum_{n=1}^{\infty}W^{q,b}_{n1,2}. \qquad (2)$$

The charge energy loss in a regular waveguide with homogeneous filling analysed in many papers is equal to [8]

$$W^q_{n1,2}=\begin{cases}2q^2\left(a^2\varepsilon_{1,2}J_1^2(\chi_{0n})\right)^{-1}, & \beta>\beta_{C1,2}\\ 0, & \beta<\beta_{C1,2}\end{cases}, \qquad (3)$$

where χ_{0n} is the n^{th} zero of the Bessel function $J_0(x)$. For analysis of the charge energy loss by TR $\left(W^b_{n1,2}\right)$ we use the exact integral representation. We investigate it with analytical and computational methods. Analytical research is an asymptotic investigation with the steepest descent technique. Computations are based on original algorithm using some transformation of the integration path.

We study two cases in detail: the bunch is flying from dielectric $(\varepsilon_1>1)$ into vacuum $(\varepsilon_2=1)$ and from vacuum $(\varepsilon_1=1)$ into dielectric $(\varepsilon_2>1)$.

THE CASE OF FLYING FROM DIELECTRIC INTO VACUUM

Numerical Approach

Efficient algorithm used for numerical calculation is based on a certain transformation of the initial integration path in the complex plane of ω. Earlier such an algorithm was used for calculating of the forced field in different dispersive media [9] and in a waveguide with the boundary between vacuum and cold plasma [3]. We demonstrate this method for the vacuum area.

*Work supported by St.-Petersburg State University

#tanya@niirf.spbu.ru

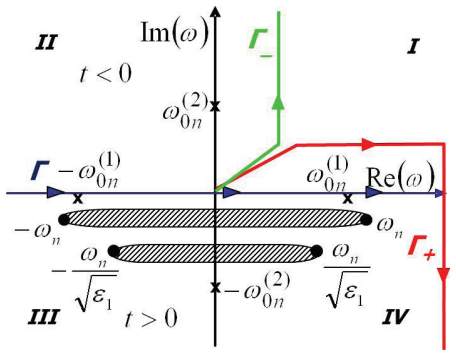


Figure 1: Disposition of the singularities of the integrands, corresponding branch cuts and integration path in a complex plane of ω for W_n^b in vacuum (at $t>0$) if $\beta>\beta_{C1}$.

The first step of the investigation consists of a study of the singularities of integrands. Disposition of these singularities is presented in Fig. 1. Four branch points $\pm\tilde{\omega}_n^{(1)}=\pm\omega_n\varepsilon_1^{-1/2}-i\delta_1$ and $\pm\tilde{\omega}_n^{(2)}=\pm\omega_n-i\delta_2$ are pointed with black circles; two pairs of poles $\pm\omega_{0n}^{(1)}=\pm\beta\omega_n\left(\varepsilon_1\beta^2-1\right)^{-1/2}-i\delta_3$ (situated on the real axis if $\beta>\beta_{C1}$) and $\pm\omega_{0n}^{(2)}=\pm i\beta\omega_n\left(1-\beta^2\right)^{-1/2}$ are shown with crosses. Here $\delta_1,\delta_2,\delta_3$ are positive infinitesimal quantities which tend to zero if we do not take into account absorption in a medium. It is convenient to have the branch cuts as it is shown in Fig. 1. The integrand of W_n^b decreases in the upper half-plane of ω at $t<0$ and in the lower half-plane of ω at $t>0$ for any value of β.

As the integration path Γ goes through the poles $\omega_{0n}^{(1)}$ this leads to rather abrupt behavior of integrands. The numerical algorithm is adapted for overcoming this difficulty. First, solution can be easily written as an integral on a half-infinite contour Γ. Further, we can transform this contour in an upper half plane (ω) into the green contour Γ_- (Fig. 1) for $t<0$ and in a lower half-plane into the red contour Γ_+ for $t>0$. The new contours should bypass all the singularities and then go parallel to the steepest descending path. The integrands have regular behavior along these new contours. Note that, in this way, the charge energy loss can be computed in the domain both near and far from the boundary. We can also optimize the parameters of contours for each computation.

Results and Discussions

Dynamics of the total energy loss per unit length for the 1st mode is presented in Fig. 2. We consider the total energy loss per the length unit for the n^{th} mode in dimensionless unities:

$$\tilde{W}_n=W_n/W_n^q, \tag{4}$$

where W_n^q is the energy loss for Cherenkov mode (3).

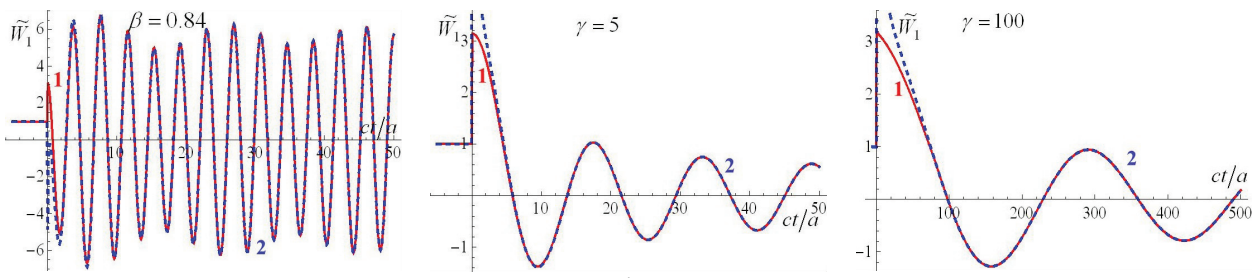
If $\beta>\beta_{C1}$, Cherenkov radiation emerges in dielectric. CR is reflected and refracted at the boundary and, as a result, so-called Cherenkov transition radiation (CTR) is generated [4,5]. In vacuum, CTR exists under conditions $\beta_{C1}<\beta<\beta_{CT1}\equiv(\varepsilon_1-1)^{-1/2}$ in the area $z<z_1=ct\sqrt{\left|1-\beta^2(\varepsilon_1-1)\right|}\Big/\beta$. The dimension of this zone z_1 increases with group velocity of waveguide waves

$$V_{g1}=\sqrt{1-\beta^2(\varepsilon_1-1)}\Big/\beta. \tag{5}$$

The limit speed β_{CT1} is connected with total internal reflection of CR from the boundary. So, if $\varepsilon_1<2$ CTR emerges for ultra-relativistic particles with $\beta\approx1$. If the group velocity (5) is more than the charge velocity, that is

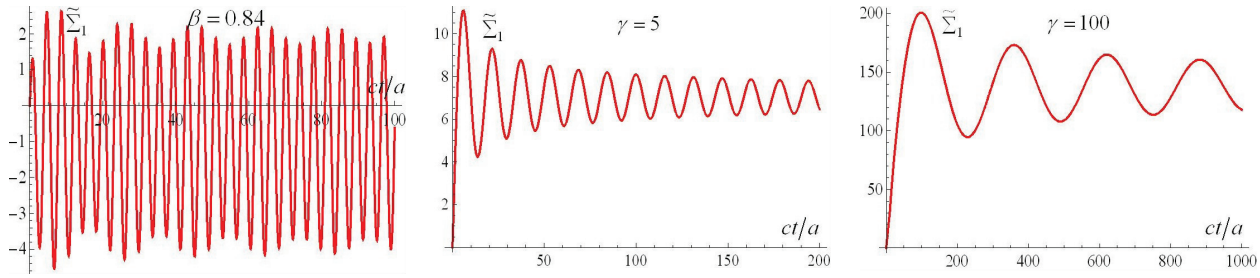
$$\beta<\beta_0=\sqrt{\left(\sqrt{(\varepsilon_1-1)^2+4}-\varepsilon_1+1\right)\Big/2},$$

the charge interacts with CTR (Fig. 2 *a*). One can see that the energy loss oscillates with approximately constant amplitude in the vacuum area. The situation changes if $\beta>\beta_0$ (Fig. 2 *b,c*), when the charge leaves the CTR behind (when the charge interacts with TR only).



(*a*) (*b*) (*c*)

Figure 2: The case of flying from dielectric into vacuum. Dependence of the normalized energy loss for the first mode of the whole field on dimensionless time ct/a for different charge velocities β (or $\gamma=1\Big/\sqrt{1-\beta^2}$). Red solid line 1 corresponds to the exact calculation. The asymptotic approximations are given as well (blue dashed line 2); $\varepsilon_1=1.5$, $\varepsilon_2=1$, $\omega_0=2\pi\cdot10\text{GHz}$, $a=5\text{mm}$, $\beta_{C1}=0.816$, $\beta_0=0.884$.



(*a*) (*b*) (*c*)

Figure 3: The case of flying from dielectric into vacuum. Dependence of the total normalized charge energy loss for the first mode on dimensionless time *ct/a* for different charge velocities; the same in Fig. 2.

In this case, the oscillation amplitude lessens as $1/\sqrt{t}$, and the oscillation period $T_0 = a\gamma(c\chi_{on})^{-1}$ is getting larger with increase in γ.

In Fig.2 one can also see a good coincidence between the exact solution (red solid line 1) and the analytical asymptotic expressions (dashed blue line 2) in some domain outside the boundary.

We investigate the total energy loss by TR in vacuum defined by the integral $\Sigma = \int_0^{Vt} W(z)dz$ as well. Figure 3 presents the total normalized energy loss for the first mode $\widetilde{\Sigma}_1 = \Sigma_1 \, a\varepsilon J_1^2(\chi_{01})(2q^2)^{-1}$. If $\beta < \beta_0$, the total energy loss has a vibrating character (Fig. 3 *a*). If $\beta > \beta_0$, this interaction decreases with t (Fig. 3 *b,c*), and Σ_1 tends to some constant which is proportional to γ.

THE CASE OF FLYING FROM VACUUM INTO DIELECTRIC

Analogous investigation of the charge energy loss can be made for the case of flying from vacuum ($\varepsilon_1 = 1$) into dielectric with ε_2. Here we discuss only physical results obtained. The dynamics of the energy loss per unit length for the first mode of the total field is presented in Fig.4, top. One can see that the energy loss is negative in some area near the boundary (where the bunch is attracted to it). The dimension of this zone decreases with increase in β and in ε_2. The total energy loss for the time interval $(-\infty, t)$ determined by the integral $\Sigma = \int_{-\infty}^{Vt} W(z)dz$ is shown in Fig.4, bottom. One can see that the total energy loss is negative up to the moment $t_0 \approx a\left[\chi_{0n} c\gamma(\varepsilon_2 - 1)\right]^{-1}$. So, this moment t_0 decreases with increase in γ and ε_2. The total energy loss is getting positive only for $t > t_0$ when the Cherenkov loss is getting dominant.

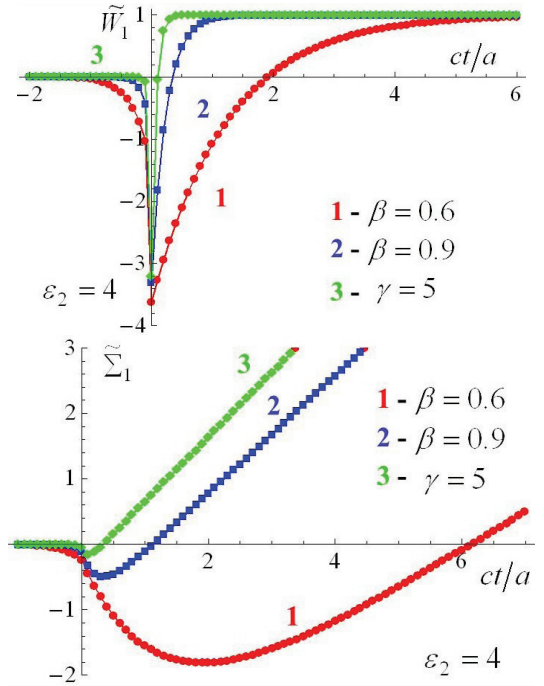


Figure 4: The case of flying from vacuum into dielectric. Dependence of the normalized energy loss per unit length (top) and the normalized total energy loss (bottom) of the first mode of the total field on dimensionless time *ct/a* for different charge velocities; $\omega_0 = 2\pi \cdot 10\text{GHz}$, $a = 5$ mm.

REFERENCES

[1] , Zh. K.A. Barsukov Tekh. Fiz. 30 (1960) 1337.
[2] V.A. Buts, A.N. Lebedev and V.I. Kurilko, *The Theory of Coherent Radiation by Intense Electron Beams*, (Berlin, Springer, 2006).
[3] T.Yu. Alekhina and A.V. Tyukhtin, Phys. Rev. E 83 (2011) 066401.
[4] T.Yu. Alekhina and A.V. Tyukhtin, Journal of Physics: Conference Series 357 (2012) 012010.
[5] T.Yu. Alekhina and A.V. Tyukhtin, IPAC'12, New Orleans, USA, May 2012, WEPPP023 (2012); http://www.JACoW.org.
[6] K.A. Barsukov, Zh. Exps. Teor. Fiz. 37 (1959) 1106.
[7] V.L. Ginzburg and V.N. Tsytovich, *Transition Radiation and Transition Scattering,* (London: Hilger, 1990) p. 445
[8] B.M. Bolotovskii, Sov. Phys. Usp. 4 (1962) 781.
[9] S.N. Galyamin, A.V. Tyukhtin, A. Kanareykin, and P. Schoessow, Phys. Rev. Lett. 103 (2009) 194802.

DESIGN AND APPLICATIONS OF THE BMAD LIBRARY FOR THE SIMULATION OF PARTICLE BEAMS AND X-RAYS*

D. Sagan, Wilson Laboratory, Cornell University, Ithaca, NY, 14853, USA

Abstract

The open-source Bmad software library has been developed for simulating both charged particle beams and X-rays. Owing to its modular, object-oriented design, Bmad has proved to be versatile, and is currently used in a number of programs at Cornell's Laboratory for Elementary-Particle Physics. This paper will discuss the design of the Bmad library and how features such as the ability to simulate overlapping elements, the ability to define the action of control-room "knobs," and the ability to select among different tracking algorithms, have all contributed to a flexible simulation environment that eases the task of both programmers and users alike. Also discussed are the uses that Bmad has been put to, including machine control and the integration of particle beam and X-ray simulations.

INTRODUCTION

Bmad, an open-source software library for simulating both charged particle beams and X-rays, has been in development at Cornell beginning in the 1990s[1]. Originally, the syntax for the lattice files for Bmad was patterned after the syntax of the MAD program[2]. Since, at that time, only a subset of the MAD language was used, the name "Baby MAD" or "Bmad" for short was chosen.

The initial purpose for developing Bmad was modest: Simply to be able to compute Twiss parameters and the closed orbit, and to provide a standard lattice description format. As Bmad evolved, the scope of Bmad expanded so that currently Bmad can simulate such things as spin, X-ray photons, coherent synchrotron radiation, intra-beam scattering, Touschek effect, etc., etc.

Bmad has proved to be versatile and, as a result, is now used in a number of programs at Cornell. Experience with Bmad has shown that there are a number of design features that have made Bmad especially useful. The purpose of this paper is to discuss this and to discuss some of the applications that Bmad has been used for. Finally, plans for the future of Bmad will be presented.

DESIGN PHILOSOPHY

The aim of the Bmad project is to:

- Cut down on the time needed to develop programs.
- Minimize computation times.
- Cut down on programming errors.
- Provide a simple mechanism for lattice function calculations from within programs.
- Provide a flexible and powerful lattice input format.
- Standardize sharing of lattices between programs.

To maximize code reuse, Bmad, written in Fortran, is designed to be object oriented from the ground up. For example, it takes only one line of executable code to parse a lattice file:

```
type (lat_struct) lat
call bmad_parser('lat.bmad', lat)
```

The call to `bmad_parser` in this example causes the file named "lat.bmad" to be parsed and the information to be stored in a variable named `lat` of type `lat_struct` (equivalent to a C++ class). For communication with C++ code, Bmad defines a set of C++ classes and there are interface routines to convey information between between the Fortran structures and the C++ classes.

From the beginning, Bmad development has been driven by the need to solve the practical problems arising from the requirements of machine simulation, lattice design and machine control. As a result, one emphasis of Bmad development has been on minimizing the bookkeeping tasks of both programmer and user. There are a number of features that have proved to be especially useful in this regard, and some are discussed below: The ability to superimpose elements on top of other elements, the ability to slice elements into pieces, the ability to define controller elements that control the attributes of other elements, and the ability to choose the algorithm used for particle tracking.

Superposition of Elements

"Superposition" is the ability to overlap lattice elements spatially. Figure 1 shows an example which is a greatly simplified version of the IR region of Cornell's CESR storage ring when CESR was an e+/e– collider. As shown in Fig. 1A, two quadrupoles named `q1w` and `q1e` are partially inside and partially outside the interaction region solenoid named `cleo`. In the lattice file, the IR region layout is defined to be

```
cesr: line = (... q1e, dft1, ip, dft1, q1w ...)
cleo: solenoid, l = 3.51, superimpose, ref = ip
```

The line named `cesr` ignores the solenoid and just contains the interaction point marker element named `ip` which is surrounded by two drifts named `dft1` which are, in turn, surrounded by the `q1w` and `q1e` quadrupoles. The solenoid is added to the layout on the second line by using superposition. The "ref = ip" indicates that the solenoid is placed relative to `ip`. The default, which is used here, is to place the center of the superimposed `cleo` element at the center of the `ip` reference element. Within a program, the representation of the lattice in the `lat_struct` structure that Bmad creates will contain two lists: One list,

* Work supported by the National Science Foundation under contract PHY-1002467.

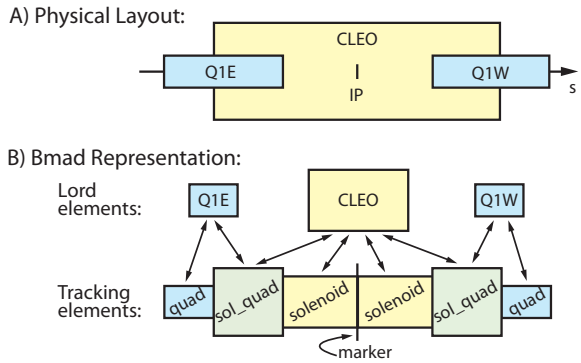


Figure 1: Superposition Example. A) Interaction region layout with quadrupoles overlapping a solenoid. B) The Bmad lattice representation has a list of split elements to track through and the undivided "lord" elements. Pointers (double headed arrows), keep track of the correspondence between the lords and their "slaves."

called the "tracking section," contains the elements that are needed for tracking particles. In the current example, as shown in Fig. 1B, the first IR element in the tracking section is a quadrupole that represents the part of `q1e` outside of the solenoid. The next element is a combination solenoid/quadrupole, called a `sol_quad`, that represents the part of `q1e` inside `cleo`, etc. This `sol_quad` element illustrates a restriction on superpositions that the combination of elements must be able to be represented by a valid Bmad element type. Bmad does have a combination solenoid/quadrupole type so superimposing solenoids with quadrupoles works.

The other list in the `lat_struct` that Bmad creates is called the "lord section." This list will contain the undivided elements which, in this case are `q1e`, `q1w`, and `cleo`. Pointers are created between the lords and their "slaves" in the tracking section so that changes in parameters of the lord elements can be transferred to their corresponding slaves by calling the appropriate Bmad routines.

Superposition has proven to be very useful. For one, since Bmad does the bookkeeping of splitting up elements and keeping the slave element attributes up-to-date, there is less chance of an error occurring. Another advantage is that a program itself does not have to "know" about how things are split since Bmad will take care of the details. Thus, for example, varying the field strength of the `cleo` solenoid in a program is trivial.

Element Slicing

Element "slicing" is similar to superposition discussed in the previous subsection in that slicing involves dividing elements longitudinally. The difference is that slicing involves crating a temporary structure within a program that represents a given section of the divided element. Slicing is hidden from the program's user.

Slicing relieves the programmer of a number of burdensome bookkeeping details when doing calculations that involve points internal to the elements. For simulations of such things like intra-beam scattering, Touschek effect, and coherent synchrotron radiation, the ability to step through an element slice-by-slice is essential.

Slicing also alleviates the need to, for example, split quadrupoles in two in the lattice file since calculating, say, the Twiss parameters at the center of any element is easily done on-the-fly using slicing.

Controllers

The ability to define controller elements that control the attributes of other elements has proved to be a very useful feature. Controllers can be used to simulate such things as the effect of control room "knobs" or simulate power supplies that power multiple magnets. A specialized `girder` controller can be used to simulate that effect of a support structure that supports a set of lattice elements. The following shows an example from a standard CESR lattice file:

```
h20w: overlay = {b20w[hkick] :0.5, &
                 b21w[hkick] :0.5}, hkick
h20w[hkick] = 0.01
```

For historical reasons, a controller is called an `overlay`. Here a controller named `h20w` controls the horizontal kick (`hkick`) attribute of two magnets named `b20w` and `b21w`. Setting the `hkick` attribute of the controller to 0.01 in the second line sets the hkick attribute of b20w and b21w to 0.005 each due to the 0.5 coefficients set in the first line.

Being able to define controllers as "just another element" simplifies program development and there is less of a maintenance burden having to maintain separate files specifying the lattice and who controls what.

Tracking Methods

Tracking and transfer map calculations (henceforth "tracking" will mean either one) are at the heart of many simulations, and different problems will have different requirements in terms of accuracy, speed, etc. To provide flexibility, Bmad implements a number of different tracking methods and what method is used can be selected on an element-by-element basis.

The commonly used "bmad_standard" method uses thick element formulas to quickly perform tracking. While not symplectic, this tracking method is useful for such tasks as computing Twiss parameters and finding the closed orbit. As an example, the program `CesrV`, which is used for data taking, machine corrections, and machine analysis at CESR, is able to calculate Twiss and coupling corrections in a few seconds using a single 2.67 GHz Xeon based computer. In this case, there are around 200 variables and 300 data values in a lattice with about 2000 elements.

For symplectic tracking, the "symp_lie_ptc" method, which uses Étienne Forest's PTC package[3], is available. Alternatively, the "taylor" method uses PTC to construct a Taylor map for tracking. Taylor tacking is faster than

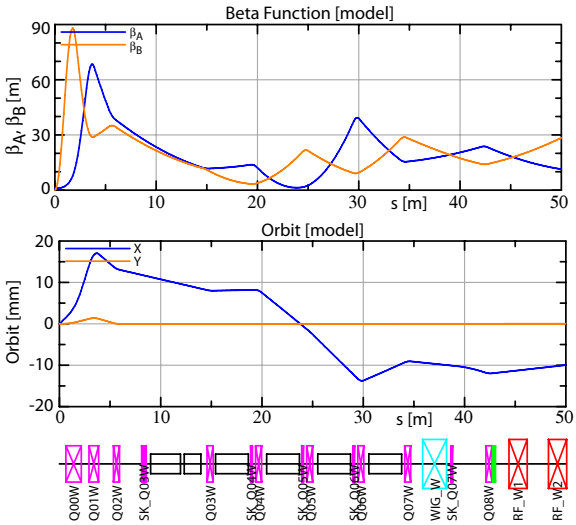


Figure 2: Example of a Tao plotting window.

symp_lie_ptc but will have a limited domain in phase space where the tracking is accurate.

There are a number of other tracking methods such as "runge_kutta" to track through a field profile, "linear" which just uses the linear part of the transfer map, and "custom" which allows for tracking with custom code.

BMAD ECOSYSTEM

The versatility of Bmad has resulted in Bmad being used as the computational engine powering a number of programs. It is thus possible to talk about the Bmad "ecosystem" Examples include

bbu_program A program to simulate the beam-beam breakup instability in Energy Recovery Linacs[4].

cesrv Data taking (orbits, phase/coupling, etc), simulation, and correction program for Cornell's CESR ring[5].

dark_current_tracker Program for simulating dark current electrons generated in a machine[6].

freq_map Frequency map program.

ibs_sim Analytic intra-beam scattering (IBS) calculation.

synrad3d Tracking of synchrotron radiation photons in the beam chamber including reflections[7].

tao A general purpose design and simulation program[8].

touschek_track Tracking of Touschek particles to determine where they are lost[9].

Three examples are illustrated below.

Tao: General Purpose Simulation Program

The disadvantage of Bmad, common to all software libraries, is that by itself it is not a program. That is, it cannot be run straight out-of-the-box. Over time, it became apparent that there was a great need for a general purpose program that could be used without any programming effort. As a result, a program called Tao[8] (Tool for Accelerator Optics) has been developed that uses Bmad as its calculational engine.

Tao has a general nonlinear optimization algorithm that can do such things as lattice design and orbit response matrix analysis. It has flexible plotting capabilities as displayed in Figs. 2 and 4. In many ways, Tao is similar to programs like MAD. A significant difference is that Tao is built from the ground up to be modular and extensible and so it is relatively easily, to add custom commands to Tao. For example by writing custom commands to read data and set magnet strengths, Tao can be become an on-line control system program to do such things as orbit flattening, etc.

The combination of Bmad and Tao thus gives the best of all possible worlds: The flexibility of a toolkit coupled with the ease of use of a program.

Cornell ERL DOOCS Interface

A control program based upon DOOCS (Distributed Object Oriented Control System), written in C++ and Java, has been developed for the Cornell Energy Recovery Linac[10] (ERL) test machine. Bmad was interfaced to this program for orbit and Twiss parameter calculations. A custom interface layer was developed to be able to translate between the C++ structures on the DOOCS side and the appropriate Bmad structures on the Fortran side. To simplify code development, use was made of the Bmad interface between Bmad Fortran structures and equivalent Bmad C++ structures. For lattice elements, the Bmad structure on the Fortran side is an `ele_struct`. On the C++ side, the equivalent Bmad class is called `C_ele`. To transfer element information from DOOCS to Bmad, for example, a `C_ele` instance is created on the C++ side of the interface and the appropriate information is transferred to it. Simplified, the code looks like:

```
C_ele quad;
quad.name = "q1";
quad.key  = QUADRUPOLE;
quad.value(B1_GRADIENT) = q_grad;
```

The `quad` instance is then passed to the Fortran side of the interface. On the Fortan side the code looks like:

```
subroutine ele_convert (c_ele)
  type (c_ele_struct) c_ele
  type (ele_struct) ele
  call ele_to_f (c_ele, ele)
```

The call to `ele_to_f` effects the transfer of information to the `ele_struct` and the element data can now be used by Bmad proper.

While not as clean as interfacing between two C++ code bases – the interface has two layers instead of one – the structure/class interface provided by Bmad greatly simplifies the integration of Bmad with C++ based code.

Dark Current Simulation

An example of the utility of Bmad in developing new simulation code is the `dark_current_tracker`

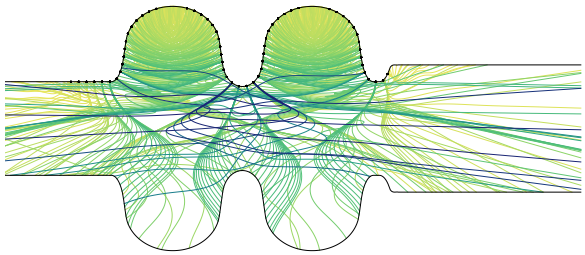


Figure 3: Dark current tracking in an RF accelerating cavity. Only part of the cavity is shown. The black circles are points of high field where particles are generated.

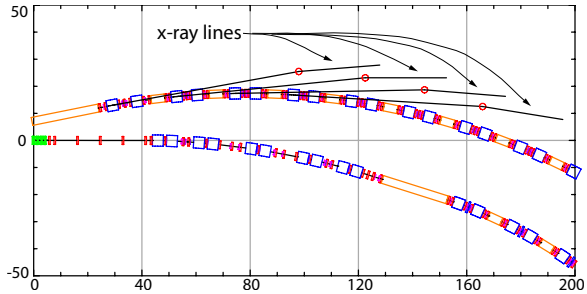


Figure 4: Simple example of the Cornell ERL with four x-ray lines.

program[6]. The problem was to simulate the trajectories of field emitted electrons from the walls of RF cavities. One of the first tasks of the program was to pull in information on the lattice including the beam chamber walls. At the time, Bmad had a model for specifying the wall for X-ray focusing capillaries, and it was decided that this model could be used for the beam chamber walls. This saved an appreciable amount programming time and effort. It also meant that the chamber wall could be specified in the lattice file which simplified the user interface.

Bmad provided many helper routines for the dark current simulation for doing such things as determining whether a particle had hit the wall, etc. Missing was suitable tracking code since the particles could reverse directions and the standard Bmad tracking methods are s-based and so assumed that particles always traveled in one direction. To overcome this, a time-based tracking routine was developed and merged into the Bmad library. Figure 3 shows a simulation from the program. Here the trajectories of a number of particles generated at high field regions in an RF cavity are shown.

With Bmad, the amount of work needed to create the simulation program was greatly reduced and, in turn, the capability of Bmad was increased by inclusion of the new tracking method which could be invoked in any program simply by setting an element's tracking method to use it. Additionally, the added ability to be able to specify a beam chamber wall will help with future simulations.

X-RAY SIMULATIONS

With the advent of Cornell's ERL project, one evolutionary track that Bmad has been following is the ability to simulate X-rays along with particle beams to provide an integrated simulation environment. One issue is the ability to define the entire machine in single lattice which contains both accelerator and X-ray components. Being able to treat the ERL as an integrated whole simplifies a number of simulations. For example, it simplifies the simultaneous simulation of the effect of electron beam movement on all the X-ray lines.

To define a line branching from a machine, Bmad uses `branch` and `photon_branch` elements. A `branch` or `photon_branch` element has zero length and marks the beginning of the branch line. `branch` elements are for describing such things as beam dumps where the charged particle being tracked can branch to. For X-ray lines, the `photon_branch` element is used. For example the following in a lattice file creates a branch:

```
br1: photon_branch, superimpose, &
                    ref = sa01, to = tg_mono
tg_mono: line = (...)
```

Here `br1` is a `photon_branch` which is placed at the center of an element named `sa01`. The associated X-ray line (not shown) is called `tg_mono`. There is no limit as to how many branches a lattice can have and branch lines can themselves contain branches ad infinitum. Figure 4 illustrates this showing four X-ray lines attached to an ERL.

For X-ray lines, besides the standard elements like `drift` and `marker`, There are a number of specialized elements that Bmad defines:

```
capillary,          mirror,
crystal,            multilayer_mirror
```

The `crystal` element supports both Bragg and Laue diffraction. The `capillary` element simulates X-ray focusing within a fine capillary[11]. The `mirror` element is a bendable mirror, and the `multilayer_mirror` is a mirror that depends upon multiple thin layers of dielectric material for reflection.

PRESENT WORK AND FUTURE PLANS

Bmad has had a steady evolution due to a steadily increasing demand for varying types of simulations. Currently, there are a number of areas of active development.

One thread is driven by the desire to be able to simulate an ERL from generation of the electrons at the gun cathode, through X-ray generation in wigglers and undulators, to propagating the X-rays to the experimental end stations. This work has two parts: Developing low energy simulation capabilities and developing X-ray simulations.

The work on low energy simulations is reaching maturity. Recent work, for example, has resulted in the creation of a `e_gun` element for simulating the cathode region of an electron gun. The major outstanding issue is

space charge. Since simulating space charge is complicated, rather then reinvent the wheel, the present goal is to integrate Bmad with existing space charge codes Impact-T[12] and OPAL[13].

X-ray simulation development is still in it's infancy. Photon generation code exists for bends and wigglers but not for undulators. The first real-world simulations are currently a work in progress. At this point in time, as mentioned in a previous section, Bmad has some locally developed code for photon tracking through crystals and a few other elements. Having locally developed code provides great flexibility for being able to simulate new effects. However, X-ray simulation can be quite complicated and it would be advantageous to be able to leverage the capabilities of existing codes. To this end, a collaboration is underway to interface Bmad with the Shadow[14] photon simulation code. In the long term, simulation of partially coherent X-rays would be useful. One possibility that is being explored is integration with the code SRW[15] developed by Oleg Chubar.

Up to now, the use of Étienne Forest's PTC code has been restricted to tracking through elements one element at a time. While this provides great flexibility in terms of being able to choose what type of tracking is used for a given element, it does not provide a way of using PTC's powerful analysis tools for calculating such things as beam emittances, resonance driving terms, etc. To remedy this, the appropriate interface code has, of late, been developed so that Bmad can now create a PTC `layout` for analysis. This makes possible the option of using PTC directly with Bmad simply providing the lattice parsing tools. It just takes two executable lines of code to construct a PTC layout from a Bmad lattice file:

```
type (lat_struct) lat
call bmad_parser (file_name, lat)
call lat_to_ptc_layout (lat)
```

In addition to the above, there are a number of minor projects on the drawing board, including development of nonlinear controllers, improved thread safety in multi-threaded environments, improved regression testing, and the interfacing to H5hut[16] – a high-performance library for storing particle position data.

ACKNOWLEDGMENTS

Many people have helped in the development of Bmad. In particular, thanks must go to David Rubin for his continued support and to Étienne Forest whose PTC package brings symplecticity to Bmad. Additionally, thanks must go to Chris Mayes for his work and for the use of Fig. 3.

REFERENCES

[1] D. Sagan, "Bmad: A relativistic charged particle simulation library," *Nuc. Instrum. and Methods in Phys Research A* **558**, 356–359 (2006). `http://www.lepp.cornell.edu/~dcs/bmad`.

[2] See: `http://mad.web.cern.ch/mad/`

[3] E. Forest, and F. Schmidt, "The FPP and PTC Libraries", Proceedings of ICAP 2006, Chamonix, France (2006)

[4] J. A. Crittenden, G. H. Hoffstaetter, M. Liepe *et al.*, "Recent progress on beam-breakup calculations for the Cornell x-ray ERL," *Proceedings of the 2009 Particle Accelerator Conference* (2009).

[5] D. Sagan, R. Meller, R. Littauer and D. Rubin, "Betatron phase and coupling measurements at the Cornell electron/positron storage ring," *Phys. Rev. ST Accel. Beams* **3**, 092801 (2000). Note: While the paper does not mention it, CesrV is the progame being used here for the analysis.

[6] C.E. Mayes, C. Chiu, G.H. Hoffstaetter, V.O. Kostroun, L.M. Nash, D.C. Sagan, "Dark Current Simulations for the Cornell ERL", Proceedings of IPAC2011, San SebastiàĄn, Spain. 2214 (2011).

[7] L. Boon J. Crittenden, T. Ishibashi, K. Harkay, "Application of the SYNRAD3D Photon-Tracking Model to Shielded Pickup Measurements of Electron Cloud Buildup at CesrTA", *Proceedings of the 2011 International Particle Accelerator Conference* (2011).

[8] D. Sagan, and J. C. Smith, "The Tao accelerator simulation program," *Proc. 2005 Part. Accel. Conf.*, 4159–61 (2005).

[9] M. P. Ehrlichman, and G. H. Hoffstaetter, "Collimating Touschek particles in an energy recovery linear accelerator," *Proceedings of ERL09* (2009).

[10] G. H. Hoffstaetter, B. Barstow, I. Bazarov *et al.*, "The Cornell ERL prototype project," *Proceedings of the 2003 Particle Accelerator Conference*, 192–194 (2003).

[11] S.A. Hoffian, D.J. Thiel, D.H.Bilderback, "Developments in Tapered Monocapillary and Polycapillary Glass Concentrators", Nucl. Instr. & methods in Phys. Res. A347, 384, (1994).

[12] J. Qiang, R. D. Ryne, S. Habib, V. and Decyk, "An object-oriented parallel particle-in-cell code for beam dynamics simulation in linear accelerators," *J. Comp. Phys.* **163**, 434 (2000).

[13] A. Adelmann, Ch. Kraus, Y. Ineichen *et al.*, "The Object Oriented Parallel Accelerator Library (OPAL), Design, Implementation and Application," *Proceedings of the 2009 Particle Accelerator Conference* (2009). `http://amas.web.psi.ch/docs/`.

[14] C. Welnak, G. J. Chena, and F. Cerrina, "SHADOW: A synchrotron radiation and x-ray optics simulation tool," *Nuc. Instrum. and Methods in Phys Research A*, 344–347 (1994).

[15] O. Chubar and P. Elleaume, "Accurate and efficient computation of synchrotron radiation in the near field region," Proceedings of the 6th European Particle Accelerator Conference p. 1177 (1998).

[16] M. Howison, A. Adelmann, E. Wes Bethel, Achim Gsell, Benedikt Oswald, Prabhat, "H5hut: A High-Performance I/O Library for Particle-based Simulations". IASDS 2010 Workshop on Interfaces and Abstractions for Scientific Data Storage, Heraklion, Crete, Greece, September 20-24 (2010).

SPACE CHARGE EFFECTS AND FOCUSING METHODS FOR LASER ACCELERATED ION BEAMS

P. Schmidt, O. Boine-Frankenheim, V. Kornilov, P. Spädtke, GSI, Darmstadt, Germany

Abstract

Within the scope of the LIGHT-Project [1] high intense beams of laser accelerated ions and co-moving electrons are produced. We are interested in methods for the controlled deneutralization of the beam, e.g. using a thin metal foil to absorb the electrons. Those beams show high initial divergence angle and velocity spread [1]. Therefore methods of focusing and collimating the beam are indispensable. Hence the focusing with a pulsed power solenoid is discussed.

INTRODUCTION

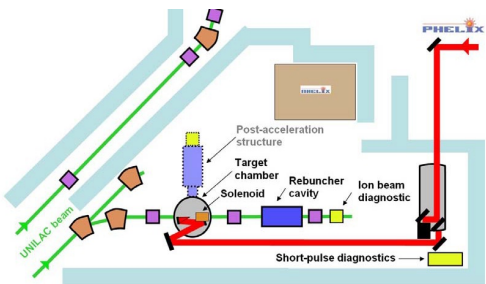


Figure 1: Setup of the LIGHT-Project.

Currently the LIGHT-Project (**L**aser **I**on **G**eneration, **H**andling and **T**ransport)[1] is performed at the Helmholtzzentrum für Schwerionenforschung Darmstadt (former GSI) (Germany). Whithin this project, high intense ion beams are generated by laser acceleration, using GSI's PHELIX-Laser. A scetch of the used setup is given in Fig. 1. Refering to the TNSA mechanism [2], one can expect, that the generated beam is quasi-neutral, composed of protons and co-moving electrons. We are therefore interested in methods for the controlled deneutralization of the beam to get clean experimental conditions. This can be done by using a thin metal foil to absorb the electrons[1]. The properties of this thin metal foil should be such that all electrons are removed, keeping the velocity -and density distribution of the protons being unaffected. The position of the foil is determined so that space charge has no effect behind the foil. A criterion in terms of beam parameters is worked out, by which one can decide how important space charge is. Following this foil, the beam is focused and carried towards a re-buncher-cavity (see Fig. 1). The focusing with a pulsed power solenoid (see Fig. 1) is discussed.

[1]The foil is not yet implemented in the setup.

BEAM MODEL

For the determination of space charge effects, we will use a rather simple beam model. Nevertheless it turns out, that this leads to general criterion, which only depends on the beam parameters. Let us assume we have a uniform cylindrical bunch with radius r_0 and length l_b, containing N_0 protons[2]. The charge number is $Z = 1$ and charge of each particle is $q = e$. Furthermore the beam has a longitudinal velocity $\vec{v}_\parallel$ with a velocity spread $\Delta v_\parallel$ and a transversal velocity $\vec{v}_\perp = v_\parallel r/r_0 \tan(\vartheta)$, where ϑ is the divergence angle. In the simplest case the bunch length is then given by the transversal velocity spread: $l_b(z) = l_0 + \Delta v_\parallel / v_\parallel z$, where z is the longitudinal coordinate and l_0 the initial length of the bunch. Neglecting edge effects, one finds the following envelope equation for the beam:

$$\partial^2_{zz}\sigma = K(z)/\sigma \tag{1}$$

Here $\sigma = r/r_0$ is the envelope radius, normalized by the initial radius. The model and geometry independent parameter

$$K(z) = \frac{Z^2\rho}{\beta_\parallel^2 c^2 \gamma^3 m_p}\frac{e^2}{2\epsilon_0} \tag{2}$$

is called perveance. With the particle density ρ, $\beta_\parallel = v_\parallel/c$, c the speed of light, γ the Lorentzian factor and m_p the ion mass. As one can see from Fig. 2, the result from eq. 1 is in good agreement with PIC Simulations done with VORPAL©[3].

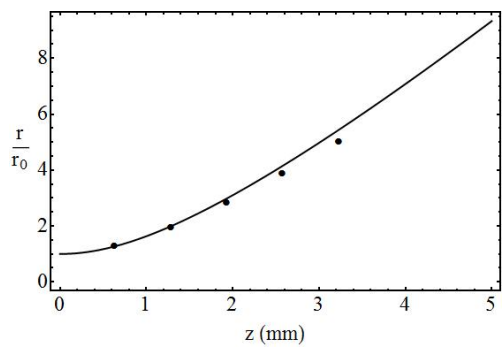


Figure 2: Beam radius from eq. 1 (black line) and simulation (dots) with VORPAL©[3].

SPACE CHARGE EFFECTS

The perveance from eq. 2 only depends on beam parameters and is a measure for the strength of space charge

[2]As one will see, the criterion can be applied for any ion.

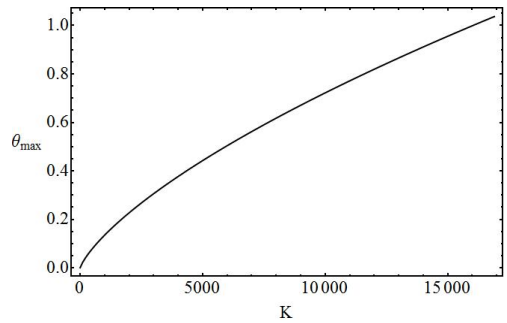


Figure 3: Maximum divergence angle as function of K for $\vartheta_0 = 0$.

effects. Due to the transversal and longitudinal expansion of the beam, K decreases while propagation. Therefore the divergence angle ϑ stays constant after a propagation length z. After this saturation, space charge has vanished. The maximum divergence angle ϑ_{max} which is reached depends on the initial perveance K_0, see Fig. 3.

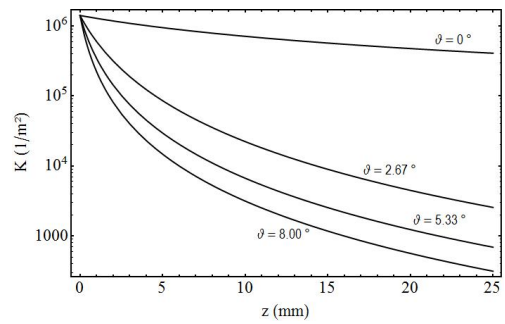


Figure 4: K as function of z for different inital angles.

Figure 4 shows the drop of K as a function of the propagation length for different initial angles in the case of space charge free drift. For example a maximum divergence angle of $\vartheta_{max} \leq 1°$ is tolerable in the setup. In the case of an initially laminar beam ($\vartheta_0 = 0$), one finds from Fig. 3 that $K \leq 15.000$ m^{-2} has to be full filled. As reference bunch we use a bunch with an average energy of 10 MeV and an energy spread of ± 1 MeV. From the measurements in [1] one gets $N_0 = 1 \cdot 10^{11}$ protons with an initial bunch length $l_0 \approx 1$ mm and radius $r_0 \approx 0,1$ mm. The primary objective of this section is to determine the position of the deneutralization foil so that space charge can be neglected after the bunch has crossed the foil. The perveance drops below the threshold value $K \leq 15.000$ m^{-2} after a space charge free drift length of at about $z \approx 1$ m (furthermore assuming $\vartheta_0 = 0$), as shown in Fig. 4. If the deneutralization foil is relocated to the position $z_{foil} = 1$ m, the divergence angle never exceeds $\vartheta_{max} \leq 1°$ and space charge can be neglected. As also shown in Fig. 4, the perveance decreases tremendously faster in the case $\vartheta_0 \neq 0$. If ϑ_0 is just $\approx 0,2°$, the foil can be placed 8 mm directly behind the proton source.

FOCUSING OF THE BEAM

The initial divergence of the laser accelerated beam is in the range of $0° \leq \vartheta_0 \leq 20°$ depending on the energy [1]. It is planned to strongly reduce this divergence by special target shapes as well as suitable laser spots. Nevertheless, the objective of the setup is to transport the proton beam into the rebuncher-cavity, which has an aperture of some mm. Therefore focusing and collimation structures are required. In the past a pulsed power solenoid was used to accomplish that aim (see Fig. 1). In the future setup it is planned to use a quadrupol-triplett instead of the solenoid.

FOCUSING WITH A SOLENOID

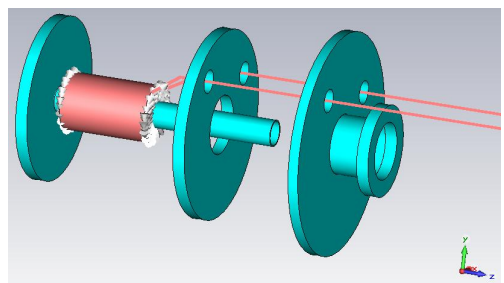


Figure 5: 3D-model of the used solenoid. Designed with CST EM Studio©[4]. Some parts of the setup are hidden for a better overview.

To verify that space charge remains negligible during the focusing with a solenoid, a CST©[4] particle tracking simulation (space charge free) is compared to a VORPAL©[3] simulation. For this purpose a 3D-model of the solenoid is designed in CST EM Studio©[4], see Fig. 5. A picture of the solenoid focal spot is given in Fig. 6 for the 10 ± 1 MeV reference bunch with $\vartheta_0 \approx 4.6°$.

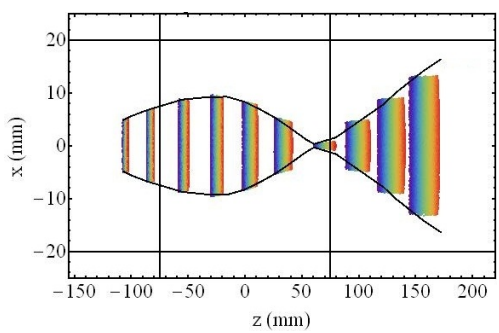


Figure 6: Stroboscopic recording of the bunch crossing the solenoid. From VORPAL©[3] simulation. The black line is the envelope trajectory from a CST©[4] particle tracking simulation (space charge free).

As can be seen in Fig. 6 both simulations show good agreement. Space charge effects can therefore be neglected during focusing. As further particle tracking simulations with CST Particle Studio©[4] showed, the power supply

wires of the solenoid (see Fig. 5) generated a magnetic dipole field, which caused a deflection of the beam, see Fig. 7. Based on this simulation, the experimental setup was improved and the wires field was removed.

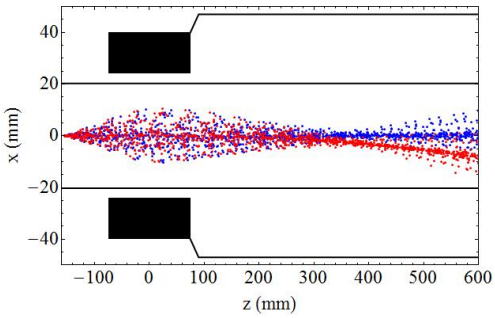


Figure 7: Particle tracking simulation with (red) and without (blue) supply wires. Done with CST Particle Studio© [4].

INDUCTIVE COUPLING AND OHMIC LOSSES

Due to the fact that the solenoid is operated in pulsed mode, eddy currents accure in all the metal parts surrounding the solenoid, especially in the beam pipe[3]. This leads to a coupling between the solenoid and the sourrounding parts, resulting in a time shift between the maximum of the current pulse and the maximum of the magnetic flux density. Particles that should be focused by the solenoid have to be triggered such that they reach the solenoid center at the maximum of the field strength. Furthermore these eddy currents cause ohmic losses, as shown in Fig. 8.

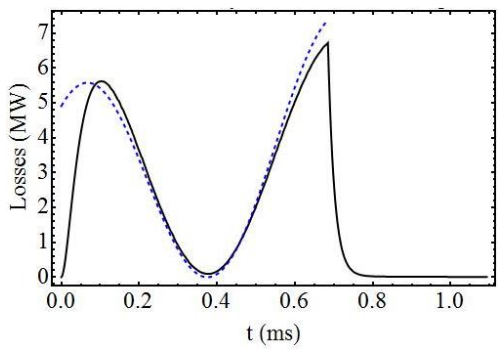


Figure 8: Ohmic losses in metal parts of the setup. Results from simulation with CST EM Studio© [4].

CONCLUSION

A method was developed with which the influence of space charge of an arbitrary particle beam can be allocated. Within that, the beam parameter depending quantity perveance K was introduced. As shown, the evolution of the perveance $K = K(z)$ is the leading quantity to rate the importance of space charge. Using this criterion, now one can determine the position of a thin metal foil for deneutralisation such that after the foil space charge can be neglected. The results from eq. 1 where compared to simulation results to verify its validity. The validity of the criterion could also be demonstrated for focusing. Moreover an improvement of the experimental setup could be achieved based on the results of particle tracking simulations.

[3] The beam pipe is centered in the solenoid.

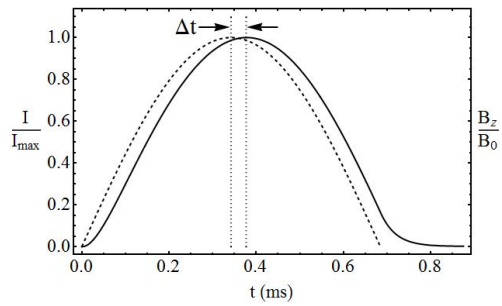


Figure 9: Time shift between maximum of current and magnetic flux density. Results from simulation with CST EM Studio© [4].

OUTLOOK

In the further work it is planned to work out a quality criterion, by which it can be decided what focusing method suits best for a given objective. Therefore further focusing structures will be taken into account (e.g. quadrupol-tripletts).

REFERENCES

[1] A. Almomani et al.: *Project Report: LIGHT - Laser Ion Generation, Handling and Transport*, GSI Darmstadt, Darmstadt (2010)

[2] T. Burris-Mog et al.: *Laser accelerated protons captured and transported by a pulse power solenoid*, Phys. Rev., Print 14 121301 (2011)

[3] Tech-X Corporation: *VORPAL*©, URL: www.txcorp.com, Boulder, USA (2012)

[4] CST Computer Simulation Technology AG: *CST Studio*©, URL: "'http://www.cst.com"', Deutschland Darmstadt (2012)

MATRIX FORMALISM FOR LONG-TERM EVOLUTION OF CHARGED PARTICLE AND SPIN DYNAMICS IN ELECTROSTATIC FIELDS

A. Ivanov*, S. Andrianov, Saint-Petersburg State University, Russia

Abstract

The matrix formalism as a numerical approach for solving of ODE equations is considered. It is a map method and has several advantages over classical step-by-step integration methods. This approach allows to present the solution as set of numerical matrices. A complete derivation of the equations this method is based on will be shown. Problems of symplectification and computing performance are discussed. We have developed an application that provides a tool for differential equations solving. The developed program allows to generate the final programming codes on C++, Fortran, MATLAB, C#, Java languages. The given approach is applied to long-term evolution of charged particle and spin dynamics in electrostatic fields.

MATRIX FORM OF ODE

Let's introduce a nonlinear system of ordinary differential equations

$$\frac{d}{dt}X = F(t, X). \tag{1}$$

Under the assumptions of $F(0, X_0) = 0$ the system (1) can be presented in the following form

$$\frac{d}{dt}X = \sum_{k=0}^{\infty} P^{1k}(t)X^{[k]}, \tag{2}$$

where $X^{[k]}$ is kronecker power of vector X, matrices P^{1k} can be calculate as

$$P^{1k}(t) = \frac{1}{(k)!}\frac{\partial^k F(t, X_0)}{\partial (X^{[k]})^T}, \quad k = 1, 2, \dots$$

Note that vector X is equal to $(x_1^{k_1}, \dots, x_n^{k_n})$, where x_i means ith component of state, $(k)! = k_1! \dots k_n!$

NUMERICAL IMPLEMENTATION

Solution of system (2) can be written in form

$$X = \sum_{k=0}^{\infty} R^{1k}(t)X_0^{[k]}. \tag{3}$$

Elements of matrices R^{1k} are depended on t and can be calculated in symbolic mode [1]. But such algorithms are quite complex. In this paper we propose a numerical implementaton of it. In this case matrices R^{1k} are evaluate in the specific time.

* 05x.andrey@gmail.com

After differentiating the equation (3) and taking into account (2) we get

$$\frac{dX}{dt} = \sum_{k=0}^{\infty} \frac{dR^{1k}(t)}{dt}X_0^{[k]},$$
$$\sum_{k=1}^{\infty} \frac{dR^{1k}(t)}{dt}X_0^{[k]} = \sum_{k=1}^{\infty} P^{1k}(t)X^{[k]}.$$

The partial derivatives of this equations with respect to $X_0^{[j]}$ are equal to

$$\begin{aligned} \frac{dR^{10}(t)}{dt} &= \sum_{k=1}^{\infty} P^{1k}(t)(R^{1k})^{[k]}, \\ \frac{dR^{1j}(t)}{dt} &= \sum_{k=1}^{\infty} P^{1k}(t)\frac{\partial X^{[k]}}{\partial (X_0^{[j]})^T}, \quad k = 1, 2, \dots \end{aligned} \tag{4}$$

and define the system of ordinary differential equations. Solution of this system is deterined matrices R^{1k}.

For integration of equations (4) numerical approach can be used. Note that step-by-step integration use only for map building. After that the solution that corresponds to the initil point X_0 can be calculate with the same map (3). In the research we use symplectic 2 stage Runge-Kutta scheme of 4 order (see [2]).

SYMPLECTICATION

The relation (3) can be presented as map transformation

$$X = R \circ X_0. \tag{5}$$

This map R is symplectic if

$$M^* J M = J, \forall X_0, \tag{6}$$

where $M = \partial X/\partial X_0$ and M^* is the transponse of M, E is identity matrix,

$$J = \begin{pmatrix} 0 & E \\ -E & 0 \end{pmatrix}. \tag{7}$$

Relation (6) in case of numerical matrices R^{1k} leads to a system of equations

$$a_0 + A_1 \mathbf{X_0}^{[1]} + \cdot + A_k \mathbf{X_0}^{[k]} = 0,$$

where A_i is a numerical vector. Note that this equation must be satisfied for any X_0. It means that the coefficients of each polynom are equal to zero and in this way appropriate corrections of the elements of the matrices R^{1k} can be found.

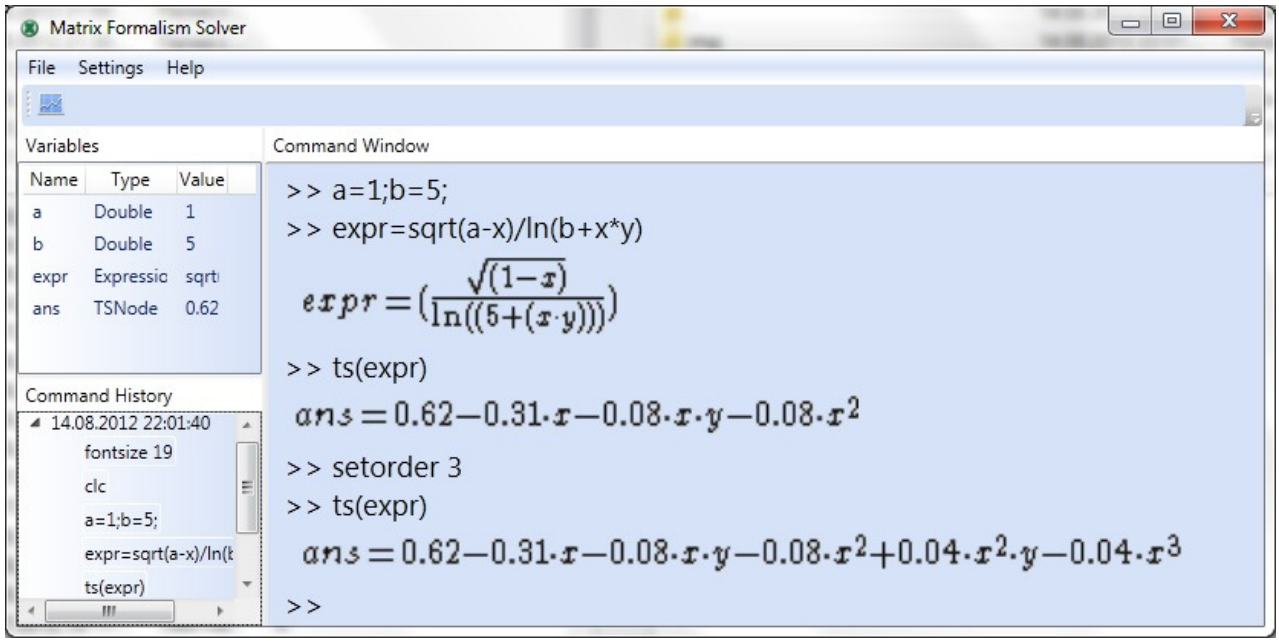


Figure 1: Command window for Taylor series expansion.

ODE SOLVING

For solving an ordinary differential eqation it is neccessary to implement following steps (see Fig. 2):

- to expanse function F in (1) in a Taylor series up to the order k;
- to build matrices $P^{1i}, i = 1 \dots k$;
- to solve the ordinary differential equations (4) with respect to the elements of matrices $R^{1i}, i = 1 \dots k$;
- calculate solution (3).

Taylor Series Expansion

We have implement libraries for automatically expansion of a nonlinear function to corresponded Taylor series up to the any order. The function may be a composition of such elementary functions as $\sin x, \cos x, \tan x, e^x, \sqrt{x}, \ln x$, and operators $+, -, *, /$.

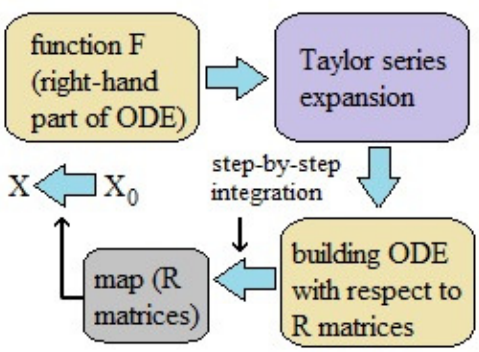


Figure 2: Map building of ODE.

In Fig. 1 developed command window is shown. For Taylor series representation symbolic polinoms are used.

SIMULATION OF ELECTROSTATIC STORAGE RING

Electrostatic storage ring consist of elements with different electric field distribution. In this research quadrupole lenses, cylindrical deflectors and drifts are used. The orbital motion and spin dynamics of the particle are described in [2]. Using such equations and library for Taylor series expansion that described above it is possible to build matrix form for each lattice element (see Fig. 3).

This software allows to build computational model and consist of pre build components. Each of this components can be added to the design area by drag and drop events.

Building of the Resulting Map

After maps are built for each elements of lattice it is possible to evaluate map for one tune of whole lattice. Such map is constructed using serial concatenation procedure. Imaging we have two numerical serial maps

$$X_1 = \sum_{k=0}^{k_1} R_1^{1k}(t) X_0^{[k]},$$

$$X_2 = \sum_{k=0}^{k_2} R_2^{1k}(t) X_1^{[k]}.$$

Substituting X_1 to the equation for X_2 we obtain

$$X_2 = \sum_{k=0}^{k_1 \cdot k_2} \tilde{R}_2^{1k}(t) X_0^{[k]}.$$

As you can see the resulting map has order $k_1 \cdot k_2$. But we can use terms of order not higher than $max(k_1, k_2)$, because other members are errors in the form of our initial in Taylor series expansions.

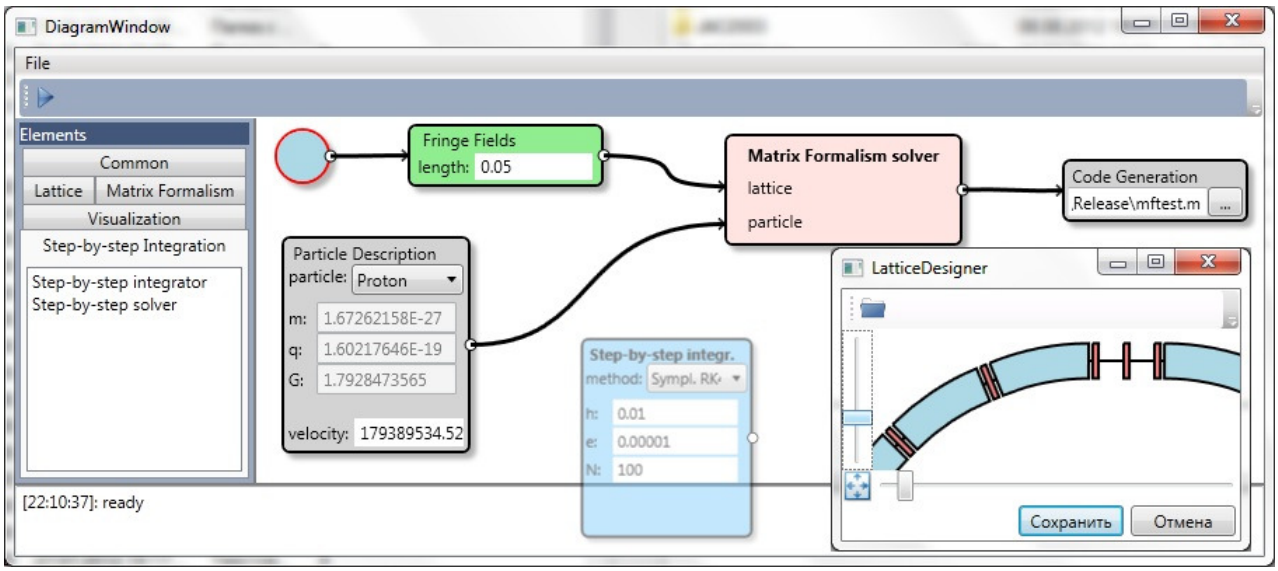


Figure 3: GUI environment for map building and code generation.

Evaluation of the Reference Orbit

Reference orbit of the map can be found from the equation (3). It is a nonlinear equation that can be solved by a numerical method. But in linear case we have:

$$X_{ref} = (E - R^{11})^{-1} R^{10}. \tag{8}$$

Symplectification Sample

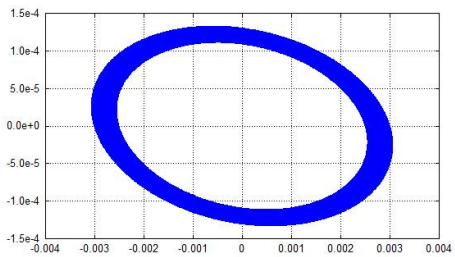


Figure 4: Map without symplectification.

Using a symplectic integration algorithm to build a map does not guarantee the symplectic map at all. For example, after 100000 turns becomes significant violation of the symplectic condition (see Fig. 4). To solve this problem we can increase the accuracy of step-by-step integration. But more accurate and efficient solution is map symplectification that described in this paper. In Fig. 5 a phase plane after additional symplectification is shown.

CONCLUSION

Matrix formalism is a high performance approach for ordinary differential equations solving. It allows present a solution as numerical matrix multiplication. This implies the possibility of implementation of the method in parallel codes [3, 4]. The result of the research is software for solving ODE in matrix form up to the neccessary order of nonlinearity. Direction of further development is the realization of matrix formalism in a symbolic form.

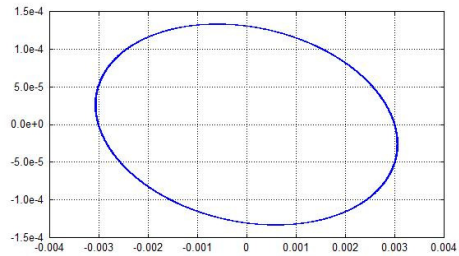


Figure 5: Map symplectification.

ACKNOWLEDGMENTS

The research is a part of JEDI[1] collaboration. The author would like to thank Yu. Senichev for advices and remarks and D. Zyuzin for comparative calculations on COSY Infinity.

REFERENCES

[1] S. Andrianov, "The Convergence and Accuracy of the Matrix Formalism Approximation," TUSCC2, these proceedings.

[2] A. Ivanov, "Particle tracking in electrostatic fields with energy conservation," WEP06, these proceedings.

[3] S. Andrianov, "Role of parallel and distributed computing in beam physics," Nuclear Instruments and Methods, 2004. Vol. 519. P. 37–41.

[4] N. Kulabukhova, "GPGPU Implementation of Matrix Formalism for Beam Dynamics Simulation," MOSDC2, these proceedings.

[1] Juelich Electric Dipole Moment Invenstigations, Spokes-persons: A. Lehrach, J. Pretz, and F. Rathmann

EM SIMULATIONS IN BEAM COUPLING IMPEDANCE STUDIES: SOME EXAMPLES OF APPLICATION

C. Zannini, CERN, Geneva Switzerland and EPFL, Lausanne, Switzerland
G. Rumolo, CERN, Geneva, Switzerland

Abstract

In the frame of the SPS upgrade an accurate impedance model is needed in order to predict the instability threshold and if necessary to start a campaign of impedance reduction. Analytical models, 3-D simulations and bench measurements are used to estimate the impedance contribution of the different devices along the machine. Special attention is devoted to the estimation of the impedance contribution of the kicker magnets that are suspected to be the most important impedance source in SPS. In particular a numerical study is carried out to analyze the effect of the serigraphy in the SPS extraction kicker. An important part of the devices simulations are the ferrite model. For this reason a numerical based method to measure the electromagnetic properties of the material has been developed to measure the ferrite properties. A simulation technique, in order to account for external cable is developed. The simulation results were benchmarked with analytical models and observations with beam. A numerical study was also performed to investigate the limits of the wire method for beam coupling impedance measurements.

INTRODUCTION

The aim of this paper is a description of the simulation studies performed in the frame of an improvement of the SPS impedance model. The simulations were performed mostly using the Wakefield solver of the 3-D commercial code CST Particle Studio.

SPS KICKER IMPEDANCE MODEL

A kicker is a special type of magnet designed to abruptly deflect the beam off its previous trajectory, for instance to extract the beam to a transfer line or to a beam dump. Among all the SPS elements, the kickers are suspected to contribute to a significant amount of the transverse impedance of the SPS.

Simple Models

In a very simple approximation an SPS ferrite loaded kicker can be modelled as two parallel plates of ferrite (see Fig. 1). For this simple geometrical model all the impedance terms (longitudinal, driving and detuning horizontal and vertical impedances) are calculated in [1, 2, 3]. Simulation results were found in very good agreement with the existing models [1, 2] and exactly predicted the detuning (also called quadrupolar) impedance of these devices [3]. The perfect agreement between analytical model and numerical simulations shown in [4] can be read as an important benchmark for the simulation code in the correct solution of electromagnetic problem involving dispersive materials such a ferrite.

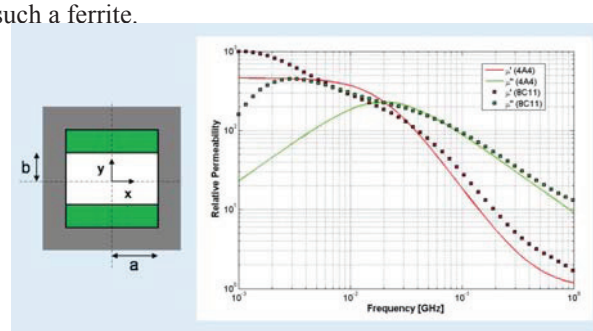


Figure 1: Tsutsui model on the left: ferrite in freen, PEC in gray and vacuum in white; complex permeability of the ferrite 4A4 and 8C11 on the right.

In the frame of an improvement of the kicker impedance model we performed a simulation study step by step to allow a good understanding of the different contributions to the kicker impedance. First, a device of finite length inserted in the vacuum tank and equipped with an inner conductor can support propagation of a quasi-TEM mode when interacting with the beam. The device behaves as a transmission line formed by the vacuum tank and the inner conductor which are continued on the external cables and closed on the appropriate circuit terminations. The TEM mode affects the impedance below a certain frequency (when the field penetration in the ferrite becomes comparable to the magnetic circuit length). This behaviour disappears as soon as we allow for 2-D geometries (infinite in the longitudinal direction) because the transverse TEM mode arises at the discontinuities which in this case are moved to infinity. In order to consider this aspect a circuit model was developed and benchmarked with CST 3-D simulations [5]. The good agreement between theoretical model and CST simulations is an ulterior proof of reliability of the code.

Realistic Models

In order to approach a more realistic model other aspects have to be considered: the longitudinal segmentation (see Fig. 2), internal circuits, non ideal terminations etc. An important simulation effort was devoted to the estimation of the impedance contribution of the SPS extraction kickers. Due to heating issues the original design of these kickers was modified. Interleaved fingers were printed by serigraphy directly on the ferrite (see Fig. 3). At the

moment among the 8 SPS extraction kickers (MKE) only one is still without serigraphy. Figure 4 shows the simulated longitudinal impedance for the MKE with and without serigraphy. As expected the effect of the serigraphy is a strong reduction of the real part of the longitudinal impedance. At the same time has to be noticed that the serigraphy introduces a resonance at 44 MHZ. This resonance (see Fig. 5) was studied in detail and recognized to be a quarter wavelength resonance on the silver fingers as shown in Figure 6. Neglecting the capacitive effect related to the finger width at the resonance frequency we obtain:

$$\lambda = \frac{c}{f\sqrt{\varepsilon_{eff}(f)\mu_{eff}(f)}} = 0.78\ m \cong 4L_{finger}$$

where f is the frequency of the resonance, and μ_{eff} and ε_{eff} are calculated as the effective permeability and permittivity of the kicker magnet approximated as an equivalent microstrip [6] and L_{finger} is the finger length. The theoretical formula confirms that we have a quarter wave resonance in agreement with the CST simulations.
The impedance model of the SPS extraction kicker shown in Figure 4 can explain the beam induced heating observed in the machine [7].

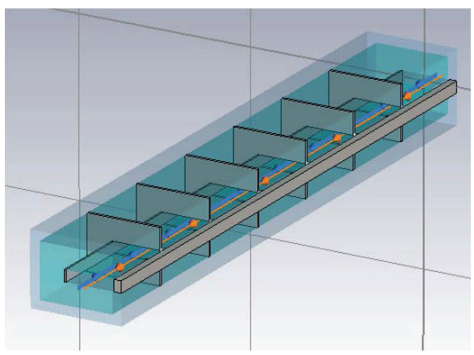


Figure 2: Simulation model of the SPS extraction kicker without serigraphy. The kicker is divided in 7 cells.

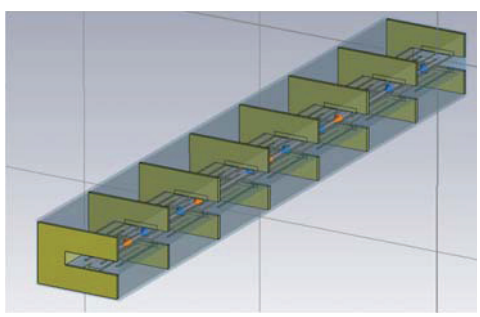


Figure 3: Simulation model of the SPS extraction kicker with serigraphy. The kicker is divided in 7 cells.

Ferrite Model

The model for the ferrite permeability μ as a function of frequency f was obtained from a first order dispersion fit on measured data by the supplier. Most SPS kickers are made of Ferroxcube ferrite type 8C11. In order to verify the ferrite properties used in the impedance model a measurement setup was implemented [8]. The measured ferrite properties have been found in good agreement with the ferrite model used up to now (see Fig. 7).

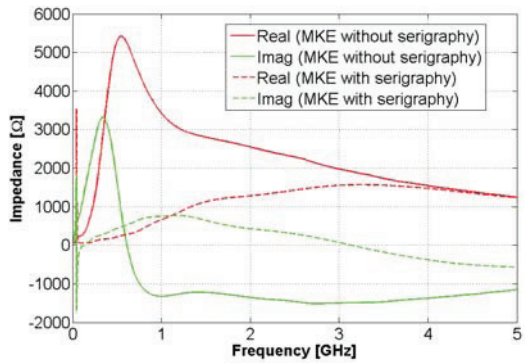


Figure 4: CST simulations of the longitudinal impedance for the SPS extraction kicker (MKE-L) with and without serigraphy.

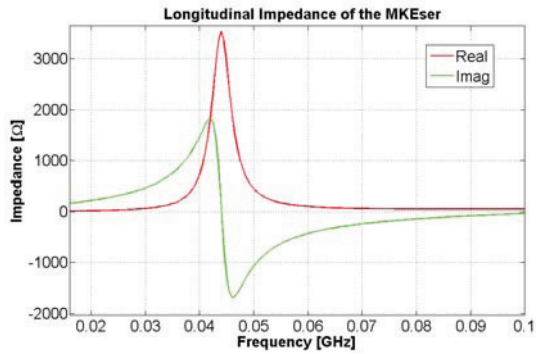


Figure 5: Zoom of the resonance due to the serigraphy.

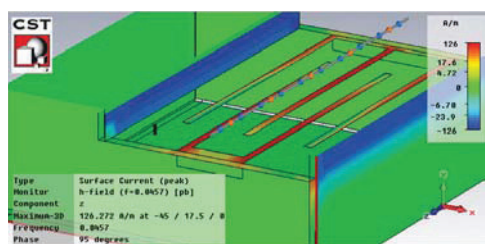


Figure 6: Surface current on the serigraphed fingers.

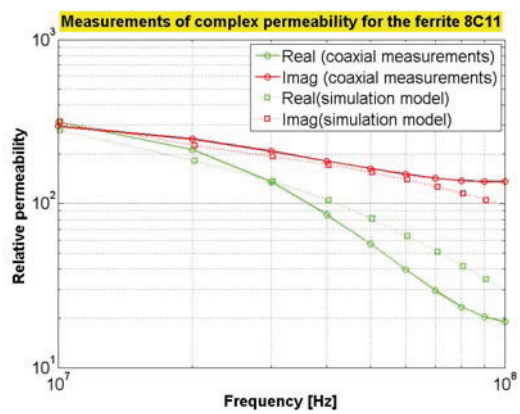


Figure 7: Measurements of complex permeability for the ferrite 8C11.

Simulation of a Kicker Loaded by a Coaxial Cable

An open ended external cable (length l, propagation constant k and characteristic impedance Z_0) exhibits an impedance given by the usual formula coming from the transmission line theory $Z = -jZ_0\,cot(kl)$. Resorting to the expansion of the cotangent function as sum of polar singularities we may reproduce the cable behavior by a lumped constant element circuit. The circuit can be approximated by a finite number of RLC parallel cells connected in series where each cell accounts for a resonance

$$Z = \frac{1}{G_0 + j\omega C_0} + \sum_{n=1}^{\infty} \frac{1}{G_n + j\left[\omega C_n - (\omega L_n)^{-1}\right]}$$

Numerical Investigation of Bench Measurements

Since many years the coaxial wire technique [9] is used to measure the beam coupling impedance of a device. Anyway the results obtained from wire measurements might be not entirely reliable because the stretched wire perturbs the EM boundary conditions, artificially introducing a TEM wave with a zero cut-off frequency among the solutions of the EM problem that will provide additional losses. For a better understanding of the limitations of the coaxial wire method, using the Transient Solver of CST Microwave Studio we performed numerical measurements, simulating the measurement setup. As example, we analyzed the case of an MKE-L with serigraphy. Difference with respect to the real setup could turn out from wire losses considered lossless in the model and from imperfect matching. In simulation the kicker is perfectly matched at both ends. Figure 8 shows a comparison of a CST Wakefield solver simulations and numerical measurements. The two curves are in good agreement for the broadband behaviour. Coherently to what expected by theory a reduction of Q and peak value for the low frequency resonance has been observed (see Fig. 9).

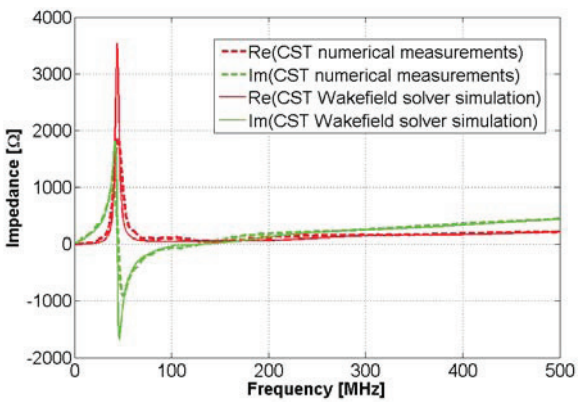


Figure 8: Longitudinal impedance of an MKE-L with serigraphy: comparing CST numerical measurements and Wakefield solver simulations.

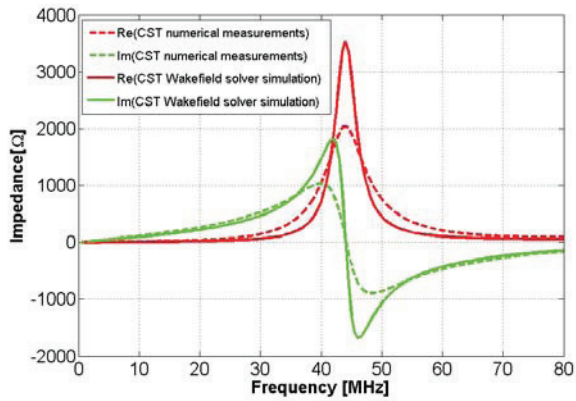


Figure 9: Low frequency behaviour of the longitudinal impedance for the MKE-L with serigraphy: comparing CST numerical measurements and Wakefield solver simulations.

CONCLUSIONS

Using the CST Wakefield Solver the SPS kicker impedance contribution has been estimated. The results were successfully benchmarked with analytical models [4, 5], and beam induced heating observations [7]. A numerical investigation of coaxial wire measurements has been presented as well.

ACKNOWLEDGMENT

The authors would like to thank M. Barnes, F. Caspers, H. Day, G. De Michele, E. MËtral, B. Salvant and V.G. Vaccaro.

REFERENCES

[1] H. Tsutsui, LHC Project Note 234 (2000).
[2] H. Tsutsui, ìTransverse Coupling Impedance of a Simplified Ferrite Kicker Magnet Modelî, LHC Project Note 234, 2000.
[3] B. Salvant et al., ìQuadrupolar Transverse Impedance of Simple Models of Kickersî, IPAC10, Kyoto, Japan, 2010.
[4] C. Zannini et al., ìElectromagnetic Simulations of Simple Models of Ferrite Loaded Kickersî, IPAC10, Kyoto, Japan, 2010.
[5] C. Zannini et al., ìEffect of the TEM mode on the kicker impedanceî, IPAC12, New Orleans, USA, 2012.
[6] D.M. Pozar, Microwave Engineering, Third Edition (2005)
[7] C. Zannini, G. Rumolo. ìMKE heating with and without seigraphyî, Presented at the SPSU meeting, 2 August 2012, http://paf-spsu.web.cern.ch/paf-spsu/
[8] G. De Michele, C. Zannini et al. ìBroadband Electromagnetic characterization of materials for accelerator componentsî, IPAC11, San Sebastian, Spain, 2011.
[9] V.G. Vaccaro, ìCoupling Impedance Measurements: an improved wire methodî, INFN/TC-94/023, 1994.

NUMERICAL CALCULATION OF BEAM COUPLING IMPEDANCES IN THE FREQUENCY DOMAIN USING FIT*

U. Niedermayer# and O. Boine-Frankenheim
Technische Universität Darmstadt, Institut für Theorie Elektromagnetischer Felder (TEMF),
Schlossgartenstraße 8, 64289 Darmstadt, Germany

Abstract

The transverse impedance of kicker magnets is considered to be one of the main beam instability sources in the projected SIS-100 at FAIR and also in the SPS at CERN. The longitudinal impedance can contribute to the heat load, which is especially a concern in the cold sections of SIS-100 and LHC. In the high frequency range, commercially available time domain codes like CST Particle Studio serve to calculate the impedance but they are inapplicable at medium and low frequencies which become more important for larger size synchrotrons. We present the ongoing work of developing a Finite Integration Technique (FIT) solver in frequency domain which is based on the Parallel and Extensible Toolkit for Scientific computing (PETSc) framework in C++. Proper beam adapted boundary conditions are important to validate the concept. The code is applied to an inductive insert used to compensate the longitudinal space charge impedance in low energy machines. Another application focuses on the impedance contribution of a ferrite kicker with inductively coupled pulse forming network (PFN).

INTRODUCTION

For the SIS100 synchrotron which will be built in the framework of the FAIR project, especially the coasting beam and the high intensity proton bunch are susceptible to impedance driven coherent transverse instabilities. Since SIS100 is a cryogenic (< 20 K) machine, the beam induced heat load is an important issue. In the relevant frequency range of several kHz to 2 GHz impedance sources are mainly given by the thin stainless steel beam pipe [1] and ferrite components. Above the cut-off frequency of the beam pipe, wake fields are traveling off-phase with the beam such that no net interaction takes place in long bunches. Additionally to the necessary ferrite kickers and their supply networks, also an inductive ferrite insertion to compensate the negative inductive longitudinal space charge impedance has been proposed. The longitudinal and transverse impedances of these objects have to be quantified in order to ensure a stable high intensity beam by designing instability countermeasures.

Usually coupling impedances are defined as the Fourier transform of the wake function. The wake function can be calculated by commercial software such as CST Particle Studio [2]. At low frequencies, which become more important for large hadron synchrotrons, this technique is inapplicable. From Küpfmüller's uncertainty principle $\Delta t \Delta f \geq 1$ with some definition of the time-duration and bandwidth [3] one finds that 300 km of wake-length have to be integrated in order to obtain a frequency-resolution of 1 kHz. Therefore at low and medium frequencies a frequency domain (FD) approach is pursued. The problem of determining coupling impedances in the frequency domain has already been addressed by Doliwa et al. [4] using a Neumann series approach with a Python implementation. The new implementation described here is supposed to carry on this work on a more general C++ platform with properly defined interfaces to the PETSc [5] framework on modern 64-bit machines.

The following will give a definition of the coupling impedances directly in FD. By convention underlined symbols emphasize complex variables. This also serves to distinguish between time domain (TD) and frequency domain (FD) fields. The beam with total charge q in a synchrotron is modeled as a disc with radius a of uniform surface charge density σ traveling with velocity v. The transverse displacement d_x of the beam (i.e. a coherent dipole oscillation) is approximated to first order by

$$\sigma(\varrho,\varphi) \approx \frac{q}{\pi a^2}\left[\Theta(a-\varrho)+\delta(a-\varrho)d_x\cos\varphi\right] \quad (1)$$

$$=: \sigma_\parallel\Theta(a-\varrho)+\sigma_\perp\delta(a-\varrho) \quad (2)$$

where Θ is the unit step and δ is its generalized derivative. The beam's volume charge density is given by $\varrho(\vec{r},t) = \sigma(\varrho,\varphi)\delta(z-vt)$ and it reads in frequency domain

$$\underline{\varrho}(\vec{r},\omega) = \int_{-\infty}^{\infty}\varrho(\vec{r},t)e^{-i\omega t}\mathrm{d}t = \frac{1}{v}\sigma(\varrho,\varphi)e^{-i\omega z/v}. \quad (3)$$

The beam current density in frequency domain is

$$\underline{J}_{s,z}(\varrho,\varphi,z;\omega) = \sigma(\varrho,\varphi)e^{-i\omega z/v} =: \underline{J}_\parallel + \underline{J}_\perp \quad (4)$$

where $\underline{J}_\parallel$ and $\underline{J}_\perp$ are the monopole and dipole components, as in Eq. (2), respectively. The coherent force due to beam induced electromagnetic fields acting back on the beam is described by the coupling impedance [6]

$$\underline{Z}_\parallel(\omega) = -\frac{1}{q^2}\int_{\mathrm{beam}}\underline{\vec{E}}\cdot\underline{\vec{J}}_\parallel^*\,\mathrm{d}V \quad (5)$$

$$\underline{Z}_{\perp,x}(\omega) = -\frac{v}{(qd_x)^2\omega}\int_{\mathrm{beam}}\underline{\vec{E}}\cdot\underline{\vec{J}}_\perp^*\mathrm{d}V. \quad (6)$$

The electric field $\underline{\vec{E}}$ is to be calculated from Maxwell's equations. Instead of the cosine distribution for dipolar excitation in Eq. (2) one can also use a twin wire approximation, as described in [7].

* Work supported by GSI
#u.niedermayer@gsi.de

MAXWELL'S EQUATIONS AND FIT IN FD

A complete description of electromagnetic fields in FD is given by Maxwell's equations

$$\nabla \times \underline{\vec{E}} = \underline{\vec{J}}_m - i\omega \underline{\vec{B}} \tag{7}$$

$$\nabla \times \underline{\vec{H}} = \underline{\vec{J}}_e + \underline{\vec{J}} + i\omega \underline{\vec{D}} \tag{8}$$

$$\nabla \cdot \underline{\vec{D}} = \underline{\varrho}_e \tag{9}$$

$$\nabla \cdot \underline{\vec{B}} = \underline{\varrho}_m \tag{10}$$

and material equations

$$\underline{\vec{D}} = \underline{\varepsilon}(\omega)\underline{\vec{E}}, \quad \underline{\varepsilon} = \varepsilon' - i\varepsilon'' \tag{11}$$

$$\underline{\vec{B}} = \underline{\mu}(\omega)\underline{\vec{H}}, \quad \underline{\mu} = \mu' - i\mu'' \tag{12}$$

$$\underline{\vec{J}} = \kappa \underline{\vec{E}} \tag{13}$$

where $\underline{\varrho}_e$ and $\underline{\vec{J}}_e$ denote the electric source charge and current densities and $\underline{\varrho}_m$ and $\underline{\vec{J}}_m$ denote equivalent magnetic source charge and current densities, respectively. Magnetic currents will be used to imprint externally calculated boundary fields. The FIT discretization of Eqs. (7)-(13) in the discrete $\mathbb{C}^{3n_p}$ space (see e.g. [8]) are the so-called Maxwell-grid-equations (MGE)

$$\mathbf{C}\underline{\widehat{\mathbf{e}}} = \underline{\widehat{\mathbf{j}}}_m - i\omega \underline{\widehat{\mathbf{b}}} \tag{14}$$

$$\widetilde{\mathbf{C}}\underline{\widehat{\mathbf{h}}} = \underline{\widehat{\mathbf{j}}}_e + \underline{\widehat{\mathbf{j}}} + i\omega \underline{\widehat{\mathbf{d}}} \tag{15}$$

$$\widetilde{\mathbf{S}}\underline{\widehat{\mathbf{d}}} = \underline{\mathbf{q}}_e \tag{16}$$

$$\mathbf{S}\underline{\widehat{\mathbf{b}}} = \underline{\mathbf{q}}_m \tag{17}$$

with the diagonal material matrices containing the first to second order numerical approximations (second order only for homogeneous grid)

$$\underline{\widehat{\mathbf{d}}} = \mathbf{M}_\varepsilon \underline{\widehat{\mathbf{e}}} \tag{18}$$

$$\underline{\widehat{\mathbf{b}}} = \mathbf{M}_\mu \underline{\widehat{\mathbf{h}}} \tag{19}$$

$$\underline{\widehat{\mathbf{j}}} = \mathbf{M}_\kappa \underline{\widehat{\mathbf{e}}}. \tag{20}$$

The grid incidence matrix $\mathbf{C}$ is real valued and consists of the partial derivative incidence operators $\mathbf{P}_x, \mathbf{P}_y, \mathbf{P}_z$. Note that in contrast to Eqs. (7)-(10), the FIT equations do not require differentiability of the the continuous fields since they are based on an integral formulation. Without the magnetic charge and current, the curl-curl linear system is obtained

$$\left(\widetilde{\mathbf{C}}\mathbf{M}_{\mu^{-1}}\mathbf{C} + i\omega\mathbf{M}_\kappa - \omega^2\mathbf{M}_\varepsilon\right)\underline{\widehat{\mathbf{e}}} = -i\omega\underline{\widehat{\mathbf{j}}}_e. \tag{21}$$

In the following, the system will be symmetrically rewritten with $\underline{\widehat{\mathbf{e}}} = \mathbf{M}_\varepsilon^{-1/2}\underline{\widehat{\mathbf{e}}}'$as

$$(\mathbf{M}_\varepsilon^{-1/2}\widetilde{\mathbf{C}}\mathbf{M}_{\mu^{-1}}\mathbf{C}\mathbf{M}_\varepsilon^{-1/2} - i\omega\mathbf{M}_\varepsilon^{-1/2}\mathbf{M}_\kappa - \omega^2 I)\underline{\widehat{\mathbf{e}}}' = -i\omega\mathbf{M}_\varepsilon^{-1/2}\underline{\widehat{\mathbf{j}}}_e. \tag{22}$$

Equation (22) will be abbreviated by

$$\mathbf{A}\underline{\widehat{\mathbf{e}}}' = b. \tag{23}$$

EXCITATION TERMS

The monopolar excitation current is given as a discretization of (4) with constant σ as

$$\underline{\widehat{\mathbf{j}}}_{e,z}^{\mathrm{mono}}(i_z) = \int \underline{\vec{J}} \cdot \mathrm{d}\tilde{\mathrm{A}}_\mathrm{z} = \mathrm{q}\mathrm{e}^{-\mathrm{i}\omega \mathrm{z_i}/\mathrm{v}} \tag{24}$$

where i_z is the z-index and z_i is its z coordinate. The dipolar excitation current is modeled by the twin wire dipole approximation as

$$\underline{\widehat{\mathbf{j}}}_{e,z}^{\mathrm{dip}}(i_z) = \underline{\widehat{\mathbf{j}}}_{e,z}^{\mathrm{mono}}(x = -d_x) - \underline{\widehat{\mathbf{j}}}_{e,z}^{\mathrm{mono}}(x = +d_x). \tag{25}$$

The beam's charge in the dual volumina is obtained using the continuity equation

$$\begin{aligned}\underline{\mathbf{q}}_e &= \frac{i}{\omega}\widetilde{\mathbf{S}}\underline{\widehat{\mathbf{j}}} \\ &= \frac{i}{\omega}(e^{-i\omega\frac{\Delta z}{2v}} - e^{-i\omega\frac{-\Delta z}{2v}})\sigma e^{-i\omega z_i/v} \\ &= \frac{2\sigma}{\omega}\sin\left(\frac{\omega\Delta z}{2v}\right)e^{-i\omega z_i/v} = \frac{q\Delta z}{v}e^{-i\omega z_i/v} + O(\Delta z^3)\end{aligned} \tag{26}$$

for a longitudinally equidistant grid. The integral in the impedance definitions in Eqs. (5) and (6) is evaluated by the functional

$$Z(\underline{\widehat{\mathbf{e}}}(\omega)) = \underline{\widehat{\mathbf{e}}} \cdot \underline{\widehat{\mathbf{j}}}^* \tag{27}$$

with normalized magnitude of the current ($q = 1$ As).

2D SIMULATIONS AND BOUNDARY CONDITIONS

In general the electromagnetic field excited by moving charges can be split into a source and a scattered part (see e.g. [9]). Below the cut-off frequency of the connected PEC beam pipe one finds only the source field (including the space charge field) in the pipe since the scattered field decreases exponentially. For longitudinally homogeneous structures like the pipe stubs the Fourier correspondence $\partial_z \to -i\omega/v$ holds [1]. Therefore, two possibilities for the beam entrance and exit boundary conditions (BC) occur: A periodic BC in which the beam entrance and exit plane are mapped to each other or an infinite beam pipe BC in which the source field is calculated separately and imprinted as boundary condition on the right hand side of (23).

Phase Corrected Periodic BC

The phase of the field within the kicker vessel cannot be known a priori since it depends on the scattered field. Nonetheless, the total phase difference between the field on the entrance and exit of the beam is a priori given by the phase advance of the source field. Therefore one can transform the field from the entrance to the exit with the operator

$$P_{z,\mathrm{exp}} = e^{-i\frac{\omega(L+\Delta z_{\mathrm{exit}})}{\beta c}}, \tag{28}$$

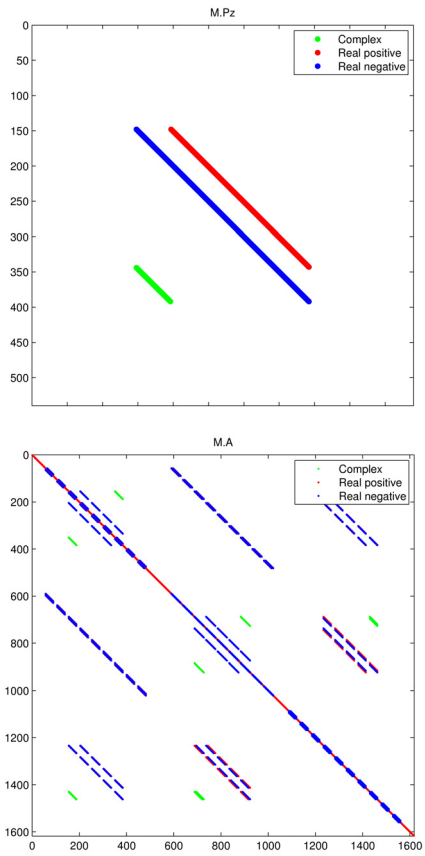


Figure 1: Sparsity pattern of $\mathbf{P}_z$ (top) and A matrix with periodic BC in z-direction and PEC-BC in x,y-direction (bottom). The green lines denote the factor $P_{z,\mathrm{exp}}$ which serves to close the curls on the z-boundary faces.

with L being the total length and Δz_{exit} being the length of the ghost edges at the exit. Drawbacks of these boundary conditions are that an additional band is introduced in the $\mathbf{P}_z$-matrix (see Fig. 1 top), leading to many additional bands in the curl-curl matrix (Fig. 1 bottom) and, more important, the main (large) curl-curl matrix is now complex.

Infinite Beam Pipe BC

For infinite beam pipe boundary conditions a supplementary grid (x,y-plane) has to be created on which the longitudinally harmonic fields are calculated and later imprinted in the right hand side of Eq. (23). We will refer to this as 2.5D solution since the longitudinal dependence is known a priori, but nonzero.

In the discrete formulation the grid incidence matrix for the longitudinal derivative is replaced by $\mathbf{P_z} = -\mathbf{1} + \exp(-\mathbf{i}\omega\triangle\mathbf{z}/\mathbf{v}) \approx -\mathbf{i}\omega\triangle\mathbf{z}/\mathbf{v}$. Note that $\widetilde{\mathbf{P}}_z = -\mathbf{P}_\mathbf{z}^\mathbf{H} = \mathbf{1} - \exp(+\mathbf{i}\omega\triangle\mathbf{z}/\mathbf{v}) \approx -\mathbf{i}\omega\triangle\mathbf{z}/\mathbf{v}$ is identical to $\mathbf{P_z}$ for small Δz. For the supplementary grid (SG) curl matrix one has $\widetilde{\mathbf{C}} = \mathbf{C}^H$. In the limit $v \to \infty$ which corresponds to $\partial_z \to 0$ the purely two-dimensional 'radial model' [1, 10] is found. Note that this simplified two-dimensional calculation serves not only as a boundary condition but also to calculate impedances of longitudinally homogeneous structures, as analytically done in e.g. [1, 7, 11].

The calculated 2D source fields are included in the main simulation by using the Source-Equivalence-Theorem. In practice this means that incomplete curls on the PEC-boundary are completed using the source field. One obtains (see also [9])

$$\begin{aligned}\mathbf{A}\underline{\widehat{\mathbf{e}}}' = &- i\omega\mathbf{M}_\varepsilon^{-1/2}\underline{\widehat{\mathbf{j}}}_e \\ &+ i\omega\mathbf{M}_\varepsilon^{-1/2}\underline{\widehat{\mathbf{j}}}_e^{\mathrm{eq}} - \mathbf{M}_\varepsilon^{-1/2}\widetilde{\mathbf{C}}\mathbf{M}_{\mu^{-1}}\underline{\widehat{\mathbf{j}}}_m^{\mathrm{eq}}. \end{aligned} \quad (29)$$

with the equivalent source electric and magnetic current densities given by

$$\underline{\widehat{\mathbf{j}}}_e^{\mathrm{eq}} = \widetilde{\mathbf{C}}_{\mathrm{R}}\underline{\widehat{\mathbf{h}}}^{\mathrm{SG}} \quad (30)$$

$$\underline{\widehat{\mathbf{j}}}_m^{\mathrm{eq}} = \mathbf{C}_{\mathrm{R}}\underline{\widehat{\mathbf{e}}}^{\mathrm{SG}} \quad (31)$$

where $\mathbf{C}_R$ and $\widetilde{\mathbf{C}}_R$ are the residual curl operators (the incidences missing in the main curl operator).

Space Charge Impedance as a Test-case

A verification of the implementation of the boundary conditions is given by the correct reproduction of analytically known space-charge fields and impedances for infinitely long beam pipes. In particular, the correct dependency on the velocity $\beta = v/c$ and mass factor $\gamma = (1-\beta^2)^{-1/2}$ of the beam show that the electric and magnetic fields cancel in the ultrarelativistic case. A mimetic correspondence from the FIT solution to the analytic calculation for a most simplified case can be found in the Appendix. Below the cut-off frequency, the correct $\beta^{-2}\gamma^{-2}$ dependence is found for both periodic and infinite beam pipe BC.

SOFTWARE

The CAD constructions and the mesh originate from CST EMS2011 [2]. It is imported via Matlab [12] where the material operators are disassembled in order to obtain the staircase material vectors. This allows to rescale the material parameters frequency dependently within the main program. Mesh data, material vectors and topological (PEC) information are transferred to the C++ main program. The actual high performance computations are carried out by the PETSc 3.2 [5] package which provides a variety of matrix structures, preconditioners and solvers for either real or complex linear systems. After computation, the fields can be visualized by transferring results back to Matlab and CST EMS2011. See Fig. 2 for the details. Equation (23) is treated with different PETSc solvers and preconditioners. So far, best results have been obtained by a symmetric successive over-relaxation (SSOR) preconditioner together with a stabilized bi-conjugate gradient (BICGSTAB), generalized minimum residual (GMRES) or a conjugate residual (CR) solver. In future, the application of algebraic multigrid (AMG) methods will be

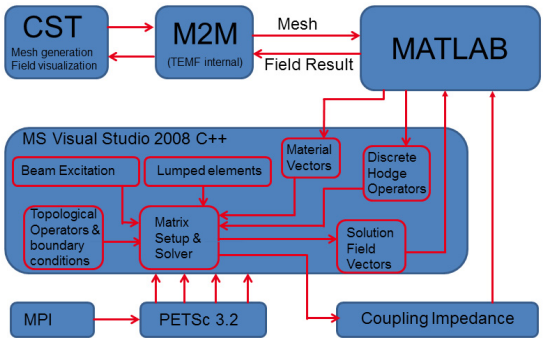


Figure 2: Implementation.

evaluated. For the supplementary grid, a direct solver (LU-decomposition) is applied since this does not constitute a bottleneck. For the main grid $3n_p = 10^6$ has been reached but the performance of the solver depends strongly on the frequency (diagonal dominance is a sufficient convergence condition for SOR). Note that the complex valued $\underline{\mu}$ deteriorates the condition number of the $\mathbf{A}$ matrix. For frequencies below the vessel cut-off (only $\approx$ 12 MHz due to the ferrite), the implementation of Doliwa's Neumann series approach [4] is considered.

APPLICATIONS

There is a variety of applications for this FD beam coupling impedance solver, especially the ones with ferrite are currently of interest. Ferrite materials provide high permeability (μ') but also high magnetization and polarization losses (μ'' and ϵ'').

Space Charge Compensation

The longitudinal space charge impedance (Eq. (40)) constitutes a major fraction of the imaginary part of the longitudinal impedance in SIS100. This leads to a RF-potential-well distortion and to a net decrease of the RF-voltage (decrease of bucket height). Since Eq. (40) behaves like a negative inductance, one has come up with the idea to compensate this with a lumped positive inductance, i.e. a cylindrical ferrite insertion. The ferrite has also the drawback of adding a major contribution to the real part of the longitudinal impedance. In the Los Alamos PSR one has found the 'Microwave Instability' occurring due to increased μ'' from the material [13]. Another issue is real-value of the transverse impedance added by the insert. With the presented simulation tool, the design and optimization of such an inductive insert can be approached. Figure 3 show the numerical calculation of a strongly simplified version of the insert. A way to avoid the instability could be to shield the insert from the beam at particular frequencies. Such shielding metal fingers are already in use in the LHC MKE-kickers [14].

Kicker with PFN

A major transverse impedance contribution in the SIS100 ring are the numerous ferrite kickers [15]. Besides

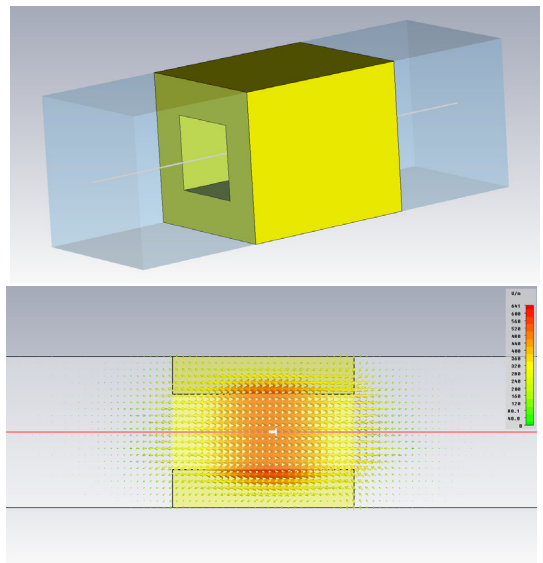


Figure 3: Simplified model of the insert with $\mu'_r = 1000, \mu''_r = 0$ (top) and electric field at $f = 100$ MHz with periodic boundary (bottom). The impedance is purely inductive.

the ferrite itself, also the supply network (pulse forming network, PFN) contributes to the impedance at lower frequencies. In the simulation the PFN can be included as a lumped admittance by adding its value to the correspondent entry in the $\mathbf{M}_\kappa$-matrix. The two frequency regimes in which the ferrite and the PFN dominate the transverse impedance are explained in [16]. There are means to shield the ferrite contribution [14] but the PFN impedance peaks are located partly within the kicker pulse spectrum, which cannot be shielded without interfering with the kicker's function (reciprocity).

CURRENT STATUS AND OUTLOOK

The fundamental part of the implementation of the code is done, the frequency range between 10 kHz and 100 MHz can be simulated with up to a few hundred thousand cells and either periodic or infinite beam pipe boundary conditions. In the near future, a divergence correction [17] using Eq. 26 will be implemented. At lower frequencies Doliwa's Neumann series approach [4] or an entirely two-dimensional simulation [1] are options. The presented software project aims for a complete database for the SIS100 longitudinal and transverse impedance below the beam pipe cut-off frequency. For the confirmation of the obtained results, bench measurements are outlined.

ACKNOWLEDGMENT

The authors wish to thank Wolfgang Ackermann for proofreading the manuscript.

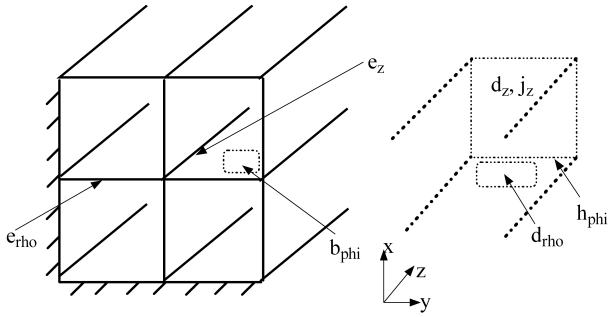


Figure 4: Simplest model mesh.

APPENDIX: SIMPLEST CASE FOR 2D CALCULATION

Applying the MGE (14)-(17) to the simplest structure in Fig. 4 one finds the following equations:

$$-i\frac{\omega}{v}\Delta z\underline{\widehat{\mathbf{e}}}_\varrho - \underline{\widehat{\mathbf{e}}}_z = -i\omega\underline{\widehat{\widehat{\mathbf{b}}}}_\varphi \tag{32}$$

$$4\underline{\widehat{\mathbf{h}}}_\varphi - i\omega\underline{\widehat{\widehat{\mathbf{d}}}}_z = \underline{\widehat{\widehat{\mathbf{j}}}}_{e,z} = q \tag{33}$$

$$-i\frac{\omega}{v}\Delta\tilde{z}\underline{\widehat{\widehat{\mathbf{d}}}}_z + 4\underline{\widehat{\widehat{\mathbf{d}}}}_\varrho = \mathbf{q}_e = \frac{q\Delta\tilde{z}}{v} \tag{34}$$

$$\underline{\widehat{\widehat{\mathbf{b}}}}_z = 0. \tag{35}$$

These three equations for the three unknowns $\underline{\widehat{\mathbf{e}}}_z$, $\underline{\widehat{\mathbf{e}}}_\varrho$ and $\underline{\widehat{\widehat{\mathbf{b}}}}_\varphi$ can be solved using the material operators (18) and (19),

$$\begin{aligned} 4\underline{\widehat{\mathbf{e}}}_z + \omega^2\mathbf{M}_\mu^\varphi\mathbf{M}_\varepsilon^z(\mathbf{I} - \frac{\Delta z\Delta\tilde{z}}{v^2}\mathbf{M}_{\mu^{-1}}^\varphi\mathbf{M}_{\varepsilon^{-1}}^\varrho)\underline{\widehat{\mathbf{e}}}_z \\ = i\omega\mathbf{M}_\mu^\varphi(\mathbf{I} - \frac{\Delta z\Delta\tilde{z}}{v^2}\mathbf{M}_{\mu^{-1}}^\varphi\mathbf{M}_{\varepsilon^{-1}}^\varrho)q \end{aligned} \tag{36}$$

where one obtains

$$(\mathbf{I} - \frac{\Delta z\Delta\tilde{z}}{v^2}\mathbf{M}_{\mu^{-1}}^\varphi\mathbf{M}_{\varepsilon^{-1}}^\varrho) = -\frac{1}{\beta^2\gamma^2}\mathbf{I} \tag{37}$$

independent of the mesh lengths. One finally finds

$$\underline{\widehat{\mathbf{e}}}_z = \frac{-i\omega}{4\beta^2\gamma^2 - \omega^2\delta_\perp^2/c^2}\mathbf{M}_\mu^\varphi q. \tag{38}$$

where $\delta_\perp = \Delta x = \Delta\tilde{x} = \Delta y = \Delta\tilde{y}$ denotes the transverse edges. Equation (36) is the mimetic correspondence to the continuous relativistic 2.5D Helmholtz equation [1]

$$\left(\triangle_\perp - \frac{\omega^2}{\beta^2\gamma^2c^2}\right)\underline{E}_z = -\frac{i\omega\mu\sigma}{\beta^2\gamma^2}e^{-i\omega z/v} \tag{39}$$

where the factor 4 in Eq. (36) originates from the difference stencil of the transverse Laplacian. From Eq. (39) one obtains after solving by a product separation ansatz and some simplifications [11] the longitudinal space charge impedance

$$\underline{Z}_\parallel^{SC} = -i\omega\frac{\mu_0 gl}{4\pi\beta^2\gamma^2}. \tag{40}$$

below cut-off. This behaviour ($\propto \beta^{-2}\gamma^{-2}$, $\propto \omega$, negative imaginary) has to be reproduced by the numerical solution, in particular by Eq. (38). Additionally, $\underline{Z}_\parallel^{\mathrm{SC}} \to 0$ must hold for $\omega \to \infty$. Note that the geometry factor g can only be poorly reproduced by staircase FIT for curved structures.

REFERENCES

[1] U. Niedermayer and O. Boine-Frankenheim, Analytical and numerical calculations of resistive wall impedances for thin beam pipe structures at low frequencies, NIM A 687, 5161, 2012.

[2] CST Studio Suite®2011/2012, `www.cst.com`

[3] K. Küpfmüller, Einführung in die theoretische Elektrotechnik, Springer, 1932.

[4] B. Doliwa et al., Optimised electromagnetic 3D field solver for frequencies below the first resonance, IET Sci. Meas. Technol., 2007

[5] Portable, Extensible Toolkit for Scientific Computation, `www.mcs.anl.gov/petsc`

[6] R. Gluckstern, Analytic methods for calculating coupling impedances, Cern Accelerator School, 2000.

[7] G. Nassibian and F. Sacherer, Methods for measuring transverse coupling impedances in circular accelerators, NIM, 1979.

[8] M. Clemens and T. Weiland, Discrete Electromagnetism with the Finite Integration Technique, PIER 32, 65-87, 2001.

[9] M. C. Balk, Feldsimulation starrer Teilchenstrahlen beliebiger Geschindigkeit und deren Anwendung in der Schwerionenbeschleunigerphysik, PhD at TU Darmstadt, 2005.

[10] H. Hahn, Matrix solution for the wall impedance of infinitely long multilayer circular beam tubes, PRSTAB 13, 012002, 2010.

[11] A. Al-Khateeb et al., Analytical calculation of the longitudinal space charge and resistive wall impedances in a smooth cylindrical pipe, Phys. Rev. E 63, 026503, 2001.

[12] MATLAB®R2011a, `www.mathworks.de`

[13] C. Beltran, Study of the longitudinal Space Charge Compensation and longitudinal Instability of the Ferrite Inductive Inserts in the Los Alamos Proton Storage Ring, PhD at Indiana University, 2003.

[14] T. Kroyer et al., Longitudinal and Transverse Wire Measurements for the Evaluation of Impedance Reduction Measures on the MKE Extraction Kickers, CERN AB-Note-2007-028.

[15] P. Spiller et al., FAIR Technical Design Report SIS100, 2008.

[16] B. Doliwa et al., Numerical calculation of transverse coupling impedances: Comparison to Spallation Neutron Source extraction kicker measurements, PRSTAB, 2007

[17] B. Doliwa and T. Weiland, Numerical Computation of Kicker Impedances: Towards a complete Database for the GSI SIS100/300 Kickers, Proc. of ICAP, 2006.

ELECTROMAGNETIC CHARACTERIZATION OF MATERIALS FOR THE CLIC DAMPING RINGS

E. Koukovini-Platia*, G. De Michele†, C. Zannini*, G. Rumolo,
CERN, Geneva, Switzerland

Abstract

The performance of the Compact Linear Collider (CLIC) damping rings (DR) is likely to be limited by collective effects due to the unprecedented brilliance of the beams. Coating will be used in both electron (EDR) and positron damping rings (PDR) to suppress effects like electron cloud formation or ion instabilities. The impedance modeling of the chambers, necessary for the instabilities studies which will ensure safe operation under nominal conditions, must include the contribution from the coating materials applied for electron cloud mitigation and/or ultra-low vacuum pressure. This advocates for a correct characterization of this impedance in a high frequency range, which is still widely unexplored. The electrical conductivity of the materials in the frequency range of few GHz is determined with the waveguide method, based on a combination of experimental measurements of the complex transmission coefficient S_{21} and CST 3D electromagnetic (EM) simulations.

INTRODUCTION

The EM characterization of the material properties up to high frequencies is required for the impedance modeling of the CLIC DR components. Layers of coating materials such as amorphous carbon (aC) and Non-Evaporable Getter (NEG) are necessary for e-cloud mitigation and ultra-high vacuum. The waveguide method is used to characterize the properties of those coating materials in a range of frequencies of few GHz. The reliability of this method is tested in the range of 9-12 GHz using a standard X-band waveguide before trying to measure in the range of 325-500 GHz using a Y-band waveguide. The electrical conductivity of the material is obtained from the measured transmission coefficient S_{21} and 3D EM simulations with CST Microwave Studio®(CST MWS) [1].

WAVEGUIDE METHOD

An X-band copper (Cu) waveguide of 50 cm length and the same Cu waveguide coated with NEG are the devices under study during the experiment. Using a network analyzer, the transmission coefficient is measured over a frequency range from 9-12 GHz. The experimental method's setup is displayed in Fig. 1. The S_{21} coefficient is related to the attenuation due to the finite conductivity of the material.

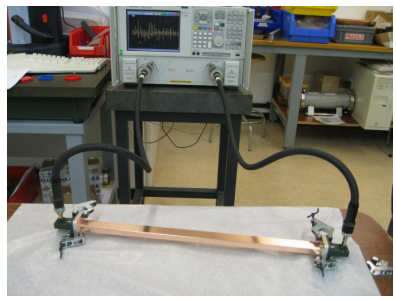


Figure 1: Measurements setup: using a network analyzer, the transmission coefficient S_{21} of an X-band waveguide is measured.

NEG Coating Procedure

NEG coating is necessary to suppress fast ion instabilities in the EDR. A Cu X-band waveguide is coated with a Ti-Zr-V thin film by magnetron sputtering using two cathodes, each of them made of elemental wires inter-twisted together [2]. The coating was targeted to be as thick as possible in order to increase the sensitivity of the measurements. From x-rays measurements, the thickness is assumed to be 9 μm. The real profile of the coating thickness will be implemented in the simulations as soon as more detailed profile measurements will be available.

SIMULATIONS WITH CST

The CST MWS is used to simulate a waveguide made from Cu and a NEG coated one with the same dimensions as the real ones (see Fig. 2). With the Transient Solver of CST, the experimental setup used for the measurements can be simulated in real time domain.

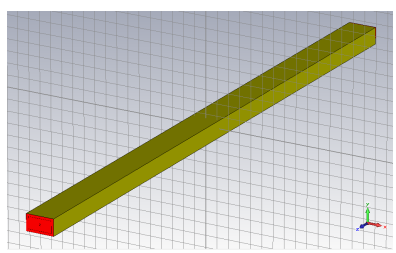


Figure 2: X-band Cu waveguide simulated with CST MWS of 50 cm length.

For each frequency from 9-12 GHz the output of the 3D EM simulations is the S_{21} parameter as a function of conductivity. The relative permittivity ε_r and permeability μ_r of the material are assumed to be equal to one while the conductivity σ is the unknown parameter which is scanned

* also affiliated at EPFL
† also affiliated at EPFL and PSI

in simulations. For CST simulations, the coating thickness is assumed to be uniform and equal to 9 μm (expected thickness from x-rays) or infinite to simulate the case where the skin depth is much smaller than the thickness (upper limit for the NEG conductivity). The intersection at each frequency (see Fig. 3) of the measured S_{21} with the CST output obtained numerically, determines the unknown electrical conductivity as a function of frequency. Repeating this method for all the frequency range the material conductivity is obtained as a function of frequency.

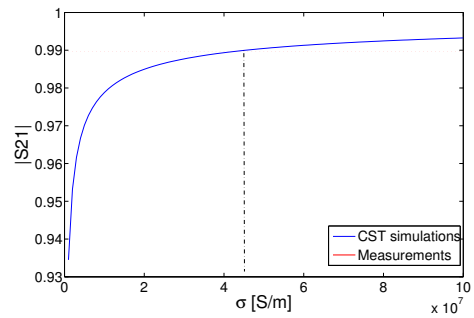


Figure 3: Example at 10 GHz: intersection of S_{21} from CST simulations with the measured value.

Conductivity Measurement of a Pure Cu X-band Waveguide

As a first step, a pure Cu waveguide is used to measure the S_{21} coefficient and test the feasibility of the method. The results from the experiment are shown in Fig. 4.

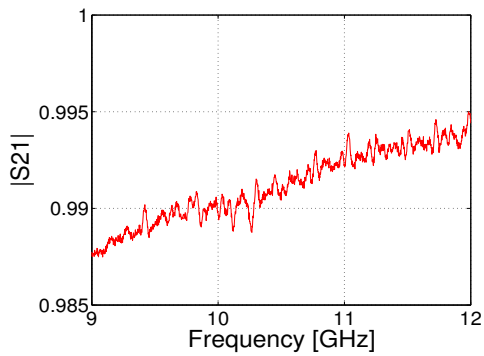


Figure 4: Measured S_{21} as a function of frequency for a pure Cu X-band waveguide.

A Cu waveguide is then simulated with CST MWS assuming conductivity equal to 5.8×10^7 S/m at 20 °C. The intersection of the measurements with the CST simulations show that the conductivity of Cu was estimated within the same order of magnitude with the known value (see Fig. 5). From the measurement, the average value of σ_{Cu} is equal to 5.91×10^7 S/m, which is in a very good agreement with the known one.

For the Cu waveguide measurements, the attenuation is very sensitive to the errors because of the small losses in Cu. Despite this fact, the results obtained were considered

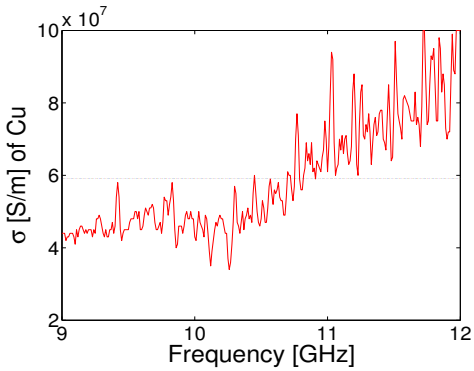


Figure 5: Copper conductivity as a function of frequency.

to be encouraging to continue with the NEG coated waveguide measurement. However, in the future it is foreseen to benchmark CST MWS and the reliability of the experimental method by using a stainless steel waveguide which will be less sensitive to the errors due to its lower conductivity. This future step will be another important benchmark of the method.

Conductivity Measurement of a NEG Coated X-band Waveguide

The same experimental method is used for the NEG coated waveguide in order to measure the S_{21} coefficient. In Fig. 6 the comparison of S_{21} between the pure Cu and the same waveguide coated with NEG is shown.

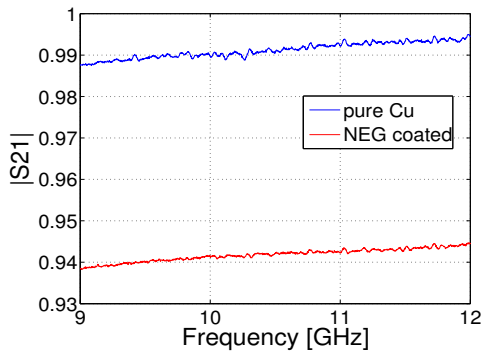


Figure 6: Measured S_{21} as a function of frequency for a Cu and a NEG coated waveguide.

Despite the unknown NEG coating thickness, the measured S_{21} coefficient indicates that the skin depth is small enough compared to the thickness allowing the EM interaction with NEG. Two possible scenarios exist; the skin depth is much smaller than the coating thickness, therefore the EM interaction is only with NEG and indeed the σ defined corresponds to σ_{NEG}. Second case is the skin depth to be comparable to the thickness and the EM interaction is with NEG and Cu. In the second case, the measured conductivity corresponds to a combination of NEG and Cu.

From the intersection of measured data with CST MWS results, the conductivity is plotted as a function of fre-

ductivity corresponds to a combination of NEG and Cu.

From the intersection of measured data with CST MWS results, the conductivity is plotted as a function of frequency. The plot displayed in Fig. 7 shows an upper limit for the NEG conductivity assuming that the NEG thickness is infinite in the simulation. The case of 9 μm thickness is also plotted in the same figure.

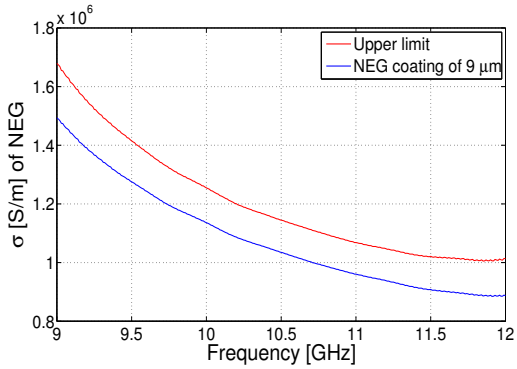


Figure 7: Conductivity of NEG as a function of frequency for infinite and 9 μm coating thickness.

Error Analysis with CST MWS

In order to check the reliability of CST MWS results on a coated waveguide, two tests are done. For the first one, different values of NEG thickness on a Cu waveguide are simulated varying from 1-20 μm. NEG conductivity is assumed to be equal to 2×10^6 S/m. The results are illustrated in Fig. 8.

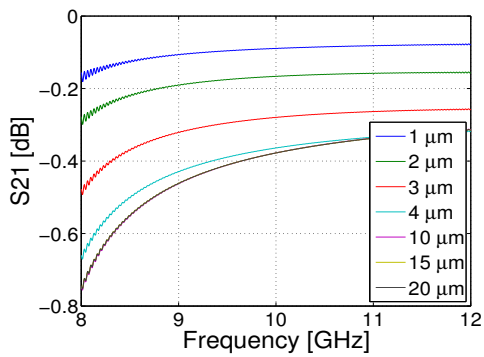


Figure 8: S_{21} results obtained from CST MWS for different values of NEG thickness.

For the lower values between 1-4 μm, the skin depth is larger or comparable to the thickness, therefore the EM fields interact with NEG and Cu. The skin depth can be estimated from Eq. 1:

$$\delta = \sqrt{\frac{2}{\mu\omega\sigma}} \approx 503\sqrt{\frac{1}{\mu_r f\sigma}} \qquad (1)$$

where δ is the skin depth in m, μ the permeability of the medium, μ_r the relative permeability, σ the conductivity in S/m and f the frequency in Hz [3]. For the range of frequencies simulated, 8-12 GHz, the skin depth varies from 3.9-3.2 μm respectively. At a certain frequency, as thickness increases, the interaction with Cu is reduced till the point where the skin depth is much smaller than the coating allowing only the interaction with NEG (thickness from 10-20 μm). CST MWS results are in agreement with the theoretical expectations.

As a second test, a waveguide with the properties of NEG (the assumption is made that the conductivity is equal to 2×10^6 S/m) is compared with a Cu waveguide coated with NEG of 100 μm thickness. Such thick coating results in the EM interaction only with NEG since the skin depth is much smaller than the thickness. Therefore the results of those two simulations should be identical. The comparison is illustrated in Fig. 9. The results are similar between those two cases as expected from theory.

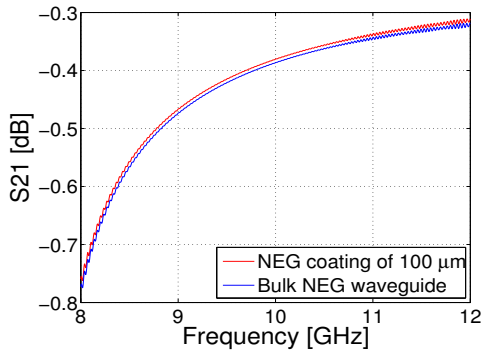


Figure 9: Comparison of S_{21} results obtained from CST MWS between a 100 μm NEG coated waveguide and a waveguide with the properties of NEG.

SUMMARY

The waveguide method combined with CST 3D EM simulations is used to determine the conductivity of NEG in a range of frequencies from 9-12 GHz. An upper limit for its value is obtained and first tests of the simulations reliability are done. In the future it is planned to benchmark in detail the CST MWS coating simulations, proceed with stainless steel waveguide measurements and conductivity measurements at higher frequencies in the range of 325-500 GHz.

ACKNOWLEDGEMENTS

The authors would like to thank A.T. Perez Fontenla, G. Arnau Izquierdo, S. Lebet, M. Malabaila, P. Costa Pinto and M. Taborelli for collaboration, preparation of the waveguide, profile thickness measurements as well as fruitful discussions.

REFERENCES

[1] CST Microwave Studio- Getting Started (2003).

[2] P. Costa Pinto, CERN-TS-Note-2005-030 (2005).

[3] K.A. Milton, J. Schwinger, *Electromagnetic Radiation: Variational Methods, Waveguides and Accelerators*, Springer, 2006.

THE DARK CURRENT AND MULTIPACTING CAPABILITIES IN OPAL: MODEL, BENCHMARKS AND APPLICATIONS

C. Wang*, Z.G. Yin, T.J. Zhang, CIAE, Beijing, China
A. Adelmann†, PSI, Villigen, Switzerland

Abstract

Dark current and multiple electron impacts (multipacting), as for example observed in radio frequency (RF) structures of accelerators, are usually harmful to the equipment and the beam quality. These effects need to be suppressed to guarantee efficient and stable operation. Large scale simulations can be used to understand causes and develop strategies to suppress these phenomenas.

We extend OPAL, a parallel framework for charged particle optics in accelerator structures and beam lines, with the necessary physics models to efficiently and precisely simulate multipacting phenomenas. We added a Fowler-Nordheim field emission model, two secondary electron emission models, developed by Furman-Pivi and Vaughan respectively, as well as efficient 3D boundary geometry handling capabilities. The models and their implementation are carefully benchmark against a non-stationary multipacting theory for the classic parallel plate geometry. A dedicated, parallel plate experiment is sketched.

INTRODUCTION

Dark current and multipacting phenomena have been observed in various accelerator RF structures, e.g. in electron guns, due to field emission caused by strong accelerating fields [1], and multipacting is also appearing in high-Q RF cavities of cyclotrons [2, 3]. These phenomena are usually harmful to the equipment and beam quality, as they will cause galvanic etching on the surface of the cavity and thus cause RF breakdown.

Multipacting in cyclotron cavities is a very disturbing phenomenon. The seed electrons will impact the cavity surface, and produce an avalanche of new electrons. Under certain conditions (material, geometry of the RF structure, frequency and level of the electromagnetic field), the secondary emission yield (SEY) coefficient will be larger than one and lead to exponential multiplication of electrons. This kind of discharge will limit the power level, until the surfaces are cleaned through a conditioning process. However, this process is very time-consuming [2, 3].

Large scale dark current and multipacting simulations based on reliable data of surface material, full size geometry model of RF structures and parallel computing capabilities, allow more thorough analysis and a deeper understanding of these phenomena, even in early design stages of RF structures.

To make OPAL [4] a feasible tool for performing large scale dark current and multipacting simulations, we implement a 3D particle-boundary collision test to minimise particle searching during the tracking process. In a subsequent step we add surface physics models including an analytic Fowler-Nordheim field emission model and two secondary emission models, developed by Furman-Pivi and Vaughan respectively. The above mentioned models and their implementation in OPAL have been benchmarked against a non-stationary theory [5]. A nano-second time resolved multipacting experiment is ongoing.

* cwang@ciae.ac.cn
† andreas.adelmann@psi.ch

MODELS

Geometry Handling

Testing particle-boundary collisions is crucial to both dark current and multipacting simulations. Since complex 3D geometries are hard to be accurately parameterized by simple functions, we use triangulated surfaces, which are extracted from a volume mesh generated by GMSH [6], to represent the complex geometries. Subsequently we can make use of efficient 3D line segment-triangle intersection (LSTI) tests to identify particle-boundary collisions.

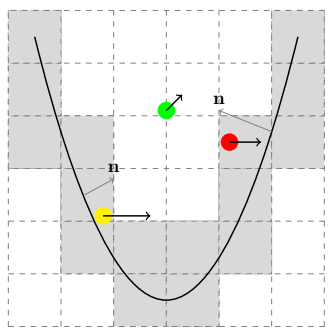


Figure 1: Schematic view of the particle-boundary early rejection strategy. The dark black line represents the boundary surface, particles are coloured dots with an attached momenta arrow. Gray arrows visualize inward normals of the boundary.

Our LSTI test algorithm is based on [7], and the detailed introduction to our implementation can be found in our previous paper [8]. Early rejection strategies skip particles far away from the boundary and with equal direction of momenta and boundary normal vectors (see Figure 1)

Surface Physics Models

Electron field emission is a major source of both dark current particles and primary incident particles in secondary emission. We employ the Fowler-Nordheim (F-N)

formula (1) to predict the emitted current density [9, 10]

$$J_{FN}(\mathbf{r},t) = \frac{A(\beta E)^2}{\varphi t(y)^2} \exp\left(\frac{-Bv(y)\varphi^{3/2}}{\beta E}\right) \quad (1)$$

where $J_{FN}(\mathbf{r},t)$ stands for emitted electric current density in position $\mathbf{r}$ and time t. The Greek letters φ and β denote the work function of the surface material and the local field enhancement factor respectively. The parameter E is the electric field in the normal direction of surface. The parameters A and B are empirical constants. The functions $v(y)$ and $t(y)$ representing the image charge effects [10] as a function of the Fowler-Nordheim parameter y with the following definition [1]

$$y = \sqrt{\frac{e^3}{4\pi\varepsilon}}\frac{\sqrt{\beta E}}{\varphi} = 3.795\times10^{-5}\frac{\sqrt{\beta E}}{\varphi}. \quad (2)$$

In our model, we have chosen simpler approximations $v_a(y)$ and $t_a(y)$ for $v(y)$ and $t(y)$ [1]:

$$\begin{aligned} v_a(y) &= a - by^2 \\ t_a(y) &\approx 1. \end{aligned}$$

where a and b are free parameters to fit the value of $v(y)$. These approximations are valid for a large range of y, corresponding to typical applied electric field ranges in RF guns.

Whenever the normal components of an electric field are strong enough the field emission current density will be limited by space charge effects [10]. To cover this situation we incorporated the 1D Child-Langmuir law

$$\begin{aligned} J_{SC}(\mathbf{r},t) &= \frac{4\varepsilon_0}{9}\sqrt{2\frac{e}{m}}\left(\frac{V^{3/2}}{d^2}\right) \\ &= \frac{4\varepsilon_0}{9}\sqrt{2\frac{e}{m}}\left(\frac{E^{3/2}}{d^{1/2}}\right) \end{aligned} \quad (3)$$

into our field emission model. $J_{SC}(\mathbf{r},t)$ denotes space charge limited emission current density in position $\mathbf{r}$ and time t, ε_0 the permittivity in vacuum, E the normal component of electric field on the surface and d the distance from the position where E is evaluated. Currently we choose d to be equal to the distance traveled by emitted particles in one time step, i.e., $d = \frac{eE\Delta t^2}{2m_0}$ where Δt is simulation time step. In each time step, the emitted current density $J(\mathbf{r},t)$ for each surface triangle will be the smaller one of $J_{FN}(\mathbf{r},t)$ and $J_{sc}(\mathbf{r},t)$,

$$J(\mathbf{r},t) = min\{J_{FN}(\mathbf{r},t), J_{sc}(\mathbf{r},t)\}. \quad (4)$$

We implemented two secondary emission models. The first one is a phenomenological model developed by M. A. Furman and M. Pivi [11], generating various generations of secondary electrons: *true secondary*, *rediffused* or *backscattered*.

The other secondary emission model is based on a secondary emission yield formula developed by Vaughan [12, 13, 14]:

$$\begin{aligned} \delta(E,\theta) &= \delta_0, \text{ for } v \le 0 && (5a) \\ \delta(E,\theta) &= \delta_{max}(\theta)\cdot(ve^{1-v})^k, \text{ for } v \le 3.6 && (5b) \\ \delta(E,\theta) &= \delta_{max}(\theta)\cdot 1.125/v^{0.35}, \text{ for } v > 3.6 && (5c) \end{aligned}$$

where

$$v = \frac{E - E_0}{E_{max}(\theta) - E_0},$$

$$k = 0.56, \text{ for } v < 1,$$

$$k = 0.25, \text{ for } 1 < v \le 3.6,$$

$$\delta_{max}(\theta) = \delta_{max}(0)\cdot(1 + k_\theta\theta^2/2\pi),$$

$$E_{max}(\theta) = E_{max}(0)\cdot(1 + k_E\theta^2/2\pi).$$

The secondary emission yield value for an impacting electron with energy E and incident angle θ w.r.t the surface normal is denoted as $\delta(E,\theta)$. Parameter k_θ and k_E denote the dependence on surface roughness. Both should be assigned a default value of 1.0, which appears appropriate for typical dull surfaces in a working tube environment. Lower values down to zero or higher values, up to about 2.0, are only valid for specific cases [12]. $E_{max}(0)$ is the impacting energy when the incident angle is zero and secondary yield reaches its maximum value. E_0 is an adjustable parameter to make the first crossover energy at which the secondary yield equals to 1 be fitted to the experiment data [14].

Implementation within OPAL

The above models are implemented in the object-oriented parallel ESPIC code OPAL[4].

Statistical data, such as particle position, momentum and particle type (primaries, field emitted electrons or true secondaries), are stored in the H5hut [15] file format. In a post processing step, the data can be converted into legacy VTK formatted files [16] and processed by standard visualisation tools.

Users can customise the dark current and multipacting models in the MAD like input file of OPAL. The input file format and examples can be found in OPAL user guide [4].

Motivated by the fact that the number of simulation particles may grow exponentially, spanning a range of ten orders of magnitude, a particle re-normalisation technique is implemented. In each electron impact event, instead of emitting the real number of simulation particles predicted by secondary emission models, this re-normalization approach emits only one particle and a scaled charge Q_s. Where $Q_s = Q_{incident}\times\delta$, the incident particle charge multiplied by the SEY value δ. This approach is an accurate representation of the secondary emission model which can be observed in the following parallel plate benchmarking cases.

BENCHMARKS

Benchmark Against the None-stationary Theory

The theory we use to benchmark the described models is restricted to the case of the simple plane-parallel model of multipactor. We consider a spatially homogeneous and time harmonic RF field in between and directed perpendicular to the plates i.e.,

$$\mathbf{E}(\mathbf{z},t) = -\hat{\mathbf{z}}E_0 \sin\omega t = -\hat{\mathbf{z}}\frac{V_0}{d}\sin\omega t \qquad (6)$$

as shown in Fig. 2. Electrons are assumed to oscillate between two parallel plates separated by a distance d. The x and y dimensions of the plates are assumed to be infinite. The impact of an electron with the plates is accompanied by a secondary emission yield, which depends on the energy and angle of the primary electron and material property of the multipactor.

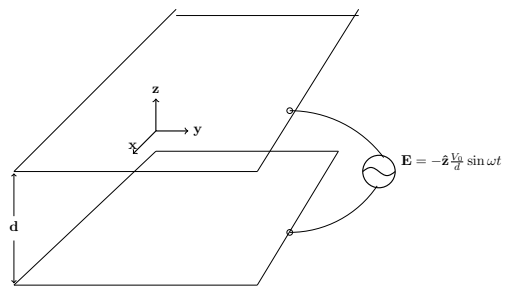


Figure 2: The geometry of the analytical model.

Assuming electrons are initially generated on the surface of the lower parallel plate, i.e., $z = 0$ at $t = t_0$ according to Fig. 2, then the equation of motion reads

$$\frac{d^2z}{dt^2} = -\frac{e}{m}E_0\sin\omega t = -\frac{e}{m}\frac{V_0}{d}\sin\omega t \qquad (7)$$

with V_0 being the peak voltage amplitude between the parallel plates.

Integrating the equation (7) w.r.t variable t, following [17], i.e., substituting the initial condition $\frac{dz}{dt}|_{t=t_0} = v_0$, $z|_{t=t_0} = 0$, and using normalised variables: $v_\omega = eV_0/m\omega d$, $\lambda = \omega d/v_\omega$, $u = v_0/v_\omega$, $\omega t_0 = \varphi_0$, we obtain the scaled velocity and absolute position of electrons before they impact as:

$$\begin{aligned} z = &-\frac{d}{\lambda}\sin\omega t + \frac{d}{\lambda}(u+\cos\varphi_0)\omega t \\ &+ \frac{d}{\lambda}\sin\varphi_0 - \frac{d}{\lambda}(u+\cos\varphi_0)\varphi_0. \end{aligned} \qquad (8)$$

If we define $\omega t = \varphi$, $\xi = \omega z/v_\omega$ and $\tau = \varphi - \varphi_0$, equation (8) can be rewritten as:

$$\xi(\varphi,\varphi_0,u) = (u+\cos\varphi_0)\tau + \sin\varphi_0 - \sin(\varphi_0+\tau). \quad (9)$$

A non-stationary statistical theory, originated from a stationary statistic theory [18], for multipactor introduced by Anza et al. [5], gives a more realistic scenario than previous multipacting theories, since it considers the random nature of the electron emission velocity and models both double and single surface impacts, as sketched in Fig. 3.

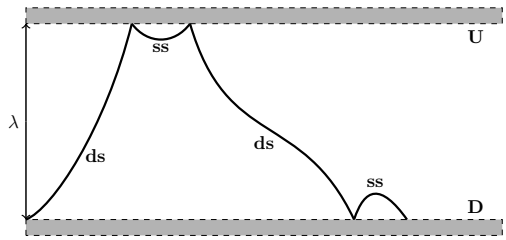


Figure 3: A full scenario of electron's trajectories between parallel plates, including both double surface (ds) and single surface (ss) impacts [5].

The basic idea of the non-stationary theory can be summarised as follows: the initial velocity u of emitted particles is a random variable, the solution of equation (9) with respect to time τ is the joint probability that an electron released at phase φ_s impacts at the opposite wall, separated by λ, in a transit time τ. As long as we know the probability density function (PDF) of the initial velocity u, which usually is a thermal distribution, then the joint PDF can be derived according to the rule of change of variable in probability theory.

Following the definitions in [5], each plate will be denoted as D and U for the boundary condition $\xi = 0$ (down) and $\xi = \lambda$ (up), respectively. Double and single surface impacts with $D-U$ or $U-D$ and $D-D$ or $U-U$ trajectories will be denoted as ds and ss respectively. Other definitions for the non-stationary theory are given in [5], and listed here in Table 1 for convenience.

Table 1: Non-stationary Theory Definitions

Impact rate (electrons/radian) in plate U/D at phase φ	$I_{U/D}(\varphi)$
Emission rate (electrons/radian) in plate U/D at phase φ	$C_{U/D}(\varphi)$
Number of electrons at time φ	$N(\varphi)$
Probability density that an electron starting at plate U/D, with starting phase φ, experiences a double/single Surface impact in a transit phase τ	$G_{ds/ss,U/D}(\tau\|\varphi)$
Secondary emission yield of an electron starting at plate U/D, with starting phase φ, experiences a double/single surface impact in a transit phase τ	$\delta_{ds/ss,U/D}(\tau\|\varphi)$

The electron emission and impact rates in each plate can be described by the joint PDF and the secondary emission yield coefficient, both of which are functions of initial velocity u, initial phase φ_s and time τ at which particles hits

the plates. Details of constructing the joint PDF and integrating the emission rates and impact rates can be found in [5]. The number of electrons between the parallel plates at phase φ then can be integrated as:

$$N(\varphi) = \int_0^{\varphi} (C_U(\varphi') + C_D(\varphi') - I_U(\varphi') - I_D(\varphi'))\,\mathrm{d}\varphi'. \tag{10}$$

We use the secondary emission yield curve of copper provided in [11] to benchmark the Furman-Pivi model and use the secondary emission yield curve of silver given in [5] to benchmark Vaughan's secondary emission model.

The initial particles are equally distributed between both plates. The velocity of emitted particles both in the theory and in the benchmark simulation follows a Maxwellian distribution [5]. We have chosen different surface material SEY curves, voltages, gap distances, frequencies to benchmark against both Furman-Pivi's model and Vaughan's model. Fristly we using $f = 200$ MHz, $V_0 = 120$ V, $d = 5$ mm and the material of the multipactor is copper. The results with Furman and Pivi's secondary emission model matches the theoretical model very well, as shown in Fig. 4.

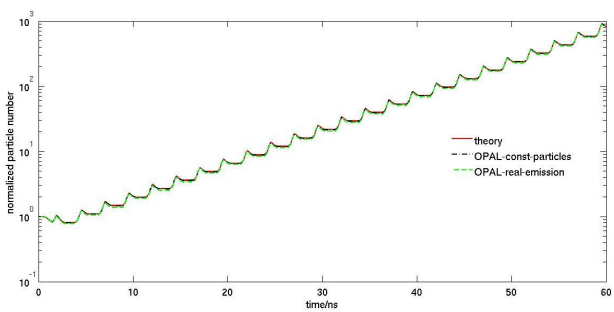


Figure 4: Time evolution of the electron density predicted by the theoretical model and the OPAL simulation. Furman-Pivi's secondary emission model with and without re-normalization is used. Model parameter are: $f = 200$ MHz, $V_0 = 120$ V, $d = 5$ mm.

The Vaughan's model has been benchmarked with silver SEY data. Again very good agreement of the model and simulation can be observed in Fig. 5.

APPLICATIONS

Multipacting phenomenas of the CYCIAE-100 H^- AVF cyclotron under construction at the China Institute of Atomic Energy (CIAE) [19] are investigated.

We have chosen two different SEY curves of copper, with and without surface treatment, both in agreement with equation (5).

Figure 6 shows the time evolution of the electron density. Multipacting has been observed in both cases, with and without surface treatment within one RF cycle. The

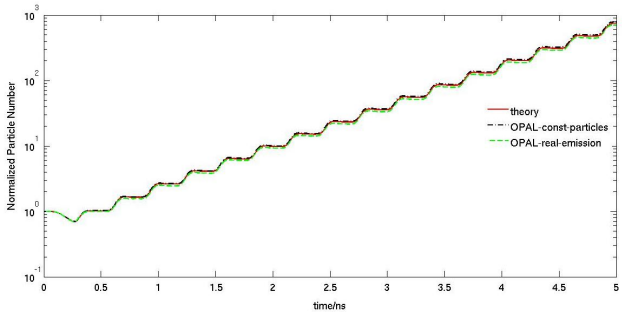


Figure 5: Time evolution of the electron density predicted by the theoretical model and the OPAL simulation. Vaughan's secondary emission model with and without re-normalization is used. Model parameter are: $f = 1640$ MHz, $V_0 = 120$ V, $d = 1$ mm.

electron multiplication without surface treatment is 5 orders of magnitude larger than in the case with surface treatment.

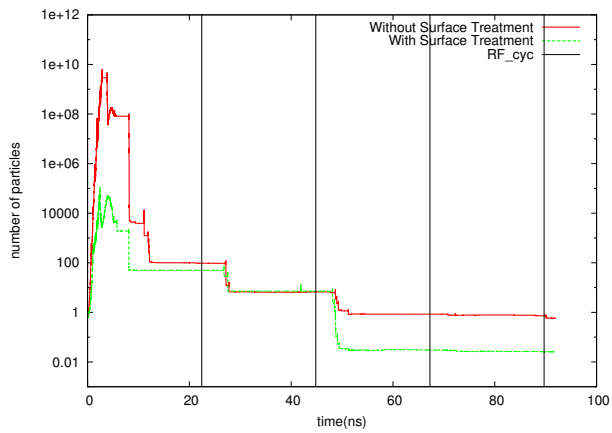


Figure 6: The time evolution of particle numbers in both with and without surface treatment cases, the phase lag of both cases are 0 degree.

For visualising the positions where multipacting happens, OPAL will dump the impact position and current of incident particles into a specified file using the efficient parallel H5hut file format [15]. The simulated hot spot of the RF cavity of CYCIAE-100 cyclotron is shown in Figure 7.

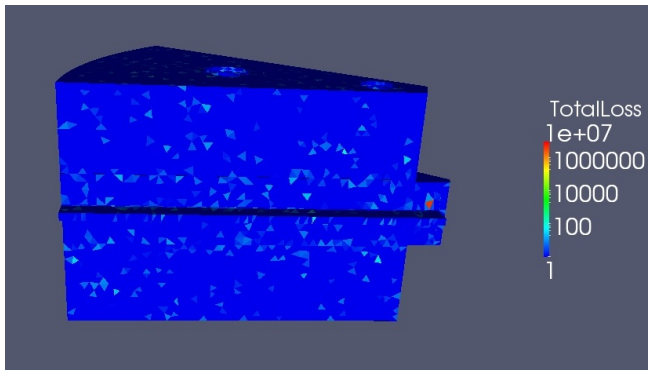


Figure 7: Simulated hot spot of RF cavity of CYCIAE-100 cyclotron.

The Dark current module of OPAL has also been used to study the Dark current of the CTF3 gun in SwissXFEL project [21].

CONCLUSION AND OUTLOOK

RF components with arbitrary, closed structures, like electron guns and cyclotron resonators can be modelled in OPAL to numerically study dark current and multipacting phenomena. Detailed benchmarks against the non-stationary multipacting theory, enables predictive numerical studies of multipacting phenomena. Post processing utilities, like the hot spot visualisation and time evolution of particle distribution, are useful to locate the multipacting zones. Further multipacting simulations are planed for the CYCIAE-100 cyclotron cavity on different RF power levels to better understand the multipacting behaviour of the cavity during the RF conditioning process. The aim is to shorten the time consuming RF conditioning process, by localised surface treatment (painting) predicted by OPAL simulations.

Time Resolved Multipacting Experiment

A dedicated nano-second time resolved multipacting experiment is ongoing to benchmark the models. This dedicated experiment is using a $\lambda/4$ transmission line resonator, which works around 73 MHz. A sinusoidal electrical field will be established between parallel plates, in the conductive end of the resonator. An electron pickup is mounted through a hole in the centre of the ground plate to collect the multipacting electrons. It is expected that the collected multipacting current will have the same pattern as predicted by the none-stationary theory and the OPAL model. The experiment is still ongoing, detailed results will be published in a forthcoming paper.

ACKNOWLEDGMENTS

The authors thank the Accelerator Modelling and Advanced Simulation (AMAS) and the members of the BRIF department of China Institute of Atomic Energy (CIAE) for many fruitful discussions. This work was performed on the *felsim* cluster at the Paul Scherrer Institut and the PANDA cluster at CIAE.

REFERENCES

[1] J.H. Han, "Dynamics of Electron Beam and Dark Current in Photocathode RF Guns," PhD thesis, Universität Hamburg, Germany, 2005, http://www-library.desy.de/preparch/desy/thesis/desy-thesis-05-045.pdf

[2] P.K. Sigg, "Reliability of High Beam Power Cyclotron RF-Systems at PSI," Workshop on Utilization and Reliability of High Power Proton Accelerators, Mito, Japan, 1998, http://rf.web.psi.ch/files/proceedings/1998/JAERI98/PaperNEA98.pdf

[3] N. Sakamoto et al., "RF-System for the RIBF Superconducting Ring Cyclotron," CYCLOTRONS'2007, Giardini Naxos, Italy, 2007.

[4] A. Adelmann et al., "The OPAL (Object Oriented Parallel Accelerator Library) Framework," Paul Scherrer Institut, PSI-PR-08-02, 2008.

[5] S. Anza et al., Phys. Plasmas. 17 (2010) 6 062110.

[6] C. Geuzaine and J.F. Remacle, International Journal for Numerical Methods in Engineering. 79 (2009) 11 1309.

[7] D. Sunday, http://softsurfer.com/Archive/algorithm_0105/algorithm_0105.htm.

[8] C. Wang et al., "A Field Emission and Secondary Emission Model in OPAL," HB2010, Morschach, Switzerland, 2010, MOPD55.

[9] R.H. Fowler and L. Nordheim, Royal Society of London Proceedings Series A. 119 (1928) 173-181.

[10] Y. Feng and J.P. Verboncoeur, Phys. Plasmas. 13 (2006) 7 073105.

[11] M.A. Furman and M.T.F. Pivi, Phys. Rev. ST Accel. Beams. 5 (2002) 12 124404.

[12] J.R.M. Vaughan, IEEE Transactions on Electron Devices. 36 (1989) 9 1963.

[13] J.R.M. Vaughan, IEEE Transactions on Electron Devices. 40 (1993) 4 830.

[14] C. Vicente et al., "FEST3D - A Simulation Tool for Multipactor Prediction," Proc. of MULCOPIM 2005, Noordwijk, the Netherlands, 2005.

[15] A. Adelmann et al., http://h5part.web.psi.ch/.

[16] "File Formats for VTK Version 4.2," http://www.vtk.org/VTK/img/file-formats.pdf.

[17] A. Kryazhev et al., Phys. Plasma. 9 (2002) 11 4736.

[18] N.K. Vdovicheva et al., Radiophysics and Quantum Electronics. 47 (2004) 8 580.

[19] T.J. Zhang et al., NIM B. 266 (2008) 19 4117-4122.

[20] I. Bojko et al., "The Influence of Air Exposures and Thermal Treatments on the Secondary Electron Yield of Copper," LHC Project Report, CERN, 376 (2000).

[21] F.L. Pimpec et al., "Dark Current Studies For SWISSFEL," arXiv:1205.3098v1 [physics.acc-ph] May 2012, http://arxiv.org/abs/1205.3098v1.

EFFICIENT MODELING OF LASER-PLASMA ACCELERATORS USING THE PONDEROMOTIVE-BASED CODE INF&RNO*

C. Benedetti†, C.B. Schroeder, E. Esarey, W.P. Leemans, LBNL, Berkeley, CA 94720, USA

Abstract

The ponderomotive force approximation enables efficient modeling of laser-plasma accelerators. It allows simulation in cylindrical geometry which captures relevant 3D physics at 2D computational cost. INF&RNO (INtegrated Fluid & paRticle simulatioN cOde) is an efficient 2D cylindrical code based on the envelope model for the laser, a PIC or fluid description for the plasma, and the ponderomotive force approximation to describe the effect of the laser pulse on the plasma. These and other features, such as an improved laser envelope solver, a dynamical resampling of the phase space distribution to reduce on-axis noise, and a Lorentz boosted frame modeling capability, allow for a speedup of several orders of magnitude compared to standard (explicit) full PIC simulations while still retaining physical fidelity. The code has been benchmarked against analytical solutions and 3D PIC simulations. In this paper we report on the latest developments of the code, focusing in particular on the improved laser envelope solver, and we discuss its performance.

INTRODUCTION

Numerical modeling of a laser-plasma accelerator (LPA) [1], where a short and intense laser pulse interacts with an underdense plasma over distances ranging from a few millimeters/centimeters (yielding ~ 0.1/1 GeV electron energy [2, 3]) up to several tens of centimeters (as in the BErkeley Lab Laser Accelerator experiment (BELLA) [4] where $\sim$ 10 GeV electrons are expected), is a computationally challenging task. A 3D "full" (*i.e.*, where the fast oscillations of the plasma electrons in the laser field are taken into account) PIC simulation requires $10^4 - 10^5$ CPU hours using today's supercomputers for a millimeter-scale plasma and $\sim 10^6$ CPU hours for a centimeter-scale plasma. The simulation of a ~ 10 GeV stage as required by BELLA would necessitate several tens of millions of CPU hours and are practically impossible to perform with standard simulation tools and today's computational resources. However, numerical modeling plays a central role in the understanding and optimization of LPAs. Simulations are required since the physics of the laser-plasma interaction is highly nonlinear and, consequently, analytical solutions are lacking.

Two approaches have been proposed to overcome these limitations and allow for the simulation of multi-GeV LPA stages: *i.* use reduced models [5, 6, 7]; *ii.* run the simulation in an optimal Lorentz boosted frame (LBF) [8] instead of in the laboratory frame. Codes based on reduced models allow for a significant speedup compared to full PIC simulations either because of dimensionality reduction (*e.g.*, 2D cylindrical instead of full 3D cartesian) or because of approximations in the physical description of the system (*e.g.*, quasi-static instead of fully dynamic plasma response, ponderomotive approximation instead of full Lorenz force, etc.). Even if they may lack important elements of the physics (*e.g.*, a quasi-static code can not describe self-injection), their use has been proven to be successful in several relevant scenarios [5, 9, 10, 11]. The use of an LBF has been strongly pursued by several groups [12, 13, 14, 15]. The advantage of running a simulation in an LBF relies on the fact that, if backward propagating waves (*e.g.*, Raman backscattering) can be neglected, and this is generally true for LPAs, then it has been shown [8] that the unbalance between the maximum and minimum physical scales involved in a simulation, which contribute to set the computational complexity of the problem, is not invariant under Lorentz transformation. It turns out that, in general, the laboratory frame is not the optimal choice to run the simulation, while running it in a boosted frame can considerably reduce the scale unbalance, shortening (also by several orders of magnitude) the simulation length.

The INF&RNO computational framework [6, 16], is a 2D cylindrical $(r - z)$ code that adopts an envelope model for the laser pulse and makes use of the (time-averaged) ponderomotive force approximation to describe the interaction of the laser pulse with the plasma. The plasma can be modeled using either a PIC or a fluid description. Both PIC and fluid modalities are integrated in the same computational framework allowing for staged simulations (*e.g.*, PIC for injection and fluid for acceleration). The adoption of the cylindrical geometry allows the description of 3D physics (laser evolution, electromagnetic field structure) at 2D computational cost. The code features an improved laser envelope solver which enables an accurate description of the laser pulse evolution deep into depletion even at a reasonably low resolution. This new algorithm overcomes some of the limitations reported in other implementations of the laser envelope solver [7]. For the PIC part, a dynamical resampling of the phase space distribution is implemented in order to reduce the on-axis noise which affects, in some cases, 2D axisymmetric codes when the computational grid is loaded with a constant number of particles per cell. Finally, an LBF modeling capability has been introduced within the (noiseless) fluid framework. The employment of the LBF by the user is transparent since

* Work supported by the Office of Science, Office of High Energy Physics, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

† cbenedetti@lbl.gov

a set of "wrapper" routines take care of all the necessary data transformations between the laboratory frame and the LBF, where the simulation is performed, during initialization and output dumps. The code has been validated and benchmarked against analytical solutions and other codes (*e.g.*, fully 3D PIC [13]). Benchmark results can be found in [6]. In this paper we provide a general overview of the features and performance of the code. With INF&RNO, detailed simulations of a ∼10 GeV LPA stage becomes feasible in a few hours or days, depending on the particular simulation settings, on small clusters with (at most) a few hundred CPUs.

OVERVIEW OF THE CODE

Physical Model and Basic Equations

The code INF&RNO works in 2D cylindrical $(r-z)$ geometry and adopts non-dimensional comoving variables defined as $\xi = k_p(z-ct)$ (longitudinal) and $\rho = k_p r$ (transverse), where $k_p = \omega_p/c$, ω_p is the plasma frequency corresponding to the chosen reference density n_0, and c is the speed of light. The time is also rescaled with $1/\omega_p$, that is $\tau = \omega_p t$. Working in 2D cylindrical geometry, unlike the 2D Cartesian, has the advantage that laser evolution (*e.g.*, self-focusing and diffraction) or wakefield structure are the same as in 3D, if non-axisymmetric features can be neglected, but the computational cost for the simulations is significantly lower (approximately 2 orders of magnitudes) compared to 3D runs.

Denoting by $a_\perp = eA_\perp/mc^2$ the normalized vector potential of the laser, the corresponding laser envelope $\hat{a}$ is defined by $a_\perp = \frac{\hat{a}(\xi,\rho)}{2} e^{i(k_0/k_p)\xi} + c.c.$ The envelope evolves according to [1]

$$\left(\nabla_\perp^2 + 2i\frac{k_0}{k_p}\frac{\partial}{\partial\tau} + 2\frac{\partial^2}{\partial\xi\partial\tau} - \frac{\partial^2}{\partial\tau^2}\right)\hat{a} = \frac{\delta}{\gamma_{\text{fluid}}}\hat{a}\,, \quad (1)$$

where $2\pi/k_0$ is the central laser wavelength, $\delta = n/n_0$ is the (normalized) electron plasma density and γ_{fluid} is the relativistic factor associated with the local plasma fluid velocity. We notice that in the derivation of Eq. 1 all high-frequency contributions from the plasma have been neglected [1].

The fully time-explicit electromagnetic wakefield generated behind the laser is described by the fields E_z, E_r, B_ϕ (normalized to $E_0 = mc\omega_p/e$, where m and e are respectively mass and charge of the electron). The wakefield evolves according to Ampère-Maxwell laws which read

$$\begin{aligned}
\frac{\partial E_z}{\partial\tau} &= \frac{\partial E_z}{\partial\xi} + \frac{1}{\rho}\frac{\partial(\rho B_\phi)}{\partial\rho} - J_z,\\
\frac{\partial E_r}{\partial\tau} &= \frac{\partial(E_r - B_\phi)}{\partial\xi} - J_r, \quad (2)\\
\frac{\partial B_\phi}{\partial\tau} &= -\frac{\partial(E_r - B_\phi)}{\partial\xi} + \frac{\partial E_z}{\partial\rho},
\end{aligned}$$

where (J_z, J_r) are the components of the (normalized) current density. The background plasma can be modeled using either a PIC or a fluid description and the laser-plasma coupling is described via the ponderomotive approximation [1]. For the PIC description the equations are

$$\begin{cases}
\frac{d\xi_j}{d\tau} = \frac{u_{z,j}}{\gamma_j} - 1 \equiv \beta_{z,j} - 1\\
\frac{d\rho_j}{d\tau} = \frac{u_{r,j}}{\gamma_j} \equiv \beta_{r,j}\\
\frac{du_{z,j}}{d\tau} = -\frac{\partial\gamma_j}{\partial\xi}|_j - E_{z,j} - \beta_{r,j}B_{\phi,j}\\
\frac{du_{r,j}}{d\tau} = -\frac{\partial\gamma_j}{\partial\rho}|_j - E_{r,j} + \beta_{z,j}B_{\phi,j}\\
\gamma_j \equiv (1 + |\hat{a}|^2/2 + u_{z,j}^2 + u_{r,j}^2)^{1/2}\,,
\end{cases} \quad (3)$$

where $(\xi_j, \rho_j, u_{z,j}, u_{r,j})$ are the phase-space coordinates (position and normalized momentum) of the $j-$th numerical particle representing one of the characteristics of the Vlasov equation for the plasma. A PIC description is also adopted to model externally injected particle beams. In the fluid description for the plasma, the electron density, δ, and momentum, $\mathbf{u} = (u_z, u_r)$, evolve according to

$$\begin{cases}
\frac{\partial\delta}{\partial\tau} = \frac{\partial\delta}{\partial\xi} - \nabla\cdot\left(\frac{\mathbf{u}}{\gamma_{\text{fluid}}}\delta\right)\\
\frac{\partial(\delta u_j)}{\partial\tau} = \frac{\partial(\delta u_j)}{\partial\xi} - \nabla\cdot\left(\vec{\beta}\delta u_j\right)\\
\quad + \delta\left[-(\mathbf{E} + \frac{\mathbf{u}}{\gamma_{\text{fluid}}}\times\mathbf{B}) - \frac{1}{2\gamma_{\text{fluid}}}\nabla\frac{|\hat{a}|^2}{2}\right]_j\,,\ j = z, r\\
\gamma_{\text{fluid}} \equiv \sqrt{1 + |\hat{a}|^2/2 + u_z^2 + u_r^2}\,.
\end{cases} \quad (4)$$

The PIC and fluid modalities are integrated in the same computational framework, enabling an easy switch from one description to the other (combined simulations).

Recently, besides the time-explicit modality, a quasi-static, linear, fluid modality has been included in the INF&RNO framework. The quasi-static approximation assumes $\partial/\partial\tau = 0$ in the equations for the wakefield and the plasma, explicit time evolution is only retained in the equations for the driver (particle or laser beam). This modality will allow for fast simulations of the interaction of a (long) particle beam in an overdense plasma [*i.e.*, where the beam density is (much) smaller than the background plasma density] [17]. Finally, the code also features a 1D fluid modality (which can be either time-explicit or quasi-static) that can be used for quick/preliminary parameter scan.

Numerical Aspects and Features of the Code

In INF&RNO all the fields are discretized into the same 2D mesh (no staggering is adopted). Longitudinal derivatives are computed using a second-order finite difference upwind scheme, namely $(\partial_\xi f)_{i,j} = (-3f_{i,j} + 4f_{i+1,j} - f_{i+2,j})/(2\Delta_\xi)$, where $f_{i,j}$ is the field value at the (i,j) node and Δ_ξ the longitudinal cell size. Radial derivatives are computed using a standard centered second-order accurate scheme. No singularity exists at the $r = 0$ boundary, and from symmetry properties we have, for instance, $\partial_\rho E_z|_{\rho=0} = E_r|_{\rho=0} = B_\phi|_{\rho=0} = 0$ and $\lim_{\rho\to 0} B_\phi/\rho = \partial B_\phi/\partial\rho|_{\rho=0}$. Second and fourth order Runge-Kutta integrators (RK2/RK4) are available for fluid plasma quantities and wakefield evolution while plasma particles and externally injected bunches can be pushed with either RK4 or the standard Boris pusher [18].

The laser envelope is updated in time solving Eq. 1 using a second order Crank-Nicolson scheme [16]. More details on the laser envelope solver are given in the next section.

In the PIC modality force interpolation, charge and current deposition are performed using quadratic shape functions. Compact low-pass digital filters [19] and standard binomial filters (with compensator) are available for field and current smoothing. The user has large freedom in loading numerical particles over the computational domain (the numerical particle distribution is controlled by a simple user-defined routine) and this freedom can be used to selectively provide a better sampling of the plasma phase space distribution within the dynamically interesting zones without greatly increasing the overall number of simulated particles [20]. When the computational grid is loaded with a constant number of particles per cell, particles loaded at large radii, because of the cylindrical symmetry, carry generally more charge than particles loaded on-axis. If/when these "heavier" particles approach the $r = 0$ axis, they may induce "spikes" in density and currents, increasing the noise level in the fields. These detrimental effects can be partially mitigated via dynamical particle splitting of the high charge particles approaching the axis. Details on the implementation of this technique to resample the phase space distribution are discussed in [6].

An LBF modeling capability is available within the INF&RNO/fluid framework. The extension of this capability to the INF&RNO/PIC modality is currently underway. In order to enable LBF simulations in a code based on a laser envelope model, the evolution equation for the laser envelope has to be Lorentz invariant. In the LHS of Eq. 1 we retain the second order time derivative (full wave operator), ensuring exact invariance of this equation under Lorentz transformation [16]. This simulation modality is transparent for the user which only has to set the velocity of the boosted frame (*i.e.*, γ_{LBF}) and initialize the system as he/she would be in the standard laboratory frame (LF). A set of "wrapper" functions automatically perform all the necessary operations to set properly the physical parameter [laser, plasma, externally injected bunch(es), *etc.*] and the numerical ones (grid settings) in the LBF. The swiping plane technique [15] is used to initialize the simulation in the LBF and to reconstruct the output data in the LF at a fixed laboratory time. In all the tests performed to validate the implementation of the LBF in the fluid framework, no evidence of the numerical instability which usually affects PIC simulations in the LBF when γ_{LBF} is large has been observed [21]. The theoretical speedup for an LPA simulation in the LBF and the optimal boost velocity ($\gamma_{\text{LBF}}^{opt}$) have been computed in [15]. It is found that, for a typical LPA experiment, $\gamma_{\text{LBF}}^{opt} \simeq \gamma_w$, where γ_w is the relativistic factor of the wake traveling behind the laser driver. A $\sim$10 GeV stage requires $n_0 \sim 10^{17}$ e/cm^3, which implies $\gamma_{\text{LBF}}^{opt} \simeq \gamma_w \sim 100$. However, in our case, the maximum value for γ_{LBF} (and so the maximum speedup) is more likely to be limited by the validity of Eq. 1 in the LBF. In fact, the envelope approximation relies on a time scale separation between the fast laser oscillations and the slowly evolving envelope. Due to the redshift of the laser light in the boosted frame, the fulfillment of the time scale separation condition becomes more difficult to achieve the higher is γ_{LBF}. We found that, for the $\sim$ 10 GeV stage, the maximum acceptable value for γ_{LBF} is around $30 - 40$. For higher values of γ_{LBF} the simulation results start to deviate significantly from those obtained in the LF.

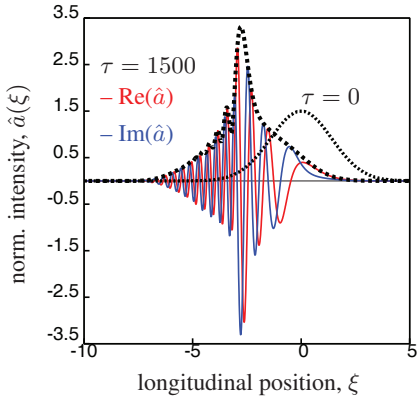


Figure 1: Evolution of the laser envelope in a 1D simulation with parameters $a_0 = 1.5, k_0/k_p = 20, L_{\text{rms}} = 1$ $[a^2(\xi, \tau = 0) = a_0^2 \exp(-\xi^2/2L_{\text{rms}}^2)]$. The dashed black lines refer to $\tau = 0$ (right) and $\tau = 1500$ (left). For $\tau = 1500$ then real (red) and imaginary (blue) part of $\hat{a}$ are also shown.

The code has been parallelized with MPI. Parallelization is achieved through 1D longitudinal domain decomposition. Studies to improve the parallel scalability are currently under consideration.

IMPROVED LASER ENVELOPE SOLVER

In a code based on a laser envelope model that adopts the ponderomotive approximation to describe the laser-matter interaction, there is no need, in principle, to resolve the laser wavelength and usually, for a resonant pulse [1], the characteristic length of the pulse (*e.g.*, the the r.m.s pulse length, L_{rms}) is the smallest relevant scale of interest. However, during propagation in the plasma, as a consequence of laser-pulse redshifting, structures smaller than L_{rms} arise in the laser envelope [22, 7, 23]. In Fig. 1, we show, as an example, the evolution of the laser envelope from $\tau = 0$ to $\tau = 1500$ in a 1D simulation with $a_0 = 1.5, k_0/k_p = 20, L_{\text{rms}} = 1$ [we assume, initially, a laser pulse of the form $a^2(\xi, \tau = 0) = a_0^2 \exp(-\xi^2/2L_{\text{rms}}^2)$]. The red and blue lines in Fig. 1 are, respectively, the real ($\Re[\hat{a}]$) and imaginary ($\Im[\hat{a}]$) part of the laser envelope at $\tau = 1500$. The presence, at later times, of structures smaller than L_{rms} (in the form of oscillations) in $\Re[\hat{a}]$ and $\Im[\hat{a}]$ (but also, to a minor extent, in the laser envelope $|\hat{a}|$) is evident. This has to be taken into account when designing a numerical scheme for the laser envelope evolution equation, or when choosing the (longitudinal) resolution to be used in the simulation.

In INF&RNO, the evolution equation for the laser envelope, Eq. 1, is discretized in time using a Crank-Nicholson scheme which reads

$$\begin{aligned} & - \quad \frac{\hat{a}^{n+1} - 2\hat{a}^{n} + \hat{a}^{n-1}}{\Delta_\tau^2} \\ & + \quad 2\left[i\frac{k_0}{k_p} + \frac{\partial}{\partial \xi} \right] \left(\frac{\hat{a}^{n+1} - \hat{a}^{n-1}}{2\Delta_\tau} \right) = \\ & - \quad \nabla_\perp^2 \left(\frac{\hat{a}^{n+1} + \hat{a}^{n-1}}{2} \right) + \frac{\delta^n}{\gamma^n_{\text{fluid}}} \frac{\hat{a}^{n+1} + \hat{a}^{n-1}}{2} , \end{aligned} \tag{5}$$

where $\hat{a}^k = \hat{a}(\xi, \rho; k\Delta_\tau)$. By using Eq. 5, knowing $\hat{a}^n$ and $\hat{a}^{n-1}$, we can evaluate $\hat{a}^{n+1}$.The first term in Eq. 5 corresponds to the (discretized) second order time derivative ($\partial^2/\partial\tau^2$) on the LHS of Eq. 1. Including this term, besides providing exact Lorentz invariance of the wave operator as required by the LBF modeling capability, it allows also for a correct description of backwards propagating waves and a better modeling of strongly depleted stages [23]. As for the spatial discretization, the transverse Laplacian operator, $\nabla_\perp^2$, takes the usual discrete form (for cylindrical coordinates) described in [18]. Extra care has to be put in the discrete representation of the longitudinal derivative $\partial/\partial\xi$. In fact, when the small structures in $\hat{a}$ develop and are not well resolved (this is the case for depleted laser pulses if the longitudinal resolution is not sufficiently high), a not optimal discrete form of the operator $\partial/\partial\xi$ might introduce significant numerical errors and prevent a correct description of the laser evolution. This issue is addressed in INF&RNO by means of a technique that involves the polar form for the representation of the complex field $\hat{a}(\xi)$, namely $\hat{a}^{(\text{polar})}(\xi) = a(\xi)\exp[i\theta(\xi)]$, where $a = |\hat{a}| = (\Re[\hat{a}]^2 + \Im[\hat{a}]^2)^{1/2}$ and $\theta = \arg(\hat{a})$, instead of the standard Cartesian representation. The polar amplitude and phase $a(\xi)$ and $\theta(\xi)$, are reasonably well behaved and less prone to show an oscillatory behavior or significant variability over small scales compared to $\Re[\hat{a}]$ or $\Im[\hat{a}]$. Evaluating the longitudinal derivative of the laser envelope field using the polar form has then some numerical advantage. More specifically, starting from the identity $\hat{a} = \hat{a}\, e^{-i\theta} e^{i\theta}$, and applying the $\partial/\partial\xi$ operator, we get

$$\frac{\partial \hat{a}}{\partial \xi} = \frac{\partial}{\partial \xi}\left(\hat{a}\, e^{-i\theta} e^{i\theta} \right) = \frac{\partial}{\partial \xi}\left(\hat{a} e^{-i\theta} \right) e^{i\theta} + i\hat{a}\frac{\partial \theta}{\partial \xi} , \tag{6}$$

where both $ae^{-i\theta} \equiv a(\xi)$ and $\theta(\xi)$ are, as previously described, reasonably smooth function, so we do not expect the numerical differentiation to introduce a significant error in the evaluation of their derivatives. We notice that in Eq. 5 the operator $\partial/\partial\xi$ acts on $\hat{a}^{n-1}$ and $\hat{a}^{n+1}$ (the unknown). To simplify the algorithm (retrieving the phase is a time consuming operation) we use the phase of $\hat{a}^n$ in evaluating both $\partial\hat{a}^{n+1}/\partial\xi$ and $\partial\hat{a}^{n-1}/\partial\xi$. The error associated with this approximation is small in all relevant cases analyzed.

An example (in 1D) of the performance of the laser envelope solver implemented in INF&RNO is shown in Fig. 2. In the simulation presented we consider the propagation of a short and intense laser pulse $a_0 = 1$, $L_{\text{rms}} = 1$ in a uniform plasma such that $k_0/k_p = 100$ (parameters of interest for a $\sim$ 10 GeV LPA stage). In Fig. 2 (a) we show the behavior of the laser energy as a function of the propagation distance (normalized to the pump depletion length which, for a resonant pulse, can be estimated as $L_{pd} \sim \lambda_p^3/\lambda_0^2 \sim 80$ cm, λ_p and λ_0 being respectively the plasma wavelength and the central laser wavelength) computed with different numerical schemes. The red dashed line is the result obtained with the full PIC code ALaDyn at high resolution, and can be considered "exact" for practical purposes. The black lines refer to results obtained with envelope code *without* the polar representation for the envelope field ("standard" solver). Different resolutions are considered, namely $L_{\text{rms}}/\Delta_\xi = 30, 100, 1000$ (see figure caption for detail). We notice that only at very high resolution the envelope results converge to the PIC result. The blue dots refer to the result computed with the envelope solver implemented in INF&RNO. Even at moderately low resolution, $L_{\text{rms}}/\Delta_\xi = 30$, the agreement with the PIC result is excellent also for a depleted laser pulse. Doubling the longitudinal resolution in this case (plot not shown) provides results in agreement with the PIC down to $L_{\text{propag}}/L_{pd} \simeq 1$. In Fig. 2 (b) we show lineouts of the longitudinal wakefield, E_z, after a propagation distance corresponding to 80% of the pump depletion length for the same numerical schemes presented before. Also in this case results obtained with the INF&RNO laser envelope model at (moderately) low resolution are in excellent agreement with the PIC simulation while, for the same resolution, results obtained with the "standard" solver show a significant (unacceptable) damping of the wakefield.

DISCUSSION

The computational framework INF&RNO is a 2D cylindrical, envelope, ponderomotive, PIC/fluid code designed to efficiently model LPAs. It features, among the others, an improved laser envelope solver, a technique to dynamically resample the phase space distribution aimed at reducing on-axis noise in the fields, and an LBF modeling capability. Compared to standard (explicit) full 3D PIC simulations it allows for a speedup of several orders of magnitude (between 2 and 5, depending on the particular problem and numerical setting) in the calculation time. In [16] it has been shown that an INF&RNO/fluid simulation of a 10 GeV stage in the quasi-linear regime, as in the BELLA experiment, requires only $\sim$ 30 CPU hours using the LBF modality with $\gamma_{\text{LBF}} = 12$. This corresponds to a speedup of approximately 5 orders of magnitude in the simulation time compared to standard simulation tools. As a second example we consider an INF&RNO/PIC simulation (in the laboratory frame) of a $\sim$ 10 GeV stage in the bubble regime where we model both self-injection and acceleration of the electron beam. Compared to the quasi-linear design the acceleration distance is shorter in this case ($\sim$ 10 cm for the bubble regime, $\sim$ 80 cm for the quasi-linear regime) but the

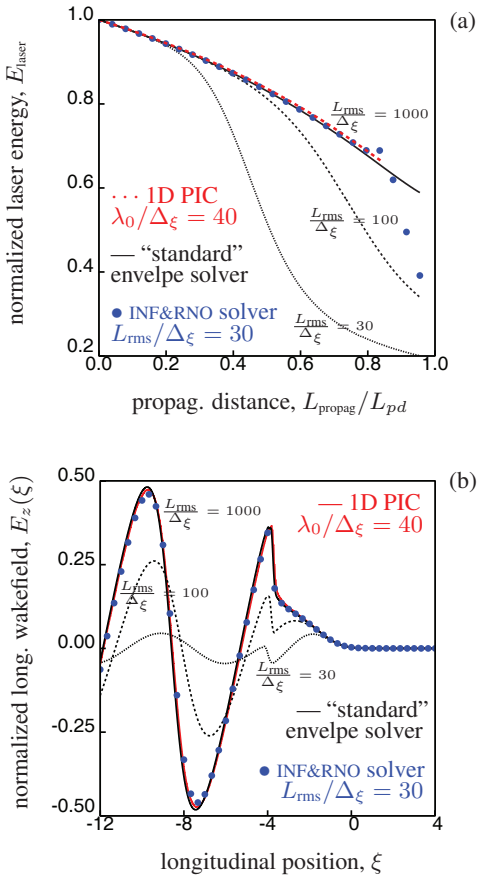


Figure 2: (a) Laser energy evolution as a function of the propagation distance normalized to the pump depletion length. (b) Lineout of the longitudinal wakefield, E_z, after a propagation distance corresponding to 80% of the pump depletion length. the simulation parameters are $a_0 = 1$, $L_{\rm rms} = 1$ and $k_0/k_p = 100$. In both (a) and (b) the red dashed line is the result obtained with the full PIC code ALaDyn. The black lines are the results obtained with the "standard" envelope solver at different resolution, namely $L_{\rm rms}/\Delta_\xi = 30, 100, 1000$. The blue points are the result obtained with the INF&RNO envelope solver for $L_{\rm rms}/\Delta_\xi = 30$.

background density is slightly higher ($\sim 3\times10^{17}$ e/cm^3 for the bubble regime, $\sim 1\times10^{17}$ e/cm^3 for the quasi-linear regime) and also the accelerating fields are higher (because of the higher background density and laser intensity) so the final beam energy is approximately the same. In order to correctly describe/represent the self-injection process and the nonlinear field structure of the bubble wake an adequate number of particles per cell and mesh resolution have to be adopted. In our test we chose $\Delta_\xi = 1/150$, $\Delta_r = 1/30$, and $\Delta_\tau/\Delta_\xi = 0.24$ with 12 particles per cell in the inner part of the computational domain (which includes the collection region for the trapped particles) and 2 particles per cell in the outer part. The size (normalized to k_p) of the computational box was ~ 20 (longitudinal) $\times$ 25 (radial). The simulation ran for ~ 5 days on 800 CPUs and the total cost was then $\sim 10^5$ CPU hours which corresponds to a speedup of about 2 orders of magnitude compared to standard tools. We notice that, even without an LBF modeling capability, accurate simulations of 10 GeV stages become feasible in a few days on relatively small computer clusters using INF&RNO.

ACKNOWLEDGEMENTS

The authors thank F. Rossi, L. Yu, J.-L. Vay, C.G.R. Geddes, and S. Rykovanov for insightful discussions.

REFERENCES

[1] E. Esarey, *et al.*, Rev. Mod. Phys. **81** (2009) 1229.

[2] S.P.D. Mangles, *et al.*, Nature **431** (2004) 535 (2004); C.G.R. Geddes, *et al.*, *ibid.* 538; J. Faure, *et al.*, *ibid.* 541.

[3] W.P. Leemans, *et al.*, Nature Physics **2** (2006) 696.

[4] W.P. Leemans, *et al.*, in: Proc. 2010 AAC Workshop, Annapolis, MD, p. 3, AIP (2010).

[5] P. Mora, T.M. Antonsen, Phys. Plasmas **4**, (1997) 217.

[6] C. Benedetti, *et al.*, in: Proc. 2010 AAC Workshop, Annapolis, MD, p. 250, AIP (2010).

[7] B.M. Cowan, *et al.*, J. Comp. Phys. **230** (2011) 61.

[8] J.-L. Vay, Phys. Rev. Lett. **98** (2007) 130405.

[9] A.J. Gonsalves, *et al.*, Nature Physics **7** (2011) 862.

[10] C. Huang, *et al.*, J. Comp. Phys. **217** (2006) 658.

[11] A.F. Lifshitz, *et al.*, J. Comp. Phys. **228** (2009) 1803.

[12] D.L. Bruhwiler, *et al.*, in: Proc. 2008 AAC Workshop, Santa Cruz, CA, p. 29 (2008).

[13] C. Benedetti, *et al.*, IEEE-TPS **36** (2008) 1790.

[14] S.F. Martins, *et al.*, Nature Physics **6** (2010) 311.

[15] J.-L. Vay, *et al.*, in: Proc. 2010 AAC Workshop, Annapolis, MD, p. 244, AIP (2010).

[16] C. Benedetti, *et al.*, Proc. of PAC2011, New York, MOP082 (2011), p. 250.

[17] C.B. Schroeder, *et al.*, Phys. Rev. Lett. **107** (2011) 145002.

[18] C.K. Birdsall, A.B. Langdon, Plasma Physics Via Computer Simulation, Adam Hilger, (1991).

[19] J.S. Shang, J. Comp. Phys. **153** (1999) 312.

[20] C. Benedetti, *et al.*, Nucl. Instr. and Meth. A **608** (2009) 94.

[21] J.-L. Vay, *et al.*, J. Comp. Phys. **230** (2011) 59083.

[22] C.B. Schroeder, *et al.*, Phys. Rev. Lett. **106** (2011) 135002.

[23] W. Zhu, *et al.*, Phys. Plasmas **19**, (2012) 033105.

MAD-X PROGRESS AND FUTURE PLANS

L. Deniau, CERN, Geneva, Switzerland

Abstract

The design efforts for the High Luminosity upgrade of the Large Hadron Collider (HL-LHC) will require significant extensions of the MAD-X code widely used for designing and simulating particles accelerators. These changes are framed into a global redesign of the MAD-X architecture meant to consolidate its structure, increase its robustness and flexibility, and improve its performance. Some examples of recent extensions to MAD-X like the RF-Multipole element will be presented. Improvement for models and algorithms selection providing better consistency of the results and a wider range of use will be discussed. The computation efficiency will also be addressed to profit better of modern technologies. In this paper, we will describe the last improvements and the future plans of the project.

INTRODUCTION

The Methodical Accelerator Design (MAD) project has a long history, aiming to be at the forefront of computational physics in the field of particle accelerator design and simulation. The MAD scripting language is de facto the standard to describe particle accelerators, simulate beam dynamics, and optimize beam optics.

MAD-X is the successor of MAD-8 and was first released in June 2002 [1]. It offers most of the MAD-8 functionalities, with some additions, corrections, and extensions [2]. The most important of these extensions is the Polymorphic Tracking Code (PTC) of E. Forest [3].

A decade after its first release, MAD-X is still the main tool used to design and simulate accelerators at CERN. But its original design was mainly focusing on the urgent needs for the LHC, and a large part of the code was inherited from old software written in the 80's. The framework of the LHC upgrade studies is a good opportunity to reorganize and upgrade the overall core of MAD-X to support recent hardware (64 bit, multicores) and new needs. In parallel, the project must continue to incorporate new functionalities in the legacy code, like the two recent optical elements added for modeling thin RF-Multipoles and thin non-linear lenses with elliptical magnetic potential.

The long-term evolution of MAD-X is an essential aspect of the project, because many users around the world consider the application as one of the most flexible and accurate for optics design, as mentioned in comparisons of optics codes [4, 5]. Moreover, MAD-X with PTC is often taken as the reference for benchmarking other codes, and seen by the community of particle accelerator physicists as a key tool that is poised to evolve.

In this paper we expose four different aspects of the project, namely the project improvements and the feature extensions performed during the past year, and the physics and application improvements that are planned for the next couple of years. Each of these aspects addresses different concerns of the project, which are of equal importance from our point of view for the future of MAD-X.

PROJECT IMPROVEMENT

Motivations

The MAD-X project falls into the category of middle-size complex projects. The size of the source code ($\approx$ 165K SLOC[1]) and the number of features provided is not very large but most of the features rely on very complex knowledge difficult to implement and support. In this kind of project, *simplicity and discipline* should have been the rule of thumb during the development process, because the complexity comes inevitably from the implementation of the provided features. Presumably due to priority issues, these rules were not strictly followed during the first decade of development of MAD-X and the drift resulting from the added complexity has led to the untidy feeling perceived by the users. The usual software metrics based on the number of SLOC under-evaluates the complexity of the application. As a consequence, a new improvement process has been set out (Fig. 1), starting from the outside layers of the project to emphasize the restoration of some cross-platform invariants (e.g. build and test system), and to simplify the development process (e.g. code organization) before important new developments is launched.

Global Redesign

The need for a global redesign of MAD-X became obvious with time, as the amount of resources required to implement new features became exponential or equivalently, the time to completion with constant resources became logarithmic. The observable behavior of the application was not always matching the users expectations, and the feedback from the active MAD community has been collected during the past decade and recorded into a project tracker.

After some attempts to improve locally the implementation, it became clear that the process would not converge to a stable solution because some undecidable and coupled bugs were found.

Undecidable bugs occur for example, when memory must be managed in the absence of proper reference handling or garbage collection: freeing the faulty object crashes the application while not freeing the object provokes significant memory leaks that will crash the application later. This latter kind of bug can be painful for ap-

[1] Source Line Of Code

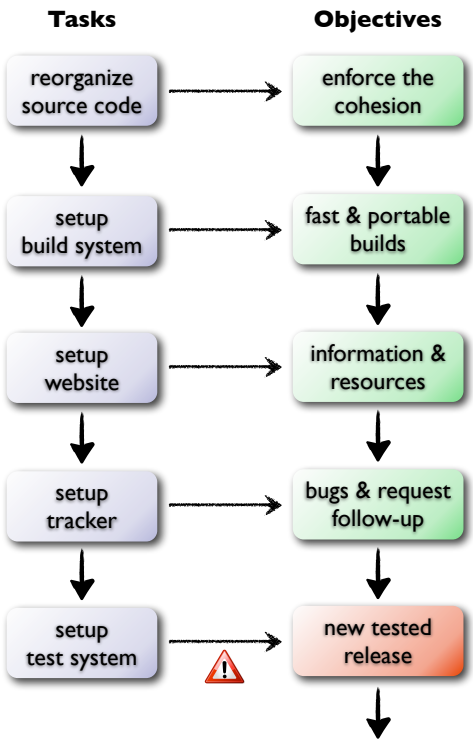


Figure 1: Sequence of tasks and corresponding objectives required before starting new MAD-X developments. The danger icon mark the difficulty to ensure that tests validate *enough* the application.

plications that run MAD-X as a subprocess, like the Java-API JMAD [6] used by the Optic and Knobs Management application, a component of the LHC online model project [7].

Coupled bugs (i.e. bugs with a life cycle) are more difficult to identify because they reappear after few debugging cycles, sometimes months later, where the final state makes the situation worst than the initial state. Some coupled bugs were identified to be a consequence of the abuse of global variables in the code. Hopefully, tools like Subversion and Trac allow keeping history and rolling back the modifications.

In the light of these attempts, it was concluded that only a significant redesign of the core could cure the situation.

Improvement Strategy

When it is realized that a large portion of the code needs to be redesigned, the option often envisaged is to start a new development from scratch and profit from the lessons learnt from the previous project. This approach offers more opportunities for strategic changes, like adopting better supported programming languages and tools, but usually requires significant time investment.

In the case of MAD-X, the option retained was to improve the project on the existing ground, making the task

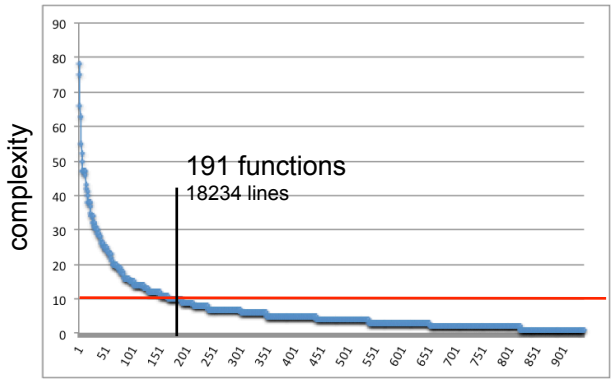


Figure 2: Cyclomatic complexity of functions written in C. The horizontal red line indicates the complexity threshold of 10 where it is recommended to split the function into smaller ones.

more difficult because it has to fit with the legacy design and code. This strategy was motivated by past experience to minimize the risk to end up with inadequate project on the mid-term (e.g. MAD9) and assuring a continuous maintainability of the code. The application had to remain operational at every step of the evolution, a very important point for satisfying the MAD community. The goal was also to avoid the problems encountered during the first attempt resulting from the coupled bugs, and to make the process convergent toward a stable solution. The list and the details of the 24 major tasks achieved during these past months can be found on the project roadmap "phase one" [8].

Code Complexity

MAD-X has hundreds of large functions involving few hundred of SLOC each. The *cyclomatic complexity*[2] [9] was considered as an appropriate software metric (i.e. measuring control flow branches) for this *structured programming*[3] style to identify the critical functions. The Fig. 2 shows the cumulated number of functions above given cyclomatic complexity for the C part of the code, which represent about 20% of the application. The Structured Testing Methodology of the National Institute of Standards and Technology recommends to split functions with *complexity above 10* into smaller units §2.5 [9]. About 20% of all functions in C, C++ and Fortran are concerned (770 over 3908). *Above a complexity of 20*, the function becomes a serious candidate for bugs because it has reached the testability limit, as too many input variables are needed to cover all the branches of the function. It is also more difficult for the developers to grasp the control flow and to identify the kind of configurations that could result in unexpected or invalid states. About 8% of the functions (298 over 3908) are concerned in MAD-X.

[2] `http://en.wikipedia.org/wiki/Cyclomatic_complexity`
[3] By opposition to *object oriented programming*.

Code Reorganization

The core of MAD-X was developed focusing on the demands for the design of the LHC and the optimization of its optics. Its architecture appears to be monolithic, weakening the cohesion and strengthening the coupling in the source code. This *single-file* approach prevented modular architecture exposing only well-defined interfaces; all functions and most variables ended in the global namespace, leading to a large state machine with unexpected side effects.

One of the first major task to reorganize the source code was to split the tens of thousands C lines of the core into 62 files following the *Separation of Concerns* principle [10], and to close the *visibility*[4] of the variables and functions as much as possible. The reorganization of the core helped to detect inconsistent function signatures and to correct a couple of stack frame corruptions. This effort was not enough to get rid of the structural problems, namely dangling pointers, reentrancy and memory leaks, which will be addressed by the future developments.

Website and Tracker

The MAD-X website [8] has been completely redesigned to provide the community an easy access to the information and materials for download: documentations, examples, accelerator optics and new releases. The online manual is undergoing important modifications during the summer 2012. The new website has been visited more than 25000 times during the past six months, testifying the activity of the community.

The Trac web application [11] was setup and fed with all the reported bugs and requests to improve the follow-up of the pending tickets and backup the solutions. An invaluable tool that allowed to detect (non-)convergence of bug corrections, classify the issues, and steer the priorities toward the users requests. The latter will be mentioned using Trac ticket numbers all along this paper in order to highlight the orientation on user requests and projects needs.

Portability and Builds

Another important aspect of the project was the ability to deliver portably and synchronously the MAD-X releases for all the main platforms — Linux, MacOS X and Windows — almost automatically, and save the resources previously allocated to this task. It was important to rely upon single methodology and technology for building and testing the application, because some identified bugs were resulting from mixing incompatible compilers settings.

For this purpose, a new portable and efficient build system has been developed, which handles all the three platforms aforementioned in 32 bit and 64 bit[5] architectures, and allows to build applications and libraries mixing code written in C, C++, Fortran 77 and Fortran 95. The new build system supports 12 compilers and provides extensible, portable, and easy-to-configure makefiles. As a demonstration of its flexibility, it took only a couple of days to extend the build system to build and distribute MAD-X and PTC as standalone libraries. The former was a request from advanced MAD-X users [11] ticket #156, and the latter was part of the collaboration with the PTC-ORBIT activities at CERN [12].

```
[ Jacobian testsuite ]
 + test-jacobian                (0.00 s) -  1/ 1 : PASSED
 + test-jacobian-2              (0.00 s) -  1/ 1 : PASSED
 + test-jacobian-knobs          (0.00 s) -  2/ 2 : PASSED
[ RF multipole testsuite ]
 + test-rfmultipole             (0.00 s) -  9/ 9 : PASSED
 + test-rfmultipole-2           (0.00 s) -  2/ 2 : PASSED
 + test-rfmultipole-3           (0.00 s) -  2/ 2 : PASSED
 + test-rfmultipole-4           (0.00 s) -  2/ 2 : PASSED
[ PTC Twiss testsuite ]
 + test-ptc-twiss               (0.00 s) -  4/ 4 : PASSED
```

Figure 3: Example showing the output of few test suites after a run of the MAD-X test system. About a hundred of tests will be setup by the end of 2012.

Test System

Before relaunching large-scale development, it was essential to equip MAD-X with an efficient and robust test system. On the long term, it will give the confidence that improvements are incremental, and it will avoid unexpected regressions or undesirable backward incompatibilities.

The new test system was implemented as an extension of the new build system, taking advantage of its portability and flexibility. In the long term, it fully automatizes hundreds of tests grouped into test suites (see Fig. 3), which could be run as often as needed to check the correct behavior of the MAD-X components; a task that was performed manually and was taking days beforehand. Another purpose is to check the stability of the numerical algorithms across combinations of platforms and builds settings.

At the time of writing, the new test system already revealed a dozen of bugs never detected before, like the asynchronous outputs between C and Fortran when executed on Windows in batch mode and built with GNU compilers (tickets #159, #160), the truncation of outputs on Windows when built with Intel compilers (ticket #161) or the implicit drifts computation during beam-lines construction that varies between platforms (ticket #155).

The test process is schematized in Fig. 4. The principle is to run a single MAD-X job for each test, and to compare all the output and data files generated (purple boxes) with their reference files (yellow boxes). This black box testing[6] process relies heavily on the `numdiff` tool, a new program developed for running test suites that need to compare unformatted files with plain numerical content, and where text

[4] `http://en.wikipedia.org/wiki/Scope_(computer_science)`

[5] The source code of MAD-X is not yet 64 bit ready, therefore this achievement must be seen as a first step toward 64 bit compliance.

[6] By opposition to white box or glass box testing methods.

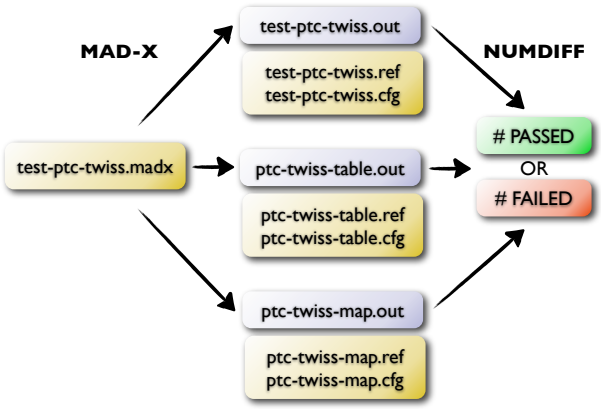


Figure 4: Schematic description of the process used to test MAD-X components. This PTC Twiss test runs a single `madx` job that generates three outputs. The yellow boxes are reference files provided to `numdiff` by the test system and backup on SVN.

```
# Test config for the Jacobian knobs
# file test-jacobian-knobs.cfg

# rows     cols     constraints
1-7        *        skip          # head banner
149-$      *        skip          # tail banner

# first matching
37-38      1-2      rel=1e-12
39         2        abs=1e-21     # from job
41         1        rel=1e-12

# second matching
109-110    1-2      rel=1e-12
111        2        abs=1e-21     # from job
113        1        rel=1e-12
```

Figure 5: Example of `numdiff` configuration file used for the test of the matching command using the Jacobian algorithm with knobs.

and numbers can be arbitrarily mixed [13].

This `numdiff` tool is able to deal efficiently with user-defined constraints applicable to numerical comparisons, and understand complex specifications with both, numerical precision and physical accuracy. The configuration files use a simple format similar to `crontab` files to specify the constraints, i.e. Fig. 5. Most of the time, table-like files (e.g. Twiss and Survey tables) can be compared directly using the default behavior of `numdiff`, which always treats input numbers as floating point values independently of their string representation. This makes the tests easy to configure and portable, and avoid spurious alarms to be reported by tests run on different platforms or with different build settings.

Future Plans

The project improvement is planned to finish by the end of 2012, once the following list of items will be achieved:

- Setup more tests
- Cleanup the examples
- Cleanup the documentation
- Produce "pro" release 5.01.00

Development releases of MAD-X are actually untested releases that include most recent extensions (see next section) and bug corrections. After the completion of these items, the baseline will be to check the development releases against all the registered tests, and the production releases against all the registered tests augmented with the CERN studies. This will put the MAD project into a position where new developments will be possible and safe.

RECENT EXTENSIONS

In this section, we present recent extensions that have been added to MAD-X. They have been implemented into the legacy code, in parallel to the project improvements aforementioned, and following the requirements of CERN projects like HL-LHC and the concerns registered on [11].

RF-multipole [14] (Tickets #104, #114)

Motivation The electromagnetic field of accelerating structures and crab-cavities can exhibit undesired transverse field components due to asymmetries in the azimuthal direction of the element geometry, for instance in the input and output power couplers. Tracking simulations are performed to evaluate the impact of such transverse RF-kicks on the beam dynamics [15].

Model The RF-field is decomposed into pulsed normal and skew components similarly to the multipolar field decomposition used to model the magnetic elements [16]:

$$C_n = B_n + i\,A_n. \tag{1}$$

The RF-multipolar coefficients oscillate at the same angular frequency ω_{RF} as the generating electromagnetic field, and vary with the relative longitudinal position $z = -c\Delta t$ of the particle experiencing the kick to the synchronous particle. Hence, the formalism of Eq. 1 can be written as:

$$\tilde{C}_n(z) = \tilde{B}_n(z) + i\,\tilde{A}_n(z), \tag{2}$$

with

$$\begin{aligned} \tilde{B}_n(z) &= \mathrm{Re}\left[B_n\, e^{i(\vartheta_n - k_{\mathrm{RF}} z)}\right], \\ \tilde{A}_n(z) &= \mathrm{Re}\left[A_n\, e^{i(\varphi_n - k_{\mathrm{RF}} z)}\right], \end{aligned} \tag{3}$$

where k_{RF} is the RF wave number ($k_{\mathrm{RF}} z = -\omega_{\mathrm{RF}}\Delta t$), and ϑ_n and φ_n are the n-th normal and skew phases. Hence, to characterize a RF-Multipole element, it is necessary to specify the RF voltage V_{RF}, frequency f_{RF} and

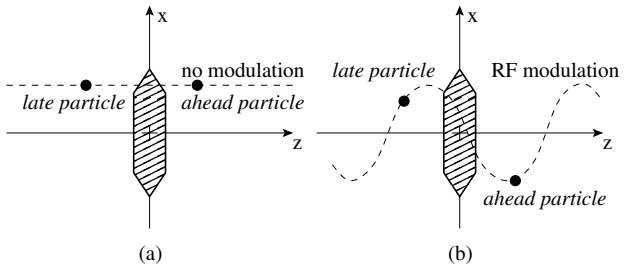


Figure 6: In an ordinary multipole (a), each particle experiences a kick that depends only on its transverse coordinates; in a RF-Multipole (b), the strength of the kick is also a function of the relative longitudinal coordinate z.

phase $\vartheta_{\rm RF}$, plus the amplitude and the phase of each normal and skew harmonic:

$$\text{RF-Multipole:} \begin{cases} V_{\rm RF} & \textit{RF voltage} \\ f_{\rm RF}, \vartheta_{\rm RF} & \textit{RF freq. \& phase}, \\ C_n = B_n + i\,A_n & 2\times n \textit{ amplitudes}, \\ \phi_n = \vartheta_n + i\,\varphi_n & 2\times n \textit{ phases}. \end{cases}$$

Within MAD-X, the convention is to provide the kicks expressed in terms of the integrated multipole strength $K_n L$ instead of the field multipoles:

$$\begin{aligned} B_n + iA_n &= \frac{1}{n!}\left[K_{N,n}L + iK_{S,n}L\right] \\ &= \frac{1}{n!}\left[K_{N,L,n} + iK_{S,L,n}\right] \end{aligned} \tag{4}$$

Physics The Hamiltonian of an ultra-relativistic particle traversing a thin RF-Multipole is similar to that of an ordinary thin multipole with RF modulation due to the rotating harmonics, plus a zero-order term for the longitudinal acceleration [17]:

$$\begin{aligned} H = &-\frac{1}{k_{\rm RF}}\frac{qV_{\rm RF}}{p_s c}\cos(\vartheta_{\rm RF} - k_{\rm RF} z) \\ &+ \sum_{n=0}^{N}\frac{1}{(n+1)!}\mathrm{Re}\Big[\Big(K_{N,L,n}\cos(\vartheta_n - k_{\rm RF} z) \\ &+ iK_{S,L,n}\cos(\varphi_n - k_{\rm RF} z)\Big)(x+iy)^{n+1}\Big] \end{aligned} \tag{5}$$

From this expression one can derive the kick experienced by a particle at coordinates (x, y, z) with respect to the reference particle (Fig. 6):

$$\begin{aligned} \Delta p_x = &-\sum_{n=0}^{N}\frac{1}{n!}\,\mathrm{Re}\Big[\Big(K_{N,L,n}\cos(\vartheta_n - k_{\rm RF} z) \\ &+ iK_{S,L,n}\cos(\varphi_n - k_{\rm RF} z)\Big)(x+iy)^n\Big] \\ \Delta p_y = &\sum_{n=0}^{N}\frac{1}{n!}\,\mathrm{Im}\Big[\Big(K_{N,L,n}\cos(\vartheta_n - k_{\rm RF} z) \\ &+ iK_{S,L,n}\cos(\varphi_n - k_{\rm RF} z)\Big)(x+iy)^n\Big] \\ \Delta p_z = &\frac{qV_{\rm RF}}{p_s c}\sin(\vartheta_{\rm RF} - k_{\rm RF} z) \\ &-k_{\rm RF}\sum_{n=0}^{N}\frac{1}{(n+1)!}\,\mathrm{Re}\Big[\Big(K_{N,L,n}\sin(\vartheta_n - k_{\rm RF} z) \\ &+ iK_{S,L,n}\sin(\varphi_n - k_{\rm RF} z)\Big)(x+iy)^{n+1}\Big] \end{aligned} \tag{6}$$

Implementation A new thin element `RFMULTIPOLE` has been added to the Twiss and Track modules of MAD-X. This command accepts all the attributes of the existing command `MULTIPOLE` (i.e. `L, LRAD, TILT, KNL, KSL`), augmented with the quantities inherited from an RF element (i.e. `RMF_XXX` parameters), plus the normal and skew phases arrays `PNL` and `PSL` specific to this element. The syntax of the command to create a RF-Multipole is [8]:

```
RFMULTIPOLE,
 L=real, LRAD=real, TILT=real,
 RFM_VOLT=real, RFM_LAG=real,
 RFM_HARMON=integer, RFM_FREQ=real,
 KNL = { knl0, knl1, ... }, ! Normal amplitudes
 KSL = { ksl0, ksl1, ... }, ! Skew   amplitudes
 PNL = { pnl0, pnl1, ... }, ! Normal phases [2pi]
 PSL = { psl0, psl1, ... }; ! Skew   phases [2pi]
```

Export to SixTrack For the tracking needs of the LHC, the SixTrack code is routinely used [18]. The data loaded for simulations are provided by the MAD-X interface, which has been extended to export the RF-Multipoles. SixTrack currently accepts only quadrupoles, sextupoles, and octupoles, the dipoles being exported as a crab-cavity element. Therefore, a RF-Multipole in MAD-X is converted into and up to three equivalent thin elements in SixTrack and higher order terms are discarded. Simulations of RF-Multipoles associated to crab-cavities will be started soon for LHC upgrade studies.

Intrabeam Scattering (Ticket #103)

The main improvement to the IBS implementation in MAD-X was to add the missing terms $6\beta_x^2\phi_x^2/(H_x\epsilon_x)$ and $6\beta_x^2\beta_y\phi_x^2/(H_x\epsilon_x\epsilon_y)$ in the expressions for a_x and b_x [19]. There were also some other minor issues; the dispersion in the MAD-X Twiss table is defined as $\Delta x/(\beta\Delta\delta)$ with $\beta = v/c$, which differs from the standard convention for β smaller than 1. We also now distinguish the r.m.s. relative energy spread $\Delta E_{\rm rms}/E$ from the r.m.s. relative momentum spread, $\delta_{\rm rms}$ according to $\sigma_\delta \equiv \delta_{\rm rms} = (\Delta E_{\rm rms}/E)/\beta^2$.

Nonlinear Lenses (Ticket #105)

The new NLLENS element models a thin nonlinear lens with the magnetic potential of 'Elliptic' type as specified in [20]. The lens is used to create fully integrable 2D nonlinear accelerator lattice with very large nonlinear tune spread/shift. The NLLENS element is recognized by the thin tracking module of MAD-X and the quadrupole term of the potential is included in the transport map computation and effects the calculation of tunes and Twiss functions.

PTC Beam-beam Element (Ticket #167)

The beam-beam element of PTC was recently connected to the equivalent MAD-X element to study single-particle source of beam losses during the squeeze of the LHC beams. The connection and the physics of the beam-beam element of PTC are undergoing validation and the scripts will be used to setup the test suite.

PHYSICS IMPROVEMENT

Motivation

Particle accelerator optics codes [4, 5], including MAD-X, are routinely used by physicists to perform the following tasks:

1. **definition**: define or modify machine parameters using the MAD language.
2. **tacking**: track particles or maps to find periodic, quasi-periodic or constrained solutions, i.e. one-turn map and closed orbit[7].
3. **analysis**: compute optics functions for the one-turn map, use normal forms for high-order terms.
4. **optimization**: optimize the design with user-defined constraints, e.g. interaction regions matching.
5. **validation**: perform single-particle tracking campaign to validate the design, e.g. check the dynamic aperture.

where each step frequently includes iterations over the preceding steps.

Step 1 is mainly related to the scripting language and its interpreter, as well as the management of the data structures representing the machine layout; this important aspect will be discussed in the next section. Steps 2 and 4 are the heart of MAD-X and may include extra matching procedures to find fixed points of transfer maps between boundary conditions or periodic conditions (i.e. closed orbit). Step 3 is the difficult part for the physicists, because the obtained results depend on the choice and the parameters of the integrators selected in step 2; this is where PTC excels and offers many options. Finally, step 5 is the validation of long-term model behavior through massive tracking, like for example dynamic aperture studies. Actually MAD-X is not really efficient for this kind of task and export the lattice attributes to SixTrack, which is better suited for massive tracking.

The main problem encountered by MAD-X users with the physics concerns the discrepancy between the models of the machine elements used by the different integrators. To quote L. Nadolski [5], *"the amazing discrepancy between codes has its origin mainly in the integrator scheme"*. In practice, MAD-X contains many tracking codes and gathers somehow most of the problems of consistency.

Modeling the Lattice

Track The Track module is a single-particle 6D drift-kick-drift symplectic integrator, which accepts only thin elements sliced by the *Makethin* module. The fixed point (i.e. closed orbit) is computed by a 6D Newton approximation. Track has symplectic integrators for drift, multipole, cavity, and few other special elements like solenoid and orbit corrector, as well as recently added elements. The splitting method of Makethin uses equidistant slices for all thick elements (e.g. dipoles, quadrupoles, sextupoles, octupoles, etc.) if the number of slices requested is greater than 4, otherwise it uses the TEAPOT placements and coefficients [21]. This method is based on the matching of the sine and cosine trajectories at their respective focal points between the thick and the thin representations of the same quadrupole, minimizing the error for the first order approximation. This gives the positions and the strength fraction of each slice relative to the thick quadrupole center and integrated strength. This geometrical method builds in practice a 2^{nd} order symplectic integrator as noted in [22], which is then applied indifferently to all kind of thick elements.

Twiss The Twiss module is an *almost*-symplectic 4D matrix-kick-matrix integrator used to track the linear terms of the transfer maps in 6D, and performs optical functions analysis from the Jacobian matrix R (1^{st} order) and the tensor T (2^{nd} order). Twiss uses a simplectification procedure described in [23] to restore the symplecticity of R after each element if the deviation becomes significant. It must be noted that (Six)Track and Twiss use different element models and methods to find the reference orbit, and that Twiss uses the lattice *as-defined*, treating thick elements directly. This means that to make the two different machine representations and models behavior equivalent between thin and thick, the users have to match some parameters (e.g. beta functions) between the two lattice models using knobs (e.g. quadrupoles strengths). Once the convergence is achieved, the (Six)Track simulations can be run on the *almost* equivalent model from Twiss. This manual procedure is somehow painful and time consuming during complex process where the matching must be performed iteratively; as for example during the squeeze that occurs simultaneously at all the interaction regions, in particular IR1 and IR5, before the two beams enter into collision in

[7] In the following of this paper, we denote for convenience the *one-turn map* as the composition of the transfer maps between two boundaries and the *closed orbit* as the fixed point of the one-turn map.

the LHC.

PTC The PTC library embedded into MAD-X provides symplectic integrators for the naive 2nd order method, the Ruth-Neri-Yoshida 4th order method and the Yoshida 6th order method [24]. It supports the drift-kick-drift, matrix-kick-matrix and delta-matrix-kick-matrix models. The latter has been introduced for speed and compatibility with the SixTrack[8] application. PTC handles exact as well as expanded version, i.e. for the square root, of the Hamiltonians. The number of integration steps (slices) NST of the splitting model is specified by the user. But if the RESPLIT flag is provided, NST is scaled by the integrated focusing strength KL divided by the parameter THIN for quadrupoles, or XBEND for dipoles. The order of the integrator is selected by the two values of the parameter LIM according to the rules:

- $\text{LIM}_1 \geq KL/\text{THIN}$: 2nd order
- $\text{LIM}_2 \geq KL/\text{THIN} > \text{LIM}_1$: 4th order
- $KL/\text{THIN} > \text{LIM}_2$: 6th order

This splitting scheme makes the final layout quite different from the layout created by the Makethin module or treated by the Twiss module. Hence, the aforementioned discrepancy between the integrator models becomes prominent because the communication between MAD-X and PTC in only one-way; results computed by PTC cannot be reused directly by MAD-X for the time being.

The study [25] makes a detailed comparison of the transfer maps computed by PTC and by Twiss for a single dipole magnet (SBEND and RBEND) under various conditions and inputs. The main discrepancies observed are relative to the variations in $\Delta p/p$ for off-momentum beams with $p_x(0) \neq 0$, and to the 6th canonical variable p_t, which is always zero in Twiss (Twiss is 4D). Hence, PTC and MAD-X outputs do not match even when tracking through a single element.

Symplectic Integrators

The Fig. 7 schematizes the process used to derive symplectic integrators, i.e. geometric integrators that preserve symplectic structures, from the Lorentz forces induced by magnets and RF-cavities in charged particles accelerator. The "danger" icons mark boxes resulting from integration steps where approximations are usually needed to solve the differential equations of the previous level. In this paper we are mainly interested by the second integration step involving the construction of symplectic integrators that solve the motion equations. When the Hamiltonian is not integrable directly (right branch in Fig. 7), one can either (1) find an approximate solution to the exact Hamiltonian, (2) find an exact solution to an approximate Hamiltonian, or (3) model the potentials and energies such that the obtained Hamiltonian becomes integrable (i.e. upward arrow) as in the RF-Multipole example.

[8] SixTrack is a matrix-kick-matrix code, this mode is also an extension.

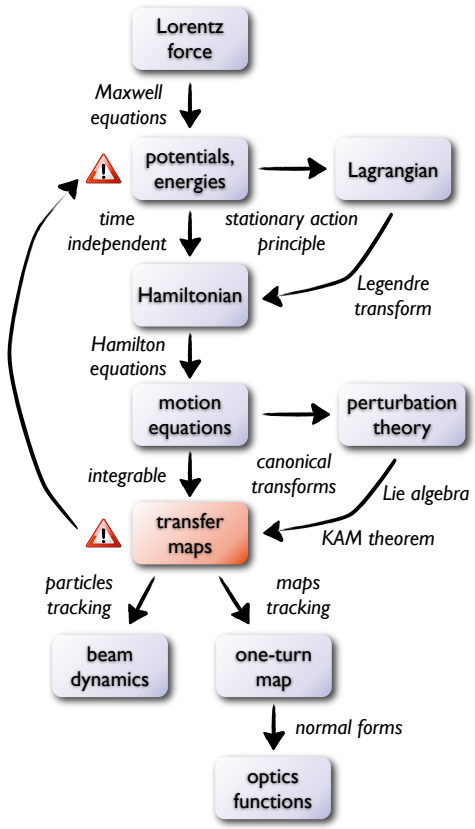


Figure 7: Schematic process to construct symplectic integrators (red box) and their application to particles beam optics. The danger icons mark integration steps where approximations may occur to solve the differential equations.

Implicit integrators The former approach (1) is used to construct implicit integrators (numerical integrators) developed to solve large class of motion equations, typically the class derived from time-independent Hamiltonian that expressed the total energy of the system as the sum of the potential and the kinetic energies.

These integrators do not preserve the phase space symplectic structure of the canonical variables $(\mathbf{q}, \mathbf{p})$; the $2n$-form $d\mathbf{p} \wedge d\mathbf{q} \equiv \{,\}$, i.e. the Poisson brackets, does not preserve the volumes [26] p. 204-207. In other words, they do not preserve the first integral invariant (i.e. incompressibility) of the Hamiltonian flow $\{, H\}$ (Liouville's theorem), at the heart of the stationary action principle. Such implicit integrators let the error of the total energy of the system grow secularly, i.e. diverging spirals appear on the Poincaré sections of the phase space. The map of such integrator does not describe the motion in an Hamiltonian system.

However, the long-term behavior can be kept under control by using backward error analysis and averaging transformations to make the system time-reversible (i.e. integrable) and restore the *near-preservation* of the symplecticity [27] (ch. XII). Nevertheless, these methods are not recommended in [22] by E. Forest, the author of PTC.

Explicit integrators The second approach (2) is used to construct symplectic integrators (i.e. transfer maps) that give exact solutions for a *perturbed* Hamiltonian admittedly close to the original Hamiltonian, with the guarantee that there is no secular trend in the error of the total energy caused by truncation error. This latter statement does not prevent symplectic integrator to be inaccurate, and yet very stable and precise. Their practical interest lies in the fact that many physical systems are perturbations of integrable systems.

A non-integrable Hamiltonian H can be seen, under regular conditions specified by the Kolmogorov-Arnold-Moser theorem, as the perturbed version of and integrable systems $\tilde{H} = H - \delta H$. Various methods and formulae are provided by the perturbation theory to find the appropriate generating functions and canonical transformations that formulate the Hamiltonian on a different configuration space and exhibit properties — stabilization, separation, cancellation, invariance, etc. — useful to simplify and solve the problem. In particular, the Baker-Campbell-Hausdorff formula is very useful to split, eliminate (the perturbative terms to some order), and combine non-commutative maps while preserving the symplectic structure of the system, as in the construction of high-order symplectic integrators [24].

There is no fundamental difference between slicing an element and splitting its Hamiltonian into integrable parts. In both cases, the goal is to build the maps of a symplectic integrator that solve the motion equations in that element. With the exception that geometric interpretation of the resulting high-order maps may leads to negative drifts, somehow unphysical.

Tracking Maps

In practice, the order of a symplectic integrator is given by the order of the error of its maps in term of the independent variable between the exact and the perturbed Hamiltonians. To model the transfer map of a single element or a fraction of this element, i.e. a slice, 2^{nd}, 4^{th} or 6^{th} order symplectic integrators are generally accurate enough for *particles* tracking. But to model a full one-turn map $M = M_n \circ \cdots \circ M_2 \circ M_1$, one needs higher orders.

The solution is to track *truncated power series* (i.e. Taylor maps) that approximate the complete composition up to some defined higher orders. The truncation breaks the symplecticity of the one-turn map, but it can be computationally restored using the same methods used to construct symplectic integrators. The error of the deviation will be anyway of higher order than the order of the truncated power series [22].

The PTC library embedded into MAD-X is able to track any order of maps. The order is fixed by the users and depends on the non-linearities supposedly present in the lattice, the off-axis off-momentum initial conditions of the simulations, and the order of the analysis that follows, i.e. bottom right branch in Fig. 7. Because the time of the computation grows exponentially with the order of the maps, some tradeoff must be found to avoid lengthy tracking time, e.g. days or weeks for large lattice and 10^{th} order.

As a comparison, Track is a 0-order map symplectic tracking code (particles orbits) and Twiss is a 2^{nd}-order map non-symplectic tracking code (orbit, R and T), both being used with some success at CERN. Hence, symplectified high order maps are adapted to model accurately one-turn maps for further analysis and to some extend perform short-time tracking, while symplectic 0-order maps are adapted for long-time tracking. Both kinds of maps are handled indifferently by PTC on the same lattice model, making PTC the only valid and unified approach for complex lattices. Unfortunately, PTC is slow comparing to (Six)Track and Twiss, even when processing with same maps order, because of its very flexible splitting model as shown in [28]. As for most components of the application, some room exists for improvement.

Matching Methods

Figure 8 shows the general principle of the matching module available in MAD-X. The optimization algorithm searches for a global minimum of the figure of merit F. If such a minimum is reached within the tolerances, the iterative process considers the problem as being solved and stops. Otherwise it updates its new position in the problem state space C (i.e. user variables), and starts a new iteration. This optimization process follows a trajectory into the state space, which depends on the properties of the figure of merit F and the nature of the problem to solve; different algorithms follow different trajectories. The figure of merit is a cost function part of the optimization algorithm, which combines the user-defined penalty functions to be minimized and the constraints on the solution domain, both defining the nature of the problem:

- local — global
- linear — quadratic — non-linear
- continuous — discrete
- smooth — irregular
- unconstrained — constrained
- constrained by equality — inequality

For a complete description of the various algorithms available for each combination of these criteria see [29]. In our case, we give a particular interest to the subclass of global, non-linear, continuous, smooth and constrained problems with both equalities and inequalities. This subclass is relevant to systems described by the Hamiltonians modeling lattice elements and to the usage made of the MATCH command. The algorithms actually available in MAD-X do not fall into this subcategory as we will see.

Simplex The Simplex minimization is an *active set method* (i.e. the algorithm class) well suited for linear programming (i.e. problem class), that performs well to solve linear problems with inequality constraints. Its strength lies

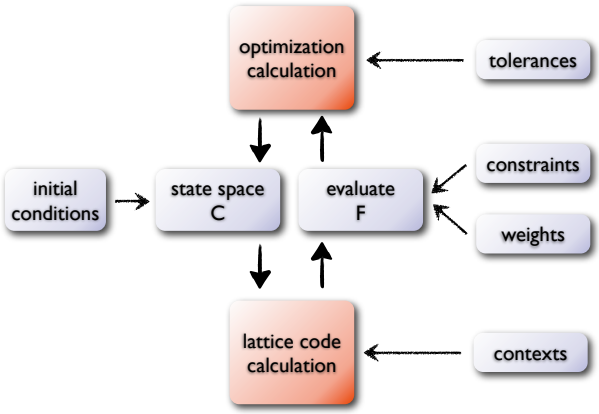


Figure 8: Optimization principle, the red boxes are the key components of the *matching* optimization process.

in its capacity to detect active constraints at the solution. In practice, it does not work well on the class of problems of interest to MAD-X users, because the features that make it so attractive for linear programming vanish in the presence of non-linearities.

Gradient The LMDIF and MIGRAD minimization are based on gradient descent method using different functions of merit. These are the simplest and fastest methods available in MAD-X, but they require more evaluations than other methods and get easily trapped by local minima, with no user-defined criteria to avoid the problem.

Newton The Jacobian minimization [30] uses the Newton algorithm and was added to deal with over-determined or under-determined problems that other methods were not addressing, using respectively QR or LQ decompositions. But this algorithm as well as the two previous ones are not supposed to deal with constraints and special tricks have been setup to bound the search in the state space and eliminate (freeze) non-active constraints. If the procedure fails, the user has to put extra weights to control the algorithms when it approaches the boundaries.

The implementation of this command was also the occasion to provide a more flexible way to define the constraints and act on the knobs with the help of the VARY command and some MAD macros. This is an important methodological aspect of the way the MATCH command might be used; some users like commands on which they get fine control. This approach works in practice because users' knowledge of beam dynamics generally let them start with good guesses and use the algorithm for fine-tuning of the expected results.

New methods On the long term, we propose to add two new methods better suited to manage large-scale non-linear problems (lattice knobs!) with large set of constraints all together. They are based on the *Augmented Lagrangian methods* to handle equality and non-equality constraints in a uniform well-defined framework. With these forefront methods in hands, MAD-X users should be able to work on complex large-scale problems efficiently.

The *active-set sequential quadratic programming* (SQP) method corresponds to the general purpose of the MATCH command where the problem is smooth and well defined, and where the inequality constraints specify soft boundaries of the solution domain, that is penalty functions can be evaluated beyond during the optimization steps. This algorithm should not trouble the users understanding since it is in the line of the Gradient and Newton methods, except that it works on the Hessian of the augmented Lagrangian quadratic form. The *active set* qualifier means that it possesses the same ability as the Simplex to select effectively the active constraints, a combinatorial problem, meaning that even large set of constraints can be handled efficiently.

The *interior point with log-barrier* method is the complementary approach of the active-set SQP essential to handle problems where the inequality constraints define strong barriers on the solution domain, i.e. user-defined penalty functions should not be evaluated beyond.

These two methods require evaluating the local Jacobian and Hessian matrices of the merit function accurately. The existing methods in MAD-X use inaccurate numerical differences to estimate the local derivatives. We propose to implement or reuse an existing *automatic differentiation* module for this purpose.

The computational power required by these new algorithms should remain equivalent to the already implemented ones, because the extra computations required by the second order derivatives should be compensated by the quadratic convergence of the algorithms. The development of this new component could be also the occasion to improve the computations performed by the differential algebra maps during high-order tracking.

Thick Elements *(Ticket #113)*

This ticket requests to add thick *dipole* and *quadrupole* to the tracking module of MAD-X. This achievement would avoid extra work currently needed to generate a thin lens version of the LHC lattice and rematch the optics during squeeze simulations. While dipoles and quadrupoles can be defined as thick elements in the lattice sequence, they actually have to be sliced to be accepted by the Track module. The request boils down to re-introduce thick elements into the Track module as in MAD8, making the code a mixed drift-kick-drift and matrix-kick-matrix code. Some care will have to be taken with guards rejecting thick elements in few places.

The architecture of Track and Twiss modules requires that transfer matrices have to be entirely recomputed each time an element is processed, including all the costly transcendental functions, which slow down significantly the simulation as observed from simulation with SixTrack [31]. In order to avoid this degradation, another flag must be added to the Makethin module to select precisely the

dipoles and the quadrupoles that must remain thick.

Teapot Slicing (Ticket #154)

This ticket proposes to apply iteratively the TEAPOT splitting scheme aforementioned until the number N of user-defined slices for the element is reached. This request should be technically easy to address, even in the legacy code and should be implemented soon. Some care will have to be taken with the memory management of the elements. For example, this extension could split N linearly leading to one of the following compositional schemes according to the remainder $\{0, 1, 2, 3\}$ of $n = \left[\frac{N}{4}\right]$ respectively:

$$T_4^{(n)}, \quad T_3 \circ T_4^{(n-1)} \circ T_2, \quad T_3 \circ T_4^{(n-1)} \circ T_3, \quad T_4^{(n)} \circ T_3 \quad (7)$$

where T_k with $k = \{2, 3, 4\}$ refers to the already implemented TEAPOT splitting schemes.

As an alternative[9], it could possible to extend the TEAPOT splitting schemes beyond 4 slices to keep the accuracy of the underlying 2nd order symplectic integrator. For examples, first estimates on T_8 show that it should improve the error by approximately a factor 2.2 with respect to a T_4^2 scheme, in the line of T_4 that improves T_2^2 approximately by a factor 2.5. With the Yoshida methods [24] and 8 steps, it is possible to construct a 6th order symplectic integrator but containing negative drifts that MAD-X is not able to handle for the time being. An issue that might be useful to solve if we like to extend the range of application of MAD-X to small or special accelerator lattices.

Symplecticity Checks (Ticket #172)

This ticket proposes to add symplecticity checks into MAD-X. In fact, these checks are already present in Twiss for the matrix R and could be improved to include the tensor T. The same checks could be added to Track to evaluate the precision and the stability of the calculations in debug mode, because Track is already symplectic.

Lattice Transformation

This ticket proposes to implement a lattice reflector to get rid of the problem of consistency between the beam direction flag `bv`, the beam 2, and the beam 4 inside MAD-X in order to simplify the LHC simulations. The details of the transformation are given in [32] and the implementation cost in the legacy code will be carefully studied to evaluate if it is not better to wait after the redesign of the core, which should be able to process complex 3D lattice topologies in line with PTC.

Other Plans

Benchmarks MAD-X users observed unexpected differences or inconsistencies in their results from time to time. While the origins are sometime the misunderstanding of correct usage, often due to the abstruseness or the incompleteness of the documentation, some of these feedbacks are valid complaints. Therefore, further studies like in [25] should be performed on other single-element models as well as on small reference lattices in order to evaluate the impact of parameters changes. Special considerations should be addressed for:

- small p_0c,
- large $\Delta p/p$,
- large aperture A versus integrated field strength KL,
- transverse offsets off-axis (x_0, y_0) and off-momentum (p_{x0}, p_{y0}) vs small length L and small radius ρ.

These points should quantify the validity of the approximations made in the models, like the small angle approximation, the error in series truncation and the radius of convergence of the series. An application like MAD-X should work for any value of the particle kinetic energy p_0c, i.e. $0 < \beta \leq 1$, and any combination of $(x_0, p_{x0}, y_0, p_{y0})$, A, L, KL, and proper ratio of transverse and longitudinal momentum.

This benchmark might be very useful to quantify the discrepancies between MAD-X and PTC when modeling small or special accelerators (e.g. Fixed-Field Alternating Gradient) studied at CERN where approximations are mostly noticeable [33]. Moreover, the conclusions could be used to draw some recommendations for the users and improve the application in that direction. On the long-term, the best strategy would be to merge these three components into a well-tested single one, most probably following the more flexible and accurate approach of PTC. Then the improvements would mainly consist to speed-up the PTC approach.

Manuals and guides The actual online documentation of MAD-X [8] has grown untidily with time, and the initial structure inherited from the MAD8 user's manual has not been preserved. The current table of content does not reflect the application logic, and users often prefer searching the information in the subject index first. The look and feel of the user's manual is under improvement but the restructuring of the document and the correction of the inconsistency is a long-term objective. In parallel, a new physics guide will be started on the model of the MAD8 physics guide, and extended to take into account the evolution of MAD-X. In particular, the Hamiltonians used in Track, Twiss and PTC for each element should be collected and carefully described with their respective assumptions and limitations. This task will go in concert with the benchmarks aforementioned to identify the approximation done by the models and the recommendations that should follow.

APPLICATION IMPROVEMENT

Motivation

In the section on project improvement, we have mentioned the untidy feeling perceived by the MAD-X users.

[9] Not part of the request ticket #154.

Table 1: Classification of MAD-X users complaints that turned out to be problematic for their studies, 77% of the tickets are related to the core of the application, 23% to the physics.

Groups	#	Tickets no
interpreter	14	56, 57, 64, 67, 73, 89, 109, 110, 124, 125, 132, 136, 147, 165, 171
table, select	10	48, 81, 85, 99, 122, 123, 130, 139, 141, 148
sequence, use	7	61, 74, 80, 89, 93, 120, 126
plots	12	42, 69, 71, 76, 85, 86, 102, 115, 116, 131, 150, 164
memory leaks	7	1, 3, 4, 92, 111, 151, 153
MAD-X physics	13	68, 77, 79, 83, 88, 106, 112, 117, 127, 135, 137, 138, 149
PTC physics	2	75, 118

Table 1 classifies the meaningful complaints of the users to spot the problematic parts of the application. The complaints relative to the core of the application (non-physics groups) represent 77% of the total, making this part of the application a priority for improvement, while the physics part (two last groups) represents only 23% of the total. In the following, we will address few of these points and make proposals to improve directly or indirectly the situation.

Data Management

Memory leaks Memory leaks appear in many occasions in MAD-X; each time a table, macro, file inclusion (i.e. call file), command, class or element are *created*, some memory leaks happen. The three latter points are equivalent in practice because classes and elements are represented as commands internally[10]. The problem becomes rapidly critical for the jobs that intermix macros and loops to create long sequences.

Data sharing The design of the application was intended to run "one-shot" jobs, where allocated memory is never reclaimed except by the operating system when the process quits. The later introduction of macros has changed the big picture, and real cases leaking up to 7 MB/s have been observed (ticket #111). These leaks cannot be cured in the present design because there is no ownership policy, that is the number of references sharing the same object is unknown, making these bugs undecidable.

Data lookup The situation is complex because all searches are based on plain strings comparison, a notoriously slow process which is also required to retrieve elements and beam parameters at each tracking step in Track and Twiss. To improve the performance, the core cashes a lot of information through pointers and indexes into large data structures and global variables (i.e. side effects) that must be manually updated upon changes. Another strategy adopted was to enforce constraints when defining new elements or data structures, like constant offsets for specific fields within structures, or hardcoded indexes to access properties stored as key-value pairs (ticket #57). These two optimizations are very error prone and make the management of data structures intrusive and very complicated.

[10] An interesting design choice that will be preserved.

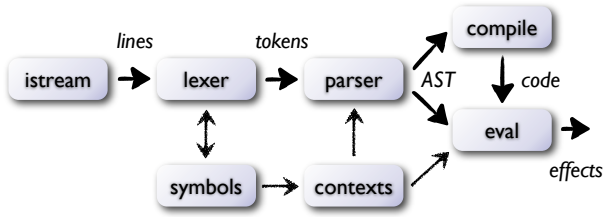


Figure 9: Components of the future interpreter, split into the analyzer (left to AST) and the evaluator (right to AST).

Side effects In many cases, the only available solution to recover from undetermined state of global variables and side-effects between modules interaction is to destroy and to recompute everything; a well known user trick used to reload sequence or call USE and TWISS commands for arbitrary reasons, hoping to restore a proper lattice (ticket #80).

Interpreter Design

A deep analysis of the situation highlighted that most design problems were coming from the development of the interpreter itself. The interpreter is not following a well-defined syntax and grammar, and the implementation grew from timely requests with little cohesion, introducing inconsistent constructions in the language.

No parser The analyzer part of the interpreter is string-based, with no lexer, no parser and no abstract syntax tree (AST), which are essential components of any interpreter, i.e. Fig. 9. As a consequence, the interpreter is spread over the entire code and processes only strings. A strategy that conditioned all further developments down to the physics computations. The slow performance of plain strings comparison generated plenty design decisions that did not scaled with the size of the core, with a strong impact on the complexity, performance and support of the application.

Ambiguous grammar There are many examples, i.e. Tab. 1, which create grammar issues that remain undetected by the analyzer. The commutative positional arguments of the `TABLE` command are at the origin of some tricks in the code and painful bugs for users. The `column` argument in `SELECT` that collects an undelimited list of names and therefore needs to know the names of the formal parameters to detect the end of the list. The unquoted

strings converted to sequence of "tokens" but still interpreted as lowercase strings, making the scripts platform dependent (e.g. file names). The extra comma in a TWISS command that cancels silently its invocation, leaving the previous Twiss table in place as if it was called, i.e. thanks to global variables and side effects. The use of unrecognized operator (e.g. ** instead of ^) that makes expressions returning zero silently. The extra space before (or after) some operators (e.g. ->) that breaks silently the expression evaluation leaving unchanged the value in assignment. The missing statement delimiter ; silently ignored, leading to concatenation of statements and strange behaviors experienced by most users.

The list is long but each of these points (and more) were reported by users, after having wasted hours if not days to figure out the problem[11], with a real cost for their projects.

Interpreter redesign We propose to redesign the interpreter, which is the cornerstone of the creation and management of most data structures, following the standard structure of interpreters schematically shown in the Fig. 9. The components will be developed step by step and introduced into the legacy code for portability check, testing and debugging purpose. The parser will be a recursive descent parser for its simplicity and flexibility and because utterly performances are not needed at this level. Moreover, it will allow parsing context dependent grammar required for backward compatibility with the MAD language that we do not want to leave out. The code will be activated once the interpreter is ready for use, and all modules adapted to its standard interface. For further details about parsing technics and interpreter implementation, see [34, 35].

The language syntax will keep almost full backward compatibility because many users are familiar with it, and like its compactness and flexibility; except for the cases aforementioned that make the grammar ambiguous or context dependent.

Application Logic

We mentioned that the design and the implementation of the interpreter has driven all MAD-X development during a decade. For the same reason, the development of a new well designed interpreter, i.e. Fig. 9, will improve all aspects of the application by introducing few concepts missing in MAD-X. We describe hereafter four concepts that we plan to introduce in the code and the implications for the application logic as shown in Fig. 10.

Ownership policy We propose to introduce intrusive manual reference counting with ownership semantic in all objects, i.e. data structures. This semantic has been defined and implemented in the Objective-C language and used with success by the software industry for two decades [36]. This will simplify the memory management drastically and should remove the undecidability of all pending memory leaks. This policy will be an integral part of the new interpreter, because all user objects declared in the scripts must be built by the parser and held by the nodes of the AST. Hence the new interpreter will be the first claimant for the ownership policy; variables, elements, sequences and tables being the next.

Runtime polymorphism A direct consequence of building an AST, is that runtime polymorphism must be supported. Indeed, the type of the nodes in an AST are known only at runtime and need polymorphic interfaces with late binding to be used or evaluated. This feature is already present in C++, easy to implement in C, and tedious to simulate in Fortran 95.

Contexts and scopes Contexts and scopes are important concepts of programming languages. A scope defines the validity and visibility of variables and identifiers, and is delimited by brackets pairs. Variables are bound to values within the scope where they are created and are destroyed on exit. A context includes the current scope and all enclosing scopes. Only variables that are part of the current context are visible, i.e. can be used. To create local variables bound to the current scope, we propose to introduce the keyword `var` and extend the semantic of `const` with the usual semantics borrowed from most programming languages.

Notification policy The last concept that we propose to introduce is the notification policy, which consists to notify registered objects about specific changes or actions through well-defined polymorphic interfaces. This will help to update the values of variables, expressions, elements, sequences, etc. consistently and only on-need. It will also help to implicitly manage the requirements of some computations, i.e. data and modules interdependencies.

Polymorphic variables With the introduction of runtime polymorphism, variables will be able to hold and deal with any kind of objects through properly defined polymorphic interfaces. Currently MAD-X is limited to deal with boolean values, integer numbers, floating point numbers, strings and expressions thereof. The evaluator (right in Fig. 9) will become polymorphic too, with the ability to evaluate any kind of objects and expressions.

Polymorphic containers By extension to variables, arrays and associative arrays will inherit for free of the same features and will be able to hold and deal with any kind of objects or group of objects. By composition, a variable will be able to store an array or table containing various kinds of elements, or an entire lattice definitions. This improvement will allow to get rid of all the global variables and most side effects present in MAD-X. Tables, sequences, files, commands, etc. will be stored in system-defined or user-defined variables, and these variables will

[11] Many thanks to all MAD-X users for their fruitful feedbacks.

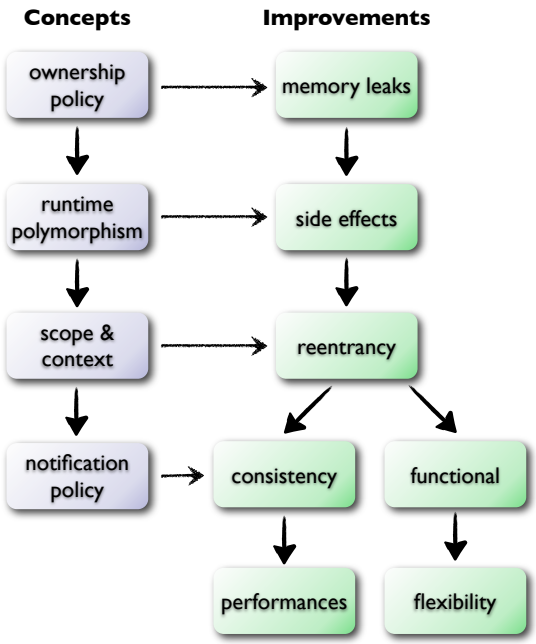


Figure 10: Schematic impact of concepts on the application logic and performance.

be passed as arguments to commands. This will restore the functional programming capability missing in the scripts.

The purpose of the simple `twiss` example hereafter is to show how hierarchical objects can be built and stored into variables and passed to the command, highlighting the future uniform data management:

```
file       = file="optics";
pattern    = pattern="^m.*";
column     = column={name,s,betx,bety};
sector_sel = sectormap=pattern;
twiss_sel  = select={pattern,column};
twiss (file, twiss_sel, sector_sel);
```

The second equal sign in the first five lines does not perform an assignment, but returns a key-value pair object, which in turn is assigned to a variable. The brackets are used to group expressions into a polymorphic array. The parenthesis of the `twiss` invocation means "evaluate in the current context augmented by these variables", i.e. the arguments. This syntax is equivalent to the first comma or space following the command name currently supported, but with a less ambiguous notation. Without going into the details, this approach standardizes the way variables, elements, sequences, commands, etc. are handled internally, with a deep simplification of the MAD-X core.

Types extensions Since the evaluator will handle polymorphic expressions everywhere[12], the language can be extended very easily to new types useful to the users. Needs have been identified for complex numbers, vectors, matrices and functions, and more may be added later painlessly.

[12]Currently expressions are only allowed in limited places.

But for the sack of simplicity, the scripting language should be kept as simple as possible, with a syntax yet far enough expressive for all needs of MAD-X users.

Sequences The following example is coming from a user request and shows an application of functional syntax to embedding sequences into other sequence:

```
seq := sequence (L=len) {
...
subseq1: seq1, at=at1;
subseq2: seq2, at=at2;
...
}
seq1 := sequence (L=len1) { ... }
seq2 := sequence (L=len2) { ... }
```

For backward compatibility of scripts, all current notations will be supported, but the new notation reflects again the uniform treatment of the information by the new interpreter.

The possibility to define subsequences of lattices stored in variables with notification of changes will allow the users to change the lattice definitions on-the-fly through the knobs of the optimization process, and avoid to rebuild the entire beam-line with costly macros at each iteration of the matching, i.e. Fig. 8 bottom red box.

Matching reentrancy A very positive consequence of the support of functional programming style is that it will allow reentrancy of commands. The reentrancy of the MATCH command is a strong user request to optimize large complex optics, e.g. interaction regions of the LHC, that need to be split into few steps to converge. This splitting method called *state space partitioning*, makes the optimization algorithms more efficient and stable.

Data Visualisation

Data visualization is an important topic for MAD-X users. As for any numerical intensive application, visualizing data is important to quickly check the validity of the results. But this is also a very time consuming programming task that cannot be addressed directly within the MAD-X project due to the limited resource. Hence, the proposed strategy for the PLOT command is to delegate the task to third party portable application, like Gnuplot [37]. The interface already exists but is actually underexploited, that is only used for 2D phase-space tracking plots (i.e. trackfile), while other kinds of plots relies on old unmaintained codes written in Fortran 77.

With the notification policy aforementioned, it will be possible to update the plots on-the-fly upon changes, e.g. Twiss table content during matching. If important limitations not compatible with MAD-X objectives are encountered, the possibility to use a portable library will be studied in the last resort, but the amount of work to support such strategy is significantly larger.

Speed Performance

Tracking Software speed is a strong point for the community of users, but the complexity of MAD-X forbids serious optimization. As a consequence many parts of the code waste CPU cycles. The old programming style based on cascades of `if-then-else` (up to few tens per function) breaks frequently the instruction prefetching of pipelines, which is a performance killer on modern CPU. With the new core design, the transfer maps will be stored in the beam-lines nodes (i.e. *fibers* in PTC) with the element attributes. Hence, Twiss and Track should experience a significant speed-up by a factor of about 4 and 4×#cpu-core respectively. As for reentrancy, the absence of global variables and side effects is mandatory to go with parallel computation on multi-core CPUs. The replacement of all identifier strings by symbols within the lexer, i.e. Fig. 9, should also have a significant speed-up on modules that relies on lattice parameters.

Interpreter One of the main drawbacks of introducing runtime polymorphism in high performance computing is the substantial loss of performance. Few programming techniques have been tested in C and C++ to overcome the limitations induced by runtime polymorphism. First attempts to test the performance of the main loop of the evaluator, i.e. Fig. 9, have shown that the choice of the technologies (not discussed in this paper) allows to implement an extremely fast polymorphic evaluator about $\times 8 - \times 10$ faster than the mainstream interpreters, e.g. Python. For example the simple loop `while(1) x=x+1;` is evaluated 1.67 10^8 times by a single core of an Intel i5 2.3 GHz CPU, where `x` is dynamically typed as an integer and all operations are performed through polymorphic interfaces. Hence this test is a direct measure of the performance of the evaluator without cheating with built-in calls to C libraries. If these performances are confirmed on the large scale, this evaluator will offer tremendous opportunities to the MAD-X users and should deeply change the way elements and lattices might be parameterized and optimized.

Math kernel Another important aspect of the future performance of MAD-X is the efficiency the matrix-matrix and polynomial-polynomial multiplications. The former is involved in many places, from geometrical transformations to orbit and linear-term tracking (i.e. R matrix and T tensor). The latter is involved in truncated power series algebra (i.e. high-order maps tracking) and differential algebra. With the new design, it will be possible to build highly specialized modules dealing with specific tasks and tuned for specific platforms and technologies; with a particular interest for the SSE2 and AVX Intel technologies present on all platforms supported by MAD-X. A first attempt as been performed to improve the matrix-matrix multiplication using SSE2 Intrinsics and exploit the topology of the data structures to vectorize the operations.

The Fig. 11 shows timing comparison in seconds between three implementations evaluating 10^8 the matrix expression $M = M_1 M_2 + M_3 M_4$, for square matrix with dynamic sizes $N = 1..9$. The results for static sizes (i.e. known at compile time) gives another $\times 10$ improvement versus dynamics sizes (i.e. known at runtime) for all implementations, leaving unchanged the performance ratio and the outcome of the study.

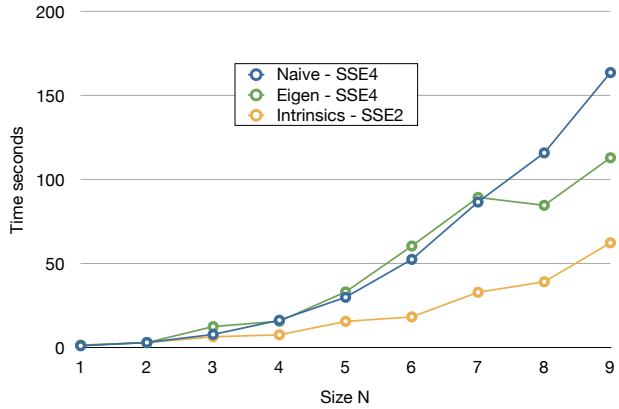


Figure 11: Timings in seconds to compute 10^8 times the matrix expression $M = M_1 M_2 + M_3 M_4$ for small dynamic matrix sizes N. For static matrix sizes N, Eigen and MAD-X equals with a $\times 10$ speed-up versus dynamic sizes.

The *naive* curve shows what C/C++ and Fortran compilers can achieve on the naive handwritten algorithm, with maximum optimizations enabled including SSE vectorization. This is the current performance status of MAD-X. The C++ Eigen library [38], one of the fastest available library with Armadillo [39], uses advanced programming technics like meta-template expressions to optimize the computations of expressions and avoid temporaries. The main goal of these C++ libraries is to provide Matlab-like syntax to programmers using operators overloading, while performing as well as plain handwritten C and Fortran code as shown by the *Eigen* curve. Finally the *Intrinsics* curve shows that handwritten code exploiting data structures topology for vectorization — an optimization beyond the capability of compilers — can perform at least $\times 2$ better than other approaches. This kind of optimization requires moderate efforts because only few functions need to be optimized, with a significant performance impact on matrix and polynomial operations at the heart of many computations in MAD-X. Estimating the performance improvement for complex computations like Track, Twiss or Match is premature and further investigations are needed.

CONCLUSIONS AND OUTLOOK

MAD-X is undergoing its first renovation process since years, a process that is meant to improve the code quality and flexibility and to restore the communication within the MAD-X community. Its maintenance and portability have been simplified and strengthened, and a bug tracking system has been setup. Some efforts are also devoted to

implement new elements in the legacy code, like the RF-Multipole necessary to evaluate the detrimental impact of the crab-cavities on the luminosity in the LHC upgrade scenarios. By the end of 2012, MAD-X should be ready to sustain new large-scale developments for the benefit of its users.

In parallel to the achievements of the past year, a structured analysis of the situation and the collected feedback from the community allowed to draw new orientations for the future development. In the present paper, we have made few proposals to deeply improve the application logic and the physics results and extend MAD-X to a wider range of uses; taking into account the needs of the future accelerator projects at CERN and the limited resource available to achieve the work over the next period of 2-3 years.

ACKNOWLEDGMENTS

We would like to thank F. Schmidt for his invaluable contribution to the project as the MAD-X custodian over the decade 2002–2011, R. De Maria for the many fruitful discussions on MAD-X, A. Latina for the implementation of the RF-Multipole, P. Skowronski for his support on PTC, and M. Giovannozzi and O. Bruning for useful feedbacks on this paper.

REFERENCES

[1] H. Grote and F. Schmidt, *"MAD-X : An Upgrade from MAD8"*, PAC'03, Portland, USA, May 2003.

[2] F. Schmidt, *"MAD-X a worthy successor for MAD8?"*, Nucl. Instrum. Methods Phys. Res., A 558, (1) 2006.

[3] E. Forest, F. Schmidt and E. Mcintosh, *"Introduction to the Polymorphic Tracking Code"*, KEK Report, 2002.

[4] W. Decking, *"Single Particle Beam Dynamics Codes*, Care HHH Workshop, CERN, November 2004.

[5] D. Einfeld, *"Optic Codes"*, OMCM Workshop, CERN, June 2011.

[6] K. Fuchsberger, X. Buffat, Y.I. Levinsen and G.J. Muller *"Status of JMAD, The Java-API for MAD-X"*, IPAC'11, San Sebastian, Spain, September 2011.

[7] G.J. Muller, *et al.*, *"Toolchain for Online Modeling of the LHC"*, ICALEPCS'11, Grenoble, 2011.

[8] MAD-X website `http://cern.ch/madx`

[9] A.H. Watson Thomas and J. McCabe, *Structured Testing: A Testing Methodology Using the Cyclomatic Complexity Metric*, NIST Special Publication 500-235, 1996.

[10] J. van Gurp, and J. Bosch, *Design, Implementation and Evolution of Object-Oriented Frameworks: Concepts & Guidelines*, Software Practice & Experience, John Wiley & Sons, March 2001.

[11] MAD-X Trac `https://svnweb.cern.ch/trac/madx`

[12] E. Forest, *et al.*, *"Synopsis of the PTC and ORBIT Integration"*, KEK Report, 2008.

[13] L. Deniau, *"MAD-X meeting"*, presentation available on [8], CERN, June 2012.

[14] L. Deniau, and A. Latina, *"Evoluation of MAD-X in the framework of LHC upgrade studies"*, IPAC'12, New Orleans, May 2012 (MOPPC074).

[15] A. Grudiev, *et al.*, *"Study of Multipolar RF Kicks from the Main Deflecting Mode in Compact Crab Cavities for LHC"*, IPAC'12, New Orleans, May 2012 (TUPPR027).

[16] S. Russenschuck, *"Field Computation for Accelerator Magnets"*, Wiley-VCH Verlag, Weinheim, 2010.

[17] R. De Maria, *"Symplectic integrator for a relativistic particle in a RF cavity"*, BE/ABP–LCU meeting, `http://cern.ch/ab-dep-abp/LCU/LCU_meetings/Minutes.html`, April 2011.

[18] F. Schmidt, *"SixTrack version 1.2"*, CERN-SL-94-56, 1994.

[19] F. Antoniou and F. Zimmermann, *"Revision of IBS with Non-Ultrarelativistic Corrections and Vertical Dispersion for MAD-X"*, Report CERN-ATS-2012-066, March 2012.

[20] A. Valishev *et al.*, *"Ring for Test of Nonlinear Integrable Optics"*, IPAC'12, New Orleans, May 2012 (WEP070).

[21] R. Talman, *"Representation of Thick Quadrupoles by Thin Lenses"*, technical report SSC-N-33, August 1985.

[22] E. Forest, *"Geometric Integration for Particle Accelerators"*, J. Phys. A: Math. Gen. **39** (2006) 5321-5377.

[23] L.M. Healy, *"Lie Algebraic Methods for Treating Lattice Parameter Errors in Particle Accelerators"*, PhD thesis, University of Maryland, 1986.

[24] H. Yoshida, *"Construction of higher order symplectic integrators"*, Physics Letters A, vol. 150, November 1990.

[25] O. Berrig, *"Comparison of transfer maps of PTC and MAD-X for the dipoles magnets: SBENDS and RBENDS"*, `http://cern.ch/cern-accelerators-optics`, CERN, 2008.

[26] V.I. Arnold, *"Mathematical Methods of Classical Mechanics"*, Springer, 2nd, 2006.

[27] E. Hairer, C. Lubich, and G. Wanner, *"Geometric Numerical Integration: Structure Preserving Algorithms for Ordinary Differential Equations"*, Springer, 2nd edition, 2006.

[28] C. Hernalsteens, *"Modeling the PS for PTC simulations"*, LIS meeting, `http://cern.ch/ab-dep-abp/LIS/Minutes/Minutes.html`, CERN, March 2012.

[29] J. Nocedal, and S.J. Wright, *"Numerical Optimization"*, Springer, 2nd edition, 2006.

[30] R. De Maria, *et al.*, *"Advances in Matching with MAD-X"*, ICAP'06, Chamonix, 2006 (WEPPP14).

[31] H. Burkhardt, M. Giovannozzi, and T. Risselada, *"Tracking LHC Models with Thick Lens Quadrupoles: Results and Comparisons with the Standard Thin Lens Tracking"*, IPAC'12, New Orleans, May 2012 (TUPPC079).

[32] S. Fartoukh, *"Proposals for changes in the MAD-X code and in the LHC sequence"*, unpublished report, CERN, December 2008.

[33] E. Keil, *"Emma in MAD-X and Comparison with Other Programs"*, CERN ATS Note 2010-044, 2010.

[34] D. Grune, and C.J.H. Jacobs, *"Parsing Technics, a Pratical Guide"*, Springer, 2008.

[35] A.V. Aho, M.S. Lam, R. Sethi, and J.D. Ullman, *"Compilers: Principles, Techniques, and Tools"*, Prentice Hall, 2nd edition, 2006.

[36] *"Advanced Memory Management Programming Guide"*, Apple Inc., 2011.

[37] P.K. Janert, *"Gnuplot in Action, Understanding Data with Graphs"*, Manning Publications, 2009.

[38] G. Guennebaud, B. Jacob, *et al.*, *"The Eigen 3 C++ Library"*, `http://eigen.tuxfamily.org`, 2010.

[39] C. Sanderson, *et al.*, *"Armadillo: An Open Source C++ Linear Algebra Library for Fast Prototyping and Computationally Intensive Experiments."*, Technical Report, NICTA, 2010.

BEAM DYNAMICS SIMULATIONS USING GPUS

J. Fitzek, S. Appel, O. Boine-Frankenheim, GSI, Darmstadt, Germany

Abstract

PATRIC is a particle tracking code used at GSI to study collective effects in the FAIR synchrotrons. Due to the need for calculation-intense simulations, parallel programming methods are being explored to optimize calculation performance.

Presently the tracking part of the code is parallelized using MPI, where each node represents one slice of the particles that travel through the accelerator. In this contribution different strategies will be presented to additionally employ GPUs in PATRIC and exploit their support for data parallelism without major code modifications to the original tracking code. Some consequences of using only single-precision in beam dynamics simulations will be discussed.

PATRIC SIMULATION CODE

The international FAIR facility with its new accelerators will be built at GSI, using the existing linac and SIS18 synchrotron as injectors. PATRIC is a particle tracking code that has been developed at the GSI accelerator physics department over many years and that is used to study collective effects in the circular accelerators within the FAIR facility. For more information on PATRIC see also [1].

Structure of the Code

Besides others, the PATRIC simulation code mainly consists of the Pic class as representation of the particles, the SectorMap class for ion optical elements like magnets including their transfer matrix, and the BeamLine class to group SectorMaps to form the accelerator (see also Figure 1). The main program takes care of the object creation, distribution of the calculation, and time measurement.

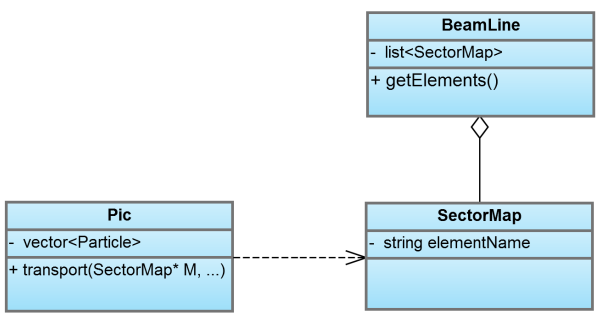


Figure 1: Structure of the PATRIC simulation code.

Existing Problem Division

Presently the tracking part of the code is parallelized using MPI, where each node represents one slice of the particles that travel through the accelerator. For tracking many thousand particles, a large number of identical calculations has to be performed on different data. Therefore the decision was to divide the problem by the data.

Dividing the data can be done in many different ways. As basis for PATRIC, the particles where diveded longitudinally in slices and assigned to MPI nodes as shown in Figure 2. Because particles that interact with each other reside on the same node, this distribution allows to include the calculation of these effects locally on the MPI node. By dividing the problem as described, PATRIC is enabled for distributed computing, which already gives a good performance gain.

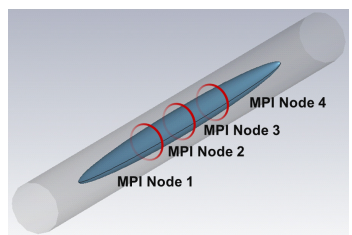


Figure 2: Structure of the existing parallelization using MPI: Each node processes one longitudinal slice of particles (figure created with CST EM Studio).

Due to the need for even more calculation-intense simulations for FAIR, parallel programming methods are being explored to optimize the calculation performance of the PATRIC simulation code beyond the possibilities of MPI. In the context of a diploma thesis different strategies are being investigated to additionally employ GPUs in PATRIC and exploit their support for data parallelism without major code modifications to the original tracking code.

Parallelization with GPUs

Today all end user computers contain powerful GPUs with remarkable floating point performance. GPUs are specialized on graphics processing and therefore support massively parallel calculations. Since they are affordable and nowadays equiped with a general purpose programming interface, they are more and more used to solve calculation-intense problems. Especially their support for data parallelism makes them suit for the calculations that have to be performed during particle simulations.

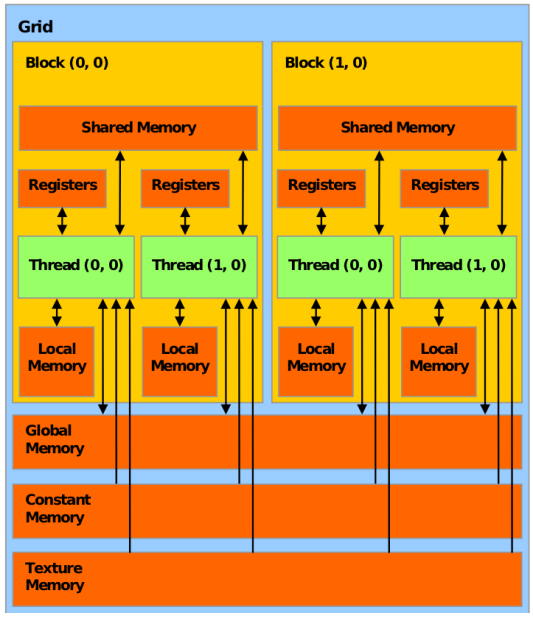


Figure 3: CUDA device memory model, NVIDIA, [2].

Mapping of the Problem

For the task to include GPUs in the existing code, an NVIDIA GPU was used together with CUDA programming. The CUDA device memory model as shown in Figure 3 allows to achieve a good structuring of the code by mapping the problem onto threads, blocks and grids. Since blocks are executed independently from each other, scalability is ensured, and the code can seamlessly be executed on e.g. a more powerful GPU without the need to rewrite it. The drawback of GPU programming is the fact that to copy data from the host to the device memory and back is very time consuming. Therefore the challenge is to find a good problem distribution onto the CPU and GPU to gain from the calculation performance benefit of the GPU.

Development of GPU Code

To develop code for NVIDIA GPUs, different programming models exist that allow either to program on a low level using CUDA for C++ [2] or to program on a high level abstraction using the template library Thrust for CUDA [3]. These possibilities are shortly presented using the example of the transfer function, i. e. a matrix-vector-multiplication as shown below.

```
for(int j=0;j<6;j++) {
    R1[j]=0.0;
    for(int l=0;l<6;l++)
      R1[j]+=T[j*6+l]*R0[l];
}
```

Figure 4: Code snippet: Original multiplication

Using the low level programming model CUDA for C++ has the advantage, that all features of the NVIDIA GPUs are available to the programmer. The API gives full control over the GPU and therefore allows to assign calculation code directly to parallel execution units by using CUDA threads, as shown in the example in Figure 5. By structuring the execution using threads, blocks and grids, the problem distribution can be handled individually. The drawback of this solution are many lines of technical code that "pollute" the original algorithm code, which leads to more complex code and reduced maintainability.

```
cudaMalloc(..);
cudaMemcpy(..);
kernel<<<dimGrid, dimBlock>>>(d_T,
  d_particleVectors, numberOfParticles);

__global__ void kernel(...) {
  (...)
  for (int i = 0; i < 6; i++) {
    value += transportMatrix[threadIdx.y * 6 + i]
      * particleVectors[particleVecStart + i];
  }
  __syncthreads();
  particleVectors[elementToCalculate] = value;
  __syncthreads();
}
```

Figure 5: Code snippet: Multiplication using CUDA for C++, each parallel execution unit calculates one element of the result vector.

Using the high level programming model Thrust for CUDA has the advantage, that an abstraction of the GPU is presented to the developer. Thrust supports differnet GPUs and programming models and is therefore used whenever interoperability is needed. Since Thrust provides many predefined functions, it also has a quick learning curve and supports rapid prototyping. The code stays clear of too many technical codelines which leads to less complex code and better maintainability. Drawback of using Thrust is that it does not provide full control over the GPU, for special purposes the low level programming must anyway be used. However, it is possible to mix these two programming models. Another aspect is that some predefined functions only support single precision floating point operations, though the underlying GPU is capable of double precision (see next section).

```
thrust::device_vector<float> x_device_vector;
x_device_vector = x_host_vector; //copy

void transport_gpu(Tmat& T, Pic_gpu& pics){
  (..)
  thrust::transform(pic_first, pic_last,
    pic_first, rotate_pic(T));
}
```

Figure 6: Code snippet: Example for using Thrust for the particle transport.

FLOATING POINT CALCULATION

Besides the challenges of code structuring, another important aspect is the fact, that older GPU models only support single precision floating point operations, which suffices for standard graphics processing but can be a problem for calculation-intense physic problems with chaotic behavior. In the past, lots of problems were not ported to the GPU because of this issue. Today, most of the recent GPU models contain double precision support with lower performance, but the performance continually gets better (up to 80% with the latest GPU models). However, not all APIs include double precision support yet. In the context of this work, this fact must also be taken into account.

FIRST APPROACHES TO EMPLOY GPUS

To employ GPUs in the PATRIC simulation code, a first approach is to simply perform the transport routine on the GPU (GPU-Version 1 in Figure 7). The static transfer matrices reside in the global memory of the device (the GPU). During each transport step, the particle vectors are copied to the device, the transfer is calculated and the result is copied back to the host. This approach is simple but as measurements show the execution time rises due to the copy overhead in each step.

A second approach is to keep also the particles in the global memory of the GPU (GPU-Version 2 in Figure 7). Transport steps can then be performed only on the GPU which eliminates most of the copy steps. However, intermediate calculations at predefined steps during the overall calculation must then either be done also on the GPU or – as one requirement is to keep existing calculation routines – intermediate particle information must be copied back to the host to perform these calculations.

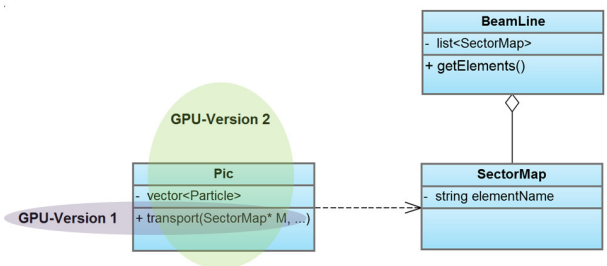


Figure 7: PATRIC code with GPU-enabled parts.

FIRST MEASUREMENTS

For first measurements, a desktop computer with six Intel 2.67GHz CPUs and one NVIDIA Tesla C2075 GPU was used. The code was executed with a constant focussing optic, 10000 macro particles, and 128 cells. The number of MPI processes was set to one, to exclude impacts on the performance due to oversubscription of the CPUs and due to the fact, that only one single GPU was available. Average execution times were taken from 20 runs. Figure 8 shows the results of the first measurements. The 10 seconds longer execution time of the GPU-Version 1 compared to the original version can be explained with the copy overhead within each transport step. This approach seems not ideal, since too many copy steps outweigh the calculation performance gain. This leads to the conclusion, that larger parts of the calculation must be moved to the GPU to achieve a performance gain. As a next step, the GPU-Version 2 approach will be implemented and measured; positive speedup is expected when the particles can be always kept on the GPU.

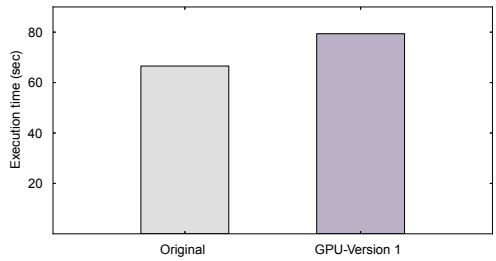


Figure 8: First measurements of the code variants.

OUTLOOK

Employing GPUs in the PATRIC code will be elaborated in the context of a diploma thesis for the GSI accelerator physics department, supervised by Prof. Dr. Keller, subject Parallelism and VLSI at the FernUniversität Hagen. As part of this diploma thesis, different variants will be implemented and measured, including different structuring of the code, usage of the low or high level API (performance versus usability), and the utilization of single or double precision (performance versus error cumulation).

Already the first tests indicate, that the copy overhead is one of the limiting factors. Structuring the code in such a way, that bigger amounts of related calculations can be performed on the GPU might be a promising approach. Therefore, the idea to completely keep the particles on the GPU will be investigated further. The constraint is that it must still be possible to include existing particle quantity calculations (e. g. emittance) at predefined points in the overall execution.

If employing of GPUs succeeds, future extensions are foreseen that include e.g. calculation of space charge effects and particle kicks locally on each node again by using the GPU do do these calculations.

REFERENCES

[1] O. Boine-Frankenheim, V. Kornilov, "Simulation of Transverse Coherent Effects in Intense Ion Bunches", this conference, TUAAI3.

[2] NVIDIA Corporation, "CUDA Programming Guide", Santa Clara USA, 2007.

[3] Thrust, open source parallel algorithms library, available at `http://code.google.com/p/thrust`

AN OpenMP PARALLELIZATION OF REAL-TIME PROCESSING OF CERN LHC BEAM POSITION MONITOR DATA

H.Renshall, L.Deniau, CERN, Geneva, Switzerland

Abstract

SUSSIX is a FORTRAN program for the post processing of turn-by-turn Beam Position Monitor (BPM) data, which computes the frequency, amplitude, and phase of tunes and resonant lines to a high degree of precision. For analysis of LHC BPM data a specific version run through a C steering code has been implemented in the CERN Control Centre to run on a server under the Linux operating system but became a real time computational bottleneck preventing truly on-line study of the BPM data. Timing studies showed that the independent processing of each BPMs data was a candidate for parallelization and the Open Multiprocessing (OpenMP) package with its simple insertion of compiler directives was tried. It proved to be easy to learn and use, problem free and efficient in this case reaching a factor of ten reductions in real-time over twelve cores on a dedicated server. This paper reviews the problem, shows the critical code fragments with their OpenMP directives and the results obtained.

THE PROBLEM

SUSSIX is a FORTRAN program for the post processing of turn-by-turn Beam Position Monitor (BPM) data, which computes the frequency, amplitude, and phase of tunes and resonant lines to a high degree of precision through the use of an interpolated Fast Fourier Transform (FFT). Analysis of such data represents a vital component of many linear and non-linear dynamics measurements.

For analysis of LHC BPM data a specific version *sussix4drive*, run through the C steering code *Drive God Lin*, has been implemented in the CERN Control Centre (CCC) by the beta-beating team. Analysis of all LHC BPMs, however, represents a major real time computational bottleneck in the control room, which has prevented truly on-line study of the BPM data. In response to this limitation an effort has been underway to decrease the real computational time, with a factor of 10 as the target, of the C and FORTRAN codes by parallelising them.

SOLUTIONS CONSIDERED

Since the application is run on dedicated servers in the CCC the obvious technique is to profit from the current multi-core hardware: 24 cores are now typical. The initial thinking was to parallelise the FFT code.

The first attempts were to try a parallelised FFT from the Numerical Algorithms Group (NAG) `fsl6i2dcl` library for SMP and multicore processors together with the Intel 64-bit FORTRAN compiler and the Intel maths kernel library recommended by NAG. This library uses the OpenMP technology.

Various NAG examples of enhanced routines were run (but not the multi-dimensional FFTs) and all slowed down in real time using more cores. This was not surprising since the examples only take milliseconds, comparable to the overhead to launch a new thread. Also it was found that the SUSSIX application calls `cfft` (D704 in the CERN program library), which maps onto NAG `c06ecf`, which had not yet been enhanced. This led to making a detailed central processing unit (CPU) profiling of the application.

Profiling the application (with gprof) showed that in fact only 7.5% of the CPU time was spent in `cfft` while 70% was spent in a function `zfunr` searching for the maximum of the Fourier spectra with large numbers of executions of an efficient inner loop of BPM data over many turns. This loop could not be improved (`maxd` turns of a BPM data is typically 1000):

```
double complex zp,zpp,zv
zpp=zp(maxd)
do np=maxd-1,1, -1
  zpp=zpp*zv+zp(np)
enddo
```

It was decided to try and parallelise SUSSIX directly using the OpenMP implementation supported by the Intel and GCC compilers. The home web site is www.openmp.org and an excellent tutorial at computing.llnl.gov/tutorials/openMP was the main reference. Examination of the code granularity revealed that the highest level of independent code execution was over the processing of individual BPM data.

The pure FORTRAN offline version was parallelised first by adding OpenMP parallelisation directives around the main BPM loop. In this version each BPMs data is in a separate physical file hence they could be opened and read in parallel:

```
!$OMP PARALLEL DO PRIVATE(n,iunit,
                          filename,nturn)
!$OMP& SHARED (isix,ntot,narm,iana,…,
               iicf)
do n=1,ntot ! Parallel loop over BPM
  call datspe(iunit,idam,ir,nt1,nt2,
             nturn,…,iana)
  call ordres(eps,narm,nrc,idam,n,nturn)
enddo
!$OMP END PARALLEL DO
```

In addition, `!$OMP THREADPRIVATE` directives were added for all non-shareable variables in the called subroutine trees of `datspe` and `ordres`.

This gave good scaling up to 10 cores on a public shared 16-core CERN computer centre *lxplus* machine so was worth extending to the target mixed C and FORTRAN version to be run in the control room. In this version the BPM data are all read into memory from a single file before being passed sequentially into the processing loop. The FORTRAN BPM processing loop is called from C and was parallelised with the similar OpenMP C-code syntax and gave the same scaling result:

```
#pragma omp parallel private(i,ii,ij,kk)
#pragma omp for
for (i=pickstart; i<=maxcounthv; i++) {
  sussix4drivenoise_(&doubleToSend[0],
                     &tune[0],&amplitude[0])
#pragma omp critical
// … I/O loop with sequential execution
}
```

The FORTRAN `datspe` and `ordres` call trees were unchanged.

The OpenMP directives multi-thread the code and the threads then map onto physical cores in a multi-core machine. The run-time environment variable `OMP_NUM_THREADS` instructs OpenMP how many threads, hence cores, it can use for an execution and enables easy measurement of the scaling.

Since the order of processing of individual BPMs is arbitrary the results file is post-processed by the Unix sort called as part of the application to give the same results output as a non-parallel execution.

A test case of real 1000 turns LHC BPM data, analyzed to find 160 lines, was run on a reserved 24 cores machine *cs-ccr-spareb7* in the LHC CCC.

A normal run of this test case takes about 50 seconds on this machine. The observed wall-time speedup of C-FORTRAN SUSSIX as a function of the number of cores (from E. Maclean,CERN,BE) is shown below in figure 1.

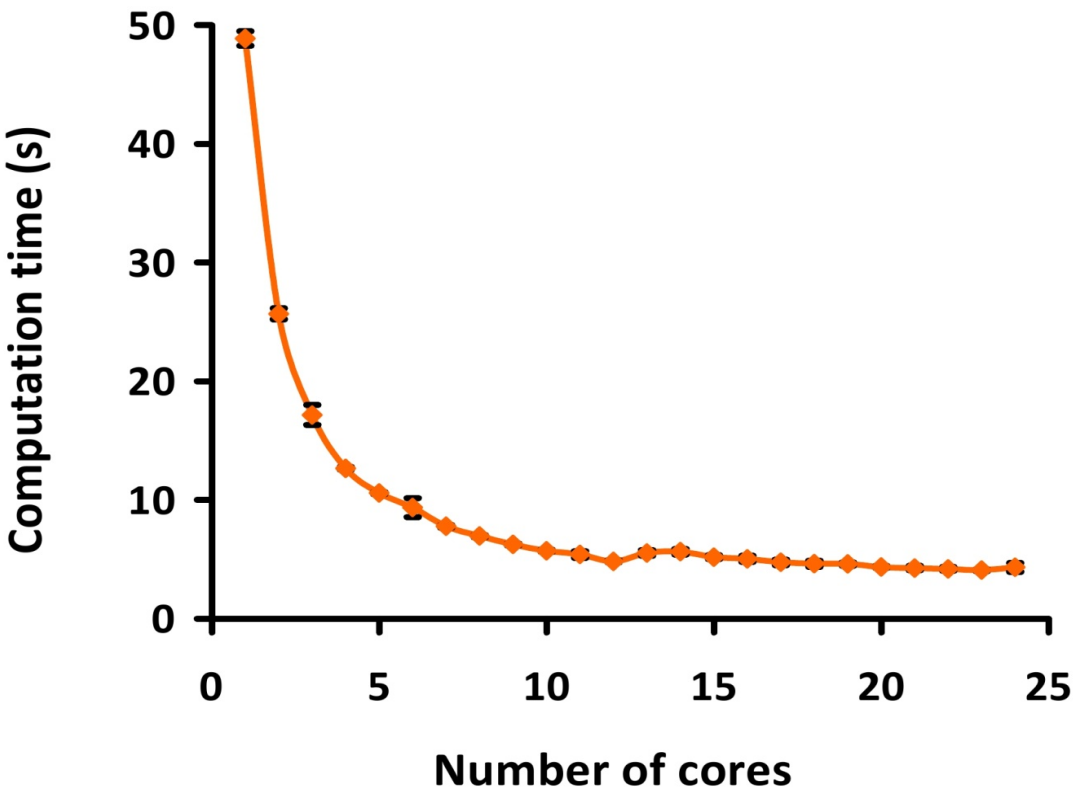


Figure 1: Real time in seconds to solution with increasing number of cores.

CONCLUSIONS

About a factor of 10 improvement in the real computation time has been realized for this test case saturating at 12 cores, probably due to memory bandwidth limits (other possible causes were investigated). For the study of amplitude detuning reported in CERN-ATS-Note-2011-052 MD [1] the parallelised C-FORTRAN SUSSIX was utilized within the LHC beta-beat graphical

user interface [2] and the target tenfold real-time reduction was verified in practice.

There was a modest learning curve for OpenMP but the Lawrence Livermore National Laboratory tutorial was easy to read and understand with its simple FORTRAN and C code fragment examples. Most of the code changes were insertion of OpenMP directives and these were correctly implemented by the Intel FORTRAN and GCC compilers.

This simple robust technique could be of interest to other real-time applications where a significant fraction of the CPU time is spent in independent calculations which have been implicitly serialized by being written in standard FORTRAN or C.

REFERENCES

[1] M. Albert et al., "Non-linear beam dynamics tests in the LHC" CERN-ATS-Note-2011-052 MD, July 2011.

[2] M. Aiba et al., "Software package for optics measurement and correction in the LHC", CERN-ATS-2010-092.

SOME COMPUTATIONAL CHALLENGES IN THE MODELING OF ACCELERATORS AND THEIR SOLUTIONS IN THE SIMULATION CODE WARP

J.-L. Vay*, C. Benedetti, LBNL, USA
D.P. Grote, R.H. Cohen, A. Friedman LLNL, USA

Abstract

This paper presents an overview of the Particle-In-Cell accelerator code Warp's capabilities, summarizing recent original numerical methods that were developed within Warp, including a large-timestep "drift-Lorentz" mover for arbitrarily magnetized species, a relativistic Lorentz invariant leapfrog particle pusher, an electromagnetic solver with tunable numerical dispersion and efficient stride-based digital filtering, Particle-In-Cell with Adaptive Mesh Refinement, and simulations in Lorentz boosted frames.

INTRODUCTION

The Particle-In-Cell (PIC) Framework Warp [1] was originally developed to simulate space-charge-dominated beam dynamics in induction accelerators for heavy-ion fusion (HIF) [2]. In recent years, the physics models in the code have been generalized, so that Warp can model beam injection, complicated boundary conditions, denser plasmas, a wide variety of accelerator lattice components, and the non-ideal physics of beams interacting with walls and plasmas. The code now has an international user base and is being applied to projects both within and far removed from the HIF community. Ongoing or recent examples of applications outside HIF include the modeling of plasma traps for the production of anti-Hydrogen [3], Paul traps [4, 5], non-conventional Penning-Malmberg micro-trap [6], transport of electron beams in the UMER ring [7], ECR ion sources [8], capture and control of laser-accelerated proton beams [9], and fundamental studies of multipacting [10]. It is also applied to the study and design of existing and next generation high-energy accelerators including the study of electron cloud effects [11], coherent synchrotron radiation [12] and laser wakefield acceleration [13].

These studies have necessitated the introduction or development of advanced numerical methods, including methods to model multiple-species effects in accelerators and chambers, efficient ensemble methods, particle advance algorithms that allow a longer time step, and adaptive mesh refinement (AMR).

NOVEL ALGORITHMS

Hybrid Drift-Lorentz

It was observed in [14] that the Boris pusher causes particles to gyrate with spuriously large radius for time steps that are large compared to the gyroperiod, albeit with the correct drift velocities (provided the gradients are still sampled adequately). A new solver that interpolates between the Boris velocity push and a drift kinetic advance was developed and implemented in Warp [15, 17]; it reproduces both the correct drift velocity and gyroradius for an arbitrarily large ratio of time step δt relative to cyclotron period τ_c, as well as correct detailed orbit dynamics in the small-timestep limit. The pusher has provided an order of magnitude or more saving in computing resources in the simulations of electron cloud effects in the HCX experiment [15, 16]. An implicit time-advance scheme incorporating drift-Lorentz interpolation has also been developed [17].

Lorentz Invariant Advance

The relativistic version of the Boris (or Hybrid Lorentz-Drift) particle pusher does not maintain strict Lorentz invariance, resulting eventually in unacceptably large inacuracies when modeling the transport of ultra-relativistic beams in accelerators. To this effect, an alternative to the Boris pusher that conserves strict Lorentz invariance (to machine precision) was developed and implemented in Warp, and its effectiveness demonstrated on the modeling from first principles of the interaction of a 500 GeV proton beam with a background of electrons [18]. The pusher has subsequently been implemented by others and has also proven useful for correctly capturing the drift speed of electrons of a highly magnetized relativistic electron-ion flow in astrophysical simulations using the code TRISTAN [19].

Electromagnetic Solver

Warp's electromagnetic solver is based on the Non-Standard Finite-Difference (NSFD) technique [20, 21], which is an extension of the Finite-Difference Time-Domain technique to larger stencils in the plane perpendicular to the direction of the finite difference. This gives the user some control on the numerical dispersion and Courant time step limits which do depend on those parameters. As shown in [22], for a given set of parameters, and for cubic cells, the Courant time step multiplied by the speed of light equals the cell size, and the numerical dispersion vanishes along the main axes. More details on the solver implementation and characteristics for several sets of coefficients are available in [23]. Also described in [23] are the implementation of Perfectly Matched Layers for the absorption of waves at grid boundaries and of Friedman's damping algorithm for noise control [24]. In the same paper, it is shown that introducing a stride in the usage of standard linear filtering allows for construction of efficient iterative sideband

* jlvay@lbl.gov

digital filters that are nonetheless compact, thus well suited for implementation on parallel computers.

Mesh Refinement

The mesh refinement methods that have been implemented in Warp were developed following the following principles: i) avoidance of spurious effects from mesh refinement, or minimization of such effects; ii) user controllability of the spurious effects' relative magnitude; iii) simplicity of implementation. The two main generic issues that were identified are: a) spurious self-force on macroparticles close to the mesh refinement interface [25, 26]; b) reflection (and possible amplification) of short wavelength electromagnetic waves at the mesh refinement interface [27]. The two effects are due to the loss of translation invariance introduced by the asymmetry of the grid on each side of the mesh refinement interface.

A cornerstone of the Particle-In-Cell method is that assuming a particle lying in a hypothetical infinite grid, then if the grid is regular and symmetrical, and if the order of field gathering matches the order of charge (or current) deposition, then there is no self-force of the particle acting on itself: a) anywhere if using the so-called "momentum conserving" gathering scheme; b) on average within one cell if using the "energy conserving" gathering scheme [28]. A breaking of the regularity and/or symmetry in the grid, whether it is from the use of irregular meshes or mesh refinement, and whether one uses finite difference, finite volume or finite elements, results in a net spurious self-force (which does not average to zero over one cell) for a macroparticle close to the point of irregularity (mesh refinement interface for the current purpose) [25, 26].

A method was devised and implemented in Warp for reducing the magnitude of spurious self-forces near the coarse-fine boundaries as follows. Noting that the coarse grid solution is unaffected by the presence of the patch and is thus free of self-force, extra "transition" cells are added around the "effective" refined area. Within the effective area, the particles gather the potential in the fine grid. In the extra transition cells surrounding the refinement patch, the force is gathered directly from the coarse grid. The number of cells allocated in the transition zones is controllable by the user in Warp, giving the opportunity to check whether the spurious self-force is affecting the calculation by repeating it using different thicknesses of the transition zones. Automatic remeshing has been implemented in Warp following the procedure described in [29], refining on criteria based on measures of local charge density magnitude and gradients.

The method that is used for electrostatic mesh refinement is not directly applicable to electromagnetic calculations. As was shown in section 3.4 of [30], refinement schemes relying solely on interpolation between coarse and fine patches lead to the reflection with amplification of the short wavelength modes that fall below the cutoff of the Nyquist frequency of the coarse grid. Unless these modes are damped heavily or prevented from occurring at their source, they may affect particle motion and their effect can escalate if trapped within a patch, via multiple successive reflections with amplification.

To circumvent this issue, an additional coarse patch (with the same resolution as the parent grid) is added, as described in [31]. Both the fine and the coarse grid patches are terminated by Perfectly Matched Layers, reducing wave reflection by orders of magnitude, controllable by the user [32, 33].

The source current resulting from the motion of charged macroparticles within the refined region is accumulated on the fine patch and is then interpolated onto the coarse patch and added onto the parent grid. The process is repeated recursively from the finest level down to the coarsest. The Maxwell equations are then solved for one time interval on the entire set of grids, by default for one time step using the time step of the finest grid. The field on the coarse and fine patches only contain the contributions from the particles that have evolved within the refined area but not from the current sources outside the area. The total contribution of the field from sources within and outside the refined area is obtained by adding the field from the refined grid $F(r)$, and adding an interpolation I of the difference between the relevant subset s of the field in the parent grid $F(s)$ and the field of the coarse grid $F(c)$, on an auxiliary grid a, i.e. $F(a) = F(r) + I[F(s) - F(c)]$. The field on the parent grid subset $F(s)$ contains contributions from sources from both within and outside of the refined area. Thus, in effect, there is substitution of the coarse field resulting from sources within the patch area by its fine resolution counterpart. The operation is carried out recursively starting at the coarsest level up to the finest.

Lorentz Boosted Frame

A method was recently proposed to speed up full PIC simulations of a certain class of relativistic interactions by performing the calculation in a Lorentz boosted frame [34], taking advantage of the properties of space/time contraction and dilation of special relativity to render space and time scales (that are separated by orders of magnitude in the laboratory frame) commensurate in a Lorentz boosted frame, resulting in far fewer computer operations. The method has been applied successfully to the modeling of laser plasma acceleration [35, 36, 37, 13], electron cloud effects [18], free electron lasers [38], coherent synchrotron radiation [12], and production of ultrabright attosecond x-ray pulses [39].

In a laser plasma accelerator, a laser pulse is injected through a plasma, creating a wake of regions with very strong electric fields of alternating polarity [40]. An electron beam that is injected with the appropriate phase can thus be accelerated to high energy in a distance that is much shorter than with conventional acceleration techniques [41]. The simulation of a laser plasma acceleration stage from first principles using the Particle-In-Cell

technique in the laboratory frame is very demanding computationally, as the evolution of micron-scale long laser oscillations needs to be followed over millions of time steps as the laser pulse propagates through a meter long plasma for a 10 GeV stage. In the laboratory frame the laser pulse is much shorter than the wake, whose wavelength is also much shorter than the acceleration distance ($\lambda_{laser} \ll \lambda_{wake} \ll \lambda_{acceleration}$). In a Lorentz boosted frame moving at a speed near the speed of light with the laser in the plasma, the laser will be Lorentz expanded (by a factor $(1 + v_f/c)\gamma_f$ where $\gamma_f = (1 - v_f^2/c^2)^{-1/2}$ and v_f is the velocity of the frame and c is the speed of light). The plasma (now moving opposite to the incoming laser at velocity $-v_f$) is Lorentz contracted (by a factor γ_f). In a boosted frame moving with the wake ($\gamma_f \approx \gamma_{wake}$), the laser wavelength, the wake and the acceleration length are now commensurate ($\lambda_{laser} < \lambda_{wake} \approx \lambda_{acceleration}$), leading to far fewer time steps by a factor $(1 + v_f/c)^2\gamma_f^2$, hence computer operations [34, 13].

A speedup of up to a million times was reported for Warp modeling of a hypothetical 1 TeV stage [42]. Control of a violent numerical instability (which nature is being investigated) that had been plaguing early attempts was obtained via the combination of: (i) the use of Warp's tunable electromagnetic solver and efficient wideband filtering [23], (ii) observation of the benefits of hyperbolic rotation of space-time on the laser spectrum in boosted frame simulations [42], and (iii) identification of a special time step at which the growth rate of the instability is greatly reduced [23]. In addition, a novel numerical method for injecting the laser pulse through a moving planar antenna was introduced in Warp [13].

CONCLUSION

The Warp code-framework has recently been augmented with various novel methods including PIC with adaptive mesh refinement, a large-timestep mover for particles of arbitrary magnetized species, a new relativistic Lorentz invariant leapfrog particle pusher, simulations in Lorentz boosted frames, an electromagnetic solver with tunable numerical dispersion and efficient stride-based digital filtering. With its new capabilities and thanks to a design that allows for a high degree of versatility, the range of application of Warp has considerably widened far beyond the initial application to the Heavy Ion Fusion Science program.

ACKNOWLEDGMENTS

Work supported by US-DOE Contracts DE-AC02-05CH11231 and DE-AC52-07NA27344, the US-DOE SciDAC program ComPASS and the US-LHC Accelerator Research Program (LARP). Used resources of NERSC, supported by US-DOE Contract DE-AC02-05CH11231.

REFERENCES

[1] D. Grote, *et al.*, AIP Conf. Proc., no. 749, 2005, pp. 55–8.
[2] B. Logan, *et al.*, Nucl. Instr. & Meth. A 577 (2007) 1–7.
[3] K. Gomberoff, *et al.*, Phys. Plasmas 14 (10).
[4] E. P. Gilson, *et al.*, Phys. Plasmas 17 (5)
[5] S. Ohtsubo, *et al.* Phys. Rev. ST-AB 13 (4).
[6] P. Folegati, *et al.*, J. of Physics: Conf. Series 262 (2011).
[7] I. Haber, *et al.* Nucl. Instr. & Methods A 606 (2009) 64–68
[8] D. Winklehner, *et al.* J. Instr. 5 (2010).
[9] F. Nürnberg *et al.*, J. Physics: Conf. Series 244 (2010).
[10] R. A. Kishek, *et al.*, Phys. Rev. Lett. 108 (2012) 035003.
[11] J.-L. Vay, *et al.*, Proc. Part. Accel. Conf. (2011).
[12] W. Fawley, *et al.* Proc. Int. Part. Accel. Conf., (2010).
[13] J. L. Vay, *et al.*, Phys. Plasmas 18 (12).
[14] S. Parker, *et al.*, 97 (1) (1991) 91–102.
[15] R. H. Cohen, *et al.*, Phys. Plasmas 12 (5) (2004).
[16] A. W. Molvik, *et al.*, Phys. Plasmas 14 (2006).
[17] R. H. Cohen, *et al.*, Nucl. Instr. & Meth. A 606 (2009) 53.
[18] J. L. Vay, Phys. Plasmas 15 (2008) 056701.
[19] L. Sironi, A. Spitkovsky, private Communication (2011).
[20] J. Cole, IEEE Trans. Microwave Th. & Tech. 45 (1997) 991.
[21] J. Cole, IEEE Trans. Ant. & Prop. 50 (2002) 1185–1191.
[22] M. Karkkainen, *et al.*, ICAP, Chamonix, France, (2006).
[23] J. L. Vay, *et al.*, J. Comput. Phys. 230 (2011) 5908–5929.
[24] A. Friedman, J. Comput. Phys. 90 (1990) 292–312.
[25] J. Vay, *et al.*, Laser & Part. Beams 20 (2002) 569–575.
[26] P. Colella, P. C. Norgaard, J. Comput. Phys. 229 (2010) 947.
[27] J.-L. Vay, J. Comput. Phys. 165 (2) (2000) 511–521.
[28] C. Birdsall, A. Langdon, Plasma physics via computer simulation, Adam-Hilger, 1991.
[29] J.-L. Vay, *et al.* Nucl. Instr. & Meth. A 544 (2005) 347–352
[30] J.-L. Vay, J. Comput. Phys. 167 (2001) 72–98.
[31] J.-L. Vay, *et al.*, Comput. Phys. Comm. 164 (2004) 171–177.
[32] J. Berenger, J. Comput. Phys. 127 (1996) 363–379.
[33] J.-L. Vay, J. Comput. Phys. 183 (2002) 367–399.
[34] J.-L. Vay, Phys. Rev. Lett. 98 (2007) 130405.
[35] D. Bruhwiler, *et al.* AIP Conf. Proc., 1086, (2009), 29–37.
[36] J.-L. Vay, *et al.*, Proc. PAC09, Vancouver, Canada (2009).
[37] S. F. Martins, *et al.*, Nat. Physics 6 (2010) 311–316.
[38] W. Fawley, J.-L. Vay, Proc. FEL 2010, Sweden (2010).
[39] H. C. Wu, *et al.* Phys. Rev. ST-AB 14 (2011).
[40] T. Tajime, J. Dawson, Phys. Rev. Let. 43 (1979) 267.
[41] W. P. Leemans, *et al.* Nat. Physics 2 (2006) 696–699.
[42] J. Vay, *et al.*, Phys. Plasmas 18 (2011) 030701.

BEAM DYNAMICS STUDIES FOR PARTICLE DRIVEN PLASMA WAKEFIELD ACCELERATION EXPERIMENTS AT PITZ

M. Khojoyan*†,M. Gross, G. Klemz, G. Koss, M. Krasilnikov, A. Oppelt, F. Stephan,
DESY, Zeuthen, Germany

Abstract

The Photo Injector Test Facility at DESY, Zeuthen site (PITZ) is developing and optimizing high brightness electron sources for linac based free electron lasers such as FLASH and the European XFEL. The high quality of the 25 MeV electron beam together with the availability of a highly flexible photocathode laser system makes the PITZ injector a perfect facility for variety of experimental studies. Two approaches are of great interest for future applications in the context of particle driven plasma wakefield acceleration experiments: self-modulation and transformer ratio studies. In both cases a high density electron beam is interacting with plasma which has a density of about $10^{15}\,\mathrm{cm}^{-3}$. ASTRA [1] simulations were done to study the e-beam density along the existing PITZ beamline, especially at two different possible longitudinal positions of the planned plasma cell, in order to reach the particle density required for occurrence of self-modulation. The results of the beam dynamics studies are presented and discussed in this paper.

MOTIVATION

Plasma is a promising medium for high gradient acceleration of charged particles. It can sustain orders of magnitude higher fields compared to conventional accelerators. Large amplitude plasma waves can be excited either by powerful lasers [2] or by charged particle beams [3]. In both cases either the pulse length of the intense laser or the temporal length of the particle beam should be in the order of the plasma wavelength. In a particle driven wakefield acceleration scheme, the fields generated by the drive bunch / bunches are used to accelerate a witness bunch. The transformer ratio $\mathrm{R} \equiv \mathrm{W}_\mathrm{w}^+/\mathrm{W}_\mathrm{d}^-$(maximum energy gain of the witness bunch divided by maximum energy loss of the drive bunch) is an important parameter which in general is limited to two [4], except in special cases when the bunch has an asymmetric current profile [5]. The most successful experiment in concept of particle driven plasma wakefield acceleration was performed at Stanford Linear Accelerator Center (SLAC), by doubling the energy of an ultra-short (50fs) 42 GeV electron beam [6]. Although two order of magnitudes higher accelerating gradients compared to RF accelerators were demonstrated, the experimental results revealed the limitation of the transformer ratio. An other idea wa s to accelerate an electron beam up to hundreds of GeV by using high energy ultra-short proton beams as a driver [7]. The plasma in this case will play the role of a medium, which transfers the energy of the protons to the energy of the electrons. As currently available high energy proton beams are tens of centimeters long, the self-modulated regime for long proton beams was afterwards suggested to excite large amplitude wakefields in plasma [8]. When a long particle beam (bunch length is long compared to the plasma wavelength) enters the plasma, it is radially modulated by the periodic focusing/defocusing forces and the beam density modulation provides a positive feedback for the instability to grow [9]. For the case of the CERN 450 GeV proton driver beam, simulations show that an injected 10 MeV witness electron beam can be accelerated to an energy of several hundred MeV after a few meters propagation in a $7 \cdot 10^{14}\,\mathrm{cm}^{-3}$ density plasma [10]. For beams which are longer than the plasma wave period, where self-modulation occurs, the wakefield properties are strongly affected by the drive beam dynamics [11]. PITZ is a known facility for generating and characterizing high brightness electron sources. Recently PITZ has become a very attractive facility for different challenging experiments due to the excellent transverse beam properties such as the transverse emittance which turned out to be unique over a wide range of bunch charges [12]. PITZ can make an important contribution to the questions related to self-modulation by measuring the longitudinal phase space properties of a relatively long 25 MeV electron beam after interaction with a $10^{15}\,\mathrm{cm}^{-3}$ dense plasma. The basic observation would be to detect electron beam energy modulation $> 1\%$, which should be possible with the recently available PITZ diagnostics [13]. In this work two possible longitudinal positions of the planned plasma cell arranged into the PITZ linac are considered. Beam dynamics studies were done predicting the possibility of high beam densities required for the self-modulation process.

INTRODUCTION: BEAM PARAMETERS FOR SIMULATIONS

An overview of the current PITZ setup is shown in Fig. 1. The highly flexible laser system at PITZ, which was developed by the Max-Born institute, produces flat-top laser pulses with sharp rise and fall times [14]. High brightness electron beams are generated using a Cs_2Te photocathode and are accelerated in an L-band RF gun cavity up to 7 MeV (low energy section). The electron beam gets its final energy of about 25 MeV by the cut disc structure (CDS) booster cavity. A dipole can be used to deflect the beam by 180° for different studies, particularly for the fi-

* martin.khojoyan@desy.de
† On leave from ANSL, Yerevan, Armenia

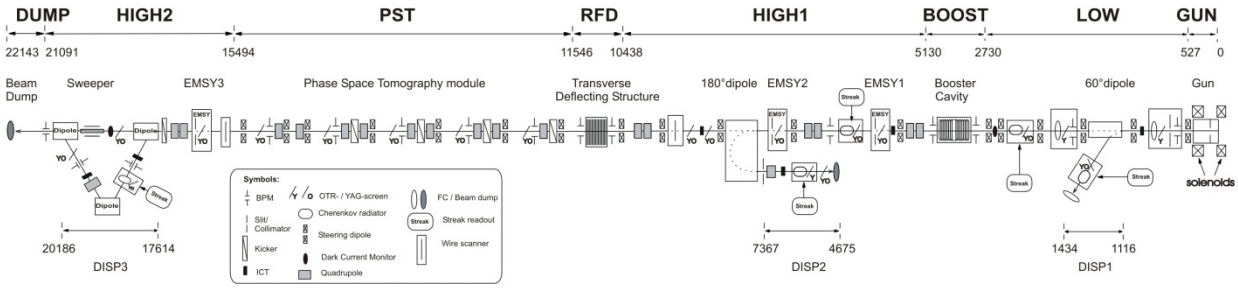


Figure 1: Schematic view of the current PITZ beamline. The electron beam propagates towards the left.

nal beam energy measurements. Two pairs of identical quadrupole magnets are located between the booster cavity and the 180° dipole. The plasma cell should be inserted before the matching section for the phase space tomography module (PST) which starts at $\sim 10\,\mathrm{m}$ after the cathode. The RF deflector (RFD) for measuring the longitudinal beam properties is situated right after the matching section. A Space charge TRacking Algorithm (ASTRA) was applied to track the electron beam to two possible central positions of the plasma chamber (which is assumed to be $\sim 1\,\mathrm{m}$ long) either directly in front of or behind the 180° dipole, i.e. until the point where the beam will enter the $\sim 30\,\mathrm{cm}$ long plasma. The goal in the simulations was to reach the highest possible electron beam density at the starting point of the beam-plasma interaction by smooth transverse focusing of the beam. In the simulation setup

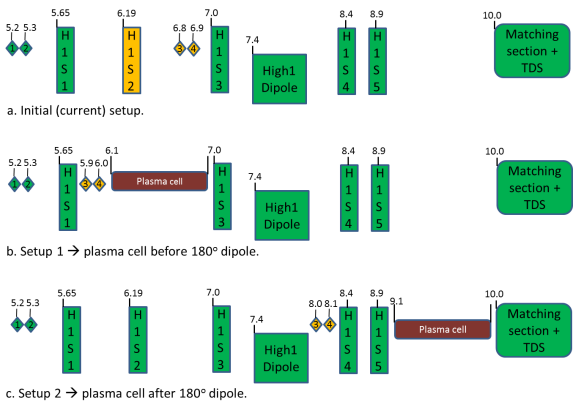


Figure 2: The scheme of possible scenarios for plasma cell insertion.

the photocathode laser temporal profile was considered a flat-top with 22 ps FWHM length and with 2 ps rise and fall times. The laser rms spot size at the cathode was chosen to be $0.3\,\mathrm{mm}$ for $100\,\mathrm{pC}$ bunch charge. The gradient of the RF gun was set to $61\,\mathrm{MV/m}$ corresponding to $7\,\mathrm{MeV/c}$ beam momentum after the gun. The peak field in the second accelerating module (CDS booster) was tuned to achieve a final beam momentum of $25\,\mathrm{MeV/c}$. 0.5 Million macro-particles have been used for the simulations. In both cases the magnetic strength of the main solenoid was chosen for minimum beam spot size at the first screen station in the high energy section (EMSY1 in Fig.1). Two pairs of quadrupoles are used to further focus the beam transversely. The simulations were carried out in two steps. Firstly, the electron beam was tracked to $5.1\,\mathrm{m}$ downstream of the cathode by using the 2D space charge routine in ASTRA. The quadrupole focusing was applied afterwards for the final beam transverse focusing by using the 3D space charge settings. The principal part of the simulation setup starting from the first quadrupole in the high section until the matching section is illustrated in Fig. 2. The components which should not be removed / replaced are colored in green. The yellow marked components can be removed / replaced. The plasma cell is presented in red. The numbers written on the components convey their distance in meters calculated from the cathode. Figure 2 (a) shows the current situation. Two possible scenarios for the plasma cell ar-

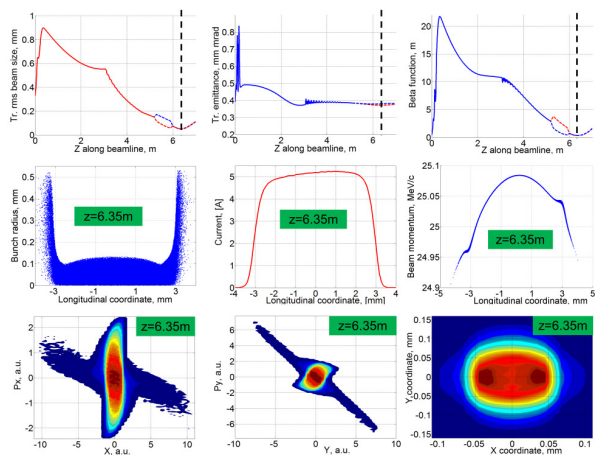


Figure 3: Beam transverse properties along the PITZ beamline and the beam overview at 6.35m downstream the cathode (setup1).

rangement in the PITZ beamline are described as setup1 (b) and setup2 (c). The quadrupoles are introduced by numbers, from one to four as indicated in Fig. 2, and the screen stations are presented as H1S#, i.e. H1S1 being the first screen in the high section. The electron beam properties for the setup1 (b), where the plasma cell is situated before the 180° dipole are given in Fig. 3. The first line shows the electron beam transverse sizes, emittances as well as

the beta functions (different colors correspond to the parameters in horizontal and vertical planes) along the PITZ beamline. The bunch radius (assuming cylindrical symmetry), the charge distribution and the longitudinal momentum vs. the longitudinal coordinate are shown in the second row. Transverse phase spaces and the beam projection are presented in the last line. The transverse electron beam sizes along the beamline for the two different setups are shown in Fig. 4. The analysis of the beam overviews for the

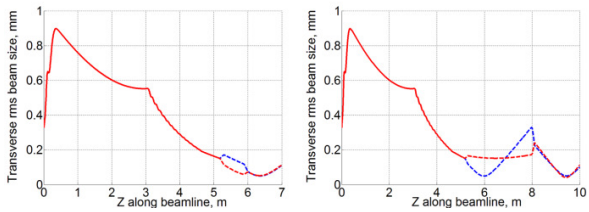


Figure 4: Transverse beam focusing for two different positions of the plasma cell. Different colors correspond to rms beam sizes in different transverse planes.

two cases (Tab.1) suggests the first setup (plasma cell before the 180° dipole) as a better candidate compared to the second case (plasma cell after the 180° dipole) which requires more complex transverse beam focusing. The peak beam density values in Tab.1 are calculated using a formula which assumes Gaussian shape of the electron beam.

Table 1: E-beam Output Parameters for Two Different z Positions of the PITZ Linac (before and after the 180° dipole)

Beam parameters	Before	After
Total charge, pC	100	100
Long. beam position, mm	6.35	9.45
Hor. rms beam size, um	49.8	59.4
Ver. rms beam size, um	51.2	59.4
Bunch length in FWHM, mm	5.92	5.92
Average kinetic energy, MeV	24.55	24.55
RMS energy spread, keV	27.5	27.8
Peak slice current, A	5.2	5.2
Hor. emittance, mm mrad	0.37	0.38
Ver. emittance, mm mrad	0.38	0.36
Long. emittance, keV mm	42.4	42.4
Beam density, $10^{12}\,\mathrm{cm}^{-3}$	9.1	6.6

SELF-MODULATION AT PITZ

The self-modulation instability is a result of a long particle beam (longer than the plasma wave period) interaction with the plasma. In this case different slices of the long beam sample different transversely focusing (defocusing) and longitudinally accelerating (decelerating) forces depending on their longitudinal position inside of the plasma. As a result both the charge density as well as energy of the long bunch is modulated with respect to its longitudinal coordinate. The period of such modulations is in the order of the plasma wave period. The self-modulation instability grows exponentially with the N number of e-foldings $\sim \exp(\mathrm{N})$. For the PITZ beam parameters $\mathrm{N} \approx 5$ and a beam energy modulation of $> 1\%$ are expected after a few centimeters propagation in $10^{15}\,\mathrm{cm}^{-3}$ density plasma [15]. However, hosing instability [16] which is a result of the interaction of the beam and plasma sheath electrons remains a critical issue. The latter can lead to a position displacement of the beam centroid and therefore have an impact on beam-plasma interaction efficiency. Ideally the sharp falling edge of the electron beam ($\sim 0.6\,\mathrm{mm}$) will initiate significant fields in the plasma and suppress the effect of hosing instability. A precise beam density calculations should be performed by applying suitable distribution functions for longitudinal and transverse projections of the simulated electron beam outputs. Further simulations on beam-plasma interactions, as e.g. in [17] are necessary for the case of the PITZ beam-plasma parameters to get better understanding of the self-modulation process.

SUMMARY AND FUTURE PLANS

Beam dynamics simulations have been performed to transversely focus the electron beam at two different longitudinal positions of the PITZ beamline. In both cases an rms beam spot size of $\sim 50\,\mu\mathrm{m}$ was demonstrated for 100 pC bunch charge at 24.55 MeV. The results are promising and meet the preliminary requirements on the electron beam for self-modulation studies. The transverse deflecting cavity, recently installed in the PITZ beamline, will make possible to accurately study the longitudinal beam properties before and during the planned self-modulation experiments. Installation of a bunch compressor for transformer ratio studies is also in the future schedule of the PITZ facility.

REFERENCES

[1] K. Flöttmann, ASTRA, `http://www.desy.de/~mpyflo`
[2] T. Tajima and J.M. Dawson, PRL, Vol. 43, No. 4 (1979).
[3] P. Chen et al., Phys. Rev. Lett., Vol. 54, No. 7 (1985).
[4] K.L. Bane et al., SLAC-PUB-3662, 1985.
[5] P. Schutt et al., Nor Amberd Conf. Proc., Armenia, (1989).
[6] I. Blumenfeld et al., Nature 445, (2007), p. 741.
[7] A. Caldwell et al., Nat. Phys. 5, (2009), p. 363.
[8] N. Kumar et al, PRL 104, 255003 (2010).
[9] Y. Fang et al, Proc. Of IPAC2012, USA, 2012.
[10] G. Xia et al., Proc. Of PAC'11, 2011.
[11] C.B. Schroeder et al., Phys. Rev. Lett. 107, 145002 (2011).
[12] M. Krasilnikov et al, submitted to PRST-AB.
[13] D. Malyutin et al., Proc. of IPAC2012, USA, 2012.
[14] I. Will et al., Optics Express, Vol. 19, (2011) p. 23770 .
[15] C.B. Schroeder and F.J. Grner, private communications.
[16] D.H. Whittum et al., PRL 67, 991 (1991).
[17] A. Martinez de la Ossa et al., to be published in AAC 2012.

GLOBAL SCAN OF ALL STABLE SETTINGS (GLASS) FOR THE ANKA STORAGE RING*

M. Streichert†, N. Hiller, E. Huttel, V. Judin, B. Kehrer, M. Klein‡,
S. Marsching, C. Meuter, M.J. Nasse, A.-S. Müller, M. Schuh, N.J. Smale
Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Abstract

The design of an optimal magnetic optics for a storage ring is not a simple optimization problem, since numerous objectives have to be considered. For instance, figures of merit could be tune values, optical functions, momentum compaction factor, emittance, etc. There is a technique called "GLobal scan of All Stable Settings" (GLASS)[1], which provides a systematic analysis of the magnetic optics and gives a global overview of the capabilities of the storage ring. We developed a parallel version of GLASS, which can run on multi-core processors, decreasing significantly the computational time. In this paper we present our GLASS implementation and show results for the ANKA lattice.

INTRODUCTION

ANKA is the synchrotron light source of the Karlsruhe Institute of Technology (KIT)[2]. It consists of 4 super periods with two double bend achromats each. Each DBA structure contains 2 bending magnets, 5 quadrupole families and two chromatic sextupole families to control the vertical and horizontal chromaticity. We want to study all linear stable settings of ANKA at 2.5 GeV by scanning all possible quadrupole settings. At the moment, ANKA is operated in three different modes, in low-emittance with the natural emittance of 50 nm rad[3], low-β_y mode with vertical beta value of 1.9 m in the straight sections[4] and low-α_c-mode with reduced momentum compaction factor.[5].

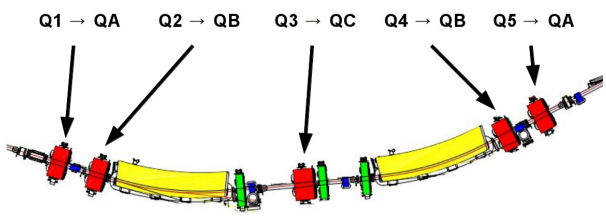


Figure 1: One half of the super period showing the reduced number of quadrupole families.

For our studies, we considered a symmetric super period by reducing the number of quadrupole families from 5 down to 3 (see Fig. 1), which allowed us to decrease the computation time. Since the ANKA storage ring consists of 4 identical super periods, we need to consider only one super period to check for linear stability of each set of quadrupole settings. The linear stability criterion is given by

$$|\mathrm{tr}(M_{\mathrm{x,y}})| < 2, \tag{1}$$

where $M_{\mathrm{x,y}}$ is the transversal transfer matrix of one super period.

We also considered fringe field integrals and quadrupole components in bending magnets to get a realistic model at 2.5 GeV[6]. To decrease the computation time, we implemented a parallel C++ code using OpenMP.

GLASS SCAN

The quadrupole strength scan was performed in the range of -2.4 to 2.4 m^{-2} with a resolution of 0.02 m^{-2} for the ANKA magnets corresponding to the maximum possible current. The compiled C++ code checked the linear stability criterion of 240^3 quadrupole settings in only 11 seconds on 12 cores and filtered out all unstable settings. Only 3% of all possible quadrupole settings are stable for the ANKA lattice at 2.5 GeV, but not all of them are feasible for the real machine. Therefore, we considered only the settings that satisfy the following constraints:

- $\beta_{x,y} < 40\,\mathrm{m}, \quad |\eta_x| < 2\,\mathrm{m}, \quad J_x, J_s > 0,$
- no tune resonance up to the 2^{nd} order,

where $\beta_{x,y}$ is the transversal beta function, $|\eta_x|$ the horizontal dispersion function and J_x, J_s are the damping partition numbers. The constraints were checked with the Accelerator Toolbox (AT)[7] for MATLAB, which took 28 hours on a single core. Then, we queried the resulting database with stable solutions and computed the corresponding emittance for each of them. In Figure 2 all stable settings are shown in the quadrupole-strength-space, classified by different colors into three emittance ranges. As shown in Figure 2, we found two islands of stable settings but only the one with the positive quadrupole strengths for QA family and negative for QB family contains emittance values lower than 100 nm rad. The other island has emittance values that are greater than 770 nm rad. Hence this region is not interesting for further studies on low-emittance optics.

We were also interested in low beta values in the straight sections. Since there are three insertion devices installed in

* This work has been supported by the Initiative and Networking Fund of the Helmholtz Association under contract number VH-NG-320.
† max.streichert@student.kit.edu
‡ now at SOLEIL, Gif-sur-Yvette, France

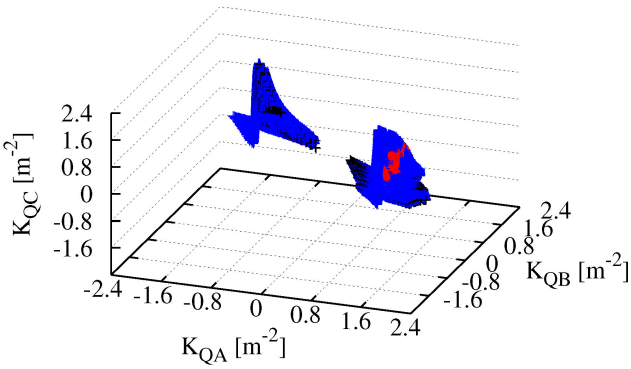


Figure 2: Natural emittance of all linear stable settings in quadrupole strengths space. Red: $10 < \epsilon < 100$, blue: $100 < \epsilon < 1000$, black: $\epsilon > 1000$ [nm rad].

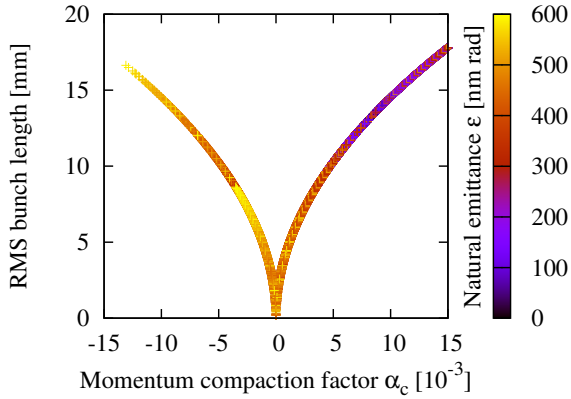


Figure 4: Relationship between momentum compaction factor α_c, natural emittance and natural bunch length RMS at infinitesimal current for 2.5 GeV optics.

the straight sections of ANKA at the moment. All of them require a low beta value. Figure 3 shows the beta function diagram and corresponding emittances in the straight section.

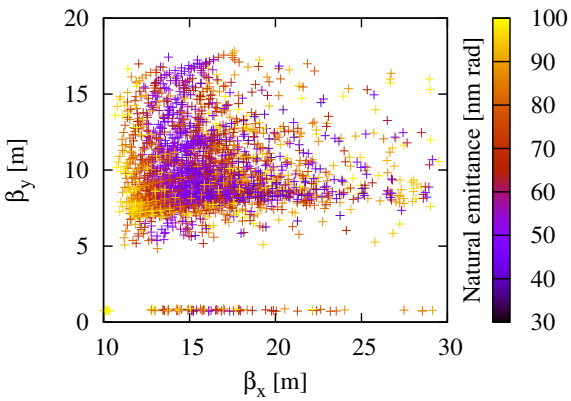


Figure 3: Beta functions in the straight sections of the ANKA for the stable optics with corresponding emittance values lower than 100 nm rad.

A low momentum compaction factor is also an interesting parameter, since the bunch length depends directly on it. At ANKA we use low-α_c optics at 1.3 GeV to decrease the bunch length and produce coherent synchrotron radiation (CSR). As Figure 4 shows, there are stable optics configurations with bunch lengths shorter than 5 mm, which can be considered for production of CSR at a beam energy of 2.5 GeV.

The GLASS scan found relatively low emittance values of 18 nm rad, but the corresponding horizontal natural chromaticity has a high negative values of -34, hence this optics is not possible for the ANKA machine with the current sextupole magnets. To decrease the emittance and compensate the natural chromaticity, the ANKA ring would need to be upgraded. Figure 5 shows the relationship between natural chromaticity and emittance.

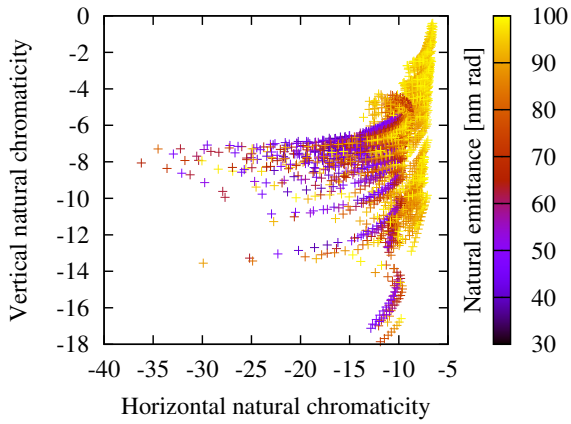


Figure 5: Horizontal and vertical chromaticity values for the stable optics with emittance values lower than 100 nm rad and beam energy of 2.5 GeV.

Furthermore, we found a stable setting with a low possible emittance of 38.6 nm rad with chromaticity values of $\nu'_{x,y} = (-15.90, -7.16)$ that can be compensated by the sextupole magnets. The corresponding optical functions are presented in Figure 6. The emittance of the existing low-emittance optics is 50 nm rad, which is 23% higher than the value found by the GLASS analysis.

DYNAMIC APERTURE

Not all linear stable settings with low emittance are feasible for the real machine, since non-linear effects lead to small dynamic aperture and the beam becomes unstable. To determine the quality of a given optics, tracking simulations need to be performed. For the GLASS scan we turned off the sextupole magnets and considered only the natural chromaticity. Then, we turned on the sextupole magnets

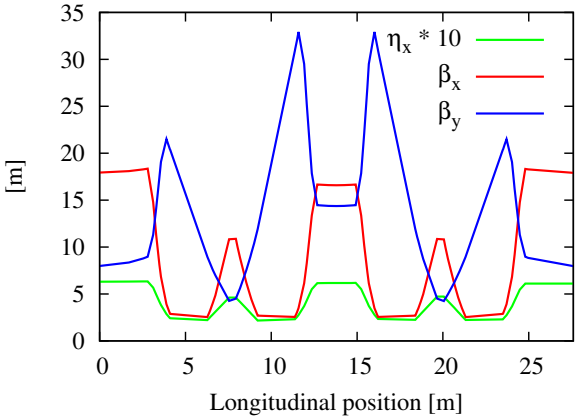


Figure 6: Low-emittance optics with a natural emittance of 38.6 nm rad and tunes h/v = (7.13/1.67). The figure shows one super period of the ANKA lattice.

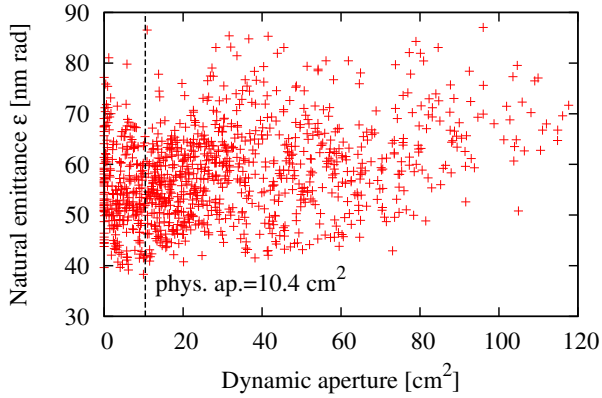


Figure 8: Dynamic apertures with corresponding natural emittance values and chromaticity values set to $\nu'_x = 2, \nu'_y = 6$ at 2.5 GeV.

and set the chromaticity to $\nu'_x = 2, \nu'_y = 6$. These values were determined by measurements to be optimal for damping of beam instabilities at ANKA. Figure 7 shows the geometric dynamic aperture for the described low-emittance optics with different momentum deviations dp with in units of the natural energy spreads σ_e. Figure 8 shows that there are lots of stable optics with emittance values lower than 90 nm rad and the squared dynamic aperture greater than the physical aperture, which can be considered as possible operating points. For the dynamic aperture simulations we considered only half of the electron beam pipe with the physical aperture of 10.4 cm^2, since the beam is symmetric to the horizontal plane.

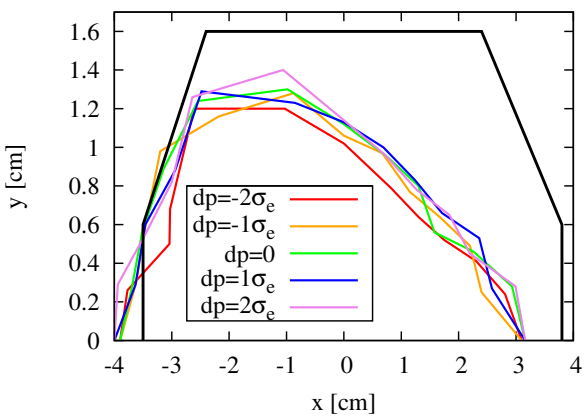


Figure 7: Dynamic apertures of the low-emittance-optics shown in Fig. 6 with chromaticity values $\nu'_x = 2, \nu'_y = 6$, for different momentum deviations dp, where $\sigma_e = 9.13463 \cdot 10^{-4}$ is the natural energy spread at the beam energy of 2.5 GeV. The black line marks the physical aperture at ANKA.

SUMMARY AND OUTLOOK

The GLASS technique allows us to explore the global linear properties of the machine and provides us with a systematic method to find stable optics. We performed GLASS at ANKA successfully and we could find a realizable optics with lower emittance than the currently used optics. However, the GLASS technique requires a lot of computational power. To decrease the computational time, we implemented a parallel code to make use of multicore computers. If we want to optimize more parameters than three, the GLASS technique is no longer an effective approach. Hence we are exploring genetic algorithms to find optimal settings in much shorter time. First studies show a good agreement with the optimum settings found with GLASS, while reducing the computational time to only several minutes[8].

REFERENCES

[1] D.S. Robin, W. Wan, F. Sannibale and V.P. Suller, Phys. Rev. ST Accel. Beams 11, 024002 (2008)

[2] `http://ankaweb.fzk.de`

[3] E. Huttel et al., "Operation with a Low Emittance Optics at ANKA", PAC 2005, p.2467

[4] E. Huttel et al., "Low Beta Structure for the ANKA Storage Ring", EPAC 2008, p.2034

[5] A.-S. Müller et al., "Far Infrared Coherent Synchrotron Edge Radiation at ANKA", PAC 2005, p.2518

[6] M. Streichert et al. "Simulations of Fringe Fields and Multipoles for the ANKA Storage Ring Bending Magnets", IPAC 2012, p.1626

[7] A. Terebilo, "Accelerator Toolbox for MATLAB", SLAC-PUB-8732, May 2001

[8] M. Streichert et al. "Global Optimization of the ANKA Lattice using Multiobjective Genetic Algorithms (MOGA)", ICAP 2012

EMISSION STUDIES OF PHOTOCATHODE RF GUN AT PITZ

J. Li*+, G. Asova‡, I.V. Isaev, M. Groß, L. Hakobyan, Y. Ivanisenko, M. Khojoyan‡+,G. Klemz, G. Kourkafas, M. Krasilnikov, K. Kusoljariyakul‡‡, M. Mahgoub, D. Malyutin, B. Marchetti, A.Oppelt, B. Petrosyan, S. Rimjaem‡‡, A. Shapovalov, F. Stephan, G. Vashchenko,
DESY, 15738 Zeuthen, Germany
G. Feng, DESY, 22607 Hamburg, Germany
L. Shang, University of Science and Technology of China, 230029 Hefei, China
D. Richter, HZB, 12489 Berlin, Germany

Abstract

The Photo Injector Test facility at DESY, Zeuthen site (PITZ), was built to develop and optimize electron sources for linac based Free Electron Lasers (FELs) like FLASH and the European XFEL. For the value of the bunch charge extracted from a photocathode, discrepancy has been observed between the data measured at PITZ and simulation results from the ASTRA code. As a factor which could explain the discrepancy, a Schottky-like effect is considered. Meanwhile, the PARMELA code was applied to the emission studies on the PITZ gun as benchmark. Since PARMELA cannot be used to simulate Schottky-like effects with its own modules, MATLAB scripts have been developed to implement this feature of the photoemission in an RF gun.

INTRODUCTION

With the aim to produce beams with high density, low transverse emittance and short bunch length, the PITZ characterizes the RF photocathode electron sources for FLASH and the European XFEL. Considering it is one of the most sensitive components for high quality beams in an XFEL facility, detailed research should be done on the electron gun.

Measurement of the accelerated charge downstream of the gun as a function of the beam launch phase, which is called phase scan or Schottky scan, is a basic experiment to study the emission properties of the PITZ gun. A Space Charge Tracking Algorithm (ASTRA) [1], originally developed at DESY and extensively used in photo injector design and benchmarking of experimental data [2], has been used to simulate the phase scans. The measured extracted charge is not consistent with the simulation results. One of the possible reasons is a Schottky-like effect. In order to justify this hypothesis, Phase and Radial Motion in Electron Linear Accelerators (PARMELA) [3], which is widely applied to linac design and dynamic analysis, has been used to simulate the emission process of the PITZ gun. Simulations have been done considering Schottky-like effects with ASTRA and PARMELA, respectively.

BACKGROUND AND MOTIVATION

PITZ Gun

The electron gun is a 1.6-cell copper cavity with resonance frequency of 1.3 GHz. The Cs_2Te photocathode is inserted into the backplane of the cavity by a load-lock system. The cavity is surrounded by a pair of solenoids: the main solenoid, used to focus the beam, counteracting its expansion due to the space charge force, and the bucking one, used to compensate the field of the main solenoid on the photocathode surface to ensure that the electron bunch leaves the magnetic focusing region without any remaining average angular momentum [4].

Discrepancy for Higher Charge

When specific machine settings were directly used in ASTRA, it was not possible to produce 1 nC at the gun operation phase, whereas 1 nC and even higher charge is experimentally detected. ASTRA phase scans have different shapes as compared to the experimentally measured data in Fig. 1, in which zero phases are the Maximum Mean Momentum Gain (MMMG) phases. Initial parameters are shown in Tab. 1.

As a factor which could explain this discrepancy, field enhancement was considered, such as Schottky-like effect.

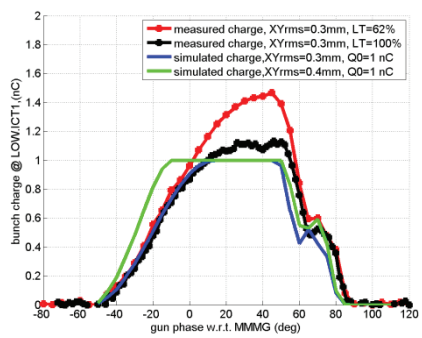


Figure 1: Measured and simulated (ASTRA) phase scans.

* ji.li@desy.de
+ On leave from USTC/NSRL, Hefei, China
‡ On leave from INRNE, Sofia, Bulgaria
‡+ On leave from ANSL, Yerevan, Armenia
‡‡ Currently at Chiang Mai University, Chiang Mai, Thailand.

Table 1: Parameters for Simulation

Parameters	Units	Value
Laser temporal profile (r.t/FWHM\f.t)	ps	2.2/21.46\2.2
RMS lase spot size (XYrms)	mm	0.3
Initial kinetic energy	eV	0.55
Maximum gradient on the cathode	MV/m	-60.58
Maximum B_z in the gun	T	0.2333
Laser transmission (LT)		62%, 100%

Schottky Effect

The Schottky effect describes the lowering of the work function or the potential barrier of a metal by an external electric field, which leads to an increased electron emission from the metal. As the Schottky effect is originally defined for a metal cathode, a similar phenomenon observed in an electron gun with semiconductor cathode is called Schottky-like effect. For a photocathode RF gun, this effect can be significant due to the strong field.

If a Schottky-like effect is considered in ASTRA, the charge of a bunch is determined at the time of its emission as [1]:

$$Q = Q_0 + SRT\_Q\_Schottky \cdot \sqrt{E} + Q\_Schottky \cdot E \,, \qquad (1)$$

where E is the combined (external and space-charge) longitudinal electric field in the centre of the cathode. The charge Q_0, is the charge of the macro particles as defined in the input distribution (eventually rescaled according to the parameter Q_{bunch}), *SRT_Q_Schottky* and *Q_Schottky* describe the field dependent emission process [1].

Motivation

When XYrms=0.3 mm is used in the ASTRA simulation with Schottky-like effect, it is not possible to produce the measured charge for any combination of Q_0, *SRT_Q_Schottky* and *Q_Schottky* [6]. If the laser spot size was slightly increased to 0.32 mm, the simulated results fit to the measurement well as shown in Fig. 2.

Considering this discrepancy between measured data and ASTRA results, we would like to use other codes to verify the data, like PARMELA. Since PARMELA could not be used to simulate the Schottky-like effect by itself, MATLAB scripts have been developed to add this feature.

An arbitrary temporal shape of the bunch can be generated by varying the number of macro particles in each separated short pulse in PARMELA. A code was written by MATLAB to obtain the longitudinal distribution considering the Schottky-like effect and run phase scans. Through Eq. 1, the emitted electron bunch can be composed by several short bunches, which are uniform in longitudinal direction and transverse direction with various heights according to different RF phases.

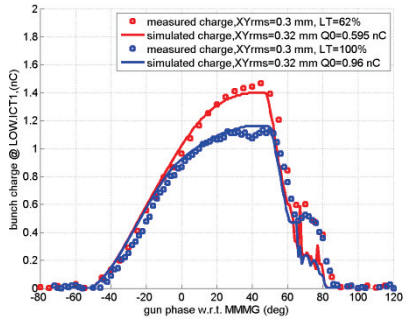


Figure 2: Phase scans with optimized parameters.

PARMELA SIMULATIONS

MATLAB Scripts

In the PARMELA input file, the charge is set by defining the length of the bunch and the average current during an RF cycle. The scripts divide the bunch into slices and calculate the increase of charge induced by Schottky-like effect in each slice separately. An increased electron emission is achieved by the increased number of macro particles, which finally leads to the increase of average current.

When Eq.1 was used in the ASTRA, only Q_0 was changed for different LT if proper *SRT_Q_Schottky* and *Q_Schottky* were set. But when it was used for PARMELA, results could not fit the measured data if just Q_0 was changed for different LT with constant *SRT_Q_Schottky* and *Q_Schottky*. A modified equation was applied in the scripts, which reads:

$$Q = Q_0(1 + \alpha \cdot \sqrt{E} + \beta \cdot E) \quad , \qquad (2)$$

where E is the combined (RF force and space charge force) longitudinal electric field in the centre of the cathode. The Q_0 is the charge produced by the laser, α and β describe the field dependent emission process which are related to the material properties [5].

The flow chart is shown in Fig. 3, where *E(i)* is the strength of the RF field on centre of the bunch slice ; N means the total number of macro particles in the bunch; *x(i)* is the number of macro particles of each slice. *X(i)* is the number of macro particles in each slice considering the Schottky-like effect; N' is the total number of the macro particles in PARMELA considering the Schottky-like effect.

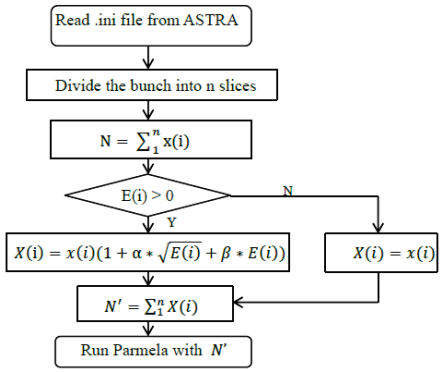


Figure 3: Flow chart of MATLAB scripts.

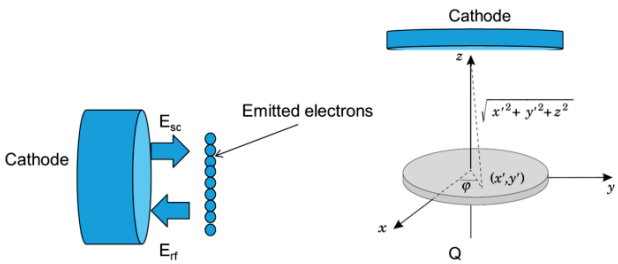


Figure 4: Space-charge force model.

For the space-charge field strength, an estimation model was used which is shown in Fig. 4. Considering that the emitted electrons exert a force on the cathode surface, each emitted slice was treated as a flat plate with uniform charge distribution. The force on the centre of the cathode from each slice was calculated using Eq. 2, 3 and 4, where Q is the charge of the flat plate and R is the radius of the flat plate.)

$$E_{sc} = \frac{Q}{2\pi\varepsilon_0 R^2}\left(1-\frac{z}{\sqrt{R^2+z^2}}\right) \quad , \tag{3}$$

$$E = E_{rf} - E_{sc} = E_{cathode}\cdot\sin(\varphi_{launch}) - E_{sc} \quad . \tag{4}$$

Results

The simulated results and measured data were compared in Fig. 5. The values of charge were simulated at the position where a Faraday cup is located along the beam line, and the measured data were obtained in the year 2011. The initial conditions were set according to Tab. 1. In the PARMELA simulation, the Q_0 is 0.57 nC in the LT=62% case, while the Q_0 is 0.77 nC corresponding to the LT=100% case.

The experimental data in the year 2012 have been used to check the Schottky model for PARMELA. When XYrms is 0.3 mm and maximum voltage on the cathode is 45 MV/m, Q_0 is set to 0.326 nC for LT=14%. The simulated results are consistent with measurement, which are shown in Fig. 6.

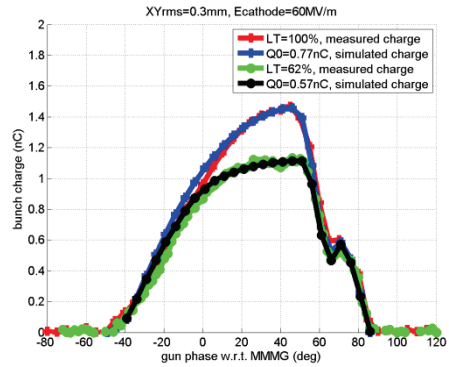


Figure 5: Measured and PARMELA phase scans (2011).

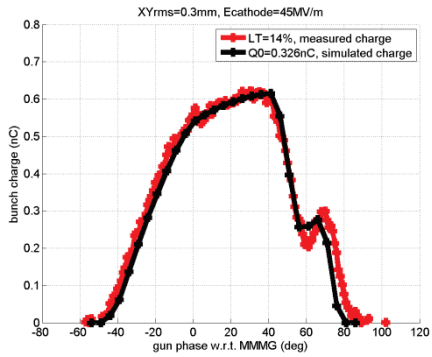


Figure 6: Measured and PARMELA phase scans (2012).

SUMMARY

Through simulations for different cases, the results from PARMELA considering Schottky-like effects fit the measurement data well. Eq. 2 can be used to explain the increase of the electron emission in the PITZ gun successfully.

In the simulations, the relationship between Q_0 and LT is not linear. Experiments should be done to check if the photocurrent is saturated at high laser transmission. Studies to accurately model the initial distribution and parameters optimization for Eq. 2 are to be done.

REFERENCES

[1] K. Floettmann, "A Space Charge Tracking Algorithm ", user manual (version 3), 2011.
[2] K. Floettmann et al. , "Recent improvements to the ASTRA particle tracking code", PAC 2007.
[3] Lloyd M.Young, "PARMELA", user manual, 2005
[4] F. Stephan et al. , "Detailed characterization of electron sources yielding first demonstration of European X-ray Free-Electron Laser beam quality", PRST-AB 13, 020704, 2010.
[5] Yongzhang Huang et al. , "Simulation on the emittance of the RF gun including the Schottky effect", LINAC 1996.
[6] J.F. Castanon, PITZ summer student report, 2011

LUMPED EQUIVALENT MODELS OF COMPLEX RF STRUCTURES*

T. Flisgen†, J. Heller, U. van Rienen, Universität Rostock, IAE, IEF, 18059 Rostock, Germany

Abstract

The prediction of RF properties of complex accelerating structures is an important issue in computational accelerator physics. This paper describes the derivation of state space equations for complex structures based on real eigenmodes of sections of the decomposed complex structure. The state space equations enable the calculation of system responses due to port excitations by means of standard ordinary differential equation solvers. Therefore, the state space equations are referred to as lumped equivalent models of such complex RF structures. Besides fast computation of system responses, the equivalent models enable the calculation of secondary quantities such as external quality factors. The present contribution discusses theoretical aspects by means of a validation example.

INTRODUCTION

The investigation of complex radio frequency (RF) structures is a crucial task in the design of particle accelerators. Often the treatment of these complex and large RF devices exceeds the capabilities of modern workstation computers. Therefore, many problems are tackled by means of cluster computers in combination with suitable parallel simulation codes (e.g. ACE3P [1]). In this contribution a proof of principle for an alternative approach is presented. Here, the complex structure is decomposed into segments likewise it is proposed by the method of Coupled S-Parameter Calculations (CSC) [2, 3]. This decomposition has the following advantages:

- The separate treatment of the segments is less time- and memory-consuming.
- The properties of equal segments need to be computed only once.
- If sections show symmetries their treatment can be further simplified.
- If Maxwell's equations can be solved analytically for sections their numerical treatment is not needed.
- If parameter studies are of interest only the properties of sections with free parameter(s) need to be recalculated.

After decomposition of the complex structure, state space equations in an impedance formulation for the segments are created by a 3D eigenmode expansion [4, 5, 6]. These impedance state space equations are also referred to as lumped equivalent models of the segments as they describe the dependency of the transient modal voltages on transient modal currents at the segment's waveguide ports. By employing a modification of the CSC scheme, the state equations of the segments are combined to obtain the state space equations of the complete structure. These equations allow for the computation of transient system responses as well as frequency domain transfer functions. Furthermore, secondary quantities such as external quality factors can be derived from the lumped model. In addition, the field distribution inside the full structure is also available as the state equations for the individual segments are obtained by a 3D eigenmode expansion.

In comparison to [6] where the coupling scheme is introduced and the focus lies on theoretical derivations, this paper presents a more sophisticated application example for the creation of lumped equivalent models of complex RF structures.

APPLICATION EXAMPLE AND CREATION OF ITS LUMPED MODEL

A third harmonic cavity with three cells (see [7] for geometric details) and simplified higher order mode (HOM) couplers[1] is chosen as an application example. The entire structure is considered to be lossless and the beam pipes are shortcut with perfect electric conducting (PEC) walls. Fig. 1 shows the structure under test and its decomposition into segments. The entire structure is comparably small and not complex such that the direct numerical calculation of its properties like the transfer function is easily feasible for the sake of comparison and validation. To model these sections, five different state space equations of the form

$$\frac{\partial}{\partial t}\mathbf{x}_r(t) = \mathbf{A}_r\,\mathbf{x}_r(t) + \mathbf{B}_r\,\mathbf{i}_r(t), \quad (1)$$

$$\mathbf{v}_r(t) = \mathbf{B}_r^{\mathrm{T}}\,\mathbf{x}_r(t). \quad (2)$$

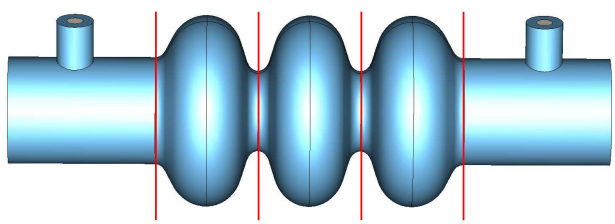


Figure 1: Lossless third harmonic cavity with two end cells, one middle cell and two simplified HOM couplers and its decomposition into five segments (left coupler, left end cell, middle cell, right end cell and right coupler).

* partially funded by EU FP7 Research Grant No. 227579
† Thomas.Flisgen@Uni-Rostock.de

[1] An antenna tip protruding into the beam pipe.

are established. The matrix $\mathbf{A}_r$ is the state matrix, $\mathbf{B}_r$ the input matrix and $\mathbf{B}_r^{\mathrm{T}}$ the output matrix of the r-th section. The excitation term $\mathbf{i}_r(t)$ embraces the modal currents at the waveguide ports of the r-th segment and $\mathbf{v}_r(t)$ combines its modal voltages, respectively:

$$\mathbf{i}_r(t) = \begin{pmatrix} i_{r,1}(t) \\ i_{r,2}(t) \\ \vdots \\ i_{r,m}(t) \\ \vdots \\ i_{r,M_r}(t) \end{pmatrix} \text{ and } \mathbf{v}_r(t) = \begin{pmatrix} v_{r,1}(t) \\ v_{r,2}(t) \\ \vdots \\ v_{r,m}(t) \\ \vdots \\ v_{r,M_r}(t) \end{pmatrix}. \quad (3)$$

Here M_r denotes the number of total waveguide port modes of the section indexed by r.

The state matrices in (1) and (2) are obtained by a real eigenmode computation for each segment with perfect magnetic conducting (PMC) boundary conditions on the cross section of the waveguide ports and PEC boundary conditions on the remaining boundaries [4, 6]. In addition to the eigenfrequencies and modes, the impedance matrix of each segment is computed on 10 discrete frequency samples to calculate a correction term which accounts for non considered higher (or lower) order modes [4, 5, 6]. All computations are performed by CST Studio Suite® [8]. Hereinafter details on the individual calculations are presented.

State Space Matrices for HOM Coupler

Fig. 2 illustrates the left HOM coupler (segment index $r = 1$) and its network theoretical counterpart with the generalized voltages and currents of the respective waveguide modes. At the coaxial port of the coupler only the TEM mode is considered (port mode index $m = 1$), whereas 12 circular waveguide modes at the beam pipe are used for the concatenation (port mode indices $m = 2, \ldots, 13$). The HOM coupler is discretized exploiting its symmetry on a hexahedral mesh with $34,200$ mesh cells. Based on this meshing 50 eigenmodes are computed in combination with 10 impedance matrices sampled on discrete frequencies for the correction term. Furthermore, the field distributions of 13 2D port modes are determined by the CST Microwave Studio® frequency domain solver. Based on these calculations the matrices $\mathbf{A}_{HOM} \in \mathbb{R}^{151\times151}$ and $\mathbf{B}_{HOM} \in \mathbb{R}^{151\times13}$ are obtained (refer to (1) and (2)).

State Space Matrices for End and Middle Cell

Fig. 3 illustrates the end cell and its equivalent circuit diagram with generalized voltages and currents of the waveguide modes (the treatment of the middle cell is analogous). On both cut planes 12 port modes are used for the 2D field expansion. The end cell is discretized on a hexahedral mesh with $58,806$ mesh cells exploiting one symmetry plane (for the middle cell $39,600$ mesh cells are used). By employing this discretization 50 eigenmodes are computed in combination with 10 impedance matrices needed for the correction term. Furthermore, the field distributions of 24 2D

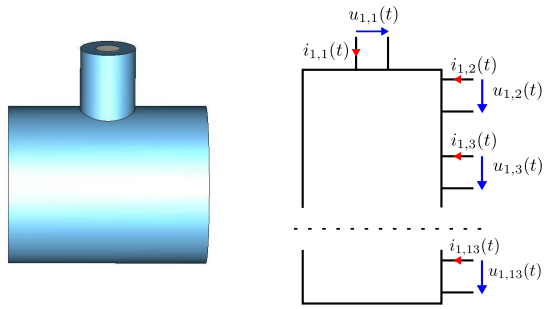


Figure 2: Lossless HOM coupler (left) and its equivalent circuit diagram (right) with indicated modal voltages (blue arrows) and currents (red arrows).

port modes (12 modes on each cut plane) are determined by the CST Microwave Studio® frequency domain solver. Based on these calculations the matrices for the endcells $\mathbf{A}_{ec} \in \mathbb{R}^{176\times176}$, $\mathbf{B}_{ec} \in \mathbb{R}^{176\times24}$ and for the midcells $\mathbf{A}_{mc} \in \mathbb{R}^{184\times184}$, $\mathbf{B}_{mc} \in \mathbb{R}^{184\times24}$ are obtained.

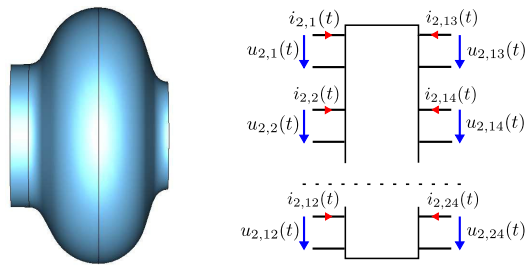


Figure 3: Lossless end cell (left) and its network theoretical counterpart (right) with the generalized voltages (blue arrows) and currents (red arrows).

Concatenation of the Sections

To obtain a lumped equivalent model of the complete structure, the state equations for the five individual segments are collected by the block system

$$\frac{\partial}{\partial t} \underbrace{\begin{pmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \\ \vdots \\ \mathbf{x}_5(t) \end{pmatrix}}_{\mathbf{x}_d(t)} = \underbrace{\begin{pmatrix} \mathbf{A}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \ldots & \mathbf{0} \\ \vdots & \ldots & \ddots & \vdots \\ \mathbf{0} & \ldots & \ldots & \mathbf{A}_5 \end{pmatrix}}_{\mathbf{A}_d} \underbrace{\begin{pmatrix} \mathbf{x}_1(t) \\ \mathbf{x}_2(t) \\ \vdots \\ \mathbf{x}_5(t) \end{pmatrix}}_{\mathbf{x}_d(t)} + \underbrace{\begin{pmatrix} \mathbf{B}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 & \ldots & \mathbf{0} \\ \vdots & \ldots & \ddots & \vdots \\ \mathbf{0} & \ldots & \ldots & \mathbf{B}_5 \end{pmatrix}}_{\mathbf{B}_d} \underbrace{\begin{pmatrix} \mathbf{i}_1(t) \\ \mathbf{i}_2(t) \\ \vdots \\ \mathbf{i}_5(t) \end{pmatrix}}_{\mathbf{i}_{can}(t)}, \quad (4)$$

where $\mathbf{A}_1 = \mathbf{A}_5 = \mathbf{A}_{HOM}$, $\mathbf{A}_2 = \mathbf{A}_4 = \mathbf{A}_{ec}$, $\mathbf{A}_3 = \mathbf{A}_{mc}$, $\mathbf{B}_1 = \mathbf{B}_5 = \mathbf{B}_{HOM}$, $\mathbf{B}_2 = \mathbf{B}_4 = \mathbf{B}_{ec}$ and $\mathbf{B}_3 = \mathbf{B}_{mc}$. In the same manner, the output equations (2) are

combined for all sections:

$$\underbrace{\begin{pmatrix}\mathbf{v}_1(t)\\ \mathbf{v}_2(t)\\ \vdots\\ \mathbf{v}_5(t)\end{pmatrix}}_{\mathbf{v}_{can}(t)} = \underbrace{\begin{pmatrix}\mathbf{B}_1^{\mathrm{T}} & \mathbf{0} & \dots & \mathbf{0}\\ \mathbf{0} & \mathbf{B}_2^{\mathrm{T}} & \dots & \mathbf{0}\\ \vdots & \dots & \ddots & \vdots\\ \mathbf{0} & \dots & \dots & \mathbf{B}_5^{\mathrm{T}}\end{pmatrix}}_{\mathbf{B}_d^{\mathrm{T}}} \underbrace{\begin{pmatrix}\mathbf{x}_1(t)\\ \mathbf{x}_2(t)\\ \vdots\\ \mathbf{x}_5(t)\end{pmatrix}}_{\mathbf{x}_d(t)}. \quad (5)$$

The resulting ordering of the currents $\mathbf{i}_{can}(t)$ and voltages $\mathbf{v}_{can}(t)$ is defined as canonical ordering. It is reordered by a permutation matrix $\mathbf{P}^{\mathrm{T}}$ (one coefficient equal to one in each row and in each column, while the remaining coefficients are equal to zero) to sort for internal and external quantities:

$$\mathbf{i}_{sort}(t) = \begin{pmatrix}\mathbf{i}_{int}(t)\\ \mathbf{i}_{ext}(t)\end{pmatrix} = \mathbf{P}^{\mathrm{T}}\,\mathbf{i}_{can}(t), \quad (6)$$

$$\mathbf{v}_{sort}(t) = \begin{pmatrix}\mathbf{v}_{int}(t)\\ \mathbf{v}_{ext}(t)\end{pmatrix} = \mathbf{P}^{\mathrm{T}}\,\mathbf{v}_{can}(t). \quad (7)$$

Internal quantities belong to waveguide ports located at the decomposition planes and external ports are ports of the complete structure. Thus, the vectors $\mathbf{i}_{ext}(t)$ and $\mathbf{v}_{ext}(t)$ embrace the port quantities which belong to the $N_{ext} = 2$ external TEM ports located at the HOM antenna tips:

$$\mathbf{i}_{ext}(t) = \begin{pmatrix}i_{1,1}(t)\\ i_{5,1}(t)\end{pmatrix} \text{ and } \mathbf{v}_{ext}(t) = \begin{pmatrix}v_{1,1}(t)\\ v_{5,1}(t)\end{pmatrix}. \quad (8)$$

It is crucial to note, that vectors holding the internal quantities are organised such that quantities belonging to connected ports are listed pairwise below each other. E.g. for the connection between the left HOM coupler and the left end cell (see Fig. 4) this reads as

$$\mathbf{i}_{int}(t) = \begin{pmatrix}i_{1,2}(t)\\ i_{2,1}(t)\\ i_{1,3}(t)\\ i_{2,2}(t)\\ \vdots\\ i_{1,13}(t)\\ i_{2,12}(t)\\ \vdots\end{pmatrix} \text{ and } \mathbf{v}_{int}(t) = \begin{pmatrix}v_{1,2}(t)\\ v_{2,1}(t)\\ v_{1,3}(t)\\ v_{2,2}(t)\\ \vdots\\ v_{1,13}(t)\\ v_{2,12}(t)\\ \vdots\end{pmatrix}. \quad (9)$$

By employment of the orthogonality of the permutation matrix $\mathbf{P}^{\mathrm{T}} = \mathbf{P}^{-1}$ the canonical quantities in (4) and (5) are replaced by the sorted quantities:

$$\frac{\partial}{\partial t}\mathbf{x}_d(t) = \mathbf{A}_d\,\mathbf{x}_d(t) + \underbrace{\mathbf{B}_d\,\mathbf{P}}_{\bar{\mathbf{B}}_d}\,\mathbf{i}_{sort}(t), \quad (10)$$

$$\mathbf{v}_{sort}(t) = \underbrace{\mathbf{P}^{\mathrm{T}}\,\mathbf{B}_d^{\mathrm{T}}}_{\bar{\mathbf{B}}_d^{\mathrm{T}}}\,\mathbf{x}_d(t). \quad (11)$$

To split these equations into internal and external quantities, the matrix $\bar{\mathbf{B}}_d$ is partitioned:

$$\bar{\mathbf{B}}_d = \begin{pmatrix}\bar{\mathbf{B}}_{d1} & \bar{\mathbf{B}}_{d2}\end{pmatrix}, \quad (12)$$

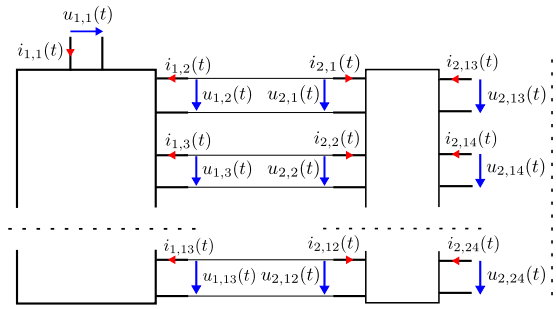


Figure 4: Equivalent circuit diagram of the concatenation between the left HOM coupler and the right end cell.

where $\bar{\mathbf{B}}_{d1}$ has $N_{int} = 96$ columns and $\bar{\mathbf{B}}_{d2}$ has $N_{ext} = 2$ columns. Utilizing this together with (10) and (11) leads to

$$\frac{\partial}{\partial t}\mathbf{x}_d(t) = \mathbf{A}_d\,\mathbf{x}_d(t) + \bar{\mathbf{B}}_{d1}\,\mathbf{i}_{int}(t) + \bar{\mathbf{B}}_{d2}\,\mathbf{i}_{ext}(t), \quad (13)$$

$$\mathbf{v}_{int}(t) = \bar{\mathbf{B}}_{d1}^{\mathrm{T}}\,\mathbf{x}_d(t), \quad (14)$$

$$\mathbf{v}_{ext}(t) = \bar{\mathbf{B}}_{d2}^{\mathrm{T}}\,\mathbf{x}_d(t). \quad (15)$$

Obviously, (15) is the output equation of the concatenated structure since it describes the relationship between the inner states and the external voltages.

To derive the corresponding state equation of the concatenated structure, the following two requirements are used. First, currents of terminals connected to each other have the same magnitude but a different sign (see e.g. $i_{1,2}(t) = -i_{2,1}(t)$ or $i_{1,3}(t) = -i_{2,2}(t)$ in Fig. 4). Second, they are listed pairwise below each other. In consequence, the N_{int} internal currents are expressed by

$$\mathbf{i}_{int}(t) = \mathbf{F}\,\hat{\mathbf{i}}_{int}(t), \quad (16)$$

where

$$\mathbf{F} = \mathrm{diag}\left(\begin{pmatrix}1\\ -1\end{pmatrix}, \dots, \begin{pmatrix}1\\ -1\end{pmatrix}\right) \in \mathbb{R}^{N_{int}\times N_{int}/2}. \quad (17)$$

The vector $\hat{\mathbf{i}}_{int}(t) \in \mathbb{R}^{N_{int}/2}$ contains the independent currents flowing through the coupled ports. Replacing (16) in (13) yields

$$\frac{\partial}{\partial t}\mathbf{x}_d(t) = \mathbf{A}_d\,\mathbf{x}_d(t) + \bar{\mathbf{B}}_{d1}\mathbf{F}\,\hat{\mathbf{i}}_{int}(t) + \bar{\mathbf{B}}_{d2}\,\mathbf{i}_{ext}(t). \quad (18)$$

The facts that voltages of connected terminals are equal (see e.g. $u_{1,2}(t) = u_{2,1}(t)$ or $u_{1,3}(t) = u_{2,2}(t)$ in Fig. 4) and that they are ordered below each other in $\mathbf{v}_{int}(t)$ yields

$$\mathbf{F}^{\mathrm{T}}\,\mathbf{v}_{int}(t) = \mathbf{F}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}^{\mathrm{T}}\,\mathbf{x}_d(t) = \mathbf{0}. \quad (19)$$

Multiplying (18) from left with $\mathbf{F}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}^{\mathrm{T}}$ gives

$$\frac{\partial}{\partial t}\underbrace{\mathbf{F}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}^{\mathrm{T}}\,\mathbf{x}_d(t)}_{\mathbf{0}} = \mathbf{0} = \mathbf{F}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}^{\mathrm{T}}\mathbf{A}_d\,\mathbf{x}_d(t) + \mathbf{F}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}\mathbf{F}\,\hat{\mathbf{i}}_{int}(t) + \mathbf{F}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}^{\mathrm{T}}\bar{\mathbf{B}}_{d2}\,\mathbf{i}_{ext}(t) \quad (20)$$

which is solved for the independent currents

$$\begin{aligned}\hat{\mathbf{i}}_{int}(t) = &- \left[\mathbf{F}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}\mathbf{F}\right]^{-1}\mathbf{F}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}^{\mathrm{T}}\mathbf{A}_d\,\mathbf{x}_d(t) \\ &- \left[\mathbf{F}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}\mathbf{F}\right]^{-1}\mathbf{F}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}^{\mathrm{T}}\bar{\mathbf{B}}_{d2}\,\mathbf{i}_{ext}(t).\end{aligned} \tag{21}$$

Substituting the internal currents in (18) leads to the state equation of the concatenated structure

$$\frac{\partial}{\partial t}\mathbf{x}_d(t) = \mathbf{K}\left[\mathbf{A}_d\,\mathbf{x}_d(t) + \bar{\mathbf{B}}_{d2}\,\mathbf{i}_{ext}(t)\right], \tag{22}$$

with the idempotent and symmetric matrix

$$\mathbf{K} = \left[\mathbf{I} - \bar{\mathbf{B}}_{d1}\mathbf{F}\left[\mathbf{F}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}\mathbf{F}\right]^{-1}\mathbf{F}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}^{\mathrm{T}}\right]. \tag{23}$$

To delete N_{int} redundant states in $\mathbf{x}_d(t)$, which result from the algebraic side constraint (19), the orthonormal basis of the kernel of $\mathbf{F}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}^{\mathrm{T}}$

$$\mathbf{Q} = \mathrm{Ker}\left(\mathbf{F}^{\mathrm{T}}\bar{\mathbf{B}}_{d1}^{\mathrm{T}}\right), \text{ with } \mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{I} \tag{24}$$

is determined. By substituting

$$\mathbf{x}_d(t) = \mathbf{K}\mathbf{Q}\,\mathbf{x}_c(t) = \mathbf{Q}\,\mathbf{x}_c(t) \tag{25}$$

in (22) and (15) the redundant states vanish and a symmetric state space system of the concatenated structure is obtained:

$$\frac{\partial}{\partial t}\mathbf{x}_c(t) = \underbrace{\mathbf{Q}^{\mathrm{T}}\mathbf{K}\mathbf{A}_d\mathbf{K}\mathbf{Q}}_{\mathbf{A}_c}\,\mathbf{x}_c(t) + \underbrace{\mathbf{Q}^{\mathrm{T}}\mathbf{K}\bar{\mathbf{B}}_{d2}}_{\mathbf{B}_c}\,\mathbf{i}_{ext}(t), \tag{26}$$

$$\mathbf{v}_{ext}(t) = \underbrace{\bar{\mathbf{B}}_{d2}^{\mathrm{T}}\mathbf{K}\mathbf{Q}}_{\mathbf{B}_c^{\mathrm{T}}}\,\mathbf{x}_c(t). \tag{27}$$

RESULTS

In the following subchapters different quantities are computed by means of the created lumped model and are compared with direct computations.

Scattering Parameters

The impedance formulation (26) and (27) can be easily transferred to a scattering formulation assuming the external port impedances to be constant [6]:

$$\begin{aligned}\frac{\partial}{\partial t}\mathbf{x}_c(t) = &\underbrace{\left(\mathbf{A}_c - \mathbf{B}_c\mathbf{D}_Z^{-1}\mathbf{B}_c^{\mathrm{T}}\right)}_{\mathbf{A}_{sc}}\,\mathbf{x}_c(t) \\ &+ \underbrace{\sqrt{2}\,\mathbf{B}_c\mathbf{D}_Z^{-\frac{1}{2}}}_{\mathbf{B}_{sc}}\,\mathbf{a}(t)\end{aligned} \tag{28}$$

and

$$\mathbf{b}(t) = \underbrace{\left(\sqrt{2}\,\mathbf{D}_Z^{-\frac{1}{2}}\mathbf{B}_c^{\mathrm{T}}\right)}_{\mathbf{B}_{sc}^{\mathrm{T}}}\,\mathbf{x}_c(t) + \underbrace{\left(-\mathbf{I}\right)}_{\mathbf{D}_{sc}}\,\mathbf{a}(t), \tag{29}$$

where the matrix $\mathbf{D}_Z$ holds the constant port impedances of the external ports on its main diagonal:

$$\mathbf{D}_Z = \mathrm{diag}\left(Z_1, Z_2\right). \tag{30}$$

The corresponding frequency domain transfer function is given by

$$\underline{\mathbf{b}}(j\omega) = \underbrace{\left[\mathbf{B}_{sc}^{\mathrm{T}}\left(j\omega\,\mathbf{I} - \mathbf{A}_{sc}\right)^{-1}\mathbf{B}_{sc} + \mathbf{D}_{sc}\right]}_{\underline{\mathbf{S}}(j\omega)}\,\underline{\mathbf{a}}(j\omega). \tag{31}$$

This transfer function is commonly known as the S-matrix[2] of the full structure. This matrix is evaluated in an interval $\Delta f = 3\ldots8\,\mathrm{GHz}$ on $1,001$ discrete frequency samples. In contrast, the scattering properties of the complete structure are computed by means of a hexahedral discretization ($192,456$ mesh cells) in combination with the CST Microwave Studio® Fast S-parameter Solver [8]. Fig. 5 and Fig. 6 show the magnitude and the phase of the transmission between the external ports obtained by the concatenation scheme and by the full calculation.

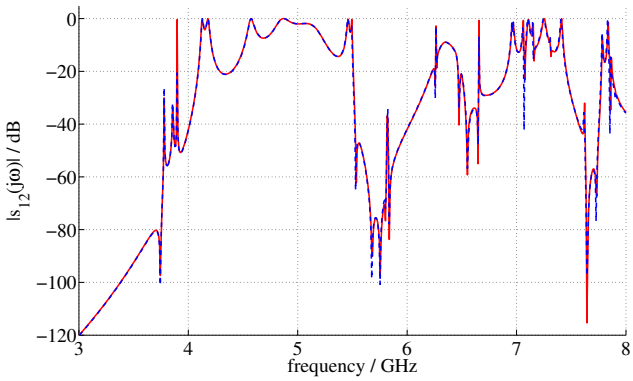


Figure 5: Absolute value of the transmission coefficient from left to right HOM coupler obtained by the concatenation scheme (red) and by direct calculation (blue).

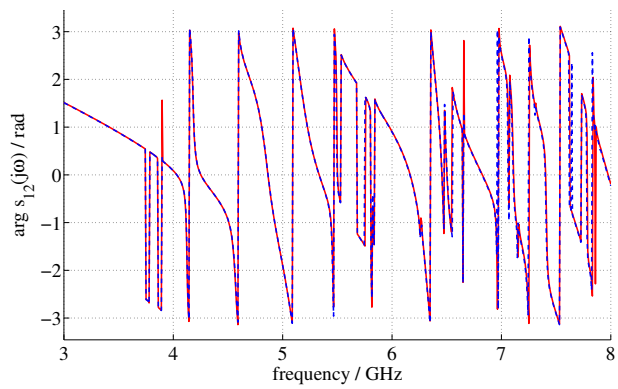


Figure 6: Phase of the transmission coefficient from left to right HOM coupler. The red curve is obtained by the concatenation scheme and the blue curve by direct calculation.

[2] In contrast to [2, 3] this matrix is not a priori sampled on discrete frequency points but is available in terms of an analytic function [6].

External Quality Factors

A typical quantity of interest in accelerator design is the external quality factor of modes inside the cavity structure:

$$Q_{ext,\nu} = \frac{\omega_\nu W_{stored,\nu}}{P_{ports,\nu}}. \tag{32}$$

Here ω_ν is the angular frequency of the ν-th mode, $W_{stored,\nu}$ the energy stored in the ν-th mode and $P_{ports,\nu}$ the averaged loss power of the ν-th mode due to energy propagation through the external ports. The external quality factor can be computed by the lumped model via

$$Q_{ext,\nu} = \frac{\mathrm{Im}(\lambda_\nu)}{2\mathrm{Re}(\lambda_\nu)}, \tag{33}$$

where λ_ν are the eigenvalues of the scattering formulation state matrix $\mathbf{A}_{sc}$. To benchmark the values computed by the concatenation scheme in combination with the upper equations, the eigenmode solver of CST Microwave Studio® is used to directly calculate external quality factors of the complete structure. In Fig. 7 the comparison for the modes in the interval Δf is given.

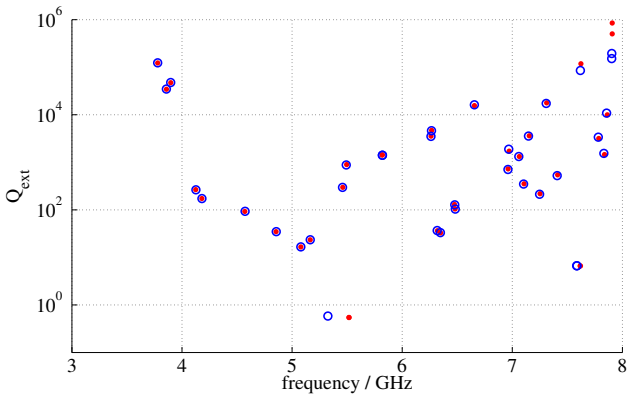


Figure 7: External quality factors of the structure under test in the interval Δf. The red dots are computed by the concatenation scheme, whereas the blue dots are computed by a direct calculation.

Transient System Responses

Often in accelerator design the response of complex structures due to different transient excitation signals is of interest (e.g. for the analysis of field filling and defilling processes). Once it is established, the state space system (28) and (29) allows for the computation of these responses using ordinary differential equation (ODE) solvers. Fig. 8 plots the transmitted signal at the right HOM coupler of the example structure due to the sinusoidal excitation

$$a_1(t) = \Big(1 - \exp(-t/T_r)\Big)\sin(2\pi f_0\, t) \tag{34}$$

at the left coupler, where $f_0 = 3.8976$ GHz and $T_r = 1$ ns. The red curve is obtained by the established lumped model in combination with an ODE solver and the blue curve is determined by the CST Microwave Studio® transient solver [8].

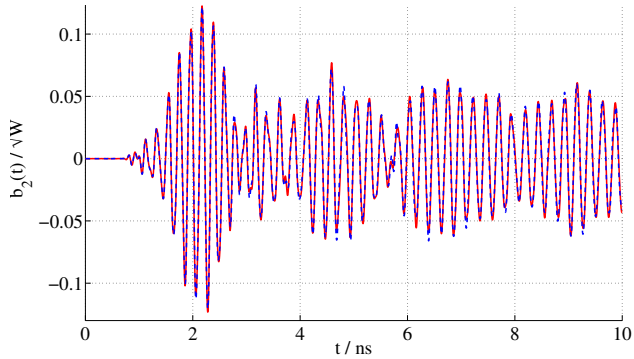


Figure 8: Transmitted signals $b_2(t)$ observed in right coupler due to excitation $a_1(t)$ at the left coupler. The red curve is obtained by the lumped model, whereas the blue curve is calculated with CST's transient solver.

CONCLUSIONS

The actual contribution presents a proof of concept for a method to create lumped equivalent models of complex RF structures. Based on a set of 3D eigenmodes of sections of the decomposed structure, state space equations are created. These equations are linked together to obtain a lumped equivalent model of the complete structure. This lumped model allows for the computation of complex structure's transfer function, transient system responses or secondary quantities. The validation example given in this paper shows that the model approximates the physical properties of the full structure in a very reasonable manner.

ACKNOWLEDGEMENTS

The authors would like to thank B. Riemann, A. Neumann, T. Galek, H.-W. Glock and C. Bahls for profound discussions on state space modelling of RF components.

REFERENCES

[1] https://www.slac.stanford.edu/grp/acd/ace3p.html

[2] K. Rothemund et al., "Eigenmode Calculation in Long and Complex RF Structures Using the Coupled S-Parameter Calculation Technique," TESLA-REPORT 2000-33.

[3] H.-W. Glock et al., "CSC - A Procedure for Coupled S-Parameter Calculations," IEEE Trans. Magn. Vol. 38, 2002.

[4] M. Dohlus et al., "Calculation of Frequency Domain Parameters using 3D Eigensolutions", International Journal of Numerical Modelling Vol. 12, 1999.

[5] T. Wittig et al., "Model Order Reduction for Large Systems in Computational Electromagnetics," LinAlgApp 415, Special Issue on Order Reduction of Large-Scale Systems, 2006.

[6] T. Flisgen et al., "Lumped Equivalent Models of Complex RF Structures based on Segment's Eigenmodes," Paper draft to be submitted to the IEEE.

[7] T. Khabibouline et al., "Higher Order Modes of a 3rd Harmonic Cavity with an Increased End-cup Iris," TESLA-REPORT 2003-01, DESY, Hamburg 2003.

[8] CST Studio Suite®, CST AG, 64289 Darmstadt, Germany.

EIGENMODE COMPUTATION FOR FERRITE-LOADED CAVITY RESONATORS *

K. Klopfer**, W. Ackermann, T. Weiland
Technische Universität Darmstadt, Institut für Theorie Elektromagnetischer Felder (TEMF)
Schlossgartenstraße 8, 64289 Darmstadt, Germany

Abstract

For acceleration of charged particles at the heavy-ion synchrotron at the GSI Helmholtzzentrum für Schwerionenforschung in Darmstadt two ferrite-loaded cavity resonators are installed within the ring. Their eigenfrequency can be tuned by a properly chosen bias current and thereby modifying the differential permeability of the ferrite material. The goal of the presented work is to numerically determine the lowest eigensolutions of accelerating ferrite-loaded cavities based on the finite integration technique (FIT). The newly developed solver includes two subcomponents: Firstly, a magnetostatic solver supporting nonlinear material for the computation of the magnetic field which is excited by the specified bias current. This enables to linearize the constitutive equation for the ferrite material at the current working point, at which the differential permeability tensor is evaluated. Secondly, a Jacobi-Davidson type eigensolver for the subsequent solution of the nonlinear eigenvalue problem. Particular emphasis is put on the implementation to enable efficient distributed parallel computing. First numerical results for biased ferrite-filled cavity resonators are presented.

INTRODUCTION

Within the heavy-ion synchrotron at GSI two ferrite-loaded cavity resonators are operated to continuously accelerate the injected charged particles. Inside the cavity housing ferrite ring cores are installed around the beam pipe. A magnetic field is established in these rings by means of two different current windings: Firstly, a field constant in time due to the bias current, and, secondly, an additional time-harmonic component induced via radio frequency coupling. During the acceleration phase the resonance frequency has to be adjusted to reflect the increasing speed of the heavy ions. This can be achieved by properly choosing a bias current and thereby modifying the differential permeability of the ferrite material. For the SIS 18 ferrite cavity, biasing enables to alter the resonance frequency in a range from about $0.6\,\mathrm{MHz}$ to $5.0\,\mathrm{MHz}$. A detailed description of the SIS 18 ferrite cavity can be found in [1].

In this paper, the main aspects of the applied numerical approach are briefly summarized followed by helpful remarks on efficient parallel computing. After that, the functionality of the solver is demonstrated based on two different simple examples.

* Work supported by GSI
** klopfer@temf.tu-darmstadt.de

COMPUTATIONAL APPROACH

The relevant fundamental relations used for the calculation of eigenmodes of biased ferrite-loaded cavity resonators as well as the applied approach for the numerical computation were already discussed in [2]. Hence, here the most important aspects are picked up and summarized.

The dependence of the eigenmodes on the differential permeability leads to several consequences: The static magnetic field generated by the bias current has to be calculated beforehand by means of a nonlinear magnetostatic solver. The constitutive equation is then linearized at the specified working point. Moreover, the frequency dependence of the permeability tensor results in a nonlinear eigenvalue problem. A dedicated solver for parallel computing has been developed to meet the tight requirements. Presently, only Hermitian eigenvalue problems are supported. Consequently, the current version of the solver is applicable for loss-less materials as well as models discretized on equidistant grids.

PARALLEL COMPUTING

The realization of the solver should allow an efficient computation on distributed memory machines. To this end, it aims at a high computation to communication ratio as well as a good load balancing. Regarding the first aspect, the degrees of freedom (DOFs) of the FIT are arranged such that all matrices have only few non-zero components in their far off-diagonal regions. This directly leads to reduced communication between the individual processes and thus to a better parallel efficiency. For illustration, the effect of the re-ordering is shown with the structure of the system matrix of the eigenvalue problem on the left of Fig. 1. Having in mind that the number of computations one processing unit has to perform is approximately proportional to the number of non-zeros of the system matrix which are owned by this process, it can also be seen from this figure that the load balancing is rather poor. The reason for this is that many variables are included that in fact are zero in the FIT because they are allocated on elements outside the computation domain (including perfect electric conductor cells) or due to boundary conditions. Consequently, in the current implementation all the pseudo DOFs are completely removed beforehand and therefore very good load balancing is obtained.

The reduction of the system size automatically results in further benefits, not only for parallel computing. In fact, the beneficial impact on the memory allocation at the time

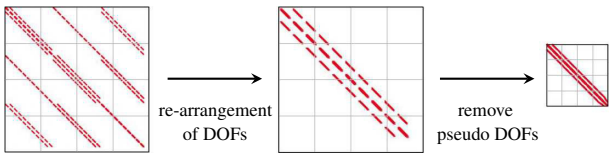


Figure 1: Structure of the system matrix of the eigenvalue problem for arrangement of the DOFs in standard FIT (left), after re-arrangement (middle) and additional removal of pseudo DOFs (right). Non-zero matrix elements are shown in red. The thin horizontal lines indicate how the data is partitioned on the different processes (four in this this example).

the preconditioner is constructed is even more crucial as the preconditioner is computed for the system matrix which is shifted by a scaled unit matrix. Since it is not possible to explicitly preallocate memory for vanishing diagonal elements when using PETSc- (Portable, Extensible Toolkit for Scientific Computation [3]) routines for sparse matrix-matrix multiplication, the consequence is that additional memory has to be allocated multiple times, which is an expensive operation. The complete removal of these elements is one way to solve this issue.

NUMERICAL EXAMPLES

Numerical results for two examples, both biased cavity resonators, are presented. Whereas the first one aims at the verification of the nonlinear eigensolver, the second one is a simple model of a cavity filled with ferrite ring cores.

Biased Cylinder Resonator

In the following, a lossless, ferrite-filled cylindrical cavity resonator longitudinally biased by a homogeneous static magnetic field is considered. Assuming that its magnetic properties can be described by the Polder tensor [4], a characteristic equation determining the resonance frequencies can be formulated analytically [5, 6]. The parameters of the test model are the same as for the one described in [2]. Yet, here the numerical computation is carried out for different orientations of the cylinder axis to the coordinate axes keeping the external magnetic field aligned with the cylinder axis. This way the construction of the permeability tensor for an arbitrarily oriented magnetic field is tested additionally. The eigenmodes for the lowest eigenfrequencies are compared with the ones obtained numerically with the implementation of the nonlinear eigensolver. Note that for a reliable verification it is crucial to assign the modes correctly. If one sorts the eigenmodes simply in ascending order with respect to the eigenvalues, by comparing the field solutions it shows that the order of the modes is different compared to the semi-analytical calculation. This is, however, only the case for discretizations on a rather coarse mesh. By a proper refinement of the grid, the expected order is retained. Taking this into account, good accordance of the numerical values with the analytical results is observed for all tested orientations, which is shown as an example in Fig. 2 for the orientation as depicted in Fig. 3. Moreover, both the accuracy at a given number of DOFs and the convergence order coincide with the results obtained for the cylinder axis aligned with the z-axis (cf. [2]).

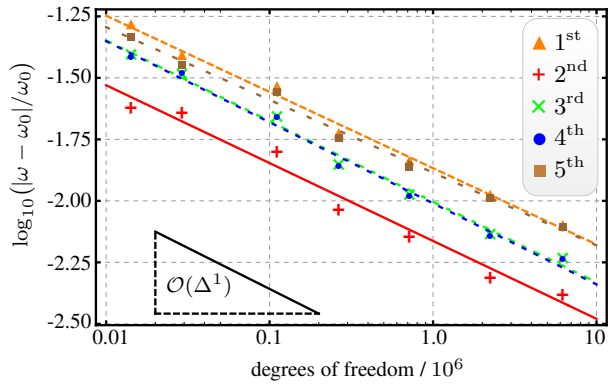


Figure 2: Relative deviation of the numerically obtained value ω to the analytical result ω_0 as a function of the DOFs for the five lowest eigenfrequencies for a lossless, longitudinally biased, ferrite-filled cylindrical cavity resonator. The cylinder is oriented as shown in Fig. 3.

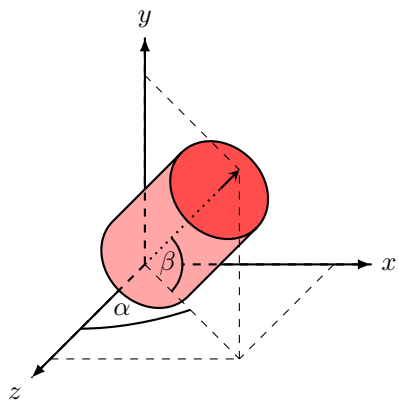


Figure 3: Orientation of the cylinder in the coordinate system. For the convergence study shown in Fig. 2, the cosine of β, i.e. the angle between the cylinder axis and the x-z-plane, is $\sqrt{2/3}$; the angle α between the projection of the cylinder axis onto this plane and the z-axis is $45°$.

For the model of the rotated cylinder the usage of the reduced system matrix as explained in the previous section is of particular importance. Since for the discretization of the model the cylinder is embedded in a larger box with the remaining space filled up with perfect electric conductor cells, the total number of mesh cells is much larger than the actual cells of the cylinder. For instance, for the orientation shown in Fig. 3 the used mesh for the finest discretization consists of approximately $22 \cdot 10^6$ cells whereas the DOFs, which are truly non-zero, amount to about only $6.2 \cdot 10^6$. Yet, thanks to the reduction of the system size not the total number of mesh cells but only the number of non-zero DOFs matters.

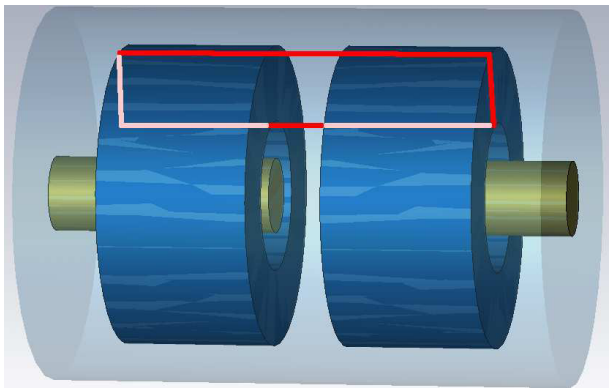


Figure 4: CAD model of a simple biased cavity with ferrite ring cores modelled with CST STUDIO SUITE® [7]. The current winding is indicated by the solid red lines.

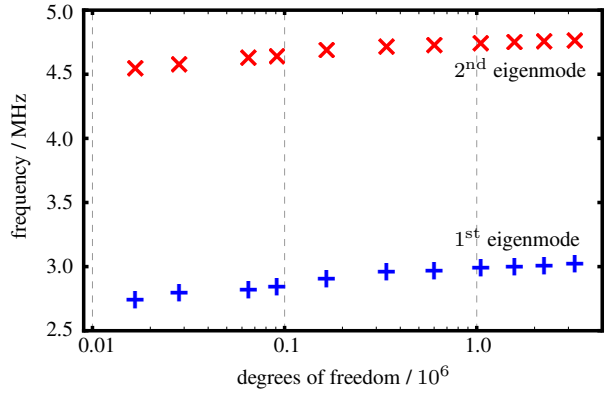


Figure 5: Eigenvalues as a function of the DOFs for the two lowest eigenmodes computed using the fully nonlinear approach for the simple cavity model.

Biased Cavity With Ferrite Ring Cores

The second example may be regarded as a simplified model of the GSI SIS 18 ferrite cavity. Inside the cavity housing ferrite ring cores are installed around the beam pipe, one of them on each side of the centric gap (cf. Fig. 4). The bias magnetic field is excited by a current winding around these rings. For simplicity, the two parts of the beam pipe are modelled as cylinders filled with perfect electric conducting material. The parameters of the model are as follows: The total length of the structure is 140 cm, its radius 50 cm; the radius of the beam pipe is 10 cm, the inner and outer radius of the ring cores are 20 cm and 40 cm, respectively; the length of each ring is 40 cm; the length of the gap is 20 cm; finally, the bias current is 2 kA. The ferrite material is characterized by the constitutive relation

$$B(H) = \mu_0\, 2.5 \cdot 10^4 \tanh\left(H \cdot 10^{-2} \frac{\mathrm{m}}{\mathrm{A}}\right) \frac{\mathrm{A}}{\mathrm{m}} + \mu_0 H \quad (1)$$

and a relative permittivity of $\epsilon_r = 1$. The accuracy of the magnetostatic solver is chosen such that the maximum change of the relative permeability between subsequent nonlinear iterations in any of the mesh cells does not exceed 10^{-4} in the final iteration, which corresponds to a relative change of the norm of the magnetic field in the order of 10^{-6}. Moreover, for the nonlinear eigenvalue solver the accuracy is set as in the previous example. With these settings the fully nonlinear computation of the nine lowest eigenmodes, which are in the range from approximately 87 MHz to 300 MHz and all below magnetic resonance, is performed. The lowest mode is found to be suitable for acceleration. In order to test the reliability of the solver, the whole computation is repeated multiple times with a more and more refined mesh. For illustration the results for the first two eigenmodes are shown in Fig. 5.

Additionally, the effect of the nonlinearity is estimated by comparing the results with those of a linear calculation. In order to obtain reasonable results for the linear approach, the value of the relative permeability of the ferrite material is set homogeneously such that the lowest eigenfrequency of the linear and the nonlinear computation is approximately identical. Despite the fact that the linear solver is a priori provided with this accurate data, the relative deviation of the obtained eigenvalues of both approaches is still up to about 5% for the first nine eigenmodes.

SUMMARY AND OUTLOOK

The computation of eigenmodes of biased ferrite-cavities based on the FIT has been demonstrated. For this purpose, a new solver is developed for the evaluation of the permeability tensor at the working point defined by the bias magnetic field and the solution of the subsequent nonlinear eigenvalue problem. Particular emphasis is put on efficient computing on distributed memory machines. Whereas first numerical results for lossless biased cavity resonators have been presented, the support of lossy materials is planned in future implementations.

REFERENCES

[1] H. Klingbeil, "Ferrite cavities", CERN Yellow Report CERN-2011-007, pp. 299-317, 2012, arXiv:1201.1154.

[2] K. Klopfer, W. Ackermann and T. Weiland, "Eigenmode Computation for Ferrite-loaded Cavity Resonators", Proceedings of 3rd Int. Particle Accelerator Conference - IPAC 2012, MOPPC058, 2012.

[3] S. Balay et al., "PETSc Users Manual", ANL-95/11 - Revision 3.2, Argonne National Laboratory, 2011.

[4] D. Polder, "On the Theory of Ferromagnetic Resonance", Phil. Mag., 40, p. 99, 1949.

[5] M. L. Kales, "Modes in Wave Guides Containing Ferrites", J. Appl. Phys. 24, 604, 1953.

[6] G. C. Chinn, L. W. Epp and G. M. Wilkins, "Determination of the eigenfrequencies of a ferrite-filled cylindrical cavity resonator using the finite element method", IEEE Transactions on Microwave Theory Techniques, 43, May 1995.

[7] CST STUDIO SUITE®, CST Computer Simulation Technology AG, Darmstadt, Germany.

PRELIMINARY STUDY OF SINGLE SPIKE SASE FEL OPERATION AT 0.26 NANOMETERS WAVELENGTH FOR THE EUROPEAN XFEL

B. Marchetti[#], M. Krasilnikov, F. Stephan, DESY, Zeuthen, Germany
M. Dohlus, Y. Kot, I. Zagorodnov, DESY, Hamburg, Germany
J. Roensch-Schulenburg,University of Hamburg, Hamburg, Germany

Abstract

The production of ultra-short (fs or sub-fs long), high power radiation pulses in the X-ray spectral region, showing a single spike spectrum, represents a challenge for many existent SASE- FELs [1] [2].

In order to realize single spike operation the length of the electron bunch after compression must be extremely small [3] (less than a micrometer) and the consequent degradation of its emittance has not to suppress the radiation production.

Several technical restrictions, such as limits of diagnostics for small charges, RF jitter and micro-bunching instabilities play an important role in the choice of the operation working point.

In this paper we are going to study the feasibility of single spike or few spikes lasing for bunches with a charge of tens of pC in the European XFEL facility [4] giving some preliminary results concerning the choice of the working point.

INTRODUCTION

Single Spike Condition

The radiation produced by Free Electron Lasers (FELs) working in the Self Amplified Spontaneous Emission (SASE) configuration is characterized by an energy spectrum constituted by many spikes. The number of the spikes depends on the longitudinal properties of the electron bunch. In fact, once the bunch is injected into the undulator, the radiation emitted by the electrons located in a certain longitudinal position can be amplified only by other electrons placed within a fixed longitudinal distance (in the forward direction) from them. This distance is proportional to the cooperation length, defined as a radiation slippage in one power gain length [3]. The cooperation length depends on the emitted radiation's wavelength λ, according to:

$$L_c = L_{c1d}(1+\eta), \qquad (1)$$

where the parameter η, defined as in [5], takes into account transverse effects and beam energy spread, and L_{c1d} is the one-dimensional cooperation length, defined as:

$$L_{c1d} = \frac{\lambda}{\sqrt{3}\cdot 4\pi\rho}, \qquad (2)$$

with ρ being the one dimensional FEL parameter.

#barbara.marchetti@desy.de

Every longitudinal slice in the bunch whose length is $2\pi L_c$ contributes to a different spike in the energy spectrum of the produced radiation, thus in order to have a single spike, the e-bunch must be shorter than $2\pi L_c$.

It is very difficult to satisfy this condition in X-rays FELs where, according to equations (1) and (2), $2\pi L_c$ becomes very small, typically a fraction of μm.

Single Spike at the European XFEL

Several methods have been proposed in order to achieve short e-bunch lengths and thus sub-fs radiation pulses (see for example [2] and its references).

In this paper we start from the very simple case of strong compression of electron bunches having a charge of tens of pC. We characterize different compression setups considering the most recent layout of the European XFEL. Our aim is to give a starting point for further optimization. The principal effects influencing the electron beam acceleration and transport, such as RF wakefields and coherent synchrotron radiation (CSR) effects inside magnetic compressors have been included. Space charge force has been fully included in the injector, while only longitudinal space charge force has been considered in the rest of the linear accelerator.

BUNCH COMPRESSION AT XFEL

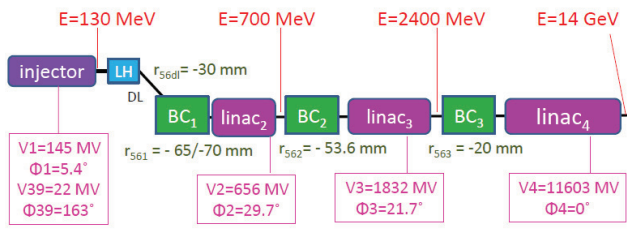


Figure 1: Scheme of the European XFEL layout including the parameters used in the following simulations. The three magnetic bunch compressors (BCs) have a variable curvature radius and thus a variable dispersion (R_{56}); in the work described in this paper the curvature radius in BC_1 has been changed in order to scan the bunch properties around the maximum compression point.

Layout Description

In Fig. 1 a scheme of the layout of the European XFEL is shown. Through the use of a third harmonic RF-cavity placed in the injector region, the correction of the non-linearity in the longitudinal phase space distribution of the e-bunch is possible. The laser heater (LH) can mitigate the effect of micro-bunching instability. The electron bunch compression starts in the dogleg and is

completed in three magnetic stages.

The total compression factor can be optimized by either changing the RF chirp before entering the magnetic compressors or the curvature radius inside the compressors (i.e. the current of the dipoles in the chicanes).

The transport and compression of the e-bunch has been recently optimized for different charges, always considering a bunch produced by a 20 ps lasting flat-top laser pulse illuminating the photo-cathode [6] [9-10].

Since for the nominal injector laser pulse duration the strong compression applied with small charges to reach a short bunch appeared to be dominated by RF-tolerances, it has been chosen here to work with shorter laser flat-top length at the cathode.

Working at extremely low charges (1 pC or less) it is possible to obtain the single spike condition while keeping the beam peak current high, the emittance small and, consequently, L_c extremely short [1] [2]. In this way attosecond lasting radiation pulses are in principle reachable but the lack of diagnostics able to fully characterize electron bunches with charge smaller than few tens of pC makes the transport and the compression of the beam practically very difficult [11]. When using higher charges, the shortest achievable bunch length increases but we gain the possibility of diagnose the electron beam in order to match it with the undulator. In this case, it is relatively easy to obtain a few spikes SASE radiation spectrum having an integrated power of the order of 10 GW, but the single spike condition can eventually be reached only by optimizing the peak current of the central slice of the bunch and partially spoiling the beam emittance in order to increase the value of L_c. We will present simulations for 20 pC and 50 pC charged bunches.

Simulations: Input and Method

The nominal electron bunch at the XFEL has 1nC charge and is generated illuminating the cathode with a laser whose RMS transverse size and flat-top duration are respectively 0.4 mm and 20 ps. In this paper the laser parameters have been scaled with respect to the nominal setup in order to maintain constant the charge density and use the machine parameters defined for the nominal setup as a starting point. For the 20 pC case a laser shape having 0.11 mm RMS transverse size and 5.43 ps flat-top duration (FWHM) has been set, while for the 50 pC case the laser pulse duration is 7.4 ps FWHM and has a transverse RMS size of 0.15 mm. In all cases the rise and fall time of the flat-top are 2 ps.

The transport in the injector has been simulated using the tracking code ASTRA [7]. The RF-wakefields have been added analytically as described in [9]. The transport through the laser heater, the dogleg and the bunch compressors has been simulated using the code CSRtrack [8]. Finally the transport in the three linac-sections has been carried out by multiplying the particles coordinates times the correspondent linac transport matrix and adding analytically the wake caused by longitudinal space charge using formulas described in [9].

The electron bunch compression has been scanned around its maximum compression point by changing the curvature radius inside the first bunch compressor (BC_1 in Fig. 1).

The longitudinal phase space distributions correspondent to the electron bunches after compression have been used as input for Genesis [12] simulations in order to evaluate the power and the energy spectrum of the emitted radiation. At the moment, longitudinal wakefields occurring in the undulator has been not taken into account and no tapering has been applied. Both of these steps are foreseen in the future.

Results

Figure 2 summarizes the results of the simulations. It shows the normalized transverse emittance (N.E.) and the energy spread of the electron bunch as a function of the curvature radius r_1 of the first bunch compressor. The maximum compression point for the 20 pC bunch lies in the region of r_1~3.14 m while for 50 pC corresponds to r_1~3.13 m. Only for the 20 pC bunch the over-compression region has been explored.

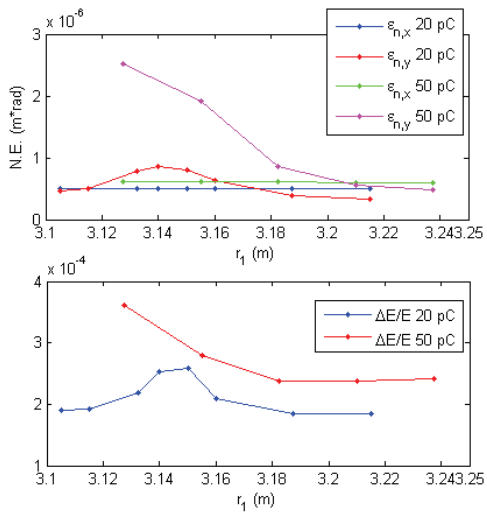


Figure 2: Summary of the compressed electron bunch characteristics as a function of the curvature radius (r_1) in BC_1. The normalized horizontal and vertical emittances ($\varepsilon_{n,x}$, $\varepsilon_{n,y}$) are plotted on the top and the total energy spread is plotted on the bottom figure.
The maximum compression point for the 20 pC bunch lies around r_1~3.14 m while for the 50 pC bunch corresponds to r_1~3.13 m.

In order to have an idea about the stability of the peak current with respect to the RF-jitter, all the simulation has been repeated changing the phase of the first accelerating cavity in the injector by ±0.005°, which represents a very conservative estimation for the RF-jitter. The jitter in this cavity has been indeed found to be the dominant one for the compression [10]. Fig. 3 shows the result of this analysis: it presents the slice peak current and the

normalized maximum deviation from its average value for each compression point. These quantities depend on the slice length that has been chosen to be close to $2\pi L_c$ at the maximum compression point, being about 0.28 μm for the 20 pC case and 0.35 μm for the 50 pC .

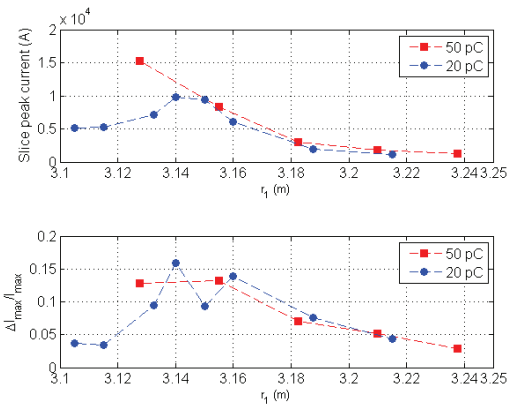


Figure 3: Slice peak current (on top) and normalized maximum deviation from its average value changing the phase of the first accelerating cavity in the injector by ±0.005° (on the bottom) as a function of the curvature radius (r_1) in BC_1. The slice peak current has been calculated considering a slice length of 0.35 μm for the 50 pC case and 0.28 μm for the 20 pC one.

SASE RADIATION

The electron bunch longitudinal phase space distributions have been used as input for Genesis simulations. In none of the simulated cases it has been possible to obtain a single spike in the spectrum nevertheless a few spikes SASE spectrum with power of ~10 GW and a radiation pulse length of the order of fs has been achieved. In Fig. 4 the longitudinal phase space corresponding to the 50 pC charge bunch compressed using r_1=3.1275 m in BC_1 at the entrance of the undulator is shown as an example. In the middle the longitudinal profile of the emitted power at saturation is plotted while on the bottom the spectrum of the emitted radiation at the saturation point is plotted.

CONCLUSIONS AND OUTLOOK

Preliminary start-to-end simulations have been performed for the optimization of single spike emission at the wavelength of 2.6 Å at the European XFEL.

Few spikes emission in the X-rays can be obtained relatively easily by compressing bunches with charges of tens of pC.

Further study is needed to decrease the number of spikes by increasing the peak current in the central slice of the bunch (tuning, for example, the energy chirp before the first magnetic compressors) and slightly spoiling the beam emittance.

The study of introducting a taper in the undulator is also foreseen.

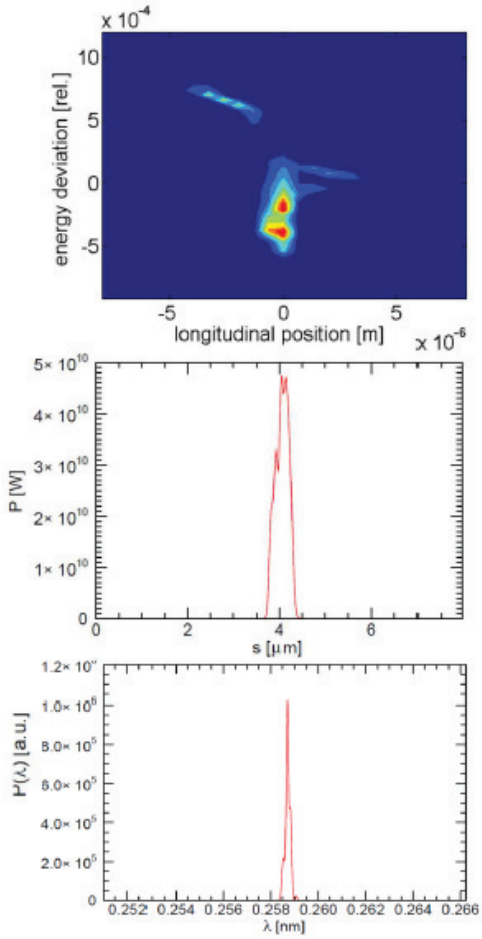


Figure 4: Longitudinal phase space distribution of the electron bunch at the entrance of the undulator, (on top), power of the emitted radiation as a function of the longitudinal coordinate within the electron bunch and spectrum of the emitted radiation at the saturation point.

REFERENCES

[1] J. B. Rosenzweig et al., NIM A 593 (2008) 39-44.
[2] S. Reiche et al., NIM A 593 (2008) 45-48.
[3] R. Bonifacio et al., PRL vol. 73 n.1 (1994).
[4] XFEL Technical design report, July 2007.
[5] M. Xie in Proc. of the 1995 PAC, IEEE, Piscataway, NJ,1995, p.183.
[6] Y. Kot, MOP003, Proceedings of LINAC 2010
[7] K. Flöttmann et al, in Proc. PAC 2003.
[8] M. Dohlus, T. Limberg, in Proc. FEL 2004.
[9] I. Zagorodnov, M. Dohlus Phys. Rev. ST Accel. Beams 14, 014403 (2011).
[10] I. Zagorodnov, Beam Dynamics Simulations for XFEL, www.desy.de/xfel-beam/s2e/data/xfel_2011/NewResults.pdf
[11] L. Wang, Y Ding, Z. Huang, THPC104, Proceedings of IPAC 2011.
[12] S. Reiche, PhD dissertation, University of Hamburg 1999.

RECONSTRUCTION OF VELOCITY FIELD*

Dmitri A. Ovsyannikov, Elena D. Kotina, St. Petersburg State University, St. Petersburg, Russia

Abstract

In this paper we suppose that the distribution density of particles in phase space is known. Using Liouville's equations the problem of finding velocity field is considered as a minimization problem. Thus the problem of determination of velocity field is reduced to solving of elliptic system of Euler-Lagrange equations.

INTRODUCTION

Different inverse problems of electrodynamics have been the subject of attention of many researchers. In particular, solving of inverse problems of electrodynamics, where by pre-assigned motions (given velocity field) electromagnetic fields were determined, had been investigated in works of G.A.Grinberg, A.R.Lucas, B.Meltzer, V.T.Ovcharov, V.I.Zubov , E.D. Kotina [1-9]. It shoud be noted, that the problem of determination of velocity field is a separate task. In particular, the task of determination of velocity field could be considered as the problem of the optimal control theory [10]. In this case it is needed to find the velocity field securing necessary beam dynamics. In this paper we suppose that the distribution density of particles in phase space is known. The problem of finding the velocity field is considered as a minimization problem. Similar problem is widely discussed in the literature for image processing based on the so-called optical flow. This approach was also used for the motion correction for radionuclide tomographic studies [11]. In this work the problem of determining the velocity field in solving the problem of charged particle beam formation in a stationary magnetic field is also considered.

PROBLEM STATEMENT

Let the dynamical system be given by a differential equation

$$\frac{dx}{dt} = f(t,x), \qquad (1)$$

where t is independent variable further referred to as time, x is $n-$vector of the phase coordinates $x_1, x_2, ..., x_n$, , f is $n-$ dimensional vector function. We assume that $f(t,x)$ is defined and, together with the partial derivates $\partial f/\partial x$ it is continuous with respect to the variables t, x on $T_0 \times R^n$, where $T_0 = [t_0, T] \subset R^1$, the number t_0, T are fixed. We assume that solution $x = x(t, t_0, x_0)$ is defined on the entire interval T_0 for any $x_0 \in R^n$.

*Work supported by SPbGU, grant 9.39.1065.2012

Let us consider together equation (1) the following equation:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial x} f + \rho \cdot \mathrm{div}_x f = 0, \qquad (2)$$

under initial condition

$$\rho(t_0, x) = \rho_0(x), \qquad (3)$$

where $\rho_0(x) -$ given function and $\rho = \rho(t,x)$.

We will consider arbitrary region $G_{t0} \subset R^n$. Suppose that, by the system (1), G_t is the image set of G_{t0}. Using the equation (2) we obtain

$$\int_{G_t} \rho(t, x_t) dx_t = \int_{G_{t0}} \rho_0(x_0) dx_0, \qquad t \in T_0. \qquad (4)$$

If for some nonnegative scalar function $\rho(t,x)$ the equality (4) holds for any $G_{t0} \subset R^n$, than let us say that the system (1) has integral invariant of order n. The function $\rho(t,x)$ is called the kernel or the density of the integral invariant [7]. From physical point of view the equality (4) may be treated as conservation of the particle (charge) mass along the trajectories of the system (1). The equation (2) is called the transport equation[10]. Note that the equation (2) is also called the generalized Liuville equation or Liuville's equation when the system (1) is given in canonical conjugate variables, here $\mathrm{div}_x f(t,x) = 0$.

Suppose that the phase trajectory family equation (1) corresponds to the set of random initial values of the coordinates with the density of probability distribution (3) and moreover $\int_{M_0} \rho_0(x_0) dx_0 = 1$. The function $\rho(t,x)$ may be treated as the variation of the density of probability distribution in time, and in the phase space of coordinates of the dynamical system (1). In this case the equation (2) is called the Fokker-Planck-Kolmogorov equation.

Further we suppose that given function $\rho(t,x)$ satisfies the transport equation (2). The problem is to restore the velocity field of the system (1), i.e. to find function $f(t,x)$. In common case, this is ill-posed problem [12]. So we will be use the method of regularization. Let us fix some moment t and formulate the problem of determination function $f(t,x)$ as a minimization problem. We will now introduce the functional

$$J(f) = \int_M (\varphi_1 + \alpha^2 \varphi_2) dx\,, \qquad (5)$$

where M – region in R^n of a nonzero measure, α^2 – parameter of regularization,

$$\varphi_1 = \left(\frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial x} f + \rho \cdot \mathrm{div}_x f \right)^2, \quad \varphi_2 = \sum_{i,j=1}^{n} \left(\frac{\partial f_j}{\partial x_i} \right)^2.$$

Thus we find the velocity field in the class of smooth enough functions.

EULER-LAGRANGE EQUATIONS

Let us introduce the following operator:

$$A(f) = -\alpha^2 \Delta f - \rho\, \mathrm{grad}\, \mathrm{div}(\rho f). \qquad (6)$$

Here $f = (f_1, f_2, ..., f_n)^T$, $\Delta f = (\Delta f_1, \Delta f_2, ..., \Delta f_n)^T$.
Euler-Lagrange equations for functional (5) can be represented as operator equation

$$A(f) = g\,, \qquad (7)$$

where $g = \rho\, \mathrm{grad}\, \rho_t$.

Further we presuppose that the set M has smooth enough boundary Γ and

$$f(x) = 0\,, \quad x \in \Gamma\,. \qquad (8)$$

Operator $A(f)$ is positively definite differential operator. Indeed, if we consider corresponding scalar product than we can obtain the following:

$$(A(f), f) = -\alpha^2 \int_M (\Delta f, f) dx - \int_M (\rho\ \mathrm{graddiv}(\rho f), f) dx =$$
$$\alpha^2 \int_M \varphi_2 dx + \int_M (\mathrm{div}(\rho f))^2 dx > 0, \quad \alpha \neq 0, \quad f \neq 0.$$

The system of equations, defined by (7) is the system of strongly elliptic differential equations. It is well-known that strongly elliptic systems behave themselves as single elliptic equation from the point of view their solvability [13, 14].

It should be noted that functional (5) is the quadratic functional. It differs only by a constant from the following functional

$$J(f) = (A(f), f) - 2(g, f)\,. \qquad (9)$$

Since operator $A(f)$ is positively definite differential operator, the solution equation (7) is the solution of minimization problem of functional (9) or functional (5) [15]. Moreover under these assumptions there exists a unique generalized solution equation (7) with boundary conditions (8). In this connection under enough smoothness coefficient of the equation (7) there exists unique classic solution of it due to the corresponding embedding theorems [15 - 17].

Let us consider the three-dimensional case, i.e., we consider the density as a function $\rho = \rho(t, x, y, z)$, and search the velocity field (1) as

$$\dot{x} = u(t, x, y, z),$$
$$\dot{y} = v(t, x, y, z),$$
$$\dot{z} = w(t, x, y, z).$$

Euler-Lagrange equations in this case have the form

$$-\alpha^2 \begin{pmatrix} u_{xx} + u_{yy} + u_{zz} \\ v_{xx} + v_{yy} + v_{zz} \\ w_{xx} + w_{yy} + w_{zz} \end{pmatrix} - \rho^2 \begin{pmatrix} u_{xx} + v_{xy} + w_{xz} \\ u_{xy} + v_{yy} + w_{yz} \\ u_{xz} + v_{yz} + w_{zz} \end{pmatrix} -$$
$$\rho \begin{pmatrix} 2\rho_x & \rho_x & \rho_x \\ \rho_y & 2\rho_y & \rho_y \\ \rho_z & \rho_z & 2\rho_z \end{pmatrix} \begin{pmatrix} u_x \\ v_y \\ w_z \end{pmatrix} -$$
$$\rho \begin{pmatrix} \rho_y & 0 & 0 \\ 0 & \rho_x & 0 \\ 0 & 0 & \rho_x \end{pmatrix} \begin{pmatrix} v_x \\ u_y \\ u_z \end{pmatrix} - \rho \begin{pmatrix} \rho_z & 0 & 0 \\ 0 & \rho_z & 0 \\ 0 & 0 & \rho_y \end{pmatrix} \begin{pmatrix} w_x \\ w_y \\ v_z \end{pmatrix} -$$
$$\rho \begin{pmatrix} \rho_{xx} & \rho_{xy} & \rho_{xz} \\ \rho_{xy} & \rho_{yy} & \rho_{yz} \\ \rho_{xz} & \rho_{yz} & \rho_{zz} \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \rho \begin{pmatrix} \rho_{tx} \\ \rho_{ty} \\ \rho_{tz} \end{pmatrix}.$$

These equations may be used if we consider, for example, the distribution density of particles in configuration space.

THE BEAM FORMATION IN MAGNETIC FIELD

Here particular case of reconstruction of velocity field for solving of problem charged particle beam formation in stationary magnetic field is considered. In this case velocity field is determined in analytical form by given density distribution of charged particles.

Further the approach suggested by V. I. Zubov is used. Let us describe charged particle motion in stationary magnetic field by the system of differential equations

$$\frac{dX}{dt} = Y, \qquad \frac{d(mY)}{dt} = eY \times B, \qquad (10)$$

where we have $X = (x_1, x_2, x_3)$ as coordinate vector of particle in the Cartesian coordinate system,

$Y=(y_1,y_2,y_3)$ as particle velocity vector, m is particle mass, e is particle charge, B is vector function, determining magnetic induction , t is time.

Let us denote η - velocity field in configuration space, then

$$\dot{X}=\eta(t,X). \tag{11}$$

In V.I. Zubov's works it has been shown, that stationary magnetic field (i.e. vector function B which satisfies Maxwell equation div $B=0$) in which charged particles motion accords to the velocities field (11), can be represented in the form $B=-\frac{m}{e}\operatorname{rot}\eta$. This means that in configuration space charged particles trajectories determined by system (10) under the same initial conditions will coincide with trajectories of the system (11).

However, this representation does not cover all possibilities, and for practical realization has been more natural to seek the magnetic field in the form [8, 9]

$$B=-\frac{m}{e}\operatorname{rot}\eta+h\eta, \tag{12}$$

where $h=h(x_1,x_2,x_3)$ is an arbitrary function satisfying the following condition $\operatorname{div}(h\eta)=0$.

Let us consider stationary axially symmetric magnetic field and cylindrical system of coordinates (r,θ,z). The radial motion of particles in this case is described by the equation

$$\frac{dr}{dz}=f(r,z). \tag{13}$$

Equation (2) in this case has the form

$$\frac{\partial\rho(r,z)}{\partial z}+\frac{\partial\rho(r,z)}{\partial r}f(r,z)+\rho(r,z)\frac{\partial f(r,z)}{\partial r}=0. \tag{14}$$

It is necessary to find a function $f=f(r,z)$, satisfying the equation (14).

We define the function of the density distribution in the form of a normal distribution

$$\rho(r,z)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{(r-\bar{r}(z))^2}{2\sigma^2}\right).$$

Then let $\bar{r}(z)=r_0\cos(z_0z)+a_0$. Solving equation (14), we can show that the velocity field in this case can be represented as

$$f=r_0z_0\sin z-r_0z_0\sin(z_0z)\times \exp\left(\frac{r^2-2(r_0\cos(z_0z)+a_0)}{2\sigma^2}\right).$$

Thus, specifying the required density distribution we can determine the velocity field on which to build further magnetic field according to the algorithm described in the works [8,9].

REFERENCES

[1] G.A. Grinberg, "Selected Questions of Mathematical Theory of Electric and Magnetic Phenomena," Izd. Ak. Nauk, pp.507-535, 1948.

[2] B. Meltzer, "Single-Component Stationary Electron Flow under Space-Charge Conditions," Journal of Electronics, v.7, N1, pp.118-127, 1956.

[3] A.R. Lucas, B. Meltzer, "A General Theorem for Dense Electron Beams," Journal of Electronics and Control, v.4, N2, pp.160-164, 1957.

[4] V.T. Ovcharov, "Axially Symmetric Electron Beam of Given Form," Dokl. Ak. Nauk, v.107, N1, pp.47-50, 1956.

[5] V.T. Ovcharov, "The Theory of the Formation of Electron Beams", Technology and Electronics, v.2, № 6, pp.696-704. 1957.

[6] V.I. Zubov, "To Control the Motion of Charged Particles in a Magnetic Field", Doklady AN USSR, v.232, № 4, pp.798-799, 1977.

[7] V.I. Zubov, *Dynamics of Control System* (Moscow: Izd. Vis. Shcola, 1982), 285.

[8] E.D. Kotina, "Formation of a Given Beam Dynamics in a Magnetic Field" Vestn. St. Petersburg. University. Ser.10, №4, pp.77-82, 2006.

[9] E.D. Kotina, "Beam Dynamics Formation in Magnetic Field" EPAC'02, Paris, June 2002, WEPRI112, p.1264 (2002).

[10] D.A. Ovsyannikov, *Modelling and Optimization of Charged Particle Beam Dynamics*, (L.: Leningrad Univ. University Press, 1990), 312.

[11] E.D. Kotina, K. M. Maximov "Motion Correction in SPECT and Planar Radionuclide Studies", Vestnik of St. Petersburg University, Ser.10, №1, pp. 29-36, 2011.

[12] A.N. Tikhonov, V.Y. Arsenin, *Methods for Solving Ill-posed Problems,* (Moscow: "Nauka", 1979), 288.

[13] A.A. Samarskii, E.S. Nikolaev, *Methods of Solving Grid Equations,* (Moscow: "Nauka", 1978), 324.

[14] M. I. Vishik, "On Strongly Elliptic Systems of Differential Equations", Mathematical Collection, v.29 (71), № 3. pp. 615-676, 1951.

[15] K. Rektorys, *Variational Methods in Mathematics, Science and Engineering,* (Prague: Dr. Reidel Publ. Comp., 1980), 590.

[16] O. A. Ladyzhenskaya, *Boundary Value Problems of Mathematical Physics,* (Moscow: "Nauka", 1973), 408.

[17] Nikol'skii, *Approximation of Functions of Several Variables and Embedding Theorems*, (Moscow: "Nauka", 1977), 456.

ON ACCELERATOR DRIVEN SUBCRITICAL REACTOR POWER GAIN

Anna G. Golovkina, Dmitri A. Ovsyannikov, SPbSU, St. Petersburg, Russia
Igor V. Kudinovich, Krylov Shipbuilding Research Institute, St . Petersburg, Russia

Abstract

The accelerator driven system (ADS) with subcritical reactor is considered. Such systems demonstrate high safety, due to the fact, that the reactor operates at sub-critical level. The problem of the reactor power rate maximisation on fixed values of effective multiplication factor and the external neutron source (neutron generating target) intensity is studied. In this paper the main attention is paid to the reactor core optimisation. Some ways of ADS power rate gain and optimised reactor core parameters are proposed.

INTRODUCTION

Since the early 1990's, accelerator driven systems (ADS - subcritical reactors with external neutron source generated by proton accelerators through a spallation target) have been proposed [1] for addressing certain missions in advanced nuclear fuel cycles. Institutes throughout the world have conducted numerous programs evaluating the role of ADS in nuclear waste transmutation and energy production. The interest is induced by a number of ADS different applications, for example:

- Transmuting selected isotopes present in nuclear waste (e.g., actinides, fission products) to reduce the burden these isotopes place on geologic repositories.
- Generating electricity and/or process heat.
- Producing fissile materials for subsequent use in critical or sub-critical systems by irradiating fertile elements.

There are needed high energy proton beams for transmutation problems. Such beams could be obtained only inlarge and expensive accelerators. So using them in ADS designed for power generation isn'teconomical effective. There should be used cheaper accelerators with lower output beam energy, so some other ways to increase reactor power rate are needed to be find.

EXTERNAL NEUTRON SOURCE AMPLIFICATION

In the general case the neutron flux $\Phi(r,E)$ in subcritical reactor is described by equation:

$$\mathbf{M}\Phi(r,E) = -\mathbf{M}_1\Phi(r,E) - q(r,E), \quad r \in V, E_T \le E \le E_f; \qquad \Phi(r_S,E) = 0, r_S \in S, \tag{1}$$

where $\mathbf{M}$ - operator describes transport, slowing-down and absorbation of neutrons,

M_1 - operator describes fission neutron source,

$q(r,E)$ - intensity of the external neutron source,

E_t - thermal neutron energy;

E_f - fission neutron energy

The multiplication factor of the reactor core k_{eff} does not depend on the external neutron source intensity and it is defined for critical fictitious system by equation:

$$\mathbf{M}\Phi_0(r,E) = -\frac{1}{k_{\text{eff}}}\mathbf{M}_1\Phi_0(r,E), \quad \Phi_0(r_S,E) = 0, r_S \in S \tag{2}$$

Eq. (2) can be written as:

$$\mathbf{M}\Phi_0(r,E) = -\mathbf{M}_1\Phi_0(r,E) - \frac{1-k_{\text{eff}}}{k_{\text{eff}}}\mathbf{M}_1\Phi_0(r,E). \tag{3}$$

Equations (1) and (3) are the same if spatial and energy distribution of the external source is:

$$q(r,E) = \frac{1-k_{\text{eff}}}{k_{\text{eff}}}\mathbf{M}_1\Phi_0(r,E), \tag{4}$$

where $\Phi_0(r,E)$ is the solution of Eq. (3).

A source with the spatial and energy distribution given by Eq. (4) is called a reference source [3].

The full intense of the external neutron source is defined by

$$Q_1 = \int_V \int_{E_T}^{E_f} q(r,E)dEdV,$$

and fission neutron generation intensity for sub-critical system:

$$Q_f = \int_V \int_{E_T}^{E_f} \mathbf{M}_1\Phi(r,E)dEdV.$$

Then thermal power rate of reactor (N_0) is:

$$N_0 = \frac{P_f Q_f}{\nu},$$

where P_f - energy, released per fission, ν - mean number of neutrons per fission.

The relation of Q_f and Q_1 for reference source is:

$$\left(\frac{Q_f}{Q_1}\right)_{\text{ref}} = \frac{k_{\text{eff}}}{1-k_{\text{eff}}} \tag{5}$$

In case of arbitrary spatial and energy distribution of external neutron source the efficiency of its amplification in the core can be estimated:

$$k_{\text{ampl}} = \left(\frac{Q_f}{Q_1}\right) \Big/ \left(\frac{Q_f}{Q_1}\right)_{\text{ref}}$$

and according to Eq.(5) :

$$k_{\text{ampl}} = \frac{1-k_{\text{eff}}}{k_{\text{eff}}} \cdot \left(\frac{Q_f}{Q_1} \right).$$

In order to maintain ADS power rate at a constant power level during reactor operation with decreasing k_{eff} it's necessary to increase accelerator current. Reactivity reduction as a result of nuclear fuel burning and fission products is about 8% for thermal-neutron reactor and 1-3% for fast-neutron reactor [2]. Thus, in ADS with fast-neutron reactor accelerator current variety during the operation period is significantly less than in ADS with thermal-neutron reactor. Consequently using fast core in ADS is more preferable.

NEUTRON-PRODUCING TARGET

Neutron yield from the target depends on parameters of the charged particles beam, target composition and dimensions.

Yields of neutrons (number of neurons which escape from the target) for finite targets depend on both beam characteristics and target dimensions. When dimensions are small a substantial part of secondary particles capable to induce fission with additional neutron production escapes form the target without interaction. At the same time in large target neutron radiative capture plays an important role.

For cylindrical target an optimal diameter with respect to neutron yield corresponds to few (2-3) characteristic inelastic interaction lengths λ_{in}. Due to anisotropy in particle production for inelastic proton scatterings in the lab frame (most of particles are produced in forward direction), the target length L should be somewhat larger than its radius, at the same time value of L has minor effect on the neutron multiplicity provided $L > D > \lambda_{in}$. Substantial part of neutrons escapes through the front butt-end of the target block, so neutron yield is maximal for comparatively small deepening of the beam injection point $z_0 \approx 0.3\lambda_{in}$.

In table 1 we present calculated values for optimal dimensions of cylindrical target, it also represents maximum neutron yields for them obtained via GEANT-4 [4].

Table 1: Optimized target diameter D (cm), target length L (cm), penetration beam deepening Z (cm) and relative neutron yield N (neutron/protons) in dependence of protons energy.

Ep,	Pb				Ta			
МэВ	**D**	**L**	**Z**	**N**	**D**	**L**	**Z**	**N**
200	16.1	20	8.1	2.12	5.20	7	2.0	1.60
300	17.6	22	7.8	4.98	6.80	10	2.0	3.37
400	18.5	23	7.4	8.38	8.50	13	2.1	5.63
500	18.9	24	6.6	12.4	9.40	15	2.0	8.04
600	19.4	26	6.5	17.0	10.3	17	2.0	10.2
700	19.6	27	6.2	21.2	10.6	17	1.8	12.3
800	20.0	28	6.0	26.1	11.0	18	1.7	14.2
900	20.3	28	5.8	30.7	11.3	19	1.6	16.2
	U				**Bi**			
200	16.1	20	8.1	7.34	16.1	20	8.1	2.09
300	17.6	22	7.8	15.36	17.6	22	7.8	4.67
400	18.5	23	7.4	25.05	18.5	23	7.4	8.01
500	18.9	24	6.6	35.61	18.9	24	6.6	12.2
600	19.4	26	6.5	47.08	19.4	26	6.5	16.2
700	19.6	27	6.2	58.27	19.6	27	6.2	20.3
800	20.0	28	6.0	68.72	20.0	28	6.0	24.9
900	20.3	28	5.8	78.27	20.3	28	5.8	29.3

SPATIAL LOCALISATION OF EXTERNAL NEUTRON SOURCE

Intensity of the electronuclear neutron source

$$Q_1 = \frac{I_p m_0}{e} \qquad (6)$$

with I_p standing for an average current of accelerator, m_0 – average neutron yield out of the target per single accelerated beam particle, e – charge of the accelerated particle.

It is possible to reduce intensity of external neutron source locating one in the center of the core and hence decrease neutron leakage.

Spatial localization of external neutron source (radius r_{sourse}) in the center of the core (radius R_{core}) can be characterized by source localization factor $\bar{a}$.

$$\bar{a} = \frac{r_{source}}{R_{core}}$$

Analytical dependence of $k_{ampl}(\bar{a})$ [5] for fast-neutron core is shown in fig.1.

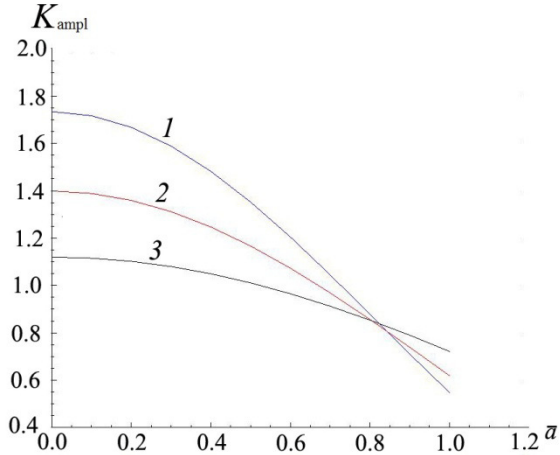


Figure 1: k_{ampl} depend on source localisation factor, 1 - spherical geometry, 2 - cylindrical, 3 - slab.

CASCADE REACTOR CORES

In works [3], [6] it's proposed to amplify external neutron source using coupling fast-thermal core zones (cascade core). Fast-thermal cascade core consists of inner subcritical section in fast neutrons with additional

neutron source and outer subcritical section in fast or thermal neutrons. The neutron feedback of inner and outer sections could be broken using neutron gate (thermal neutron absorber).

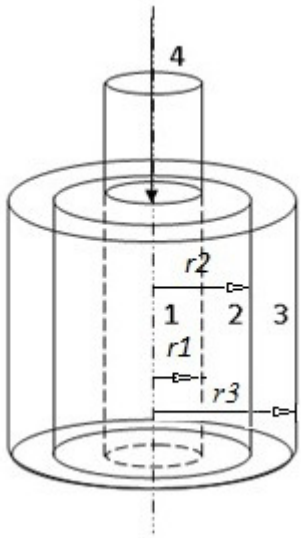


Figure 2: Cascade reactor core scheme. 1- inner section, 2- neutron gate, 3- outer section, 4 – charged particles beam.

In real cascade core it's impossible to break neutron feedback completely, because there is a lot of high energy neutrons in the outer zone, which couldn't be absorbed by the "neutron" gate and come into the inner section [2]. It can be avoided by using the second gate type: "geometrical" gate in cylindrical core (void gap $\Delta r = r_2 - r_1$). In this case both coupling zones could be fast.

In fig. 3 it's shown dependence $k_{ampl}(\Delta r)$ for fast cascade core with r_3=*150 cm.*

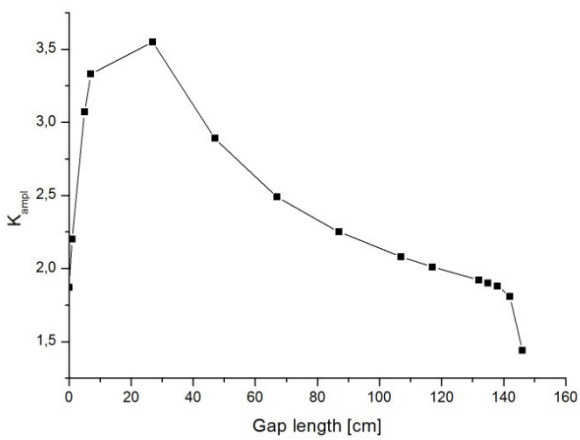


Figure 3: Dependence of k_{ampl} on gap length for cylindrical geometry.

RESULTS

Thus, ADS power rate could be increase by several ways:

- localization of spatial distribution of external neutron source in the reactor core;
- optimization of spallation target size;
- using of coupling fast-neutron reactors (cascade core with "geometrical" gate).

REFERENCES

[1] C. Rubbia "Conceptual design of a fast neutron operated high power energy amplifier". European organization for nuclear research. Geneva, 29th September,1995.

[2] L.N. Gerasimov, I.V. Kudinovich, V.P. Struev, Yu.A. Svistunov "Low-energy electronuclear power plant: possible technical decisions" // RAS. Energetics. 2005. №2 P. 3–16.

[3] V.V. Seliverstov "External neutron source multiplication in cascade subcritical reactor cores with one way neutronic feedback"// Atomic power. 1996. V. 81, Issue. 5. P. 378–390.

[4] Yu. A. Svistunov, M.F. Vorogushin, I.V. Kudinovich "The feasibility of low-energy electronuclear power plant" LINAC'08, Victoria, BC, Canada, 2008 MOP061, P. 217 http://www.JACoW.org

[5] A.G. Golovkina, I.V. Kudinovich, D.A. Ovsyannikov "Subcritical homogeneous reactor power rate subject to space distribution and energy of the external neutron source" Vestnik St. Petersburg University. Ser. 10. 2012. Issue 2. P. 13–24.

[6] G.V. Kiselev "The possibility of ADS characteristic improvement". Atomic energy, 2001, V. 91., Issue. 1. P.54-63.

SIMULATION OF MULTIBUNCH INSTABILITIES WITH THE HEADTAIL CODE

N. Mounet, E. Métral and G. Rumolo (CERN, Geneva, Switzerland)

Abstract

Multibunch instabilities due to beam-coupling impedance can be a critical limitation for synchrotrons operating with many bunches. To study these instabilities, the HEADTAIL code has been extended to simulate the motion of many bunches under the action of wake fields. All the features already present in the single-bunch version of the code have remained available, in particular synchrotron motion, chromaticity, amplitude detuning due to octupoles and the ability to load any kind of wake fields through tables. The code has been then parallelized in order to track thousands of bunches in a reasonable amount of time, showing a linear scaling with the number of processors used. We show benchmarks against Laclare's theory in simple cases, obtaining a good agreement. Results for bunch trains in the LHC and comparison with beam-based measurements are also exhibited.

INTRODUCTION

Transverse coupled-bunch instabilities occur in general when several bunches interact with their surroundings, creating wake fields that act back on the bunch train in such a way as to give rise to an exponentially growing oscillation. To evaluate the rise times of such instabilities, several theories exist, such as Sacherer's [1], Laclare's [2] and Scott Berg's [3], as well as a macroparticle simulation code, MTRISM [4]. All have some limitations: MTRISM and Scott Berg's theory do not take into account quadrupolar impedance which is quite significant in e.g. the LHC due the high-impedance flat collimators [5]; Sacherer's and Laclare's formalisms assume a machine entirely filled with equidistant bunches, which is not the case in e.g. the LHC and even less in the SPS where only about 30% of the machine is filled to produce the nominal LHC beam. Therefore, to be able to study multibunch instabilities in the LHC and the SPS, we chose to extend the single-bunch macroparticle code HEADTAIL [6]. A first simplified version, accounting only for rigid bunch oscillations with no longitudinal motion, was already developed in Ref. [7]. We present here a new extension that can handle both multibunch and intrabunch motion, benchmarking it with respect to theory, and showing various results in the case of the LHC. Finally, we compare HEADTAIL simulations using the LHC impedance model with beam-based experiments.

DESCRIPTION OF HEADTAIL MULTIBUNCH

HEADTAIL is a macroparticle simulation code where each individual macroparticle i is tracked through a ring subdivided into several kick sections. After initialization with Gaussian (or uniform) distributions, with the possibility to enforce longitudinal matching, macroparticles are tracked in mainly three steps per kick section: 1) the bunches are sliced longitudinally, 2) wake fields kicks are applied to each macroparticles, and 3) their transverse phase space coordinates are linearly transported to the next kick section. Once per turn, the synchrotron motion update is applied, separately for each bunch. For the second step, the kicks $\Delta x'_i$, $\Delta y'_i$ and $\Delta\delta_i$ are computed as

$$\Delta x'_i = \mathcal{C} \sum_{z_S > z_{S_i}} n_S W_x\left(z_{S_i} - z_S, x_S, y_S, x_{S_i}, y_{S_i}\right),$$

$$\Delta y'_i = \mathcal{C} \sum_{z_S > z_{S_i}} n_S W_y\left(z_{S_i} - z_S, x_S, y_S, x_{S_i}, y_{S_i}\right),$$

$$\Delta \delta_i = \mathcal{C} \sum_{z_S \geq z_{S_i}} n_S W_{||}\left(z_{S_i} - z_S\right),$$

where $\mathcal{C} = -\frac{e^2}{E_0 \beta^2 \gamma}$, γ being the Lorentz factor, $\beta = \sqrt{1-\gamma^{-2}}$, E_0 the rest mass of the elementary particles (protons or electrons) and e the elementary charge. S_i is the slice containing the macroparticle i, and n_S, x_S, y_S, z_S are the number of particles, and the transverse and longitudinal positions of each slice S (z decreases when going toward the tail of the bunches). In the above expressions the sums run over all slices and bunches before the slice of the macroparticle considered, neglecting thus any wake emitted in the forward direction. The sums continue up to a certain number of turns, i.e. the wakes of preceding turns are taken into account. $W_{||}(z)$ is the longitudinal wake function, while $W_x(z)$ and $W_y(z)$ are given by

$$W_x(z, x_S, y_S, x_{S_i}, y_{S_i}) = W_x^{dip}(z) x_S + W_{xy}^{dip}(z) y_S + W_x^{quad}(z) x_{S_i} + W_{xy}^{quad}(z) y_{S_i}, \quad (1)$$

$$W_y(z, x_S, y_S, x_{S_i}, y_{S_i}) = W_y^{dip}(z) y_S + W_{xy}^{dip}(z) x_S + W_y^{quad}(z) y_{S_i} + W_{xy}^{quad}(z) x_{S_i}, \quad (2)$$

where *dip* stands for "dipolar" and *quad* for "quadrupolar". Note that coupled terms – i.e. linear wakes in the x direction but proportional to the y position and vice versa, are taken into account. The wake functions above ($W_x^{dip}(z)$, $W_x^{quad}(z)$, etc.) are provided in a table given in input.

The code has been parallelized over the bunches, which is quite efficient since all bunches can be treated almost independently, the only requirement being that after each slicing the processors exchange for all the bunches the positions and number of particles of each slice, such that the wakes can be computed in all bunches. This represents a limited amount of data since the number of slices usually does not exceed a few hundreds. Indeed, Fig. 1 shows

that the computational time is inversely proportional to the number of processors used, i.e. the parallelization is linear with the number of processors.

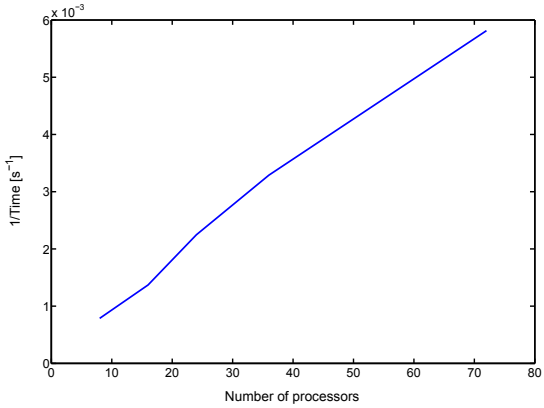


Figure 1: Inverse of the computational time vs. number of processors used, for 72 bunches with 10 slices per bunch.

COMPARISON WITH THEORY

The code has been compared to Laclare's formalism [2], in which the complex angular frequency shifts $\delta\omega$ of all possible modes are found as the eigenvalues of an infinite matrix. The most critical instability is then the one whose $\delta\omega$ has the lowest (negative) imaginary part. The formalism has been implemented in a code that automatically checks that the necessary matrix truncation still gives accurate eigenvalues (within 0.1%) by testing convergence with respect to the matrix size. The formalism assumes no linear coupling, a completely filled machine with equidistant bunches, as well as dipolar transverse impedances and a linear longitudinal bucket without distortion. We used therefore the same conditions in HEADTAIL. Also, longitudinal Gaussian distributions cut at $2\sigma_z^{rms}$ were used in both the theory and HEADTAIL. Note that in the simulations the longitudinal parameters were initially matched.

To obtain the rise times from the simulations, we compute thanks to SUSSIX [8] the highest spectral lines of the beam average transverse position and momenta, on a sliding window along the simulation, and fit the amplitude of the highest spectrum line as a function of time by an exponential.

We study here the case of the CERN SPS filled with a 25 ns beam. The impedance of its vacuum pipe is computed from Zotter's theory for an infinitely long axisymmetric cylindrical structure [9], assuming the pipe has 2 cm radius, 2 mm thickness and is made of stainless steel surrounded by vacuum [10]. The wake function is then obtained thanks to a Fourier transform with an uneven sampling [11]. Since the beam pipe is actually of elliptical cross section with the horizontal semi-axis significantly larger than the vertical one, to obtain the final dipolar impedances and wake functions we multiply by the Yokoya factors [12] for a flat chamber, i.e. $\frac{\pi^2}{24}$ in x and $\frac{\pi^2}{12}$ in y.

Table 1 shows the beam parameters used in both the simulation and the theory, and in Fig. 2 we can see the excellent agreement on the rise times, with a slight discrepancy only for the highest positive chromaticity considered.

Table 1: Beam Parameters for the SPS

Nb. of bunches	N_b	924
Bunch population	N	$5 \cdot 10^{10}$
RMS bunch length	σ_z^{rms}	0.19 m
Momentum spread	σ_δ	0.002
RF voltage	V_{rf}	3 MV
Harmonic number	h	4620
Bunch spacing	ΔT_b	25 ns
Circumference	C	6911 m
Tunes	$Q_{x,y,s}$	26.13, 26.16, 0.0073
Beta functions	$\beta_{x,y}$	42, 42 m
Norm. emittances	$\varepsilon_{x,y}$	4, 4 μm.rad
Mom. compaction	α_p	$1.92 \cdot 10^{-3}$
Lorentz factor	γ	27.7

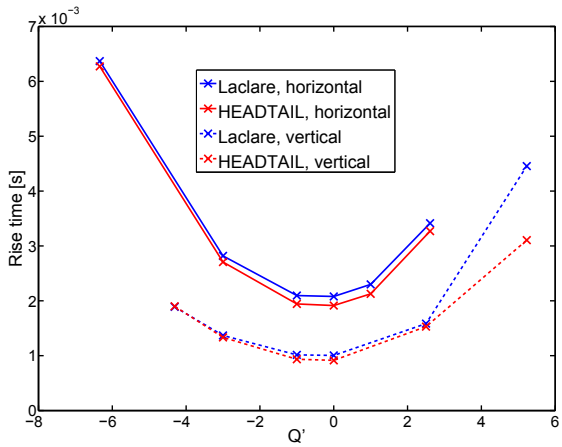


Figure 2: Rise times vs. Q' for 924 bunches in the SPS: Laclare's theory and HEADTAIL multibunch.

SIMULATIONS RESULTS FOR THE LHC

For these simulations, we use HEADTAIL with a non linear longitudinal bucket and the impedance model presented in Ref. [11], which presently includes the resistive-wall impedance of the 44 collimators (some being in graphite), of the copper-coated beam screens covering 86% of the ring, and of the copper vacuum pipe for the remaining 14%, together with a broad band impedance model to account for most of the smooth transitions around the ring [13]. Wake functions are obtained thanks to a Fourier transform on a uneven sampling [11]. Dipolar, quadrupolar and coupled terms are all taken into account, and weighted by the corresponding beta functions in order to be lumped into a single kick.

Comparison Between Short and Long Trains

During current physics operation in the LHC, the bunches are not all equidistant. Indeed, due to the limited capacity of the injectors, the rise time of the injection kickers and that of the LHC dump system kicker, the bunches are grouped in batches of 36 bunches all separated by 50 ns, these batches being separated between them by empty gaps of various sizes (from 225 ns to 3 μs) [14]. Note that due to an additional constraint on the injection system, a low intensity pilot bunch plus a batch of 12 bunches are currently always injected before the batches of 36 bunches.

The multibunch HEADTAIL code presented above now allows to study any filling pattern. This is particularly interesting for small trains which are in principle far from the situation of a completely filled machine accessible to theories such as the ones from Sacherer [1] or Laclare [2]. We study here the case of a train of 36 bunches spaced between them by 50 ns and compare it to the case of 1782 equidistant bunches with the same 50 ns spacing. We show in Table 2 the parameters used in the simulations, which are close to those of normal operation at 3.5 TeV/c in 2011. In Fig. 3 the average x positions of the full beam is shown. The beam with 1782 bunches gets unstable more quickly than the other one, but the rise time is only about a factor 2.5 higher (in the vertical plane - not shown here, the factor is 4). To analyse the coupled-bunch nature of the instabilities, a singular value decomposition (SVD) [15, 16] was performed on the bunch centroid data (vs. bunch number and turn). In Fig. 4 it appears that the spatial pattern of the most critical "mode" from the SVD has a larger wavelength with 1782 bunches than with 36 bunches.

Table 2: LHC Parameters for the Comparison Between 36 and 1782 Bunches

N	$1.2 \cdot 10^{11}$	Q_x	64.28	Q_y	59.31
σ_δ	$1.6 \cdot 10^{-4}$	σ_z^{rms}	0.09 m	V_{rf}	12 MV
α_p	$3.2 \cdot 10^{-4}$	h	35640	ΔT_b	50 ns
C	26659 m	γ	3730.3	Q_s	0.0025
ε_x	3.75 μm.rad	Q'_x	0	β_x	66 m
ε_y	3.75 μm.rad	Q'_y	0	β_y	71.5 m

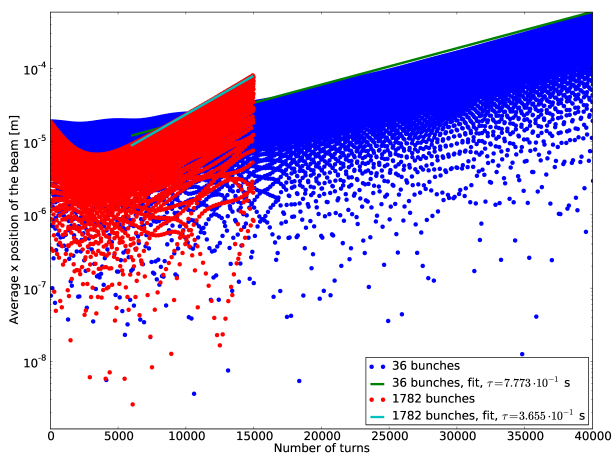
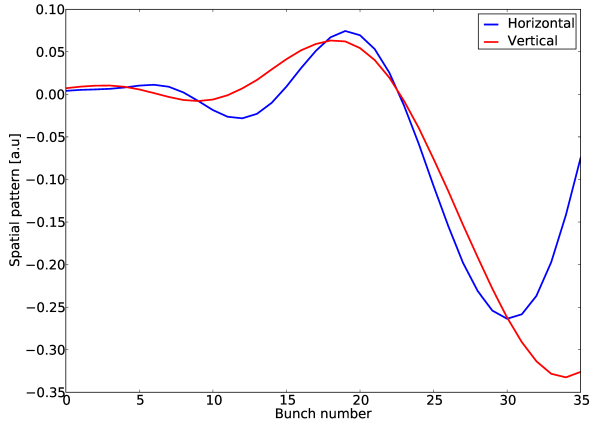
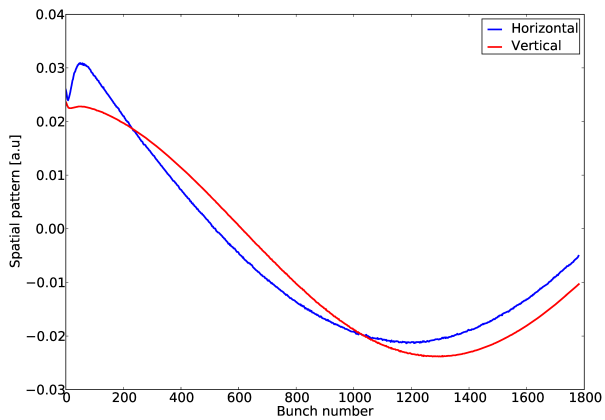


Figure 3: Average horizontal beam position vs. turns for 36 and 1782 bunches.



(a) Case with 36 bunches.



(b) Case with 1782 bunches.

Figure 4: SVD spatial pattern along the bunches for short and long bunch trains.

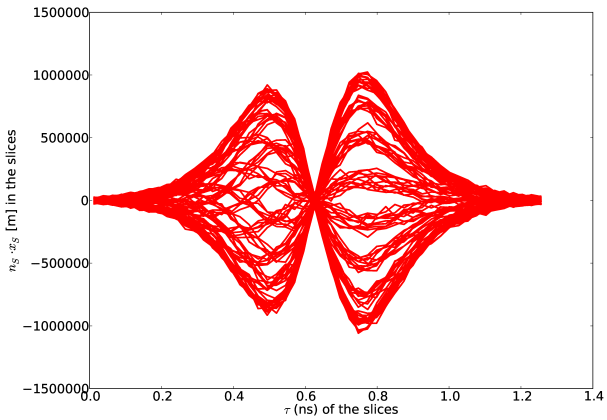
Coupled-Bunch Headtail Instabilities

Single-bunch headtail modes are well known instabilities occurring in synchrotrons, and they are currently damped in the LHC thanks to Landau damping provided by octupoles [17]. When many bunches are in the machine, coupled-bunch headtail modes (i.e. coupled-bunch modes with intrabunch motion) could be stronger than the single-bunch headtail modes. Using the parameters of Table 2 except for $Q'_x = Q'_y = 6$ and $N = 3 \cdot 10^{11}$ protons per bunch, we compare in Table 3 the tune shifts (with respect to the synchrotron sideband) and the rise times of the $m = -1$ headtail mode, for the single-bunch and 36-bunches cases. The values were obtained thanks to a spectral analysis on complex frequencies described in Ref. [11], performed on the time pattern from the SVD decomposition (see above) for the 36-bunches case. Clearly, the rise times are lower and the tune shifts significantly higher in the 36-bunches case with respect to the single-bunch ones. For the 36 bunches case, an headtail mode with one node is

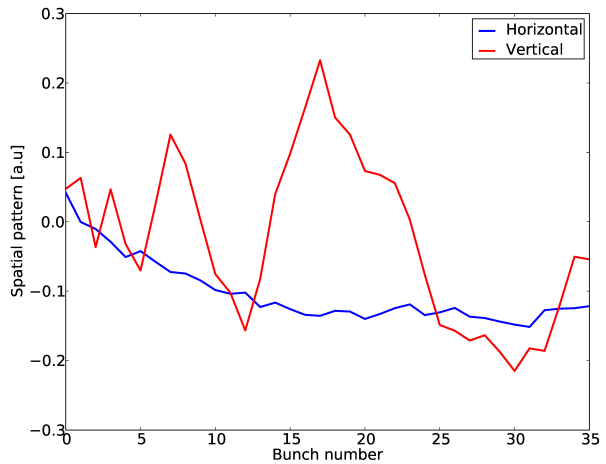
clearly visible in Fig. 5a, and the coupled-bunch nature of the spatial pattern from the SVD in Fig. 5b.

Table 3: Rise Times and Tune Shifts of the $m = -1$ Mode

	τ_x [s]	τ_y [s]	ΔQ_x	ΔQ_y
1 bunch	2.25	2.45	$4.1 \cdot 10^{-6}$	$4.1 \cdot 10^{-5}$
36 bunches	1.74	1.95	$2.9 \cdot 10^{-5}$	$5.7 \cdot 10^{-5}$



(a) Horizontal bunch profile of the last bunch.



(b) SVD spatial pattern along the bunches.

Figure 5: Intrabunch and coupled-bunch motion for 36 bunches.

COMPARISON WITH LHC BEAM-BASED MEASUREMENTS

To check the accuracy of both the LHC impedance model presented above and the HEADTAIL code, a dedicated experiment [18] was carried on to measure transverse instability rise times of the rigid-bunch modes, as well as the loss of Landau damping threshold in terms of octupole current at top energy. This was done on a beam of 48 nominal bunches (a batch of 36 50 ns spaced bunches preceded by 12 bunches also 50 ns spaced) and a low intensity pilot bunch.

The idea of the experiment was to switch the feedback off during a long enough time window, in order to allow transverse coupled-bunch instabilities to develop. At top energy, it was also necessary to reduce Landau damping by decreasing the current in the octupoles, which was done in steps. Note that the defocusing octupoles were set to a positive current and the focusing ones to its opposite. Bunch centroid data were acquired during the time window when the feedback was off, and all beam and machine parameters (intensity, bunch length, collimator half gaps and emittances) continuously monitored and used in the simulations.

Results at Injection

In Fig. 6 the average instability rise times (obtained with two different methods) of the 8 last bunches of beam 2, are compared to HEADTAIL simulations for various chromaticities. The agreement between the model and the measurements is remarkable for Q' close to zero or negative. The only significant discrepancy appears in vertical when $Q'_y = 2$. Note that in this latter case only one set of data and one fitting method could be used, hence the absence of error bar. In the same figure single-bunch simulations results are also shown and seem to rule out the possibility that the instabilities observed were actually single-bunch. This is confirmed by the spatial pattern from the SVD (see above) along the batch of 36 bunches in Fig. 7.

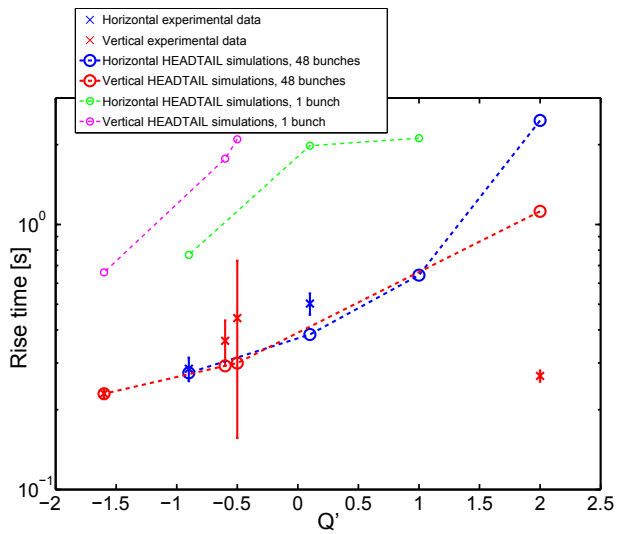


Figure 6: Beam 2 rise times vs. Q' at injection, from measurements and HEADTAIL (single-bunch and multibunch).

Results at Flat Top

At 3.5 TeV/c some instabilities were observed in the vertical plane only. On the contrary, HEADTAIL simulations exhibit instabilities only in the horizontal plane for non zero octupole currents, which can be explained by the fact that $Q'_x = 0$ while $Q'_y \geq 1$ (the impedances in both planes being quite similar [11]). Also, a discrepancy of factor $2-3$ between simulations and measurements is visible on the rise times of the 8 last bunches of the train, as shown in Fig. 8. Note the quite large error bars on the simulation data, due to the different fitting methods used, which probably indicates that a higher number of simulated turns would be better for a higher accuracy of the fit.

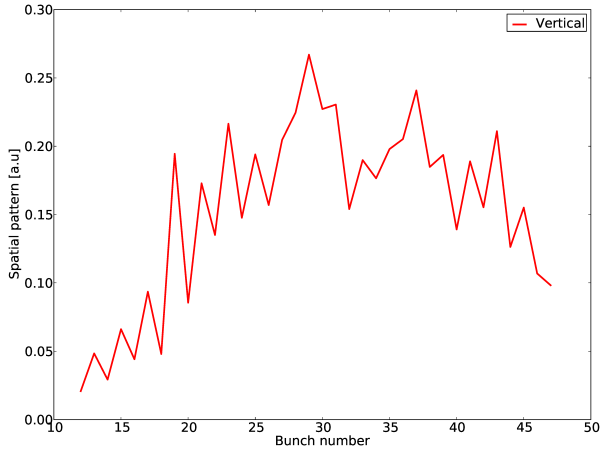


Figure 7: Vertical spatial pattern of the highest "mode" from the SVD of 36-bunches batch, for beam 1 with $Q'_y = 0.3$, at injection.

In Fig. 9 we compare beam 1 vertical rise times from the measurements with HEADTAIL simulations results when the octupoles are switched off. The discrepancy between HEADTAIL and the measurements already mentioned appears clearly. On the other hand, the accuracy of the measurement of Q' at 3.5 TeV/c is rather large (of the order of one unit), which could explain the discrepancies observed: if $Q'_y = 0$, these are much reduced.

Finally, measurements show that when the damper is off 60 A (resp. 70 A) in the octupoles are enough to stabilize beam 1 (resp. beam 2), which is less than foreseen in the model (resp. 120 A and 100 A). A possible explanation of the discrepancy is that some sources of non-linearity have been neglected in the simulations, in particular Q'' (second derivative of the tune with respect to the momentum deviation), which is quite high when octupoles are on. The effect of Q'' on beam stability is currently under study.

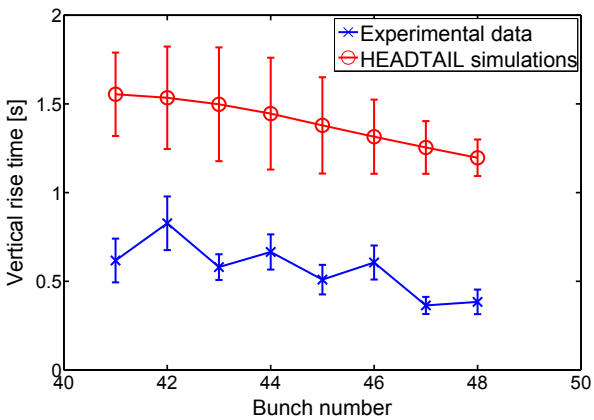


Figure 8: Vertical rise times of the last 8 bunches of beam 2 when octupoles are off at 3.5 TeV/c, compared to HEADTAIL simulations.

CONCLUSION

The wake fields code HEADTAIL is now able to simulate multibunch trains, and was successfully benchmarked with theory in simple cases, for instability rise times. Using the LHC impedance model, the code was also compared to an LHC measurement, giving reasonable agreement.

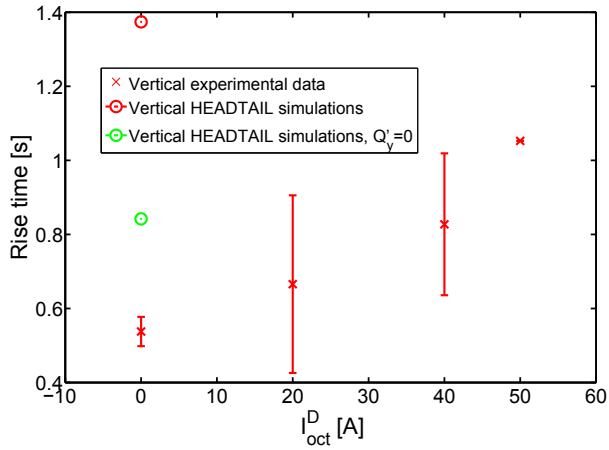


Figure 9: Measured vertical rise times vs. octupole current for beam 1 at 3.5 TeV/c, compared to HEADTAIL simulations, both with the measured Q'_y (≈ 2) and with $Q'_y = 0$.

ACKNOWLEDGMENTS

The authors warmly thank V. Kain, D. Valuch, L. Ponce and W. Höfle for helping with the LHC measurements, R. Assmann, A. Rossi, D. Wollmann and R. Bruce for providing collimator information, and B. Salvant for helpful discussions. This work is part of the EPFL PhD thesis 5305, supervised by L. Rivkin (EPFL-LPAP).

REFERENCES

[1] F.J. Sacherer, 9th Int. Conf. on High-energy accelerators, Stanford, USA, p. 347 (1974).

[2] J.L. Laclare, CERN-87-03-V-1, p. 264.

[3] J. Scott Berg, Stanford PhD Thesis, SLAC-R-478 (1996).

[4] A. Koschik, Tech. Univ. Graz PhD Thesis (2004).

[5] F. Roncarolo et al, Phys. Rev. ST AB 12 (2009) 084401.

[6] G. Rumolo and F. Zimmermann, Phys. Rev. ST AB, 5 (2002), 121002.

[7] G. Rumolo and E. Métral, PAC'09, Vancouver, Canada.

[8] R. Bartolini and F. Schmidt, CERN-SL-Note-98-017.

[9] B. Zotter, CERN-AB-2005-043.

[10] B. Salvant, EPFL PhD Thesis 4585 (2010), p. 85.

[11] N. Mounet, EPFL PhD Thesis 5305 (2012).

[12] K. Yokoya, Part. Acc., 41 (1993), p. 221.

[13] O. S. Brüning et al, CERN-2004-003-V-1, Ch. 5, p. 101.

[14] G. Arduini et al, CERN LHC Project Note 401 (2007).

[15] G. Arduini et al, ECLOUD'04, CERN-2005-001.

[16] Y. Ohnishi et al, EPAC'00, Vienna, Austria.

[17] E. Métral et al, IPAC'11, San Sebastian, Spain.

[18] N. Mounet et al, CERN-ATS-Note-2011-035 MD.

[19] G. Rumolo et al, Chamonix 2012 Workshop, France.

CALCULATION OF LONGITUDINAL INSTABILITY THRESHOLD CURRENTS FOR SINGLE BUNCHES *

P. Kuske, Helmholtz-Zentrum Berlin für Materialien und Energie, Germany

Abstract

Based on the publication by M. Venturini, et al. [1] a computer program has been developed that solves the Vlasov-Fokker-Planck equation numerically on a two dimensional grid. In this code different types of longitudinal interactions and their combinations are implemented. Calculations have been performed for the 1.7 GeV storage ring BESSY II and the 600 MeV ring MLS using the wake created by the shielded coherent synchrotron radiation (CSR). The results are compared with measurements on both rings which were based on the observation of the onset of time dependent fluctuations ("bursts") of CSR [2, 3]. Fair agreement is found between theory and experiment. The theoretical results complement calculations performed by Bane, et al. [4] for this type of interaction. The new results emphasize the resistive nature of the CSR-interaction, especially in regions where shielding effects are small or the resonance-like features of this wake are important. It is found that in these regions the instability is weak and thresholds depend on the damping time.

INTRODUCTION

The longitudinal stability of a single bunch of charged particles circulating in a storage ring is still a domain of active experimental and theoretical research. On the experimental side the observation of CSR has supplemented the conventional RF-techniques and extended the region of observation frequencies into the THz part of the bunch spectrum. Thus the time resolution reaches the fs-region. Already the first observations of the time dependent CSR at BESSY II created the desire to compare theoretical and experimental results [2]. On the theoretical side the most advanced semi-analytical approach has been introduced by Oide and Yokoya [5], however, predictions based on their approach are ambiguous. Multi particle tracking codes suffer from the rather high shot noise and therefore a numerical solution of the Vlasov-Fokker-Planck (VFP) equation was chosen for the theoretical modelling [2,6].

THEORETICAL MODEL

The ensemble of electrons is described by a distribution function, $\psi(q,p,\tau)$, which can be written in terms of the normalized phase-space coordinates: $q=z/\sigma_z$, $p=(E_0-E)/\sigma_E$, and τ, in units of the synchrotron period, $1/\omega_s$. The low current rms bunch length, σ_z, and the natural energy spread, σ_E, are related by: $\omega_s\sigma_z/c=\alpha\sigma_E/E_0$ with c, the speed of light, and α, the momentum compaction factor. The VFP equation describes the evolution of the distribution which depends on the radiation excitation and damping and the interaction of the particles [1]:

$$\frac{\partial\psi}{\partial\tau}+p\frac{\partial\psi}{\partial q}-[q+F_c(q,\tau,\psi)]\frac{\partial\psi}{\partial p}=\frac{2}{\omega_s t_d}\frac{\partial}{\partial p}\left(p\psi+\frac{\partial\psi}{\partial p}\right)$$

The r.h.s. is the Fokker-Planck-term which introduces damping and diffusion and t_d is the longitudinal damping time. The coefficient in the square brackets is the approximated linear restoring force of the RF-potential and F_c is the collective force from the interaction of the charge distribution with itself and its metallic surrounding. This force or the induced voltage is obtained by the convolution of the longitudinal charge density with the wake function. The charge density, $\lambda(q,\tau)$, is given by $\lambda(q,\tau)=\int dp\,\psi(q,p,\tau)$. The wake function depends on the types of interactions studied.

Numerical Solution of the VFP-Equation

Warnock, et al [7] and Venturini, et al [1] proposed two different numerical algorithms to find the distribution function on a two dimensional grid as a function of time. In both cases the required extrapolation of the distribution function to off-grid points can lead to unrealistic negative values for $\psi(q,p,\tau)$. Therefore, a kind of wave function, $g(q,p,\tau)$, is introduced so that the probability to find electrons is given by $\psi(q,p,\tau)=g^2(q,p,\tau)$ and always larger than zero. As a result only the Fokker-Planck-term needs to be modified in the VFP equation for $g(q,p,\tau)$:

$$\frac{\partial g}{\partial\tau}+p\frac{\partial g}{\partial q}-[q+F_c]\frac{\partial g}{\partial p}=\frac{2}{\omega_s\tau_l}\left(\frac{g}{2}+p\frac{\partial g}{\partial p}+\frac{1}{g}\left(\frac{\partial g}{\partial p}\right)^2+\frac{\partial^2 g}{\partial p^2}\right)$$

This equation is solved by splitting the time step, $\Delta\tau$, of the temporal evolution of the wave function, $g(q,p,\tau)$, into smaller steps [1]:

$$g_1(q,p,\tau)=g_0(q-p\cdot\Delta\tau/2,p,\tau)$$
$$g_2(q,p,\tau)=g_1(q,p+(q+F_c(q,g_0))\cdot\Delta\tau,\tau)$$
$$g_3(q,p,\tau)=g_2(q-p\cdot\Delta\tau/2,p,\tau)$$
$$g_4(q,p,\tau+\Delta\tau)=g_3(q,p,\tau)+diffusion+damping$$

The code uses a fourth order polynomial in order to interpolate to off-grid points. The interpolation is performed between the grid points $g(x-2\Delta)$, $g(x-\Delta)$, $g(x)$, $g(x+\Delta$ and $g(x+2\Delta)$ and thus is more symmetric relative to the central point compared to a cubic Hermite interpolation [1]. x stands for the p or q coordinate and Δ is the distance between grid points. The first step coded in BASIC looks like this:

```
For iq = -iqmax To iqmax: For ip = -ipmax To ipmax
  g0 = gold(iq, ip)
  If Abs(g0) > .000001 Then
    dQ = - ip *deltaP/deltaQ* dtau / 2
    gmm = gold(iq - 2, ip): gm = gold(iq - 1, ip)
```

*Work supported by the BMBF and the Land Berlin

```
    gp = gold(iq + 1, ip): gpp = gold(iq + 2, ip)
    a1=(gmm-8*gm+8*gp-gpp) / 12*dQ
    a2=(-gmm+16*gm-30*g0+16*gp-gpp)/24*dQ^2
    a3=(-gmm+2*gm-2*gp+gpp)/12*dQ^3
    a4=(gmm-4*gm+6*g0-4*gp+gpp)/24*dQ^4
    gnew(iq, ip) = (g0 + a1 + a2 + a3 + a4)
  End If
Next ip: Next iq
```

The second step looks very similar. In all steps terms are only included for sufficiently large values of g, i.e. $|g|>10^{-6}$. The final diffusion and damping step is coded in the following way:

```
For iq = -iqmax To iqmax: For ip = -ipmax To ipmax
  g0 = gold(iq, ip)
  If Abs(g0) > .000001 Then
    gp = gold(iq, ip + 1): gm = gold(iq, ip - 1)
    g1= (4*gp-6*g0+4*gm-2*gp*gm/g0)/deltaP ^ 2
    g1=g1+ ip*(gp-gm) + (g0+gp)/2
    gnew(iq, ip)= g0+g1*beta*dtau
  End If
Next ip: Next iq
```

Just a few lines of code are required in order to follow the temporal evolution of the distribution function. The calculation could be speeded up by storing and updating the two partial derivatives, $\partial g/\partial q$ and $\partial g/\partial p$, of the amplitude function. This is not done since the mesh size can become quite large [6] for certain calculations. For the stability studies presented here usually a 256 by 256 mesh (ipmax=qmax=128) with a spacing of σ/10 is used. Stability of the algorithm requires a certain number of steps per synchrotron period in proportion to the damping and excitation given by $\omega_s\tau_l=1/\text{beta}$. For the cases studied here 512 to 1024 steps with $\Delta\tau=2\pi/512$ or $\Delta\tau=2\pi/1024$ are sufficient. The temporal evolution of the particle distribution is followed over as many synchrotron periods as required to cover at least 6 and up to 10 damping times. This can be thousands of periods. If the resolution is increased also the number of steps must go up. Doubling the resolution leads to an 8-fold increase in the computing time. Instability thresholds and features of the instability do not change significantly if the resolution is increased.

RESULTS

In Table 1 the relevant parameters for the longitudinal single bunch dynamics of the two storage rings are collected. The simulations begin with the potential well distorted distribution function which implies that the beam is longitudinally stable and has a Gaussian energy distribution with a normalised rms width, σ_E, equal to 1. In the threshold calculations the current is increased or decreased by small amounts and simulations continue with the previously found distribution. If approached from below the current threshold is reached as soon as $\sigma_E>1$. The code was compared successfully to the weak instability created by a resistive δ-function wake, $w(q)=R\cdot\delta(q)$ [8]. The results of this simulation showed more strongly increasing threshold currents as $\omega_s\tau_l/2\pi$ approaches 1. On the other hand, simulations with an inductive wake, $w(q)=L\cdot T_0/2\pi\,\partial\delta(q)/\partial q$, as expected showed no instability at all. In addition, results using the wake of a broad band resonator (BBR) have been compared to predictions of other calculations [4]. The present detailed simulation reproduces these results and certain jump-like changes in parameters like the frequency of the first unstable mode similar to what can be observed in experiments [9] and where the semi-analytical approach [5] fails.

Table 1: Parameters Used in the Simulations

Parameter	BESSY II	MLS
Energy, E_0/MeV	1700	629
Bending radius, ρ/m	4.35	1.528
Momentum compaction, α	$7.3\ 10^{-4}$	$1.3\ 10^{-4}$
Cavity voltage, V_{rf}/kV	1400	330
Accelerating frequency, ω_{rf}/MHz	2π 500	2π 500
Revolution time, T_0/ns	800	160
Natural energy spread, σ_E	$7.0\ 10^{-4}$	$4.36\ 10^{-4}$
Zero current bunch length, σ_0/ps	10.53	1.549
Longitudinal damping time, τ_l/ms	8.0	11.1
Synchrotron frequency, ω_s/kHz	2π 7.7	2π 5.82
Height of the dipole chamber, *2h/cm*	3.5	4.2

The impact of the CSR-wake with parallel plate shielding [6] has been simulated for both rings. The threshold currents are displayed in Figure 1 together with some experimental results and using a scaling introduced by Ries, et al. [3]. At the MLS and for bunches longer than 1 mm the findings in this publication agree quite well with the theoretical predictions. For bunches shorter than 2 mm the simulated thresholds depend strongly on the synchrotron tune in relation to the damping time. This characterises the weak instability mechanism. The simulation for V_{rf}=500 kV and for 1 - 2 mm long bunches (not shown for clarity) would overlap the curve for BESSY II. At higher cavity voltages and with shorter bunches the predictions are approaching the results of

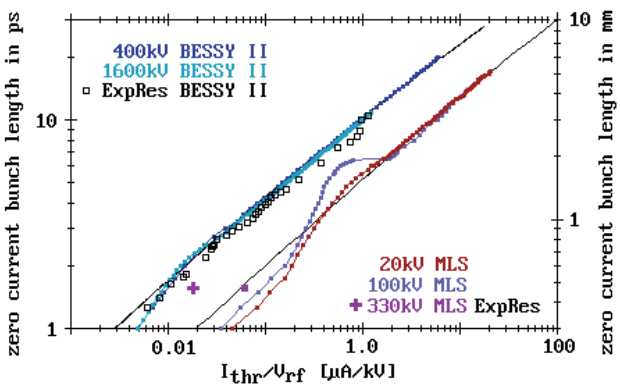


Figure 1: Comparison of theoretical (colour lines) and experimental results for the CSR-instability threshold currents. The black solid line is the theoretical curve from Bane, et al. [4]. They already predicted a deviation from that line around a certain bunch length. Clearly visible in the MLS results and much less pronounced for BESSY II.

Bane, et al. [4]. In Figure 1 only one of the experimental results of the MLS is included: the case with a cavity voltage, V_{rf}, of 330 kV and a momentum compaction factor of α=1.3·10^{-4}. In general, for the shorter bunches the observations at the MLS and simulations disagree.

Similar measurements performed at BESSY II do agree much better with the general trend of the theory [4]. A typical result of a CSR measurement at BESSY II for a short bunch can be seen in Figure 2. The instability sets in at around 11.5 μA with a fast growing line between the 3rd and 4th harmonic of the synchrotron frequency. For these short bunches, where the impact of the shielding is small, the CSR distorted bunch shape is very similar to the distortion created by a resistive wake. With such a wake and with $1/\omega_s\tau_l$ approaching 1 the threshold currents are shifted to higher values in agreement with expectations [8] for a weak instability. This happens in both rings, however, could not yet be verified experimentally. At the MLS the transition to a strong instability occurs at higher cavity voltages (i.e. 330 kV).

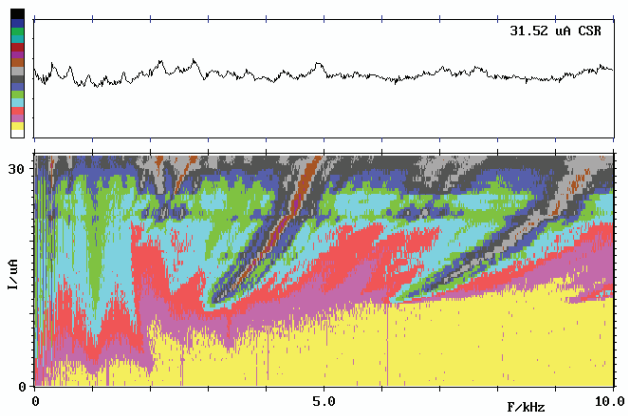


Figure 2: FFT spectra of the temporal THz-signals recorded as a function of increasing current showing remarkable features of the synchrotron sidebands. The zero current synchrotron frequency is 1 kHz and the rms bunch length is ~1.4 ps. The logarithmic plot of the spectrum at 31.5 μA is shown in the top and the colours span 6 orders of magnitude.

Figure 3 shows a comparison of the frequency of the first unstable mode at BESSY II. The slope of the experimental [9] and the theoretical data agree quite well. The slope depends on the bunch length and the peak frequency of the resonator-like CSR-impedance, given by $F_{res}/c=(\pi\rho/24h^3)^{1/2}$ [10]. Significant deviations are found for the shortest bunches. There the shielding is less important and one would expect that the m=2, quadrupole mode, has the lowest threshold [8], it is rather the m=3...4 mode which becomes unstable first, see Figure 2.

SUMMARY

In conclusion, calculations with a shielded CSR-wake are generally in surprisingly good agreement with the measurements. This wake alone does not reproduce some of the observed features of the longitudinal single bunch dynamics like the strong "inductive" lengthening or the "resistive" shift and shape distortion observed with the streak camera [11]. However, the observations on two different storage rings in accord with the simulations demonstrate the importance of the shielded CSR-wake and thus the dipole chamber geometry: the slope of the first unstable mode frequency observed at BESSY II in agreement with a broad band resonator with F_{res}~100GHz and the pronounced threshold behaviour at the MLS with a bunch length around 2 mm with F_{res}~44GHz. Simulations with a more realistic model of the vacuum chamber and a more elaborate CSR-wake [12] should improve the agreement with the observations. The final goal still is the theoretical modelling of measured spectra similar to Figure 2, and which are usually much more complex at higher intensities.

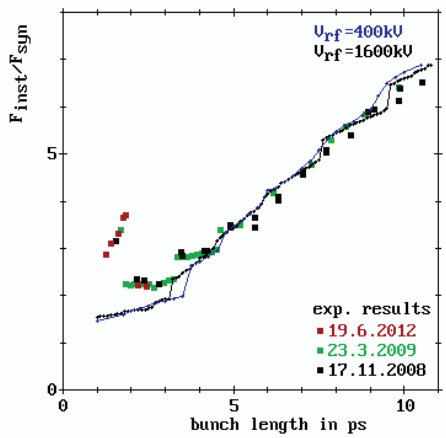


Figure 3: Frequency of the first unstable mode in units of the synchrotron frequency as a function of the zero current bunch length. The simulation was performed for two different cavity voltages and, like in the experiment with V_{rf}=1.4 MV, the momentum compaction factor was varied in order to change the length of the bunches.

ACKNOWLEDGMENT

The support of K. Holldack with the THz-detector is acknowledged. G. Wüstefeld is thanked for making the experimental results of the MLS available prior to publication.

REFERENCES

[1] M. Venturini, et al., Phys. Rev. ST-AB **8**, 014202 (2005)
[2] M. Abo-Bakr, et al., PAC'03, RPPB006, www.JACoW.org
[3] M. Ries, et al., IPAC'12, WEPPR046, www.JACoW.org
[4] K.L. Bane, et al., Phys. Rev. ST-AB **13**, 104402 (2010)
[5] K. Oide and K. Yokoya, KEK-Preprint-90-10, April 1990
[6] P. Kuske, IPAC'11, THPO025, http://www.JACoW.org
[7] R. Warnock and J.A. Ellison, SLAC-PUB-8404 (2000)
[8] K. Oide, Part. Accel. **51**, 43 (1995)
[9] P. Kuske, PAC'09, FR5RFP063, www.JACoW.org
[10] R.L. Warnock, PAC'91, PAC1991_1824, http://www.JACoW.org
[11] M. Abo-Bakr, et al., PAC'93, RPPB005, www.JACoW.org
[12] D. Zhou, et al., IPAC'11, WEPC108, www.JACoW.org

COMPUTATIONAL NEEDS FOR RF DESIGN OF SUPERCONDUCTING CAVITIES*

A. Lunin#, T. Khabiboulline, S. Yakovlev, Fermilab, Batavia, IL 60510, USA

Abstract

The computational approaches assure essential guidance and order for the design of superconducting cavities and cryomodules. The nature of superconductivity requires the precise computation of surface electromagnetic fields in order to design the cavity shape with a maximum accelerating gradient. At the same time, the thickness of the cavity shell is limited by the ability to cool it down to the temperature of liquid He, which makes the mechanical stability of a cavity and a liquid He vessel assembly extremely important. Hence, a self-consistent electro-mechanical optimization is required in order to minimize microphonics and/or Lorentz force detuning phenomena. Some specific challenges are estimation of the amount of RF losses caused by the interaction of the passing beam with SC cavity and a multipactor analysis in the SC cavity and RF coupler. Finally, the irregular time structure of a beam train with its own dense spectra may stochastically induce High Order Mode (HOM) fields in a cavity which results in beam emittance dilution.

The study of these effects leads to specification of the SC cavity and cryomodule and can significantly impact the efficiency and reliability of superconducting linac operation.

INTRODUCTION

SRF cavities are widely used today for various applications in high energy physics, nuclear physics and material science requiring an acceleration of charged particles [1][2][3]. New projects such as Accelerator Driven Subcritical systems (ADS), high intensity proton accelerator (Project X), Facility for Rare Isotope Beams (FRIB) and Next Generation Light Source Facility (NGLS) are under development [4][5][6][7]. Because of such diversity, designs of SRF cavities cover wide range of particles velocities (beta ~ 0.05 - 1), operating frequencies (0.072 - 4 GHz), beam currents (1mA - 100mA) and suitability for pulsed and continuous operation regimes. Besides, the SRF cavity is a complicated electro-mechanical assembly and consists of: bare cavity shell with power and HOM couplers, stiffening elements (ring, bars), welded LHe vessel, slow and fast frequency tuners and vacuum and coupler ports. A mechanical design of SRF cavity is illustrated in Figure 1 [8]. Therefore, the development of SRF cavity requires a complex, self-consistent electro-mechanical analysis in order to minimize microphonics and/or Lorentz force detuning phenomena and preserve good cavity tenability simultaneously.

* Operated by Fermi Research Alliance, LLC under Contract No.DE-AC02-07CH11359 with the U.S. Department of Energy.
#lunin@fnal.gov

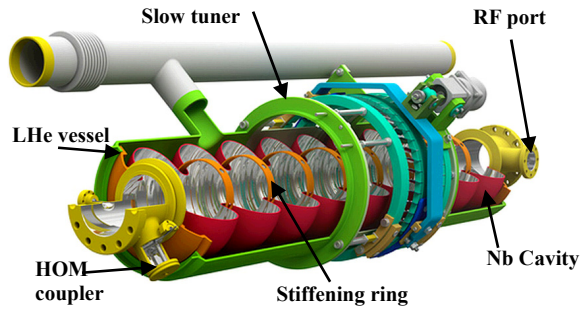


Figure 1: ILC 1.3 GHz 9-cell cavity.

The following aspects need to be taken into account during the design process:

- Cavity acceleration efficiency, including minimization of surface field enhancement factors (both electric and magnetic) and choosing optimum beta value for multiple cavities in the linac section.
- High gradient pulsed operation and minimizing of Lorentz force detuning coefficient.
- Operation with small beam current and narrow cavity bandwidth, particularly suppressing microphonics and preserving good cavity tenability simultaneously.
- HOM dumping in order to minimize cryogenic losses, preserving good beam emittance and excluding any transverse or longitudinal beam instabilities.
- Analyzing field emission effects including multipactor and dark current simulations.

We discuss all these issues and possible solutions below

SRF CAVITY EM ANALYSIS

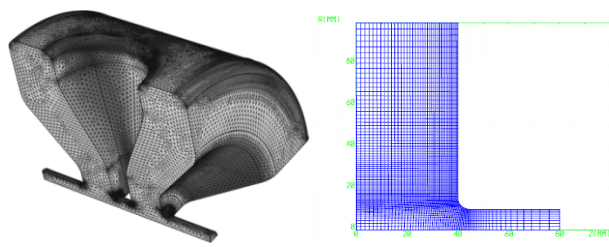


Figure 2: Meshing approaches in 3D (left) and 2D (right) domains.

Superconducting resonators suffer from both high magnetic field, which causes a thermal quench if it goes above critical value, and high electrical field, because it initiates surface field emission which produces additional cryolosses and may result in strong x-rays as well. Thus, the SRF cavity EM design requires precise surface

electromagnetic fields computation in order to optimize the cavity shape and achieve a maximum accelerating gradient.

Despite the fact that modern software for eigenmode EM analysis (ANSYS, COMSOL, CST, OMEGA3P [9,10,11,12]) provide very good accuracy in eigen frequencies, the precise calculation of surface fields is still a challenge. It requires a high quality surface mesh, which results in either the dense tetrahedral mesh with curved boundary elements in 3D domain or running the simulation in 2D space for axisymmetric structures (see Figure 2). The simulations with dense mesh require extra CPU time, while solving a problem in 2D has an obvious limitation in representing all the details of complex cavity geometry. The remedy is creating artificial mesh structure with high density area near the cavity outer surface only. The result of precise surface fields' simulation with less than 1% error is presented in Figure 3 for HFSS eigenmode solver [13].

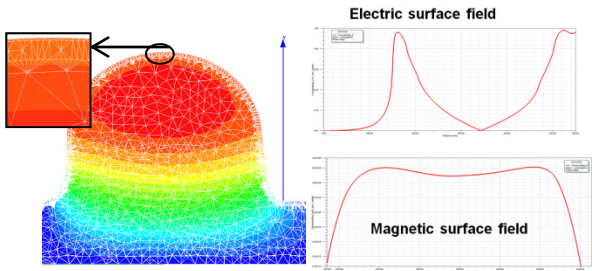


Figure 3: HFSS artificial surface mesh a) and calculated fields b).

The goal of the cavity shape optimization is to minimize both surface electric E_s and magnetic H_s fields. Hence, the result is not just a single point but the series of limiting curves in the E_s versus H_s coordinates. The idea is illustrated in Figure 4. The recent results of a cavity shape optimization for Project X linac are presented in [14], [15], [16] for low beta half-wave, medium beta spoke and high beta elliptical cavities respectively. The final choice of the cavity geometry is a trade-off between the requirements on cavity mechanical stability and surface processing issues.

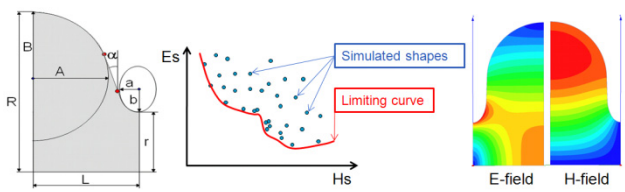


Figure 4: The conception of a cavity shape optimization.

LOSS FACTOR SIMULATION

Incoherent losses introduced by radiated wakefields might be an essential part of the total cryolosses in the SC accelerating structure. The effect of beam velocity on generated wakes can be demonstrated using the following consideration (see Fig. 5). The spectrum of wakefields excited by a beam bunch, passing an SRF structure, depends on σ_{field}, the characteristic size of the EM field distribution on the wall of the beam pipe at the cavity entrance. In linacs with relativistic beams, such as ILC or NGLS, field distribution is essentially disk-like and $\sigma_{field} = \sigma_{bunch}$, frequency spectrum up to $f_{max} \sim c/\sigma_{bunch}$. For the bunch size of 50 μm frequencies up to 6 THz can be excited. In a case of non-relativistic beam, field lines diverged much more, $\sigma_{field} \sim a$ and $f_{max} \sim c/a$. If the beam pipe radius is 50 mm, only HOMs below 6 GHz are exited. Thus, the cavity loss factor depends strongly on beam velocity.

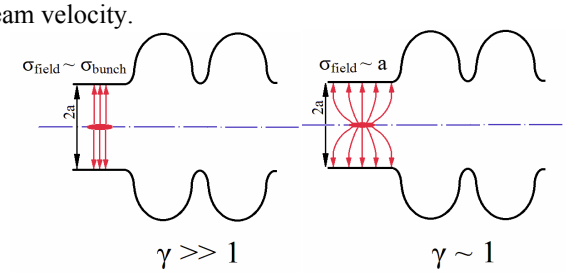


Figure 5: Effect of beam velocity on wakes excitation.

There are two approaches on how to calculate cavity loss factors, the integration of a wake potential over the bunch profile and the addition of loss factors for individual cavity modes. The time-domain computation of beam energy loss factors is very common and well developed for relativistic beams. It can be done by a variety of codes like ACE3P, GdfidL, CST or ABCI [17, 18, 11, 19]. Nevertheless, the wake calculation for a periodic system with the short bunch presents serious difficulties because the mesh size should be small enough in order to provide calculation stability and accuracy. There are several methods (e. g. indirect wake potential calculation and moving mesh technique) which allow overcoming the problem and calculating loss factors for the relativistic beam with sub-micron bunch size.

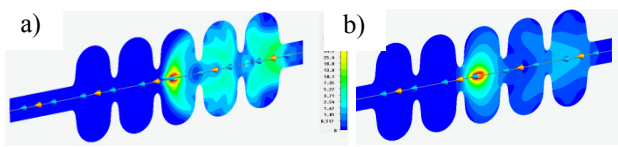


Figure 6: Time domain simulations of wakefields in 5-cell elliptical cavity for Project X in a case of ultra-relativistic beam (a) and high-relativistic beam (b).

For non-relativistic beams passing through a cavity one needs to take into account the static Coulomb forces, and indirect methods and moving mesh technique are not applicable. The direct method requires a meshing of the full structure volume, and the beam pipe length needs to be longer than a wake's catch-up distance. The alternative is a simulation in the frequency domain and finding the loss factor of individual cavity modes. This method is preferable when the beam is low-relativistic, its spectrum is limited and the number of modes is not too large. The example of a time domain loss factor simulation in 5-cell elliptical cavity made by CST Studio is presented in Figure 6 for both ultra-relativistic and high-relativistic beams [20].

MULTIPACTOR AND DARK CURRENT SIMULATIONS

Secondary electron emission, RF discharge or multipactor (MP) might be a serious obstacle for normal operation of SC cavities and couplers. When the electron trajectories oscillate in a resonance with the RF field, it causes exponential growth in the number of electrons by secondary emission.

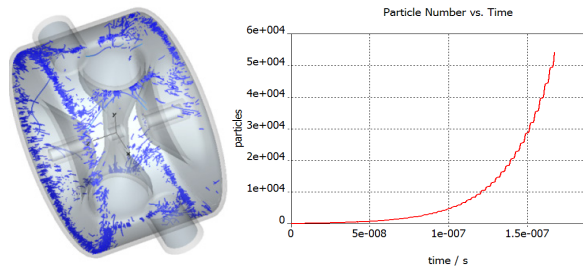


Figure 7: Multipactor simulation in the SSR1 cavity for Project X.

The main challenges in reliable multipactor prediction are accurately applying a probabilistic mechanism of secondary emission model given in [21], adjusting the tracking mesh with precise surface representation and obtaining a smooth distribution of electromagnetic fields in the cavity. Most codes for multipactor simulation tend to simplify secondary emission model and to use so-called counter function or impact function to predict a multipactor instead of a straightforward stochastic analysis [22, 23, 24]. Such assumptions usually give good results but predict a multipactor only within relatively narrow separated bands of field levels. The most advanced model of secondary emission [21] is implemented in CST PS solver, which makes it possible to detect the multipactor within a wider range of field levels and see the overlapped (merged) multipactor zones. The recent multipactor simulations made with CST PS are shown in Figures 7 and 8 and demonstrate the close agreement with experimental observations [25, 26].

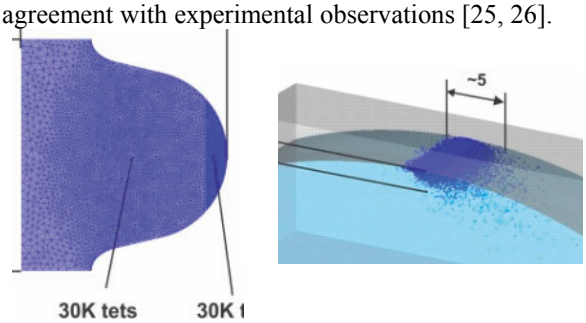


Figure 8: Multipactor simulation on the equator of the 9-cell ILC cavity.

Operation of SC cavity at high gradient may initiate the electron emission or dark current from the cavity surface; electrons can be captured by RF fields and accelerated along the linac to hundreds of MeV before being kicked out by quadrupoles. The goal of dark current simulation is to estimate the amount of additional losses in the cavity due to high field emission and the percentage of electrons that might be accelerated along the cryomodule. Because the surface distribution of electric field is highly non-uniform, the field emission cannot occur in an arbitrary location, and realistic model of emission is very important in the simulation. There are two codes, ACE3P and CST PS, which realize the advanced Fowler-Nordheim field electron emission model.

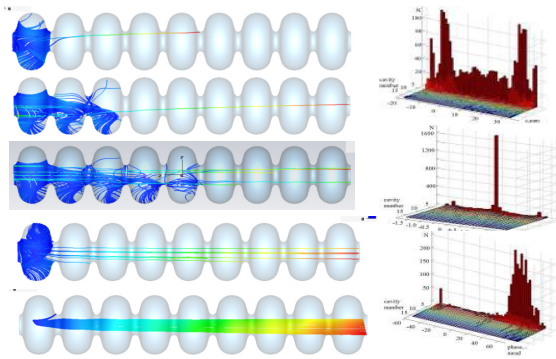


Figure 9: Dark current simulation in the ILC 1.3 GHz, 9-cell cavity and captured particle distributions: a) radial, b) angular and c) phase.

In order to calculate the effects of dark current, heat and RF loading of the cavity, influence of secondary emission and the properties of accelerated particles, the program has to provide the detailed statistics on lost and captured electrons, the number of impacts and their locations across the cavity surface. The example of dark current simulation with CST PS is shown in Figure 9 for the ILC, 1.3 GHz 9-cell elliptical cavity [27].

HOM STATISTICAL ANALYSIS

Excitation of High Order Modes (HOM) in SRF cavities is always a concern. Heating caused by beam power lost to HOMs adds to the cryogenic losses and increases the operational cost of the linac. The interaction of the beam with excited HOMs deteriorates the quality of the beam. In order to accurately estimate the probability of resonance HOM excitation in the Project X linac by the beam, one has to do statistical analysis, which requires data for HOM frequency, impedance and quality factor spreads.

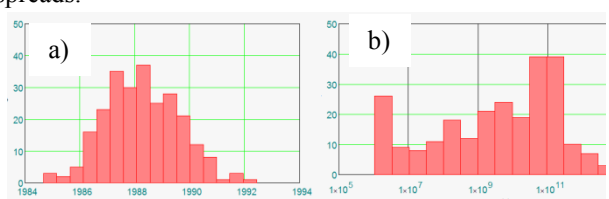


Figure 10: The frequency (a) and Q_{ext} (b) spreads of monopole HOMs for the 650 MHz, β_{geom}=0.9 Project X structure.

Because of the fabrication tolerances and further surface processing the actual cavity shape never matches with the theoretical shape. Thus, there is a natural spread of the HOM parameters from cavity to cavity. The simulation of HOM parameter spreads is a challenging problem because it requires eigenmode analysis of many random structures and HOM spectrums. Nevertheless, the

use of scripts or macros allows easy running of such statistical analysis together with modern eigenmode solvers. The result of HOM statistical analysis is presented in Figure 10 for high-beta elliptical cavity for Project X [28].

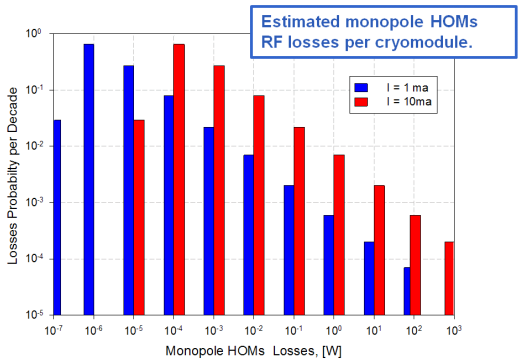


Figure 11: Probability of monopole HOMs RF losses per cryomodule in Project X linac for 1 mA and 10 mA beam currents.

Based on the predicted deviations in monopole HOM frequencies, R/Q and Q_{ext} it is possible to estimate the probability of RF losses per cryomodule as well as the probability of the beam longitudinal emittance growth. The outcome of HOMs influence on RF losses and beam emittance for the high energy part of Project X CW linac is summarized in Figures 11 and 12 respectively.

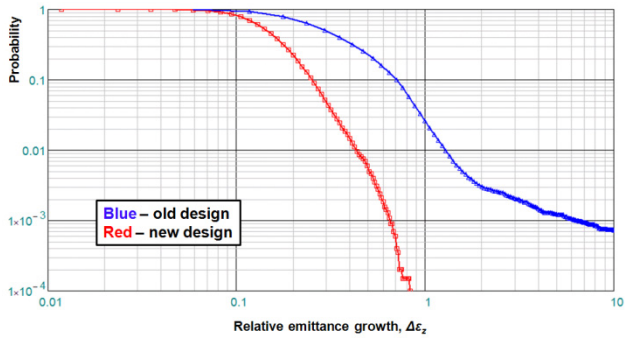


Figure 12: The probability of longitudinal emittance growth in the Project X linac for the 10mA beam current for two different designs of the same cavity.

The HOMs statistical analysis is a powerful tool to predict the SRF cavity behavior in the presence of a beam with complex time structure, especially for high current operations and CW regimes.

MULTIPHYSICS ANALYSIS

The study of coupled multiphysics problems is extremely important during the SRF cavity design, because it defines the technical feasibility to build, assemble and run the cavity under the operating parameters. Among the common problems which needed to be solved, are the microphonics effect, Lorentz Force Detuning (LFD) and coupled RF-thermal analysis.

Microphonics is a cavity mechanical deformation driven by external forces such as the pressure fluctuations in a He-vessel and results in unfavourable RF phase errors. In order to calculate the frequency dependence on external pressure fluctuation, named *df/dp* coefficient, we have to run eigenmode simulation to find the nominal frequency f_0 , find the cavity deformation under the given pressure load and repeat the eigenmode analysis again to find the resonant frequency, f_p, after the deformation. Thus, microphonics study requires self-consistent electro-mechanical simulations. The main difficulty here is transferring the deformed cavity model back to the EM solver, because it requires re-meshing of the cavity volume and, thus, introduces undesirable mesh errors, which cannot be neglected for small deformations. The clever solution was proposed and realized in COMSOL multiphysics environment. The idea is to use a common mesh for all solvers and transfer only the boundary deformation. Therefore, the conception of moving mesh allows to exclude the mesh errors and to achieve excellent accuracy (~1Hz/Torr) in *df/dp* analysis. Examples of required mesh for microphonics simulation are presented in Figure 13, a) and b) for COMSOL and ANSYS programs respectively [29, 14].

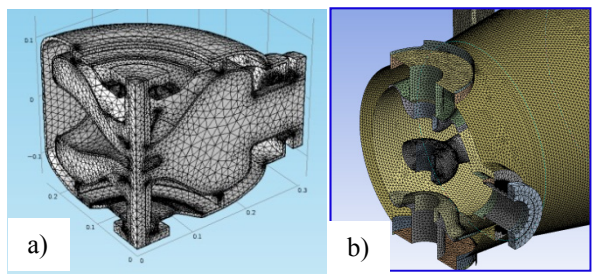


Figure 13: The concept of common mesh in COMSOL (a) and independent mesh in ANSYS (b) multiphysics simulations.

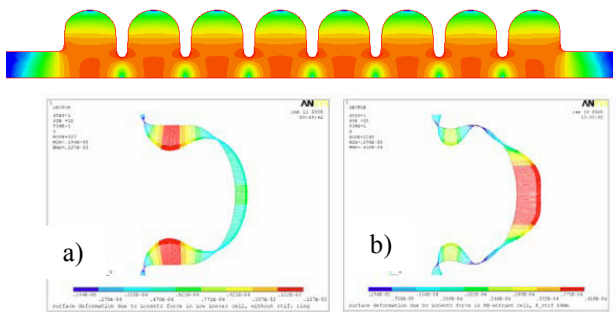


Figure 14: Lorentz force detuning simulation of SRF β=0.81 elliptical cavity and cavity walls deformation with no ring (a) and with stiffening ring (b).

Lorentz Force Detuning (LFD) is the cavity mechanical deformations caused by a pressure load of the Lorentz forces from EM-fields in the cavity. It significantly distorts the pulsed operation of SC cavity at high accelerating gradient. The LFD analysis requires transferring the solution for EM-fields to mechanical solver, finding the cavity wall deformation and then transferring it back to EM solver. Thus, LFD simulation suffers the same difficulties as microphonics, and introduces additional ones because it requires the consideration of the complete cavity mechanical model

including He-vessel and slow tuner. The example of LFD analysis, made in ANSYS, is shown in Figure 14 for SRF β=0.81 elliptical cavity [30]. The simulations show that stiffening rings allow the compensation of deformations due to electric and magnetic fields and, thus, the reduction of the LFD coefficient.

CONCLUSION

The design of SRF cavity has to satisfy a complex relationship between accelerator requirements, cryogenic effects and cryomodule structure. Modern state-of-the-art software for multiphysics analysis allows the simulation of various problems related to the design of SRF cavity and the attainment of reliable engineering solutions.

The study of these effects leads to specification of SC cavity and cryomodule and can significantly impact the efficiency and reliability of the superconducting linac.

REFERENCES

[1] D. Boussard, et al., "The LHC superconducting cavities", Proceeding. of PAC1999, New York, USA

[2] K. Shepard, at al., "Superconducting RF Development at ATLAS", Proceedings of the 6th Workshop on RF Superconductivity, 1993, VA, USA

[3] http://www.xfel.eu/

[4] P. Singh, "SRF accelerator for Indian ADS Program: Present & future prospects", Proceeding. of SRF2011, Chicago, IL, USA

[5] Project X: Accelerator Overview, http://projectx.fnal.gov/.

[6] The Facility for Rare Isotope Beams (FRIB) http://www.frib.msu.edu/

[7] J. Corlett, "A Next Generation Light Source Facility at LBNL", Proc. of PAC 2011, New York, NY, USA

[8] http://newsline.linearcollider.org/

[9] http://www.ansys.com

[10] http://www.comsol.com

[11] http://www.cst.com

[12] Lie-Quan Lee, et al, Omega3P: A Parallel Finite-Element Eigenmode Analysis Code for Accelerator Cavities, SLAC-PUB-13529, 2009.

[13] A. Lunin, "Design of alternative β=0.92 cavity for Project X", Project X Document 926-v1

[14] P. Ostroumov, et al., "Development of a Half-Wave Resonator for Project X", WEPPC039, IPAC'12, New Orleans, USA

[15] L. Ristori, et al., "Design of SSR1 Single Spoke Resonators for PXIE", WEPPC057, IPAC'12, New Orleans, USA

[16] A. Lunin, et al., "Alternative Cavity for H E Part of the Project X linac", WEPPR029, IPAC'12, New Orleans, USA

[17] Cho Ng, et al, Design and optimization of large accelerator systems through high-fidelity electromagnetic simulations, Journal of Physics: Conference Series 125, 2008

[18] http://www.gdfidl.de/

[19] Y. Chin, "User Guide for ABCI", LBL-35258

[20] A. Lunin, et al., "Cavity Loss Factors of Non-relativistic Beams for Project X", TUP075, PAC'11, NY, USA

[21] M.A. Furman, M.T.F. Pivi, "Probabilistic model for the simulation of secondary electron emission", PRSTAB, 5, 124404, (2002)

[22] S. Molloy, et al., "Multipacting Analysis for the Superconducting RF Cavity HOM Couplers in ESS", ESS Docs Document 115-v1

[23] J. DeFord, et al., "Improved Algorithms for Multipacting Simulation in the Analyst Code", WEP159, PAC'11, NY, USA

[24] P. Ylä-Oijala,et al., "MultiPac 2.1 - Multipacting simulation toolbox with 2D FEM field solver", User's manual, Rolf Nevanlinna Institute, Helsinki (2001).

[25] G. Romanov, "Simulation of Multipacting in HINS Accelerating Structures with CST Particle Studio", MOP043, LINAC'08, British Columbia, Canada

[26] S. Kazakov, et al., "Multipactor Simulation in SC Elliptical Shape Cavities", WEPPC051, IPAC'12, New Orleans, USA

[27] S. Solyak, et al., "Dark Current Model for ILC Main Linac", MOPP036, EPAC'08, Genoa, Italy

[28] S. Yakovlev, et al., "Higher Order Modes in the Project-X Linac", International Workshop on Higher-Order-Mode Diagnostics and Suppression in Superconducting Cavities, HOMSC12, UK

[29] L. Ristori, et al., "Design of Single Spoke Resonators for Project X", MOPO024, SRF2011, Chicago, USA

[30] I. Gonin, et al., "Development of the SCRF ß=0.81 Cavity for Proton Driver", TUP35, SRF2005, NY, USA

ARBITRARY HIGH-ORDER DISCONTINUOUS GALERKIN METHOD FOR ELECTROMAGNETIC FIELD PROBLEMS*

K. Papke†, C. R. Bahls, U. van Rienen, Universität Rostock, IAE, IEF, 18059 Rostock, Germany

Abstract

In this paper, we present a time integration scheme applied to the Discontinuous Galerkin finite element method (DG-FEM, [1]) for the computation of electromagnetic fields in the interior of three-dimensional structures. This approach is also known as Arbitrary High-Order Derivative Discontinuous Galerkin (ADER-DG, [2, 3]). By this method, we reach arbitrary high accuracy not only in space but also in time. The DG-FEM allows for explicit formulations in time domain on unstructured meshes with high polynomial approximation order. Furthermore, the Discontinuous Galerkin method in combination with the arbitrary high order time integration scheme is well suited to be used on massively parallel computing architectures. Moreover the method can be extended for local time stepping to become more efficiently by reducing the computation time [4].

INTRODUCTION

For the design and optimization of Higher-Order-Mode Coupler (Fig. 1), used in RF accelerator structures, numerical computations of electromagnetic fields as well as scattering parameter are essential. These computations can be carried out in time domain. In this work the implementation and investigation of a time integration scheme based on the Discontinuous Galerkin Finite-Element Method (DG-FEM) with arbitrary order in space and time is demonstrated for solving 3-D electromagnetic problems in time domain.

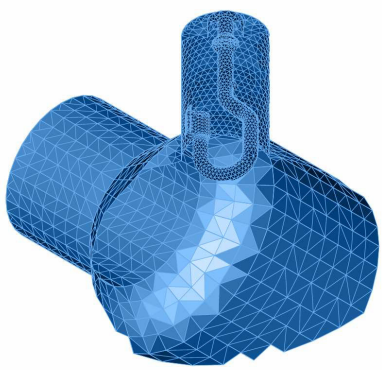


Figure 1: Tapered beam pipe with Higher-Order-Mode Coupler. The aspect ratio of such a grid is relatively high, due to the filigree structure of the antenna. Therefore, such a model is well suited for the application of local time stepping schemes.

THE NUMERICAL SCHEME

With known initially and boundary conditions it is sufficient to describe classic electromagnetic phenomena only by AMPERE's and FARADAY's law of MAXWELL equations. The partial differential equations can be written in the following general form:

$$\frac{\partial \boldsymbol{u}}{\partial t} + \mathcal{A}_1 \frac{\partial \boldsymbol{u}}{\partial x} + \mathcal{A}_2 \frac{\partial \boldsymbol{u}}{\partial y} + \mathcal{A}_3 \frac{\partial \boldsymbol{u}}{\partial z} = 0, \qquad (1)$$

where

$$\boldsymbol{u}(x, y, z, t) \quad = \quad (E_x, E_y, E_z, H_x, H_y, H_z)^T. \quad (2)$$

The space-dependent Jacobian matrices $\mathcal{A}_i$ determine the physical behavior of the equations. They contain material properties as well as the curl operator applied to E- and H-field. To solve this ordinary partial differential equation, a physical initial condition as well as boundary conditions are still needed.

For the numerical scheme the computational domain $\Omega \in \mathbb{R}^3$ will be partitioned into conforming tetrahedral elements D^k. The approximate solution $\boldsymbol{u}_h^k$ of (1) inside each tetrahedron D^k is given by:

$$\boldsymbol{u}_h^k = \sum_{i=1}^{N_p} \hat{\boldsymbol{u}}_i^k(t) \cdot \Phi_i^k(\boldsymbol{x}). \qquad (3)$$

Here $\Phi_i^k(\boldsymbol{x})$ are the nodal ansatz functions and $\hat{\boldsymbol{u}}_i^k(t)$ are the time-dependent degrees of freedom which are allocated at the nodes of the element D^k.

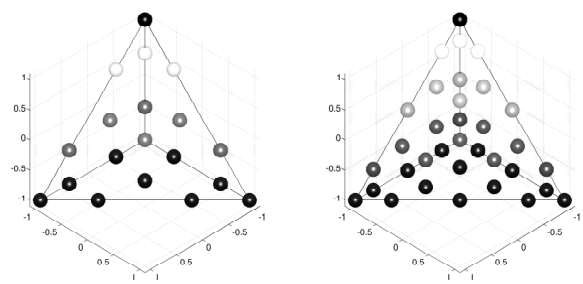


Figure 2: Reference element with the nodes on which the time-dependent degrees of freedom are defined. (left: 3rd order ansatz functions, right: 4th order ansatz functions)

* Work supported by Federal Ministry for Research and Education BMBF under contract 05K10HRC

† kai.papke@uni-rostock.de

Multiplication of Equation (1) by test functions $\Phi_j(\boldsymbol{x})$ and integration over the element D^k followed by integration by parts results in the following variational formulation [1, 2]:

$$\int_{D^k} \Phi_j \frac{\partial \boldsymbol{u}_h^k}{\partial t} dV + \int_{\partial D^k} \Phi_j \boldsymbol{F_h} dS - \int_{D^k} \left(\frac{\partial \Phi_j}{\partial x} \mathcal{A}_1 \boldsymbol{u}_h^k + \frac{\partial \Phi_j}{\partial y} \mathcal{A}_2 \boldsymbol{u}_h^k + \frac{\partial \Phi_j}{\partial z} \mathcal{A}_3 \boldsymbol{u}_h^k \right) dV = 0, \tag{4}$$

where we introduced the numerical flux that provides the coupling between the solutions of adjacent elements. This weak formulation allows for solutions in Ω which are discontinuous at the borders between the elements D^k. The surface and volume integrals over the ansatz- and test function, respectively their derivatives will be pre-computed for the reference element (Fig. 2) and adapted onto each element D^k by conformal mapping. Equation (4) has to be fulfilled for every test function Φ_j, which leads to a linear system of equations with the well known mass matrix $\mathcal{M}^k$ and stiffness matrices $\mathcal{S}_x^k$, $\mathcal{S}_y^k$ and $\mathcal{S}_z^k$:

$$\mathcal{M}_{(i,j)}^k = \int_{D^k} \Phi_i(\boldsymbol{r}) \Phi_j(\boldsymbol{r}) d\Omega = \langle \Phi_i | \Phi_j \rangle$$

$$\mathcal{S}_{x(i,j)}^k = \int_{D^k} \frac{\partial \Phi_i(\boldsymbol{r})}{\partial x} \Phi_j(\boldsymbol{r}) d\Omega = \langle \frac{\partial \Phi_i}{\partial x} | \Phi_j \rangle.$$

The resulting semi-discrete formulation corresponds to homogeneous differential equations with time derivations, which can be written for every element D^k as:

$$\frac{d\boldsymbol{u}_h^k}{dt} = \mathcal{L}^k(\boldsymbol{u}_h^k, t), \tag{5}$$

with the operator $\mathcal{L}^k$ that summarizes the mass- and stiffness matrices as well as the numerical flux. To get a fully discrete scheme in space and time we redraft Equation (5) in the integral form:

$$\boldsymbol{u}_h^k(t_{n+1}) = \boldsymbol{u}_h^k(t_n) + \int_{t_n}^{t_n+\Delta t} \mathcal{L}^k(\boldsymbol{u}_h^k, \tau) d\tau. \tag{6}$$

This leads to time integration schemes, where the problem lies now in the approximation of the integral in Equation (6). In the case of the MAXWELL equations the operator $\mathcal{L}^k$ behaves linear for $\boldsymbol{u}_h^k$ and is independently on the time. This enables the possibility to change the order of integration and application of the operator $\mathcal{L}^k$, what is advantageous but not necessary for the ADER approach. For the following formulas we remove the superscript k in $\boldsymbol{u}_h^k$ and all matrices to simplify the expressions. This means also that the following procedure for the time integration is carried out separately for each element.

ARBITRARY HIGH-ORDER DERIVATIVE TIME DISCRETIZATION

The basic idea of the ADER approach is the TAYLOR-expansion of the numerical solution in time, which allows for an analytical time integration.

$$\boldsymbol{u}_h(x, y, z, t) = \sum_{p=0}^{N} \frac{t^p}{p!} \frac{\partial^p}{\partial t^p} \boldsymbol{u}_h(x, y, z, t_0) \tag{7}$$

The foundation for this approach is the CAUCHY-KOWALEWSKAYA theorem that asserts a unique analytic solution for the considered boundary problem in the neighborhood of t_0. Here, N determines the convergence order of the resulting time integration scheme.

The time derivatives will be replaced by spatial derivatives based on the problem (1) itself.

$$\frac{\partial^p \boldsymbol{u}_h}{\partial t^p} = (-1)^p \left(\mathcal{A}_1 \frac{\partial}{\partial x} + \mathcal{A}_2 \frac{\partial}{\partial y} + \mathcal{A}_3 \frac{\partial}{\partial z} \right)^p \boldsymbol{u}_h \tag{8}$$

Since the ansatz functions Φ_i are only defined for a reference element, which is given in the $(\xi\eta\zeta)$- coordinate system, the spatial derivatives have to be carried out with respect to ξ, η and ζ. Equation (8) can be rewritten with a linear transformation from the reference element to every individual tetrahedral element D^k as:

$$\frac{\partial^p \boldsymbol{u}_h}{\partial t^p} = (-1)^p \left(\mathcal{A}_1^* \frac{\partial}{\partial \xi} + \mathcal{A}_2^* \frac{\partial}{\partial \eta} + \mathcal{A}_3^* \frac{\partial}{\partial \zeta} \right)^p \boldsymbol{u}_h. \tag{9}$$

For the MAXWELL equations, for example the transformed Jacobi matrice $\mathcal{A}_1^*$ offers the following structure:

$$\mathcal{A}_1^* = \begin{pmatrix} 0 & \mathcal{C}_\xi \\ \mathcal{C}_\xi^T & 0 \end{pmatrix}, \text{ with } \mathcal{C}_\xi = \begin{pmatrix} 0 & -\frac{\partial \xi}{z} & \frac{\partial \xi}{y} \\ \frac{\partial \xi}{z} & 0 & -\frac{\partial \xi}{x} \\ -\frac{\partial \xi}{y} & \frac{\partial \xi}{x} & 0 \end{pmatrix}. \tag{10}$$

Clearly, these matrices differ from element to element. Together with Equation (9), the analytical expression for the approximate solution $\boldsymbol{u}_h$ in Equation (7) can be rewritten as:

$$\boldsymbol{u}_h(\xi, \eta, \zeta, t) = \sum_{p=0}^{N} \frac{t^p}{p!} (-1)^p \left(\mathcal{A}_1^* \frac{\partial}{\partial \xi} + \mathcal{A}_2^* \frac{\partial}{\partial \eta} + \mathcal{A}_3^* \frac{\partial}{\partial \zeta} \right)^p \boldsymbol{u}_h(\xi, \eta, \zeta, t_0) \tag{11}$$

Similar to finite element methods, this approximation will be projected onto each test function in order to get an approximation of the evolution of the degrees of freedom $\hat{\boldsymbol{u}}_j(t)$ during one time step from t_n to t_{n+1}.

$$\hat{\boldsymbol{u}}_j(t) \langle \Phi_n | \Phi_j \rangle = \langle \Phi_n | \sum_{p=0}^{N} \frac{t^p}{p!} (-1)^p \times \left(\mathcal{A}_1^* \frac{\partial}{\partial \xi} + \mathcal{A}_2^* \frac{\partial}{\partial \eta} + \mathcal{A}_3^* \frac{\partial}{\partial \zeta} \right)^p | \Phi_i \rangle \hat{\boldsymbol{u}}_i(t_0) \tag{12}$$

On the left hand side of Equation (12) we find again the elements of the mass matrix $\mathcal{M}_{(n,j)}$, which is already used for the DG-FEM. Furthermore, after expansion of the right hand side we obtain also the following projections:

$$\mathcal{S}^{\lambda\mu\nu}_{(n,i)} = \langle \Phi_n | \frac{\partial^p}{\partial\xi^\lambda \partial\eta^\mu \partial\zeta^\nu} \Phi_i \rangle, \tag{13}$$

where $\lambda + \mu + \nu = p$ and $\lambda, \mu, \nu \geq 0$. All these projection matrices can be pre-computed for the reference element by a computer algebra system once and then be stored. In the case of nodal ansatz functions, that result from an interpolation on orthonormal polynomials it is not necessary to compute the projections. Instead, we can use the following relation from [1]:

$$(\mathcal{M}^{-1}\mathcal{S}^{\lambda\mu\nu})_{(i,j)} = \frac{\partial^p}{\partial\xi^\lambda \partial\eta^\mu \partial\zeta^\nu} \Phi_j \Big|_{\xi_i,\eta_i,\zeta_i}. \tag{14}$$

By using only these derivation matrices we directly obtain an explicit expression for the time dependent degrees of freedom $\hat{\boldsymbol{u}}_j(t)$ in Equation (12), due to the contained multiplication with the inverse mass matrix.

Finally, the integral of the time dependent degrees of freedom from t_n to t_{n+1} can be written as:

$$\int_{t_0}^{t_0+\Delta t} \hat{\boldsymbol{u}}_j^k(t)dt = \mathcal{M}^{-1}_{(n,j)} \langle \Phi_n | \sum_{p=0}^{N} \frac{\Delta t^{(p+1)}}{(p+1)!} (-1)^p \left(\mathcal{A}_1^* \frac{\partial}{\partial\xi} + \mathcal{A}_2^* \frac{\partial}{\partial\eta} + \mathcal{A}_3^* \frac{\partial}{\partial\zeta} \right)^p |\Phi_i\rangle \hat{\boldsymbol{u}}_i^k(t_0). \tag{15}$$

With the $\mathcal{L}$ operator applied on these resulting matrices, the fully time integration in Equation (6) is performed with arbitrary high order accuracy N.

RESULTS

The numerical method for the spatial discretization is implemented in NUDG++, an open source framework written in C++. The time integration procedure is currently in progress, but runs already on CPU. It follows the implementation for GPU's, which accelerates the algorithm extensive. At this time the algorithm is tested only for a first simple example, the lossless rectangular cavity (Fig. 3).

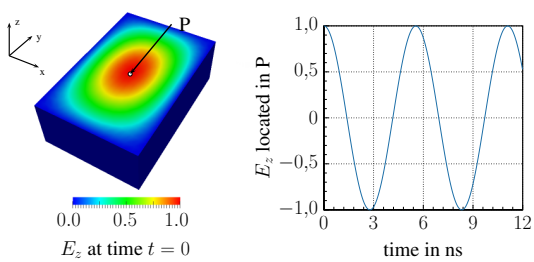


Figure 3: Left: Electric field distribution of the E_z component in a rectangular resonator. Right: The corresponding time evolution of E_z in the center of the cavity.

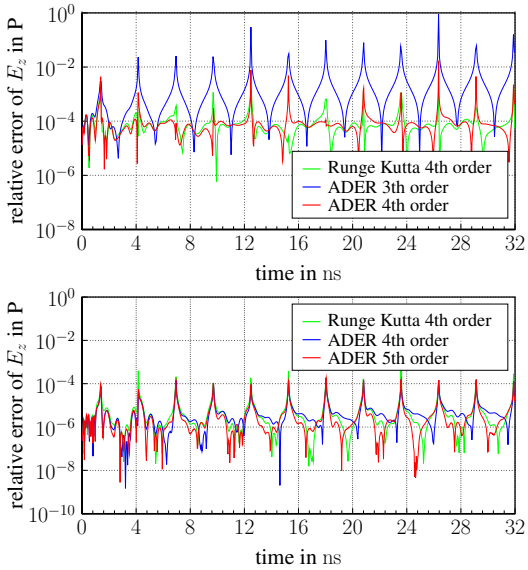


Figure 4: Comparison between ADER time integration scheme and a 4th order RUNGE-KUTTA scheme. Top: 3rd order ansatz functions, Bottom: 4th order ansatz functions.

In Figure 4 the relative error is shown for the ADER scheme compared with a 4th order RUNGE-KUTTA scheme. For the same convergence order the errors of both algorithms lies as expected in the same order of magnitude.

CONCLUSIONS AND OUTLOOK

By using only one CPU core the 4th order RUNGE-KUTTA scheme is $\sim$ 3 times faster then the ADER approach in the same convergence order (3rd order ansatz functions). This relation factor reduces in the case of 4th order ansatz functions to 2. An implementation on parallel architectures probably improves this ratio related to the ADER aproach. Therefore we are currently focused on an implementation for GPUs with the Cuda or OpenCL framework.

REFERENCES

[1] J. S. Hesthaven and T. Warburton, *Nodal Discontinuous Galerkin Methods*, (New York: Springer, 2007).

[2] M. Dumbser and M. Käser, "An arbitrary High-Order Discontinuous Galerkin Method for Elastic Waves on Unstructured Meshes – II. The Three-Dimensional Isotropic Case," Geophys. J. Int., 2006.

[3] A. Taube, M. Dumbser et al. "Arbitrary High-Order Discontinuous Galerkin Schemes for the Magnetohydrodynamic Equations," Journal of Scientific Computing Vol. 30, 2007.

[4] N. Gödel, S. Schomann et al. , "GPU Accelerated Adams-Bashforth Multirate Discontinuous Galerkin FEM Simulation of High Frequency Electromagnetic Fields," IEEE Transactions on Magnetics Vol. 46, 2010.

STATUS OF THE HOM CALCULATIONS FOR THE B*ERL*inPro MAIN LINAC CAVITY*

A. Neumann†, W. Anders, J. Knobloch, Helmholtz-Zentrum Berlin, Berlin, Germany
B. Riemann, T. Weis, TU Dortmund University, Dortmund, Germany
K. Brackebusch, T. Flisgen, T. Galek, K. Papke, U. van Rienen
Rostock University, Rostock, Germany

Abstract

The Berlin Energy Recovery Linac Project (B*ERL*inPro) is designed to develop and demonstrate CW LINAC technology and expertise required to drive next-generation Energy Recovery Linacs (ERLs). Strongly higher order mode (HOM) damped multicell 1.3 GHz cavities are required for the main linac. The cavity under study is an integrated design of the Cornell base cell with JLab HOM waveguide couplers. Modifications to the end group design have also been pursued, including the substitution of one waveguide by a HZB-modified TTF-III power coupler. In this paper the progress in HOM calculations to avoid beam-breakup instabilities for the favored cavity structure will be presented.

INTRODUCTION: THE B*ERL*inPro PROJECT

The B*ERL*inPro ERL will be a CW driven machine accelerating a 100 mA beam to 50 MeV while preserving a normalized emittance of better than 1 mm mrad at a pulse length of 2 ps [1]. For the different sections of the superconducting accelerator -SRF photo-injector, booster module and main linac in the recirculator- the operating boundary conditions for the cavity design vary from high current, high beam loading to a high current, zero net beam loading environment. Thus the requirements for quality factor and peak fields are quite different for these three cavity types as well as the HOM damping technique applied. Main emphasis of this paper will be the calculations for the main linac cavity.

The main linac cavity has to have strong HOM damping as it accelerates the beam during the first passage and decelerates the recirculated beam for energy recovery thus interacting with two 100 mA beams. The beam may excite a transverse acting mode, e.g. a TM_{110} dipole mode, which deflects a following bunch. After recirculation, this bunch will arrive within the same structure affected now by an offset and, depending on the phase advance, possibly further exciting this dipole mode.This effect may add up from bunch to bunch until the beam will be lost - the so called beam break up (BBU) [2]. Equation 1 describes the dependance of the threshold current I_{th} for one cavity HOM on the mode's transverse shunt impedance $R/Q_{\perp}$, its external quality factor Q_{ext}, frequency ω and the beam's energy γ, recirculating time t_{r} and the transfer matrix of the linear beam optics M_{12}:

$$I_{\mathrm{th}} = -\frac{2\gamma}{e}\frac{c}{\frac{R}{Q}_{\perp} Q_{\mathrm{ext}}\omega M_{12}}\frac{1}{\sin(\omega t_{\mathrm{r}})}. \qquad (1)$$

From the viewpoint of cavity design it means to minimize the HOM's $R/Q_{\perp}$ and Q_{ext}. The HOM's spectrum and dispersion relation is mainly influenced by the mid-cell design, especially the cell-to-cell coupling via the iris diameter. By tuning the end-cells one achieves first of all field flatness for the fundamental TM_{010}-π mode and may avoid trapped modes within the center cells. The end-cell is mainly responsible to couple the HOMs to the loads or damping structures.

In this paper an update of the HOM calculations of the main linac cavity is given which was already introduced in [3]. Further a short outlook on ongoing activities is given of which more is described in [4].

STRUCTURE DESIGN AND EIGENMODE CALCULATIONS

The baseline design so far consists of a seven cell structure using the Cornell's ERL design [5] and combining it with JLab style symmetric assembled waveguide HOM couplers [6] (see Figure 1). To allow for a flexible coupling, one group of HOM couplers is broken in symmetry by replacing one waveguide with a TTF-III type coaxial fundamental power coupler. This design combines the good peak field properties of the Cornell design with the advantage of waveguide couplers having a natural cutoff above the fundamental and further limiting the possibility of dust from ferrite beam tube absorbers propagating in the SC cavity.

Table 1 summarizes the figures of merit of the field-flatness tuned seven cell structure. The main task of this adapted design was to calculate its performance with respect to BBU, identify limiting HOMs and find means to tune them to lower Q_{ext} while preserving the fundamental's RF properties. Because of the influence of the couplers on the field distribution, most calculations done so far were carried out in 3D with CST MWS [7] making use of the given symmetry plane at the coaxial coupler and using both mesh types, tetrahedral and hexahedral, where the latter had to be used to calculate the Q_{ext}. In all results shown here mainly the Jacobi-Davidson-Method (JDM) was used. Figure 2 shows the calculated spectrum up to 3 GHz, mainly limited by mesh size versus computing power and time. Figure 3 depicts the corresponding Q_{ext}, demonstrating a rather good damping of the lowest order dipole

*Work supported by Federal Ministry for Research and Education BMBF under contract 05K10HRC and 05K10PEA
† Axel.Neumann@helmholtz-berlin.de

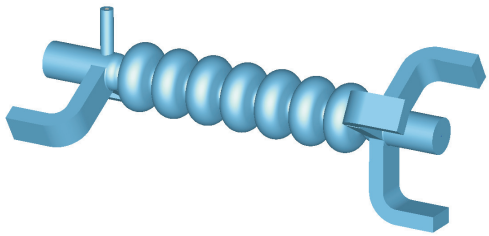


Figure 1: Drawing of the cavity design under consideration for B*ERL*inPro. It combines the Cornell ERL base cell with JLAB style waveguide HOM dampers and a tuneable TTF-III fundamental power coupler (FPC). This design features one symmetry axis in the FPC plane.

band, but some weak coupling to some of the quadrupole modes.

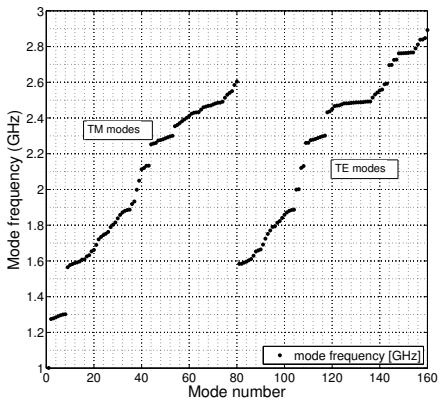


Figure 2: Frequency of all TE/TM modes up to 2.85 GHz, about the lower range of the 4th dipole band. All the modes were calculated using the 3-D eigenmode solver of CST MWS™.

The final goal will be to minimize the product $R/Q_\perp \cdot Q_{\text{ext}}$ for all HOMs. The $R/Q_\perp$ was directly evaluated making use of the Panofsky-Wenzel theorem by integrating the transverse Lorentz forces along the beam path. The resulting $R/Q_\perp$ is shown in Figure 4. Again the dipole modes seem sufficiently damped while some quadrupole modes due to their high Q_{ext} and some unexpected non-zero shunt impedance on axis show to be the most harming modes concerning BBU. An example of a strongly localized HOM with high Q_{ext} is given in Figure 5. The question arised whether these remnant on-axis fields are due to meshing and thus numerical inaccuracies of the method used or due to some shift of the quadrupole fields by the influence of the FPC.

TRANSVERSE SHUNT IMPEDANCE

To investigate the possible shift of quadrupole modes a calculation of $R/Q_\perp$ along circular distributed integration paths at an arbitrary offset of 5mm was done using the fol-

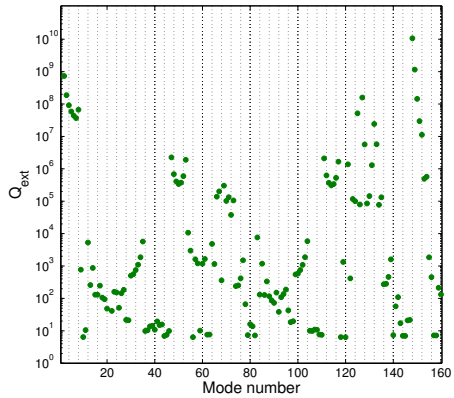


Figure 3: External Q of all TM/TE modes up to 2.85 GHz. While the dipole modes are sufficiently damped, some quadrupole modes exhibit large $Q_{\text{ext}} > 1 \cdot 10^6$.

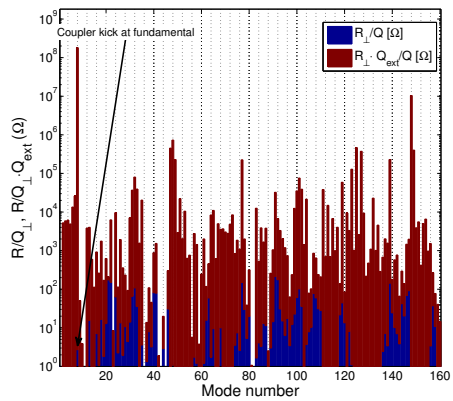


Figure 4: Transverse $R/Q_\perp$ and $R/Q_\perp \cdot Q_{\text{ext}}$ for all TM and TE modes up to 2.85 GHz (beginning of 4th dipole band). The arrow marks the coupler kick by the TTF-III power coupler at the fundamental mode.

lowing approximation to mitigate numerical meshing errors:

$$R_{\perp x} \approx \left| \frac{c}{\omega} \frac{\sqrt{R_\parallel(x_2)} - \sqrt{R_\parallel(x_1)}}{x_2 - x_1} \right|^2 . \qquad (2)$$

Figure 6 shows the distribution of $R/Q_\perp$ versus the polar angle of the integration paths for a quadrupole mode. Obviously the mode is shifted along the vertical axis (FPC axis).

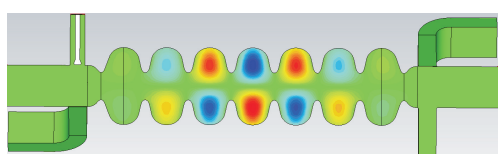


Figure 5: Example of strongly localized quadrupole mode at 2.3 GHz. This mode exhibits a high external Q with only low coupling to the waveguide dampers.

Table 1: Cavity Figures of Merit for the TM_{010}-π Mode of the Tuned Structure

Parameter	Value
$R/Q(\Omega)$	754
E_{surf}/E_{acc}	2.07
$\Delta\nu$ (MHz)	1.54
Q_{ext}	$5 \cdot 10^7$
Geometry factor $G(\Omega)$	272.7

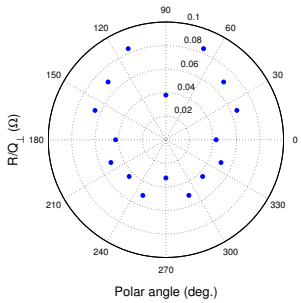


Figure 6: Polar plot of transverse R/Q determined on circularly arranged integration paths at 5mm offset from beam axis for a quadrupole mode.

Figure 7 shows the shift of the first dipole and quadrupole mode bands versus the beam axis. Most modes of these bands are shifted more than the expected vertical beam size in the linac and thus quadrupole modes may exhibit a dipole component. This effect needs to be investigated in

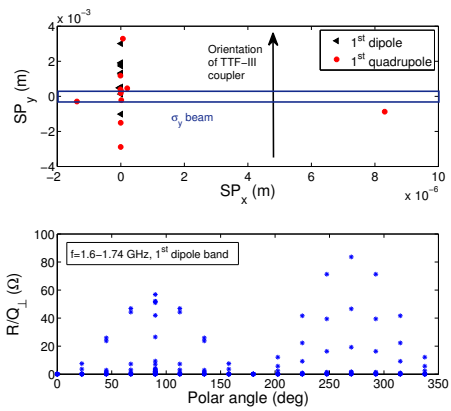


Figure 7: The upper plot shows the shift of the dipole and quadrupole modes of the first bands mainly in direction of the FPC. The blue square denotes the transverse beam size. The lower plot depicts the $\frac{R}{Q}_{\perp}$s of the first dipole band versus polar angle at a distance of 5mm from the beam axis.

more detail and it has to be understood how much the coupler's position influences this effective dipole component of the quadrupole modes.

OUTLOOK

The calculations using 1.5-3M mesh cells are too time consuming and cumbersome to become the core of an optimizer routine for tuning the end-cells to lower the Q_{ext} for the dangerous HOMs. In the next steps the authors will combine Coupled S-parameter Calculations [8] of the segmented structure with 2-D eigenmode calculations of the rotationally symmetric cavity cells to create a set of field-flat cavity designs and run an optimizer loop for minimizing $R/Q_{\perp} \cdot Q_{ext}$ [4]. Once a sufficiently performing design is found, these calculations will be extended to the full three cavity string, as already the lowest order dipole modes propagate along the beam tube and thus also couples to the waveguides of the next cavity.

Also a more detailed investigation of the FPC's influence of the HOM's center shift is planned by changing the coupler's position. As a side-effect the coupler kick of the fundamental mode will be calculated for different Q_{ext}.

Recently the bead pull and RF measurement of an aluminium model of the base cell design was started at TU Dortmund. After fabrication of the end groups, this model may be used to validate some of the findings of the calculations presented within this paper.

ACKNOWLEDGMENT

The authors acknowledge the help, advise and fruitful discussions with F. Marhauser formerly at JLab, now at μons Inc.

REFERENCES

[1] M. Abo-Bakr et al., "B*ERL*inPro - A Compact Demonstrator ERL for High Current and Low Emittance Beams, " LINAC'10, Tsukuba, September 2010, http://www.JACoW.org.

[2] Y. Petenev et al., "Modeling of the Beam Break Up Instability for B*ERL*inPro ," IPAC'11, San-Sebstian, September 2011, MOPS051, p. 718, http://www.JACoW.org

[3] B. Riemann et al., "First Considerations Concerning an Optimized Cavity Design for the Main Linac of B*ERL*inPro ," IPAC'11, San-Sebstian, September 2011, THPC109, p. 3143, http://www.JACoW.org

[4] T. Galek et al., "Traveling Poles Elimination Scheme and Calculations of External Quality Factors of HOMs in SC Cavities," WEP07, these proceedings.

[5] N. Valles and M. Liepe, "Seven-cell Cavity Optimization for Cornell's Energy Recovery Linac," SRF'09, Berlin, September 2009, THPPO008, p. 538, http://www.JACoW.org

[6] F. Marhauser, "JLab High Current Cryomodule Development," ERL'09, Ithaca NY, September 2009, WG307, http://www.JACoW.org

[7] Computer Simulation Technology AG, Microwave Studio®, 64289 Darmstadt, Germany http://www.cst.com

[8] K. Rothemund, H.-W. Glock, U. van Rienen, "Eigenmode calculation of complex rf-structures using S-parameters," (2000) IEEE Transactions on Magnetics, 36 (4 I), pp. 1501-1503.

ASTRA BASED SWARM OPTIMIZATIONS OF THE B*ERL*inPro INJECTOR

M. Abo-Bakr*, B. Kuske, Helmholtz-Zentrum-Berlin, Germany

Abstract

The Berlin Energy Recovery Linac Project B*ERL*inPro is a compact ERL to develop the accelerator physics and technology required to generate and accelerate a 100-mA, 1-mm mrad normalized emittance beam. One of the project challenges is to generate a beam of this kind in the injector line of the machine. Extensive injector optimization studies have been done over the last years. A deep insight in the physics of high brilliance, low energy beams together with single parameter scans allowed for an efficient optimization, resulting in a layout, capable to deliver bunches of the needed charge and dimension. However, changes in the injector components' technical layout, as they are unavoidable in the current stage of the project, may require re-optimizations at any time, if necessary of the whole injector part. To support these work an ASTRA based 'swarm optimization' tool for massive parallel calculations on the institute's Linux computing cluster has been developed. Once the optimization wrapper code is written, results come for free and can help to extend the understanding of the underlying physics. Strategy, procedure and results of the 'swarm optimizations' will be presented in this paper.

INTRODUCTION

B*ERL*inPro [1] is a Energy Recovery Linac (ERL) Project of the "Helmholtz-Zentrum Berlin für Materialien und Energie", funded in 2011. ERLs combine the advantages of linear accelerators (linac) and storage rings: since in principle like linacs, the excellent beam properties of photo-injector electron sources, as they became available in the last years, can be used in an ERL in contrast to storage rings, where the beam parameters arise from an equilibrium state of excitation and damping processes. In addition adiabatic damping while acceleration as well as beam manipulation techniques (e.g. bunch compression) further improve the ERL's beam quality.

In contrast to storage rings, the complete beam energy is dumped in a linear accelerator, limiting the maximum average currents to small values. In ERLs the invested energy is recovered by re-passing the initially accelerating RF structures a second time on a decelerating phase, so that the beam energy restores the cavity fields. Average currents of cw operated ERLs, using super conducting RF technology, become thus comparable to those of storage rings.

B*ERL*inPro's primary function is to demonstrate a stable and reliable low emittance, high-current operation, proving the ERL to be suited for the variety of future applications, including 4^{th} generation X-ray and Compton sources, EUV lithography and nuclear physics. The main B*ERL*inPro parameter are listed in Table 1.

* Michael.Abo-Bakr@helmholtz-berlin.de

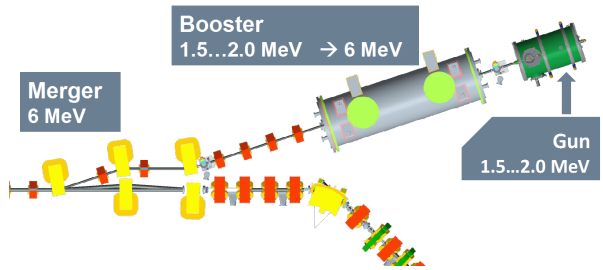


Figure 1: Layout of the B*ERL*inPro Injector, including the modules of the photo cathode gun and the booster linac as well as a dipole dogleg, merging the injector beam onto the main linac / recirculator beam path.

Table 1: Main B*ERL*inPro Parameters in "Standard Mode" Operation

Parameter	Value	Unit
Beam energy	50	MeV
Beam current @ 1.3 GHz	100	mA
Bunch charge	77	pC
Bunch length	~ 2	ps
Energy spread	0.5%	-
Emittance (norm.)	$\simeq 1$	mm mrad
Beam loss	$< 10^{-5}$	-

As the ERL bunch parameters a determined by those of the electron source the injector is one of an ERL's key aspects. To reach normalized emittances of about 1 π mm mrad and below not only a high performance gun is required but also a sophisticated injector setup for emittance conserving acceleration and transport of a strongly space charge (SC) dominated electron beam.

Injector beam line design aims for a trade-off in the bunch dimensions, minimizing the beam distortions due to SC effects, aberrations and RF-nonlinearities. In addition an effective emittance compensation scheme is mandatory.

Figure 1 shows a sketch of the B*ERL*inPro injector, as presented in the Conceptual Design Report (CDR) [2]. Main components are

- a 0.6 cell gun cavity, followed by a sc solenoid also integrated in the gun module, the photo cathode laser has a uniform transverse and a Gaussian longitudinal profile,
- a three 2-cell booster cavities (cavity and module are based on a Cornell design), due to limited RF power at high currents one of the cavities will be operated only at "zero crossing",
- a three bend, 18° dogleg, with quadrupole magnets, both in front of the merger to control the transverse beam size and inside the merger to control the dispersion.

A setup, satisfying the project requirements of $\varepsilon_{n,xy} < 1.0$ mm mrad, has been found as result of extensive injector studies. These are mostly done using ASTRA [3], a "Particle in Cell" code including SC effects. Numerous single parameter scans and experience based "trial and error" parameter settings are usually required to optimize the injector beam line. In the current project phase changes of parameter limits (e.g. max. field gradients), geometry or even of hardware (e.g. gun cavity design) are unavoidable. To support the time consuming injector setup work an ASTRA based automatic optimizer has been developed as an additional tool. The program is written to run on the HZB high-performance-computing cluster with 31 knots / 820 CPU-cores and a 64-bit OpenSuse Linux OS. The optimizer acts as a wrapper for ASTRA, which is used in its serial version but with many instances running at the same time.

THE OPTIMIZER PROGRAM

The program is written in Fortran and compiled with a freely available Intel compiler. Before running the program a set of reference ASTRA files must be created, where all parameters to be varied are replaced by variable names. The general beam line design is included in these files.

In the beginning of the program the parameters to be varied are read together with their start values and initial variation ranges. The parameters of the initial particle distribution as well as position and field strength of cavities (RF-fields), solenoids and quadrupoles can be varied. For RF fields of course also the phase variations are possible. Due to the implications for the geometry of the downstream beam line dipole magnets may not be varied at the moment. With randomly varied parameters the program generates for each run all required ASTRA files including the particle input distribution. It starts the according batch job, running the distribution generator and one ore more ASTRA runs. When the job has finished the optimizer collects the results, namely the target beam parameters like transverse and longitudinal bunch sizes, divergences and emittances but also energy and energy spread. From these values a weighted goal function is calculated. This function value is to be minimized by the program in an iterative approach: based on the best result found in the previous runs, parameters with small variations are generated for the next iteration. Convergence is reached by lowering the variation range every time no improvement has been found during a defined number of last runs. Since the program runs many ASTRA jobs in parallel a kind of "swarm" search is performed and accordingly we name this kind of optimization.

Special care had to be taken to correctly handle:

- variations in position: overlap of hardware has to be avoided. This is done by defining some reserved space up- and downstream of each elements.
- fixed distances of elements: e.g. for modules containing RF cavities the module position might be varied, while the distance between cavities is fixed. For those cases a flag can be set, freezing the distance to the previous element.
- ASTRA long runs: starting with large variation from a stable solution or even from an unstable one extremely long ASTRA runs may occur. Therefore the run time of each job is controlled and jobs, exceeding a defined maximum run time, will be canceled.
- jobs with two or more successive ASTRA runs: splitting a beam line is convenient, whenever a permanent change in the particles direction of motion occurs. With a pure motion in z, the geometry description in the input file significantly simplifies. To do so the output phase space must be rotated and correct job assignment of input files has to be guaranteed.

SWARM OPTIMIZATION OF THE B*ERL*inPro INJECTOR

Although very recently completed a variety of general program tests, mostly concerning the setup and interaction of ASTRA runs, its result analysis and the job control on the cluster have already been completed. Moreover first optimizations of a full beam line have been started for the B*ERL*inPro injector. Aim of these studies is to verify and - if possible - to improve the injector performance of the setup described in detail in the CDR.

Figure 2 shows the B*ERL*inPro injector elements, relevant for beam dynamics simulations and optimization. The following parameters were used within the swarm optimization: the laser spot size and pulse length on the cathode, the gun cavity's field and phase, the solenoid field strength, the booster cavity fields and phases (phase fixed to "zero crossing" for the first cavity) and the quadrupole magnets upstream, inside and downstream the merger. No parameters of the main linac cavities are varied, but since the SC driven plasma oscillation freezes only at sufficiently high energies the simulations need to be performed until the linac's end. Neither positions nor dimensions of elements have been used for optimization.

The goal function is basically calculated from the transverse projected emittances at the main linac end. An energy

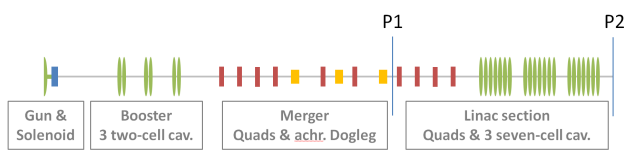


Figure 2: B*ERL*inPro Injector elements used for the optimization: gun, booster and main linac cavities (light green), the gun solenoid (dark green), the three merger bends (yellow) and ten quadrupoles (red). The beam line and thus the ASTRA run is split at P1, right after the last merger bend, where the beam has changed its direction.

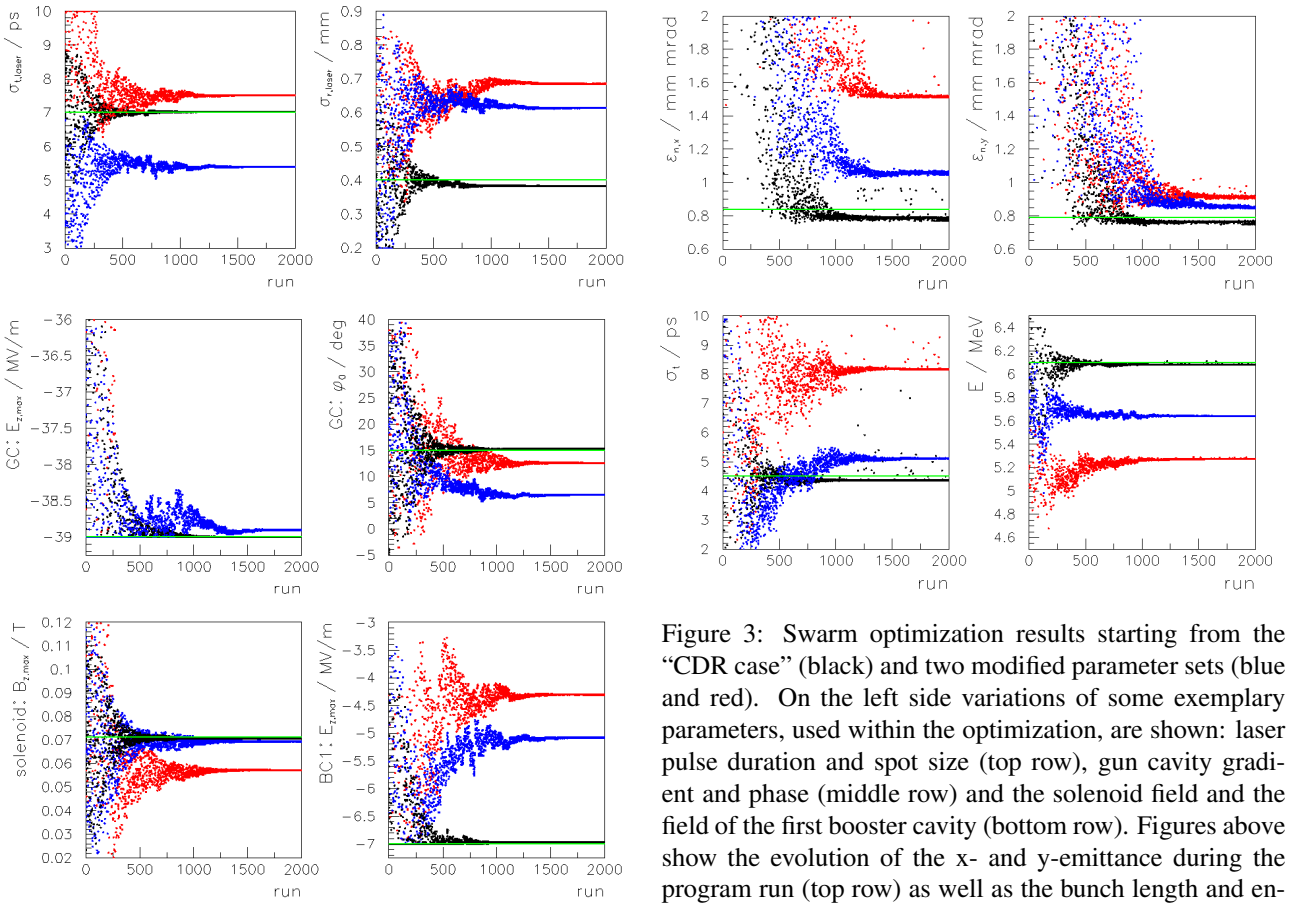


Figure 3: Swarm optimization results starting from the "CDR case" (black) and two modified parameter sets (blue and red). On the left side variations of some exemplary parameters, used within the optimization, are shown: laser pulse duration and spot size (top row), gun cavity gradient and phase (middle row) and the solenoid field and the field of the first booster cavity (bottom row). Figures above show the evolution of the x- and y-emittance during the program run (top row) as well as the bunch length and energy from booster (bottom row). The reference values from the "CDR case" are represented by the green line in each plot.

$\geq$ 6 MeV from the booster as well as a final bunch length $\leq$ 4.5 ps are also implemented in the goal function. A typical single ASTRA run with 2000 particles takes about 15 minutes. Using the 64 cores (max. number per user) the optimization is an "over-night" run of about 6-12 hours.

Starting with the "CDR case" the optimizer found a solution with both emittances reduced by a few percent, still satisfying the energy and bunch length requirements. Starting with modified parameters the optimizer does not reach the performance of the CDR case, neither in emittance nor in energy and bunch length. Results of selected optimizations are presented in Fig. 3. They indicate that the quality of the optimization result sensitively depends on the optimization time, the definition of the goal function but especially on the start setup together with the applied variation ranges. More tests of the program are needed to clarify the requirements of a successful optimization. To increase the number of jobs per time unit ASTRA run tests with less particles will be executed, maybe also more than 64 cores might become available. If this is successful the convergence causing reduction of each parameters variation range needs to be investigated in more detail. Maybe more sophisticated procedures like e.g. "simulated annealing" will be required.

CONCLUSION & OUTLOOK

Although not generally able to find the CDR solution, the optimizer proved its capability to improve from a given start setting. Even starting with the extensively optimized "CDR case" a setup with a slightly lower emittance could still be found. More advanced analysis tools of the results will be necessary to uncover correlations between parameters and performance or to evaluate distribution features like shape distortions or beam halo.

A next application will be a comparison of an 0.4 / 1.4 cell gun cavity based injector for B*ERL*inPro .

REFERENCES

[1] J. Knobloch et al., "Status of the B*ERL*inPro Energy Recovery Linac Project", IPAC 2012, New Orleans, July 2012, MOPPP015, p. 601, http://www.JACoW.org.

[2] "Conceptual Design Report B*ERL*inPro ", editors: A. Jankowiak, J.Knobloch, B. Kuske, N. Paulick, 2012.

[3] K. Flöttmann, ASTRA - A Space Charge Tracking Algorithm, Version 3.0, http://www.desy.de/~mpyflo/astra.

HYBRID PROGRAMMING AND PERFORMANCE FOR BEAM PROPAGATION MODELING

Misun Min*, Jing Fu, Azamat Mametjanov
Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, USA

Abstract

We examined hybrid parallel infrastructures in order to ensure performance and scalability for beam propagation modeling as we move toward extreme-scale systems. Using an MPI programming interface for parallel algorithms, we expanded the capability of our existing electromagnetic solver to a hybrid (MPI/shared-memory) model that can potentially use the computer resources on future-generation computing architecture more efficiently. As a preliminary step, we discuss a hybrid MPI/OpenMP model and demonstrate performance and analysis on the leadership-class computing systems such as the IBM BG/P, BG/Q, and Cray XK6. Our hybrid MPI/OpenMP model achieves speedup when the computation amounts are large enough to compensate the OMP threading overhead.

INTRODUCTION

Multicore architectures are current trends for gaining improvement in computing power, instead of increased clock speed. To achieve scalable performance with minimal parallelization overhead on such platforms, we have explored incorporating multithreading frameworks into our existing MPI-only code NekCEM for beam propagation modeling. Specifically, we are using hierarchical parallelization frameworks based on MPI/OpenMP schemes for intranode operations of MPI programs using OpenMP directives around time-consuming loops that do not contain data dependencies, while leaving the source code unchanged.

NekCEM [1, 2, 3] is a freely available, massively parallel, scalable high-order code for electromagnetic device simulations. NekCEM has great potential for meeting the future computational needs of experimental and theoretical research at exascale, by using a fast communication kernel for efficiency and body-fitted hexahedral meshes that allow significant gains in accuracy. We previously conducted wakefield calculations using the spectral-element discontinuous Galerkin (SEDG) scheme [4, 5] with fourth-order Runge-Kutta time stepping, with favorable results in comparison with those from low-order methods.

Future-generation supercomputing systems will be memory-limited relative to the raw computational performance. Currently, per processor memory requirements for NekCEM scale roughly as 600 8-byte words per allocated gridpoint. The total memory requirements are $n = EN^3$ points x 600 (words/point) x 8 (bytes/word). For example, with E=800K and N=16, the total memory requirements are $800K \times 16^3 \times 600 \times 8$. Assuming that 200 MB of memory per core are available to the user, one can run this simulation with P>786,432 cores by setting the number of local elements more than 10. In the current parallel context, however, there will be an increasing memory penalty associated with two variables, the maximum number of cores and the upper bound on the total number of elements, as the problem size becomes very large at extreme scale. A hybrid MPI/share-memory framework can reduce the memory dependency on these two parameters.

* mmin@mcs.anl.gov

IMPLEMENTATION

Ultrarelativistic beam propagations are governed by Maxwell's equations,

$$\mathbf{Q}\frac{\partial \mathbf{q}}{\partial t} + \nabla \cdot \mathbf{F}(\mathbf{q}) = \mathbf{S}, \tag{1}$$

where we define the field vector $\mathbf{q} = [\mathbf{H}, \mathbf{E}]^T$ and the flux $\mathbf{F}(\mathbf{q}) = [\mathbf{F_H}, \mathbf{F_E}]^T$ with $\mathbf{F_H} = \mathbf{e_i} \times \mathbf{E}$ and $\mathbf{F_E} = -\mathbf{e_i} \times \mathbf{H}$, and the source term $\mathbf{S} = [\mathbf{0}, \mathbf{J}]^T$. The electric, magnetic, and current fields are represented by $\mathbf{E} = (E_x, E_y, E_z)^T$, $\mathbf{H} = (H_x, H_y, H_z)^T$, and $\mathbf{J} = (0, 0, J_z)^T$, respectively. The material properties are defined as $\mathbf{Q} = \text{diag}(\mu, \mu, \mu, \epsilon, \epsilon, \epsilon)$ with the free space permittivity ϵ and free space permeability μ. Initial fields in the presence of the Gaussian beam are obtained by solving the Poisson equation in transverse direction at the beam location in the longitudinal direction.

Numerical Approach

We consider the computational domain Ω with nonoverlapping hexahedral elements Ω^e such that $\Omega = \cup_{e=1}^{E} \Omega^e$, and we define a weak formulation, introducing the numerical flux $\mathbf{F}^*$ as in [4, 5]:

$$\left(\mathbf{Q}\frac{\partial \mathbf{q}}{\partial t} + \nabla \cdot \mathbf{F}(\mathbf{q}) - \mathbf{S}, \phi\right)_{\Omega^e} = (\hat{n} \cdot [\mathbf{F} - \mathbf{F}^*], \phi)_{\partial\Omega^e}. \tag{2}$$

The local solutions of the fields can be written as

$$q^N(\mathbf{x}, t) = \sum_{i,j,k=1}^{N} q_{ijk}\psi_{ijk}(\mathbf{x}), \tag{3}$$

where q_{ijk} is the solution at $\mathbf{x}$=(x_i, y_j, z_k) on Ω^e and ψ_{ijk}=$l_i(r)l_j(s)l)l_k(t)$ using the one-dimensional Legendre Lagrange interpolation polynomial l_i of degree N-1 associated with the N Gauss-Lobatto-Legendre (GLL) quadra-

ture nodes [6]. Plugging (3) into (2) with a local discontinuous test function $\phi = \psi_{ijk}$ and applying the GLL quadrature for the spatial integration, we obtain a semi-discrete scheme with the mass and stiffness matrices defined as

$$\mathbf{M} = (\psi_{ijk}, \psi_{\hat{i}\hat{j}\hat{k}})_{\Omega^e}, \mathbf{D}_x = \left(\frac{\partial \psi_{ijk}}{\partial x}, \psi_{\hat{i}\hat{j}\hat{k}}\right)_{\Omega^e}, \quad (4)$$

$$\mathbf{D}_y = \left(\frac{\partial \psi_{ijk}}{\partial y}, \psi_{\hat{i}\hat{j}\hat{k}}\right)_{\Omega^e}, \mathbf{D}_z = \left(\frac{\partial \psi_{ijk}}{\partial z}, \psi_{\hat{i}\hat{j}\hat{k}}\right)_{\Omega^e} \quad (5)$$

and the surface integration on the right-hand side in (2) (we omit the detailed form in this paper).

Algorithms

The algorithm can be simplified as

$$U^{n+1} = U^n + \texttt{mxm}(U^n) + \texttt{comm}(F(U^n)), \quad (6)$$

where `mxm` and `comm` respectively represent the matrix-matrix product and communication operation for the surface integration, $F(U^n)$. The main operation count in our algorithm is dominated by the `mxm` routine: the curl operator takes 30% of total cost and 15% of total cost for communication. We focus on speedup by additional threading on the `mxm` routines.

Within each subdomain, derivatives are evaluated in a tensor product fashion by using a one-dimensional differentiation matrix on a reference domain $[-1, 1]^3$. A derivative of u^e ($e = 1, ..., E$) in a subdomain with respect to $(r, s, t) \in [-1, 1]^3$ is expressed by

$$\mathrm{u}_r^e \equiv D_r u^e = (I_t \otimes I_s \otimes \hat{D}) u^e = \Sigma_{l=1}^N \hat{D}_{il} u_{ljk}^e, \quad (7)$$

$$\mathrm{u}_s^e \equiv D_s u^e = (I_t \otimes \hat{D} \otimes I_r) u^e = \Sigma_{l=1}^N \hat{D}_{jl} u_{ilk}^e, \quad (8)$$

$$\mathrm{u}_t^e \equiv D_t u^e = (\hat{D} \otimes I_s \otimes I_r) u^e = \Sigma_{l=1}^N \hat{D}_{kl} u_{ijl}^e, \quad (9)$$

where I_t, I_s, and I_r are the $N \times N$ identity matrices and $\hat{D}$ is the one-dimensional differentiation matrix of $N \times N$ defined in [6].

D_x, D_y and D_z in Eq. (4) require three `mxm` operations for each. Thus the cost per timestep involving the `mxm` operations for a curl operator scales as $O(18EN^4)$ for six field components. For the MPI-only model, we use the `mxm` routine, written in Fortran and assembly code, which includes the inner-product dimension completely unrolled into a single statement, allowing a short-nested loop for more work per iteration and a hardcoded address increments into memory read instructions by the compiler [6]. We consider the OMP routine as an alternative for possible speedup.

OpenMP is a set of APIs for writing multithreaded programs on shared-memory machines. It can help the compiler parallize applications at the highest possible level through explicit compiler directives yet not involve application programmers in low-level details.

For our hybrid MPI/OpenMP design, we use the following instruction for the `mxm` routine:

```
c$OMP PARALLEL DEFAULT(PRIVATE)
      SHARED(A,B,C,N1,N2,N3)
c$OMP DO
      do j=1,N3
      do i=1,N1
         c(i,j) = 0
      do k=1,N2
         c(i,j) = c(i,j) + a(i,k)*b(k,j)
      enddo
      enddo
      enddo
c$OMP END DO
c$OMP END PARALLEL
```

PERFORMANCE

We performed our tests on the IBM BG/P and BG/Q and the Cray XK6 for different problem sizes with varying N on a hexahedral mesh with E=1152 for an undulator. The features of the systems are described below.

IBM BG/P The Blue Gene/P Intrepid consists of 40,960 compute nodes (40 racks and 1,024 nodes per rack, including 640 I/O nodes) with 850 MHz quad-core processor and 2 GB RAM per node, for a total of 163,840 cores, 80 TB of RAM, and a theoretical peak performance of 557 teraflops.

IBM BG/Q The Blue Gene/Q Mira consists of 49,152 compute nodes (48 racks and 1,024 nodes per rack, including 384 I/O nodes) with 1.6 GHz 16-core processor and 16 GB RAM per node, for a total of 786,432 cores, 786 TB of RAM, and a theoretical peak performance of 10 petaflops.

Cray XK6 The Cray XK6 Jaguar consists of 18,688 compute nodes. Each compute node consists of 16-core 2.2 GHz AMD Opteron processors and 32 GB of RAM, for a total of 299,008 cores, 598 TB of RAM, and a theoretical peak performance 2.63 petaflops.

We measured the CPU time and wallclock time for 100 timestep runs. For the CPU time measure, we used `dclock` and got an average time over the total number of MPI ranks. For the wallclock time, we used `mpi_wtime` and `omp_get_wtime` for MPI ranks and OMP threads, respectively. Here we demonstrate the CPU time using the average value over all MPI ranks.

Figure 1 shows the CPU time for a fixed number of MPI ranks (=1024) with an increasing number of OMP threads for the granularity per core n/P=576 and 4608. In the top figure, adding additional threads (i.e., more computing resources) decreases the performance because the overhead of creating threads offsets the benefit of parallelizing the loops for such coarse granularity in computation. As the work amount increases in the bottom figure, we observe more OMP threads for speedup. Compared with the MPI-only case, however, we do not gain much speedup.

Figure 2 demonstrates the CPU time for a fixed number of total threads (=1024) with varying numbers of MPI

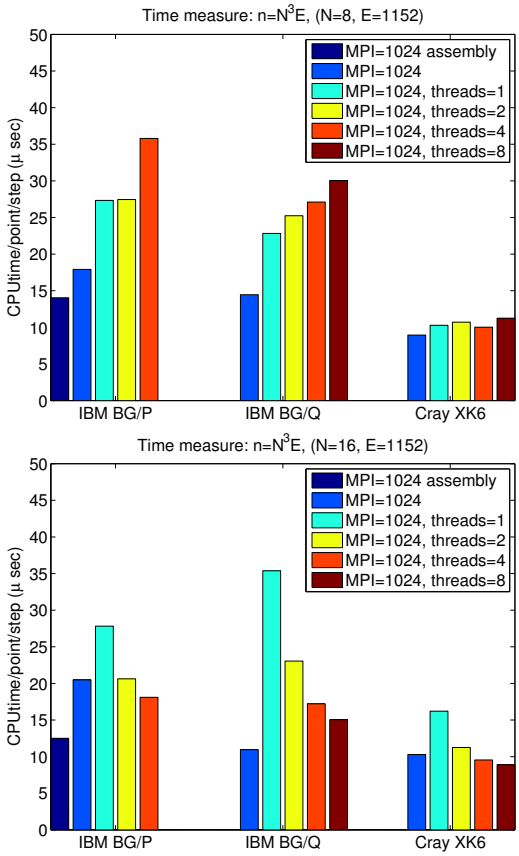


Figure 1: Scaling on MPI/multithreading algorithms on BG/P, BG/Q, and Cray XK6.

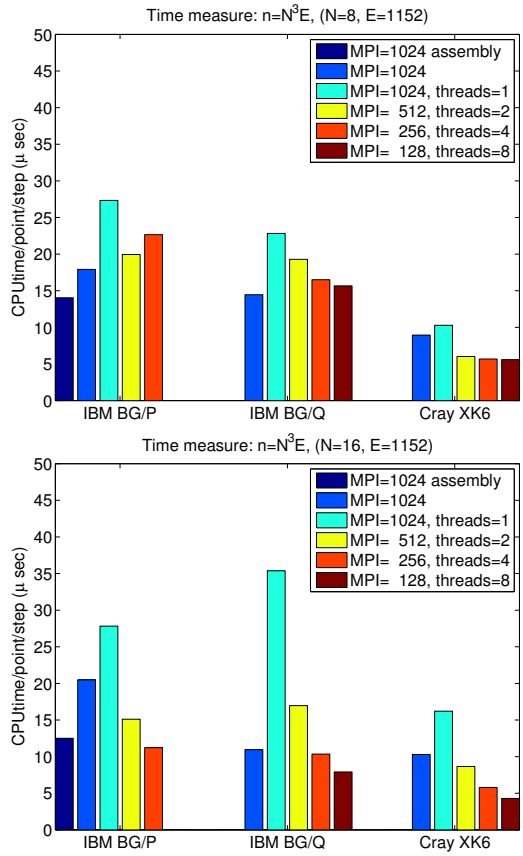


Figure 2: Scaling on MPI/multithreading algorithms on BG/P, BG/Q, and Cray XK6.

ranks and OMP threads. The granularity per core changes as n/P=576, 1152, 2304, 4608 with P =1024, 512, 256, 128, respectively, for n=0.58M (top) and n/P=4608, 9216, 18432, 36864 with P =1024, 512, 256, 128, respectively, for n=4.7M (bottom). By increasing the number of OMP threads but reducing communication with a smaller number of MPI ranks, the results show increased speedup consistently. However, superior performance compared with the MPI-only case is possible when there is enough work for the OMP loops, as shown in the bottom figure.

CONCLUSION

We have conducted performance studies on leadership-class computing systems and have demonstrated the speedup for a hybrid MPI/OpenMP model in comparison with the MPI_only case when the computation amounts are large enough to hide the OMP threading overhead. Future work includes expansion to MPI/GPU threading and comparison with further optimized MPI/OpenMP model at larger scale.

ACKNOWLEDGMENT

This work was supported by the Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy, under Contract DE-AC02-06CH11357.

REFERENCES

[1] "NekCEM," http://wiki.mcs.anl.gov/nekcem.

[2] M. S. Min, J. Fu, P. F. Fischer, "Performance analysis of the spectral-element discontinuous Galerkin method for electromagnetic modeling on the IBM BG/P and Cray XK6," Preprint ANL/MCS-P1802-0111.

[3] J. Fu, M. S. Min, R. Latham, C. D. Carothers, "I/O threads to reduce checkpoint blocking for an electromagnetics solver on Blue Gene/P and Cray XK6," in International Workshop on Runtime and Operating Systems for Supercomputers (ROSS), in conjunction with International Conference on Supercomputing (ICS), June 2012.

[4] M. S. Min, P. F. Fischer, "Spectral-element discontinuous Galerkin simulations with a moving window algorithm for wakefield calculations," in Proc. of PAC09, TH5PFP03, 2009.

[5] M. S. Min, P. F. Fischer, Y. C. Chae, "Wake fields for TESLA cavity structures: Spectral element discontinuous Galerkin simulations," in Proc. of SRF07, TUP34, 2007.

[6] M. O. Deville, P. F. Fischer, E. H. Mund, *High Order Methods for Incompressible Fluid Flow,* Cambridge University Press (2002).

COMPARISON OF EIGENVALUE SOLVERS FOR LARGE SPARSE MATRIX PENCILS *

F. Yaman**, W. Ackermann , T. Weiland, Technische Universität Darmstadt,
Institut für Theorie Elektromagnetischer Felder (TEMF),
Schlossgartenstraße 8, 64289 Darmstadt, Germany

Abstract

Efficient and accurate computation of eigenvalues and eigenvectors is of fundamental importance in the accelerator physics community. Moreover, the eigensystem analysis is generally used for the identifications of many physical phenomena connected to vibrations. Therefore, various types of algorithms such that Arnoldi, Lanczos, Krylov-Schur, Jacobi-Davidson etc. were implemented to solve the eigenvalue problem efficiently. In this direction, we investigate the performance of selected commercial and freely available software tools for the solution of a generalized eigenvalue problem. We choose characteristic setups by considering spherical and billiard resonators next to a TESLA nine-cell cavity in order to test the robustness, accuracy, and computational speed as well as memory consumption issues of the recent versions of CST, Matlab, Pysparse, SLEPc and CEM3D. Simulations were performed on a standard personal computer and on a cluster computer to enable the handling of large sparse matrices in the order of hundreds of thousands up to several millions degrees of freedom (DOF). We obtain interesting comparison results with the examined solvers which is useful for choosing the appropriate solvers for a given practical application.

INTRODUCTION

In this paper, we consider the numerical solution of the generalized eigenvalue problem

$$A\,x = \lambda\,B\,x \quad \text{for} \quad B > 0 \tag{1}$$

for a given symmetric, real valued, large and diagonally dominant sparse matrix pencils A and B. The problem can be derived from Maxwell's equations for a source free bounded domain having perfectly conducting boundary condition on its surface. In principle, such a mathematical model coincides with a superconducting cavity which enclosures excited electromagnetic fields. Therefore, accurate calculation of eigenvalues λ has a critical importance and it leads to find the eigenfrequencies of several modes which are of fundamental importance for the acceleration of charged particles. In this context, performance of eigensolvers are always of interest in industry and scientific communities which constitutes the main motivation of our study.

Jacobi-Davidson method [1] which is based on iterative expansions of subspaces is recommended to solve eigenvalue problems dealing with diagonally dominant large matrices in the literature [2, 3, 4]. Therefore, we select CEM3D, Pysparse and SLEPc solvers which all can employ Jacobi-Davidson implementation in their simulations. Here, we just mention the idea behind of the method shortly. Let $V_k = \text{span}\{v_1, \cdots, v_k\}$ be a subspace where $v_k^T B v_j = \delta_{kj}$. Then a Ritz pair $(\theta_j,\ u_j)$ in V_k, can be obtained from the projected eigenvalue problem choosing the closest value to a given target τ. In the next step, a convergence criteria

$$\| r \|_2 := \| (A - \theta B)u \|_2 < \epsilon\,. \tag{2}$$

is checked for θ if it could be an eigenvalue for a selected ϵ. Afterwards, the so-called correction equation

$$(I - Buu^T)(A - \theta B)(I - uu^T B)z = -r \tag{3}$$

is solved iteratively with *tfqmrs* in CEM3D, with *bcgsl* in SLEPc and with *qmrs* in Pysparse for the unknown z where $z \perp u$ and u^T is the transpose of u. Then we expand the subspace $V_{k+1} = \text{span}\{v_1, \cdots, v_{k+1}\}$ for the next iteration which starts finding approximate Ritz pairs for the updated subspace [5].

NUMERICAL EXPERIMENTS

In the numerical tests, spherical and billiard resonators and a 9-cell TESLA cavity are considered. Structures are drawn and meshed with tetrahedrons having planar elements. Corresponding meshes are imported to CEM3D [6] in order to generate input sparse matrices for the eigenvalue solvers Pysparse [4], SLEPc [7] and Matlab [8]. Here, CEM3D is a parallel and a higher order FEM code which was implemented in our institute for the accurate calculations of eigenfrequencies for a given structure. Furthermore the same structures are used for the eigenvalue simulations in CST [9]. It is also noted that among these solvers only CEM3D and SLEPc have the capability to run on a distributed memory machine with multiprocessors in parallel.

For the comparison of different solver results, we first compute the eigenvalues of a spherical resonator from analytical expressions given in [10] by employing a root-finding algorithm which is simply explained in [11]. Eigenvalues are also calculated with SLEPc for different DOF and a relative error is calculated as,

$$\text{relative error} = \max_{i \in \text{DOF}} \frac{|\,\lambda^{\text{analytical}} - \lambda_i^{\text{numerical}}\,|}{\lambda^{\text{analytical}}} \tag{4}$$

by considering the worst computed degenerated mode eigenfrequency, see Figure 1.

* Work supported by the DFG through SFB 634
**yaman@temf.tu-darmstadt.de

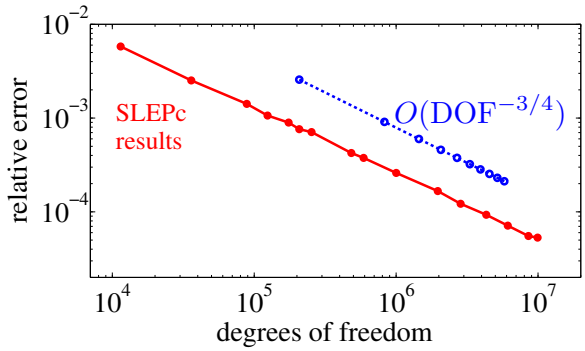


Figure 1: Convergence rate of SLEPc eigencomputations.

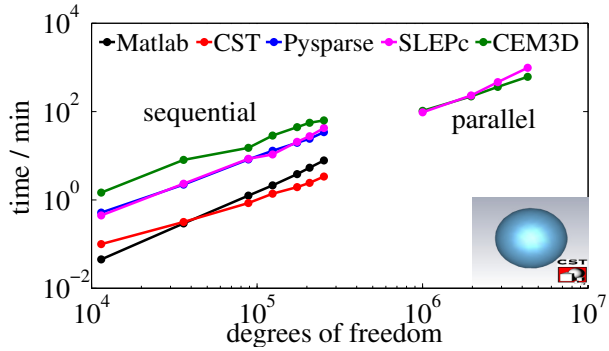


Figure 2: Time Consumption for a Spherical Resonator.

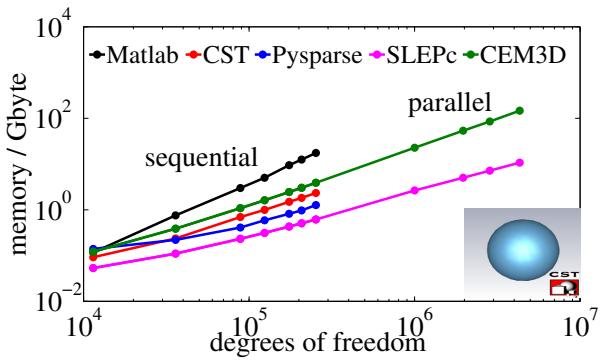


Figure 3: Memory Consumption for a Spherical Resonator.

In our simulation studies, we experienced that it is sufficient to use a single personal computer e.g. 2.27 GHz (2 processors) having 24 GB memory to compute 20 eigenvalues with the same accuracy for all solvers up to 10^5 degrees of freedoms but for the larger amount of DOF it is necessary to run solvers in parallel. Along this line results presented in this paper for approximately DOF $\geq 2.5 \times 10^5$ are obtained using parallel run on a cluster with SLEPc / CEM3D and this issue is indicated in time consumption illustrations in Figures 2, 4, 6 with two seperated regions by disconnected line plots.

In all simulations, we first mesh related geometries in CST with tetrahedrons and extract first 20 eigenvalues of the spectrum choosing accuracy 10^{-9} for different DOF with different solvers. Basically, structures have different eigenvalue spectrum's depending on their geometries and this phenomena may effect the performance of the solvers. Therefore we also test solvers with a chaotic billiard resonator which has a clustered eigenvalue distribution [12].

Spherical Resonator

As a first experiment, we consider a spherical resonator with a radius 1m. We run several simulations to obtain time and memory consumption results which are shown in Figure 2 and Figure 3, respectively. It can be concluded from the simulation results that iterative solvers needs long time but less memory as to be expected. Therefore, CST or Matlab can be recommended for the applications having DOF up to 10^5 in the case of providing enough computation sources. However, it should be noted that Matlab requires more memory as compared to the other solvers, so that SLEPc and Pysparse can also be considered a very good option in this DOF region. Especially, SLEPc has a very good characteristic from memory usage point of view for the very large DOF. On the other hand, a desired portion of the eigenvalues in the spectrum or the electromagnetic field distribution can be computed with CEM3D or CST straightforwardly. It should be noted that these tools are robust in nature even to unproperly chosen target values.

Nine-cell TESLA Cavity

For the second experiment, a 9-cell TESLA cavity with approximately 1m length is chosen. Here, Matlab and CST are clearly separated from the other solvers in time consuming, see Figure 4. On the other hand, as an important difference from the previous experiment now except SLEPc, all solvers memory consumption plots are more close to each other which actually emphasizes Matlab's good perfomance for this case. Furthermore, a linear $O(\mathrm{DOF})$ behavior is observed for DOF $\geq 10^6$ in the memory consumption of SLEPc, see Figure 5.

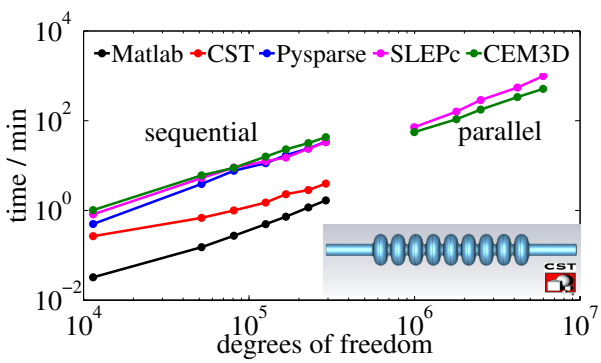


Figure 4: Time Consumption for a Tesla Cavity.

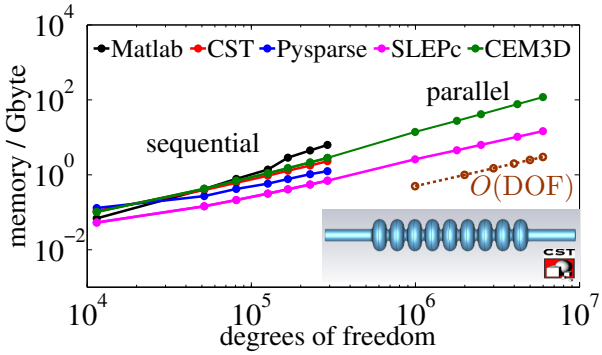


Figure 5: Memory Consumption for a Tesla Cavity.

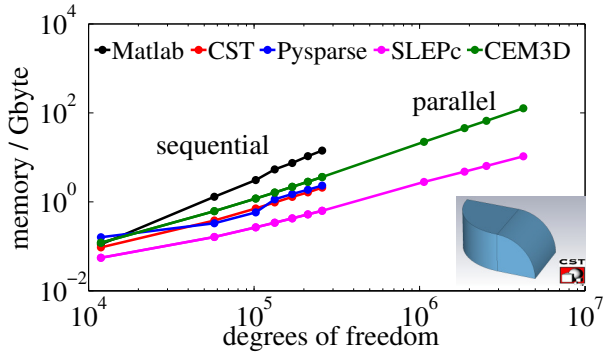


Figure 7: Memory Consumption for a Billiard Resonator.

Billiard Resonator

In the billiard resonator experiment, it is initially observed that the time and memory consumption behaviors of solvers for billiard resonator is more close to spherical resonator than a TESLA cavity. This may be a consequence of the similarity of the shapes. On the other hand, it should be emphasized that the nature of the solvers change only slightly even for a very different eigenvalue spectrum. It is shown in Figure 6 that time consumption of Matlab and CST are very similar in a region $10^5 \leq \text{DOF} \leq 2.5 \times 10^5$. However, curve of best fit of Matlab is more steep than CST for larger DOF. Moreover Pysparse and SLEPc have a similar time consumption characteristics but as it is similar in the previous simulations SLEPc requires the least memory source for billiard resonator calculations.

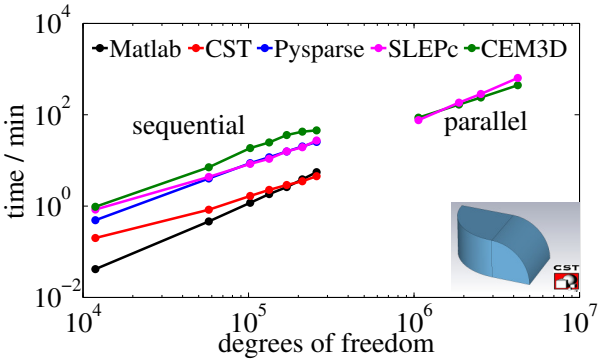


Figure 6: Time Consumption for a Billiard Resonator.

CONCLUSIONS

As a result of all experiments, we immediately conclude that solvers change their behavior only slightly depending on the geometry of the problem or the type of the eigenvalue spectrum. From time consuming point of view, Matlab and CST can be considered as one group and SLEPc, Pysparse, CEM3D can be considered as a second group according to their similar values. The flexibility of choosing target value in CST and CEM3D increases the applicability of these solvers. In the case of knowing the target value for single personal computer applications Matlab has an impressive performance in time. Free tools Pysparse and SLEPc has similar characteristics in time consuming but SLEPc uses the least memory in all three experiments and it can be used either on a single computer or on a cluster for very large amount of DOF.

REFERENCES

[1] G. L. G. Sleijpen and H. A. van der Vorst, " A JacobiDavidson iteration method for linear eigenvalue problems," SIAM J. Matrix Anal. Appl., 17 (1995) 401425.

[2] P. Arbenz and R. Geus, "A comparison of solvers for large scale eigenvalue problems occuring in the design of resonant cavities," Numer. Linear Algebra Appl. 6 (1999) 3-16.

[3] E. Romero and J.E. Roman, "A parallel implementation of Davidson methods for large-scale eigenvalue problems in SLEPc," ACM Trans. on Math. Software, 11 (2012) 1.

[4] R. Geus, "The Jacobi-Davidson algorithm for solving large sparse symmetric eigenvalue problems with application to the design of accelerator cavities, " ETH Zurich, PhD Thesis, 2002.

[5] Peter Arbenz, Urban Weber, Ratko Veprek, Bernd Witzigmann, "A JacobiDavidson Algorithm for Large Eigenvalue Problems from Opto-Electronics," Recent Adv. in Num. Meth. for Eigenvalue Problems (RANMEP2008), talk.

[6] W. Ackermann and T. Weiland, "High Precision Cavity Simulations," , MOADI1, these proceedings.

[7] V. Hernandez, J. E. Roman and V.Vidal, "SLEPc: A Scalable and Flexible Toolkit for the Solution of Eigenvalue Problems," ACM Trans. on Math. Software 31 (2005) 3.

[8] MATLAB R2011a, The MathWorks Inc., Natick, MA, 2011.

[9] CST AG, CST 2012, Darmstadt, Germany. MICROWAVE STUDIO

[10] S. Gallagher and W.J. Gallagher, "The Spherical Resonator," IEEE Trans. on Nuclear Sci. Vol NS-32 5 (1985).

[11] David M. Pozar, "Microwave Engineering, " 2nd ed, John Wiley & Sons, Inc., 1998.

[12] C. Dembowski, B. Dietz, H.-D. Graf, A. Heine, T. Papenbrock, A. Richter and C. Richter, "Experimental Test of a Trace Formula for a Chaotic Three-Dimensional Microwave Cavity ", Physical Review Letters, Vol 89 (2002) 6.

PIC SIMULATIONS OF LASER ION ACCELERATION VIA TNSA

Z. Lecz, O. Boine-Frankenheim, TU Darmstadt and GSI Darmstadt, Germany
V. Kornilov, GSI, Darmstadt, Germany

INTRODUCTION

The acceleration of ions with lasers up to energies of 60 MeV has been successfully demonstrated at different laser systems worldwide, including our facility: PHELIX [1]. The undergoing mechanism is understood as Target Normal Sheath Acceleration (TNSA) [2]. Due to the small transverse emittance and low installation cost laser-ion acceleration is a promising alternative to RF accelerators with the possible application in ion cancer therapy [3]. This contribution is devoted to the numerical investigation of the proton acceleration via the TNSA mechanism using 1D and 2D particle-in-cell electro-magnetic simulations. We employ the plasma simulation code VORPAL [4] and focus on the proton and electron phase-space distribution at the rear side of the target. The lack of knowledge about the thickness of hydrogen-rich contamination layer requires a detailed parameter study, which has been done in 1D. In this work we investigate the expansion and divergence of a thick proton layer from the rear side of the target using 2D simulations. In order to make sure that the results are physically correct and reliable we performed a detailed convergence study over the grid size and macro-particle charge.

NUMERICAL HEATING IN 2D PIC SIMULATIONS

As we know, in 1D it is really easy to resolve the physical scale lengths (Debye-length) and we can use many macro-particles per cell in order to get better statistics and smoother plasma density. With resolution high enough we obtain results comparable to the fluid-hybrid codes, but the PIC code always gives us more data and realistic velocity phase-space distribution of particles. Therefore the PIC simulations are used to get insight into the smallest details of physical processes in plasma physiscs.

Very often in 2D we can not use the same high resolution. Because of the discretization of space and time we always have to be aware of the numerical errors which occur during field calculations and integration of equation of motion of particles. These errors can lead to increasing field amplitudes, consequently artificial particle acceleration. The main equations which are solved in the loop of a PIC code are the Faraday's and Ampere-Maxwell law:

$$\frac{\partial B}{\partial t} = -c\nabla \times E \tag{1}$$

$$\frac{\partial E}{\partial t} = c\nabla \times B - J \tag{2}$$

These equations are solved in the frequency domain by calculating the Fourier transform of the fields and then they are transormed back, because the equation of motion is integrated in the time domain. This additional step speeds up the calculations significantly. Due to the granulated representation of the density, the Fourier transform of J can contain very high frequencies, which can not be resolved by the spatial grid. The high frequency noise can be reduced by increasing the grid resolution or increasing the number of particle per cell (PPC). However these two options are very expensive in terms of computing power and CPU time. A much better method is the implementation of higher order particle shapes and field interpolation which will smooth out the fields and current density in the simulation. Thus the noise disappears and the simulation will be more stable with correct energy conservation, but it does not mean that result is closer to the reality or that it is more reliable [5]. Very often the high frequency modulation of the density has a physical reason and in this case we exclude a part of the physical processes by applying a smoothing technique.

Interaction of high intensity laser with over-dense plasma requires a very high number of grid-cells in order to resolve the very small cold electron Debye-length (λ_{Dc}). The usual length in both directions is a few tens of micrometers, while the Debye-length inside of the high density cold plasma is on the order of nanometers. In this case it is inevitable to use smoothing interpolations, which allows us the define grid cells larger than λ_{Dc}. In order to check the grid heating we performed several simulations with dense plasma (n_0=10^{27} m^{-3}) filling up completely the simulation box, which has periodic boundaries. The initial electron temperature was 1 keV. In Fig. 1 the simultaion results are shown for 4 different grid resolution. Two simulations were done with linear interpolation and particle shape, which results in a very strong heating of the electrons. This numerical effect is not present if we switch to the cubic (third order) smoothing even if the grid size is 40 times larger than λ_{Dc}. A weak heating can be observed only for very small PPC, brown line.

Another important effect which can increase the electron temperature is the numerical Cherenkov emission. It is related to the Courant criteria [6], which defines the time step for a stable simulation. Discretizing the space automatically leads to a dispersion relation which says that the phase velocity of electro-magnetic waves propagating on the grid is smaller than the speed of light. If we have high energy electrons in the system they can have velocity close to the speed of light which means that they can travel fatser than maximum phase velocity in the system. This situation results in the same effect what we observ in the reality: radiation into the medium. The emitted fields accelerate the electrons, thus their temperature is increasing. This effect

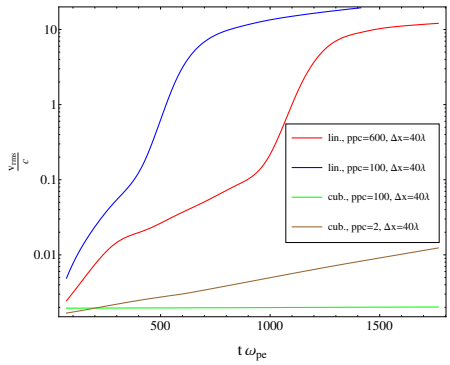


Figure 1: Normalized RMS momentum of elecrons for different simulation parameters. $\omega_{pe} = \sqrt{n_0 q_e^2/m_e\epsilon_0}$.

we can see in Fig. 2, where the black line shows that for this high resolution the grid heating is very small even with linear interpolation. The red line shows the result from the same simulation, but the electron temperature is relativistic and the blue line prooves again that the cubic interpolation suppresses the unvanted effect. In order to understand this we have to know that the Cherenkov emission is efficient for high wave numbers. For low frequencies it vanishes, that is why by eliminating the high frequency noise from the system we do not observe the heating.

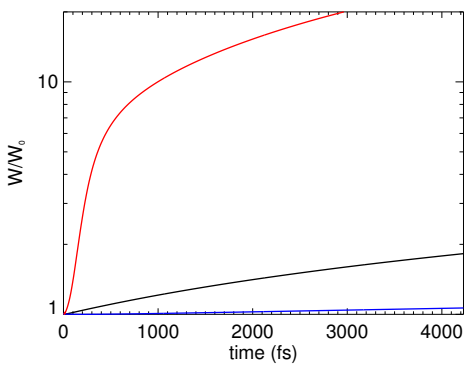


Figure 2: Total energy of the electrons over time in a simulation with $dx = 4\lambda_{Dc}$ and PPC=50. Legend: black - $T_h = 1$ keV, linear interp.; red - $T_h = 500$ keV, linear interp. and blue - $T_h = 500$ keV, cubic interp. ω_{pe}=1.78·10^{15} Hz.

SIMULATIONS INCLUDING THE LASER PULSE

In Fig. 3 the simulation setup is presented with specific boundary conditions. The simulation size is 50×50 μm and the target (red column) is at 7 μm from the left boundary. For this type of simulation the total energy of the system should be constant or slightly decreasing (due to the PML regions) after the laser pulse is off. It makes very easy to check if the grid heating is significant or not. The laser pulse duration was 200 fs and the peak intensity 6· 10^{23} W/m^2. The simulations were carried out with third order field interpolation and particle shape. The average CPU time of these simulations is 1 day and only 20 nodes were used.

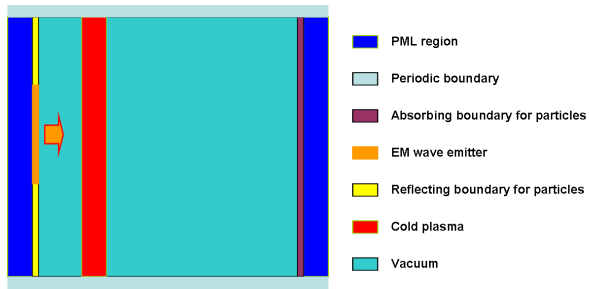


Figure 3: The simulation setup for a TNSA simulation.

The most important part of the laser ion acceleration via TNSA is the hot electron production. The electron heating up to relativistic temperatures happens on the fron side of the target and the laser skin depth (penetration depth, δ_E) becomes a crucial parameter. We found that the resulting hot electron temperature (T_h) is very sensitive on the grid resolution, because if δ_E is not well resolved then the energy absorption from the laser is not correct. Unfortunatelly this length is approximatelly 10 times λ_{Dc}, therefore the grid size should not be larger than this. It will reduce strongly our freedom in choosing the grid size. Choosing PPC is more related to the rear side of the target, where the plasma expansion happens. The simulations and an isothermal model [7] show that the expanding plasma has an exponential density profile in longitudinal direction and the density at the proton front decreases in time as $\sim 1/t^2$. If we want to resolve the density we have to use a very high PPC ($\approx$ 100 or more) which makes the simulation very slow. Anyhow we are interested in a very short time window: 200-300 fs, which is enough to understand and model the 2D expansion of a thick proton layer. In the case of the thin layer the grid size should be equal to the layer thickness, which can be smaller than λ_{Dc}. Therefore our 2D study has not been extended yet to this regime.

In order to check the effect of the grid resolution on the results we performed several simulations with different grid size and PPC. In Fig. 4 the results from simulations with the same PPC are shown. As we can see by changing the grid size the T_h is affected significantly and it does not converges yet. It is important to mention that the grid heating does not appear in any of the simulations, because the total energy is constant after the laser pulse, see fig. right. In Fig. 5 we keep the grid size constant and increase the PPC. Here a more clear convergence can be observed in the proton fron velocity and the T_h starts from the same value. An interesting feature is noticeable in the time evolution of energy: the remaining field energy (total minus particle energy) is larger if the PPC is smaller. The reason is that for large macro-charge the charge separation results

in a higher field at the proton front. With other words the energy goes into the electric field, it is not transfered to the protons.

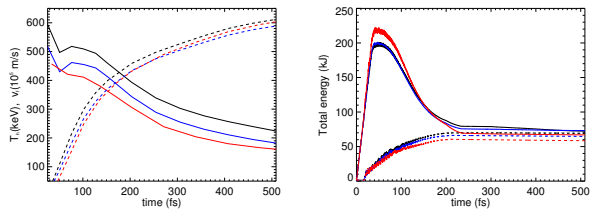


Figure 4: Left: Hot electron temperature (full line) and proton front velocity (dashed line). Right: Total energy in the system (full line) and total energy of all particles (dashed line). The color legend for both: $dx = 5\lambda_{Dc}$ (black), $dx = 10\lambda_{Dc}$ (blue) and $dx = 20\lambda_{Dc}$ (red). PPC=50.

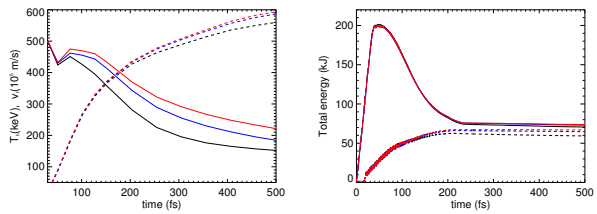


Figure 5: The same physiscal quantites as in Fig. 4. The color legend for both: PPC=20 (black), PPC=50 (blue) and PPC=100 (red). $dx = 10\lambda_{Dc}$.

After making sure that our simulations are correct we can study the physics of the two-dimensional plasma expansion. The used simulation parameters: $dx = 10\lambda_{Dc}$ and PPC=50. The only remaining issue is the hot electron temperature, which does not have its correct value at this resolution. This we do not consider a problem because by modelling the plasma expansion we take the hot electron parameters (measured temperature and density) from the simulation. Before we present the model we have to mention that in the next section non-periodic boundaries were used in the transverse direction. For the fields we used PML regions, just like in the longitudinal direction, and for the particles the so-called diffuse boundaries, which reflects back the electrons with a given thermal velocity distribution. For the thermal speed of the reflected electrons we take the initial one (cold). In this way we can mimic an infinite long transverse extension of the simulation box. With this trick we are closer to the reality, where the hot electrons spread out on the target surface and are replaced by cold electrons (return current), while the fields also disappear propagating in each direction.

ENVELOPE ANGLE OF THE PROTON BEAM

Our 2D simulations show that the hot electron density in transverse direction can be approximated with a Gaussian, which has a sigma (σ) comparable to sigma of the laser pulse and it increases slowly in time. Although the intensity profile is also Gaussian the hot electron temperature becomes isotrop in a few electron plasma periods, which is a very short time compared to the time scales of the expansion. Therefore we can assume a uniform temperature and just for sake of simpliciy we assume that it is contstant during the laser pulse. The peak value of the density (n_{h0}) is at the laser axis and it is measured from the simulation. One can calculate from the energy absorption, if it is known. We rely on the simulation and do not use other semi-empirical estimations to obtain the hot electron parameters. In the following x denotes the longitudinal and y the transverse direction.

After laying down the initial conditions and assumptions we can derive a transversal acceleration by using the results of P. Mora [7]. The model is based on the fact that electric field vector (E_y) which points into the transverse direction is proportional to the tangent of the surface angle which is determined by the proton front surface. As a simplification we calculate E_y just by multiplying E_x with the surface angle, because it is a small value and the electric field is always perpendicular to the surface. All we need to know is the time-dependent derivative of the proron front as a function of y. The starting equation gives the position of the ion front:

$$x_f(y,t) = 2\eta(y)t[\ln(\eta(y)t + \sqrt{(\eta(y)t)^2+2}) - \ln(\sqrt{2})] - 2(\sqrt{(\eta(y)t)^2+2} - \sqrt{2}) \quad (3)$$

where $\eta = \sqrt{n_h(y)/n_{h0}}$ and $n_h(y) = n_{h0}\exp(-y^2/2)$. Here y is normalized to σ, the time t is normalized to $1/\omega_{p0}$ and the longitudinal coordinate x is measured in the units of λ_{D0}, where $\omega_{p0} = \sqrt{n_{h0}q_e^2/m_p\epsilon_0}$, $\lambda_{D0} = \sqrt{T_h/m_p}/\omega_{p0}$. We have to mention that the Euler number is missing from the expression, because Mora derived the model for one-temperature model, but in the two-temperature case φ_0 appears in the exponent. It depends on the hot-to cold electron pressure ratio [8], also on y, which would make the equations way more complicated. Assuming that the cold electron density is much higher than the hot one, we can use the approximation $\varphi_0 \approx 0$ and $e^{\varphi_0} = 1$. If we calculate $\partial x_f/\partial y$, we get the following envelope angle:

$$\alpha(y,t) = -\frac{yt}{\Sigma}\left(\sqrt{\frac{1}{k(y,t)}} - \sqrt{1+\frac{1}{k(y,t)}}\right) \quad (4)$$

where $k(y,t) = \exp(-y^2/2)t^2/2$ and $\Sigma = \sigma/\lambda_{D0}$. Now we can simply say that $E_y(y,t) = \alpha(y,t)\cdot E_x(y,t)$, where [7]

$$E_x(y,t) = \frac{2\eta(y)}{\sqrt{2+\eta(y)^2t^2}} \quad (5)$$

The v_y of the protons at the front can be calculated if we integrate $E_y(y,t)$ over time, but it can not be done analytically. A very rough approximation gives us an analytical

expression: $v_y(y,t) = v_x(y,t)\alpha(y,t)$, where v_x is the time derivative of Eq. 3. However we found that the exact solution is always roughly 3.3 times smaller than the analytical expression. Thanks to this interesting observation we have a very simple function which predicts the transverse velocity of the protons:

$$v_y(y,t) = v_x(y,t)\alpha(y,t)/3.3 \quad (6)$$

The next step is the validation of our simple model with correct PIC simulations. For this we compare the result of one simulation with the analytical predicitions. The parameters of the laser: τ =500 fs, I_L=0.8·10^{23} W/m^2 and σ_L=8.5 μm. The hot electron density and temperature were measured at the peak intensity of the laser and the corresponding normalization quantities: $\lambda_{D0} = 6.51 \cdot 10^{-8}$ m and $\omega_{p0} = 8.09 \cdot 10^{13} s^{-1}$.

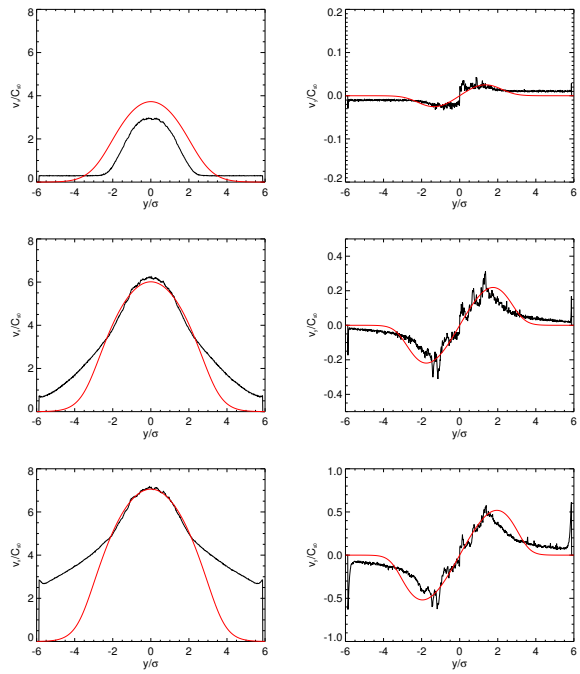


Figure 6: Longitudinal (left column) and transversal (right column) velocities of the proton front at the time moments: 80 fs ($\omega_{p0}t = 6.4$, top row), 200 fs ($\omega_{p0}t = 16$, middle row) and 320 fs ($\omega_{p0}t = 25.6$, bottom row).

In Fig. 6 the comparison is presented for different time moments. As we can see the longitudinal velocity does not match with the model at the beginning because the T_h and n_{h0} did not reach yet their saturation value. The sigma used in the model was $\sigma = 8\mu$m. The transverse velocity fits quite well, but only close to the top of protons front. Farther from the center the hot electron density does not follow the gaussian shape, it is hagher than the one assumed in the model. The reason is that in 2D the particles can move only in two transversal directions, but in the reality they spread out in each direction and the density is proportional to $1/r$, where r is the radius (distance from the center). It means that the plasma expansion has a cylindrical simmetry, which can not be handled in our 2D planar simmetry. The consequence is that the electric field is trong also far from the axis of simmetry of the laser pulse and the proton front surface will be different from the one predicted by the model. Therefore the surface angle becomes smaller and the transverse field weaker. This is why we see smaller v_y in the simullation at the distances larger than 2σ. This effect can be seen in Fig. 7, where we compare the electric fields from 2D and 3D simulations. Other simulations show that the distortion of the front surface becomes stronger if the Σ is smaller, but our model is deduced for large laser spots, where the surface angle is still small.

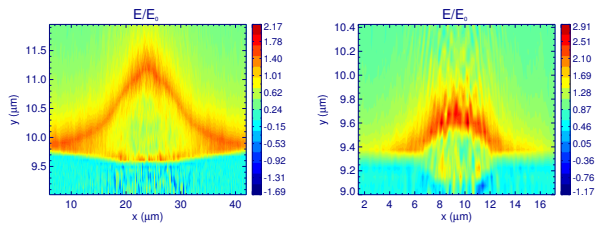


Figure 7: Contour plots of the longitudinal electric field in a 2D (left) and 3D (right) simulation. $E_0 = 1$ TV.

CONCLUSIONS

We could find the optimal numerical parameters for a 2D TNSA simulation, which ensures that the results are correct if the simulation is not too long. This type of simulations can be preformed with limited computing resources as well. The hot electron temeprature does not converege yet for our parameters, because the laser-plasma interaction has very small scale length and it is a complicated stochastic process which is difficult to resolve spatially. We have to rely on our measured values from the simulation and use them as input parameters in the plasma expasnion model. The model, derived in this contribution, is valid for large hot spots and for relatively short time (200-300 fs), but in principle it can be used as initial condition for other simulations where the grid size is much larger and the investigation of proton and electron velocity phase-space on a longer time scale is possible. The full description of the proton distribution is currently ongoing work, but it seems to be possible by using the presented model.

REFERENCES

[1] http://www.gsi.de/phelix/

[2] S. C. Wilks et al., Phys. Rev. Lett., 69, 1383 (1992)

[3] I. Hofmann, et al., PRST-AB, 14, 31304 (2011)

[4] http://www.txcorp.com/products/VORPAL/

[5] E C-Michel et al., Phys. Rev. E **78**, 016404 (2008)

[6] C K Birdsall and A B Langdon, Plasma Physics via Computer Simulations (1991)

[7] P. Mora, Phys. Rev. Lett., **90**, 185002 (2003)

[8] M. Passoni et al., Phys. Rev. E, **69**, 026411 (2004)

INDEPENDENT COMPONENT ANALYSIS (ICA) APPLIED TO LONG BUNCH BEAMS IN THE LOS ALAMOS PROTON STORAGE RING*

J. Kolski†, R. Macek, R. McCrady, X. Pang, LANL, Los Alamos, NM, 87545, USA

Abstract

Independent component analysis (ICA) is a powerful blind source separation (BSS) method[1]. Compared to the typical BSS method, principal component analysis (PCA), which is the BSS foundation of the well known model independent analysis (MIA)[2], ICA is more robust to noise, coupling, and nonlinearity[3, 4, 5]. ICA of turn-by-turn beam position data has been used to measure the transverse betatron phase and amplitude functions, dispersion function, linear coupling, sextupole strength, and nonlinear beam dynamics[3, 4, 5]. We apply ICA in a new way to slices along the bunch, discuss the source signals identified as betatron motion and longitudinal beam structure, and for betatron motion, compare the results of ICA and PCA.

INTRODUCTION

We apply BSS in a new way to slices along the bunch[6]. We digitize beam signals from the Los Alamos Proton Storage Ring (PSR) for a full injection-extraction cycle, $\geq$ 1800 turns. We divide the digitized signal into time slices of equal length using the 0.5 ns digitization bin length. The long digitized signal vector is stacked turn-by-turn to form the data matrix

$$\mathbf{x}(t) = \begin{pmatrix} x_1(1) & x_1(2) & \dots & x_1(N) \\ x_2(1) & x_2(2) & \dots & x_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ x_M(1) & x_M(2) & \dots & x_M(N) \end{pmatrix}, \quad (1)$$

such that each row of $\mathbf{x}$ is data from a single time slice for all N turns, and each column of $\mathbf{x}$ is data from all M time slices for a single turn. The slices are located at a fixed longitudinal phase and position along the bunch. The last turn beam profile plotted in Fig. 1 is an example of the last column of $\mathbf{x}$.

ICA and PCA model $\mathbf{x}$ as a linear combination of independent components (ICs) and principal components (PCs) respectively and yield respectively modes and patterns, describing the source signal's strength in space (along the bunch) and time.

Principal Component Analysis (PCA)

PCA is the simplest of the BSS methods. Many other BSS methods use PCA for preprocessing and noise reduction. PCA was originally applied for beam analysis in the well known MIA.

* Work supported by in part by United States Department of Energy under contract DE-AC52-06NA25396. LA-UR 12-24396

† Electronic address: jkolski@lanl.gov

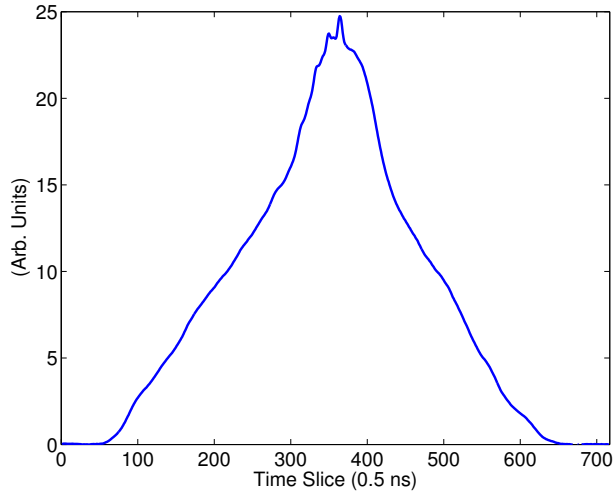


Figure 1: The last turn beam profile as an example of the last column of the data matrix $\mathbf{x}$ in Eq. (1). The PSR's revolution period is 358 ns corresponding to 716 time slices.

PCA identifies patterns in data and expresses the high-dimensional data by highlighting the underlying structures represented as PCs. The PCs are used to compress data by reducing redundant dimensions without much loss of information. PCA minimizes the redundancy measured by covariance, maximizes the signal measured by variance, and results in uncorrelated PCs. Two random variable vectors $\vec{y}_1$ and $\vec{y}_2$ are uncorrelated if their covariance is zero,

$$cov(\vec{y}_1, \vec{y}_2) = \langle \vec{y}_1, \vec{y}_2 \rangle - \langle \vec{y}_1 \rangle \langle \vec{y}_2 \rangle = 0, \quad (2)$$

where $\langle \dots \rangle$ is the expectation value. The core of PCA is singular value decomposition (SVD). SVD of the data matrix $\mathbf{x}$ ($M \times N$) yields eigenvectors $\mathbf{U}$ ($M \times M$) in column-space and $\mathbf{V}$ ($N \times N$) in row-space connected by a diagonal matrix of singular values (SVs) $\mathbf{\Lambda}$ ($M \times N$),

$$\mathbf{x} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T. \quad (3)$$

The columns of $\mathbf{U}$ span column-space, the M-dimensional space of time slice number, and are called spatial patterns. The columns of $\mathbf{V}$ span row-space, the N-dimensional space of turn number, and are called temporal patterns. The PCs are ordered by their SVs, which represent their strength.

Independent Component Analysis (ICA)

The objective of ICA is to calculate the L source signals $\mathbf{s}$ ($L \times N$) given the data matrix $\mathbf{x}$ ($M \times N$), but the mixing

matrix $\mathbf{A}$ $(M \times L)$ is unknown

$$\mathbf{x} = \mathbf{A}\mathbf{s}. \tag{4}$$

ICA assumes independent source signals, a stricter requirement than PCA. Two random variable vectors $\vec{y}_1$ and $\vec{y}_2$ are independent if the covariance of any function of $\vec{y}_1$ and any function of $\vec{y}_2$ is zero,

$$\langle f(\vec{y}_1), g(\vec{y}_2)\rangle - \langle f(\vec{y}_1)\rangle\langle g(\vec{y}_2)\rangle = 0. \tag{5}$$

For time series data, source signal independence is related to diagonality of covariance matrices[7]. The auto-covariance of a signal is $cov(\vec{y}_i(t), \vec{y}_i(t-\tau))$, where τ is a time lag, $\tau = 0, 1, 2, \ldots$ Similarly, the covariance between two signals is $cov(\vec{y}_i(t), \vec{y}_j(t-\tau))$ where $i \neq j$. Applying these two results to mean-zero signals for reasons to become evident later, we write the time-lagged covariance matrix

$$\mathbf{C}_y(\tau) = \langle \mathbf{y}(t)\mathbf{y}(t-\tau)^T\rangle. \tag{6}$$

Source signal independence requires the time-lagged covariance matrices $\mathbf{C}_s(\tau) = \langle \mathbf{s}(t)\mathbf{s}(t-\tau)^T\rangle$ be diagonal,

$$\langle \mathbf{s}_i(t)\mathbf{s}_j(t-\tau)^T\rangle = 0, \quad i \neq j, \quad \tau = 0,\ 1,\ 2,\ \ldots. \tag{7}$$

It follows that $\mathbf{A}^{-1}\mathbf{x}$ must also possess diagonal time-lagged covariance matrices. The BSS problem is solved by obtaining a demixing matrix that diagonalizes the time-lagged covariance matrices of $\mathbf{x}$.

The zero time-lagged covariance matrix $\mathbf{C}_x(\tau = 0)$ does not contain enough information to obtain the mixing matrix $\mathbf{A}$. The key is to utilize the additional information contained in the time-lagged covariance matrices $\mathbf{C}_x(\tau)$. Including more than one time lag improves ICA's performance by resolving degenerate SVs, but it introduces an additional complication of simultaneously diagonalizing many $\mathbf{C}_x(\tau)$. A numerical technique for simultaneously diagonalizing several matrices with Jacobi angles is discussed in Ref. [8]. Typically, 20 - 50 time lags are required to separate source signals with close SVs[4, 5]. We use the ICA algorithm Second Order Blind Identification (SOBI)[7], which accommodates multiple time lags, in our analysis and outline the algorithm below:

1. Whitening

The data matrix $\mathbf{x}$ is preprocessed to obtain mean-zero, whitened ($\mathbf{y}\mathbf{y}^T = \mathbf{I}$) data. Mean-zero data, which simplifies the covariance matrix calculation, is calculated by subtracting the average over the temporal variation. SVD is applied to the zero time-lagged covariance matrix of the mean-zero data matrix $\bar{\mathbf{x}}$

$$\begin{aligned}\mathbf{C}_{\bar{x}}(0) &= \langle \bar{\mathbf{x}}(t)\bar{\mathbf{x}}(t)^T\rangle \\ &= \left(\mathbf{U}_1, \mathbf{U}_2\right)\begin{pmatrix} \mathbf{\Lambda}_1 & 0 \\ 0 & \mathbf{\Lambda}_2 \end{pmatrix}\begin{pmatrix} \mathbf{U}_1^T \\ \mathbf{U}_2^T \end{pmatrix},\end{aligned} \tag{8}$$

where $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are diagonal matrices of SVs separated by a cutoff threshold λ_c such that $\min(diag[\mathbf{\Lambda}_1]) \geq \lambda_c \geq \max(diag[\mathbf{\Lambda}_2])$. The cutoff threshold λ_c is determined by the number of SVs L included in the analysis. $\mathbf{U}_1$ and $\mathbf{U}_2$ are eigenvectors corresponding to $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ respectively. The mean-zero, whitened data is calculated

$$\mathbf{z} = \mathbf{Y}\bar{\mathbf{x}}, \tag{9}$$

where $\mathbf{z}\mathbf{z}^T = I$, $\mathbf{Y} = \mathbf{\Lambda}_1^{-1/2}\mathbf{U}_1^T$, and $\mathbf{\Lambda}_1^{-1/2}$ indicates the inverse square root of the diagonal elements individually.

2. Joint diagonalization

The time-lagged covariance matrices of the mean-zero, whitened data matrix $\mathbf{z}$ are calculated for a set of time lags $(\tau_k,\ k = 0,\ 1,\ \ldots,\ K)$

$$\mathbf{C}_z(\tau_k) = \langle \mathbf{z}(t)\mathbf{z}(t-\tau_k)^T\rangle. \tag{10}$$

Modified time-lagged covariance matrices $\bar{\mathbf{C}}_z(\tau_k)$ are constructed from $\mathbf{C}_z(\tau_k)$

$$\bar{\mathbf{C}}_z(\tau_k) = \left(\mathbf{C}_z(\tau_k) + \mathbf{C}_z(\tau_k)^T\right)/2. \tag{11}$$

SVD is well defined, since $\bar{\mathbf{C}}_z(\tau_k)$ is real and symmetric

$$\bar{\mathbf{C}}_z(\tau_k) = \mathbf{W}\mathbf{D}_k\mathbf{W}^T, \tag{12}$$

where $\mathbf{W}$ is the unitary demixing matrix and $\mathbf{D}_k$ is a diagonal matrix. The Jacobi angle technique discussed in Ref. [8] is used to find the demixing matrix $\mathbf{W}$, which is a joint diagonalizer for all $\bar{\mathbf{C}}_z(\tau_k)$. The mixing matrix $\mathbf{A}$ and the source signals $\mathbf{s}$ are calculated

$$\mathbf{A} = \mathbf{Y}^{-1}\mathbf{W} \quad \text{and} \quad \mathbf{s} = \mathbf{W}^T\mathbf{Y}\bar{\mathbf{x}}. \tag{13}$$

The columns of $\mathbf{A}$ span column-space, the M-dimensional space of time slice number, and are called spatial modes. The rows of $\mathbf{s}$ span row-space, the N-dimensional space of turn number, and are called temporal modes.

INDEPENDENT COMPONENTS

We digitize beam signals from a short stripline beam position monitor (BPM) with 400 MHz peak frequency response. We digitize the BPM's vertical sum and difference signals for BSS.

We typically ran our SOBI analysis for $L = 30$ SVs, $K = 50$ time lags, and as many turns as possible.

ICs are presented in several graphs; see Fig. 2:

Top left spatial mode (blue), last turn beam profile (green).

Bottom left fft of spatial mode (top left), peak integer revolution harmonic, resolution $(1/1)$.

Top center integrated spatial mode (top left) (blue), last turn beam profile (green).

Bottom center correlation of IC and $\mathbf{z}(t-\tau)$, SV.

Top right temporal mode (-1 = last turn), fractional revolution harmonic from sinusoid fit.

Bottom right fft of temporal mode (top right), peak fractional revolution harmonic, resolution $(1/N)$.

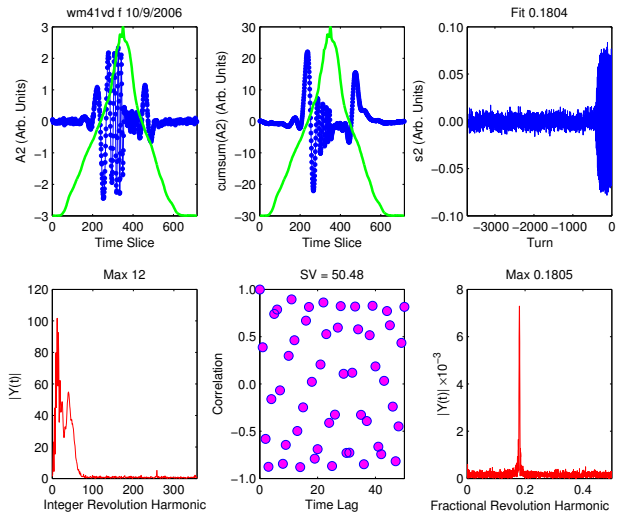


Figure 2: Typical betatron IC resulting from ICA of the difference signal for a single-turn kick.

Betatron ICs

In previous applications, ICA obtains quality ICs representing betatron motion [3, 4, 5], so the betatron IC result from slices along the bunch is of interest. We induce coherent betatron oscillation with a single-turn kick. The beam is stored for 420 turns (150 μs) after the kick.

The IC in Fig. 2 is identified as betatron motion because the fractional revolution harmonic (bottom right) is close to the operating fractional betatron tune value 0.1805 and because the temporal mode (top right) is only nonzero for turns after the single-turn kick.

The spatial mode (top left) indicates the strength of the 0.1805 fractional betatron tune oscillation along the beam bunch. However, in this case it is easier to interpret the integrated spatial mode (top center) because it has units proportional to current, whereas the spatial mode has units proportional to the derivative of the current. The integrated spatial mode maximums indicate that the majority of particles undergoing coherent betatron motion with a fractional tune of 0.1805 are located symmetric about the bunch center at time slices 237 and 474 and represent the coherent space charge tune shifted beam. The fast oscillation slightly forward of the bunch center describes mixing of betatron tunes for the central time slices.

We wish to compare the 15 identified betatron ICs with each other and with the FFT($\mathbf{x}$). We define the greatest strength location to be the fractional revolution harmonic and time slice where an IC is strongest. The time slice coordinate is determined by the integrated spatial mode (top center) maximum. We calculate a leading and trailing edge greatest strength location for each IC and compare with FFT($\mathbf{x}$) in Fig. 3. The greatest strength locations of the 15 betatron ICs reproduce the tune distribution along the bunch. The spatial mode (top left) in Fig. 2 replicates the 0.1805 fractional tune contour along the bunch of the tune

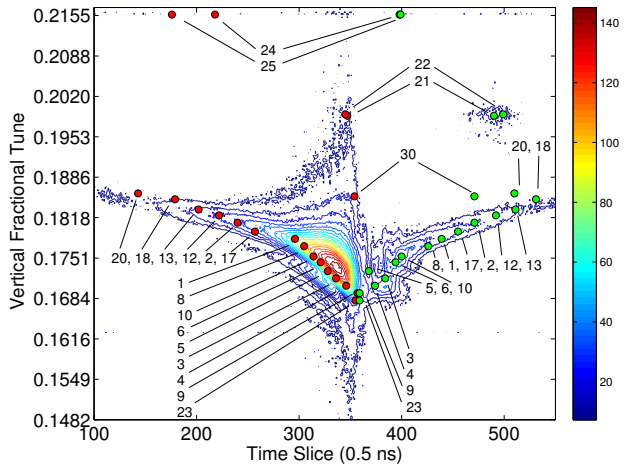


Figure 3: Betatron tune along the bunch and the greatest strength location for the first 30 ICs, some of which lie outside the plot boundary. The contour plots FFT($\mathbf{x}$) along turn for each time slice. The red and green circles mark the leading and trailing edge greatest strength locations respectively with IC number indicated by numeric labels. The beam profile peak is located at time slice 350.

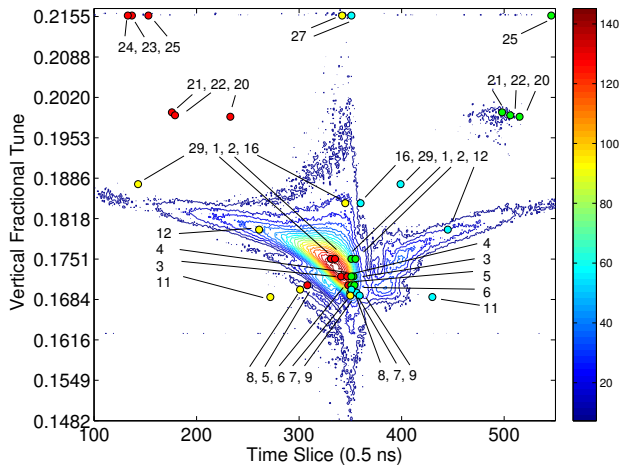


Figure 4: Betatron tune along the bunch and the greatest strength location for the first 30 PCs, some of which lie outside the plot boundary. The contour plots FFT($\mathbf{x}$) along turn for each time slice. The red (yellow) and green (cyan) circles mark the leading and trailing edge greatest strength locations respectively for PCs with dominant betatron motion (dominant source signal other than betatron motion). A numeric label indicates PC number. The beam profile peak is located at time slice 350.

distribution in Fig. 3. The betatron ICs must be viewed in concert as in Fig. 3 to obtain the full picture of the coherent space charge tune shift along the bunch.

We compare the PCA result with FFT($\mathbf{x}$) in Fig. 4. We include in Fig. 4 all PCs with mixed betatron motion. Most

betatron PCs have peak strengths located near the bunch center where the difference signal is largest because PCA is unable to diagonalize the frequency continuum beyond its peak strength and average location. It is clear from Fig. 4 that PCA is unable to recover the coherent space charge tune shift along the bunch.

201.25 MHz IC

The 201.25 MHz IC represents the longitudinal structure of beam newly injected in the PSR, Fig. 5. The 201.25 MHz IC is identified by the character of its temporal mode (top right), which has constant amplitude for the first 1400 turns and suddenly reduces to noise for the last 500 turns. The reduction coincides with the end of accumulation.

The total revolution harmonic is 72.07136 as expected because the PSR design revolution frequency is the 72.07 subharmonic of the 201.25 MHz linac frequency. Multiplying the total revolution harmonic by the revolution frequency yields exactly 201.25 MHz. The 0.07 fractional revolution harmonic describes the 14 turn rastering period of the longitudinal phase space painting executed during injection.

The power of ICA, which lies in the spatial mode (top left), is most prominent in this example. The spatial mode describes the longitudinal phase of the injected beam. The spatial mode is constant amplitude across the injection region and zero outside. The spatial mode also describes the injection length each turn (pattern width). The spatial mode is a sinusoidal oscillation with period 5 ns, a 201.25 MHz period, and 58 peaks. Each peak represents a single linac pulse (linac RF bucket) injected into the PSR. The injection pattern width is 290 ns, which divided by the 201.25 MHz period yields 58 linac pulses injected per turn as indicated by the spatial mode peaks.

We now examine additional examples of the spatial

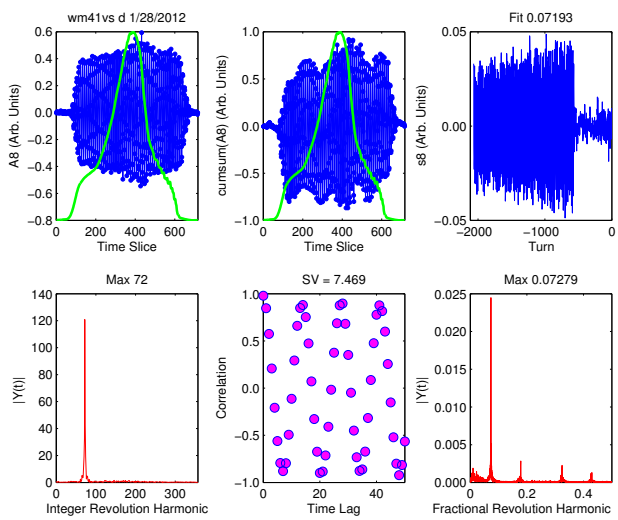


Figure 5: Typical 201.25 MHz IC resulting from ICA of the sum signal.

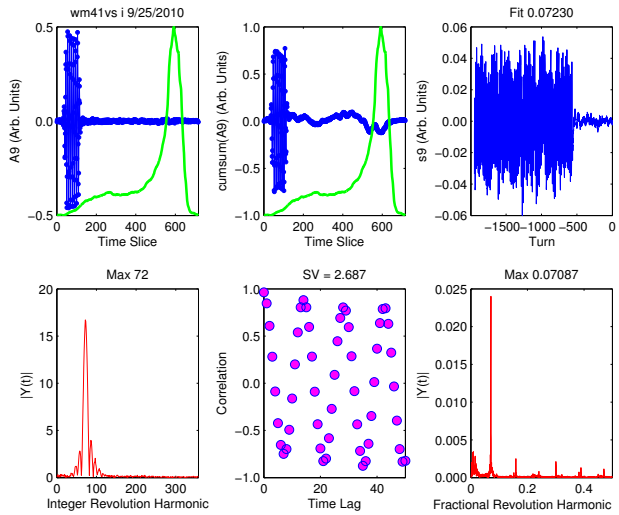


Figure 6: The 201.25 MHz IC resulting from ICA of the sum signal when injecting a 50 ns pattern width beam 120° early in the PSR's RF bucket.

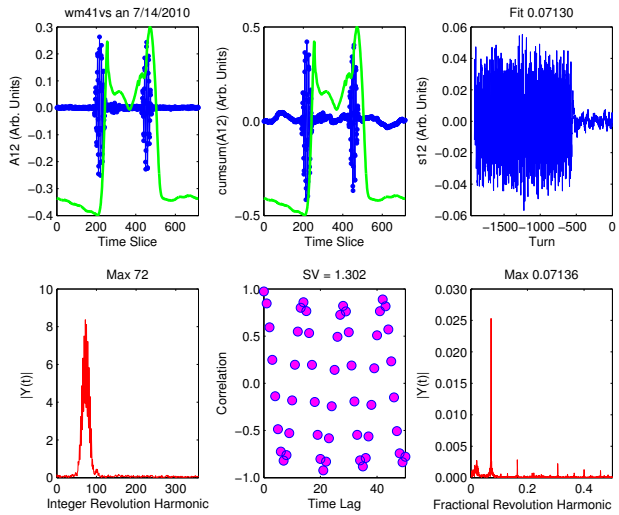


Figure 7: The 201.25 MHz IC resulting from ICA of the sum signal when injecting beam with a notch.

mode (top left) produced from beam signals collected under different injection schemes. The IC plotted in Fig. 6 was produced from beam signal collected while injecting a 50 ns pattern width beam 120° early in the PSR's RF bucket. The injection region of the spatial mode is 100 time slices wide, corroborating the pattern width. Additionally, the spatial mode's position has shifted forward, confirming early injection. The last turn profile indicates that after 200 μs of storage the majority of beam is located opposite in the RF bucket of where it was injected.

The IC plotted in Fig. 7 was produced from beam signals collected while injecting with a "notch" in the beam, illustrated by the spatial mode (top left). The injection region

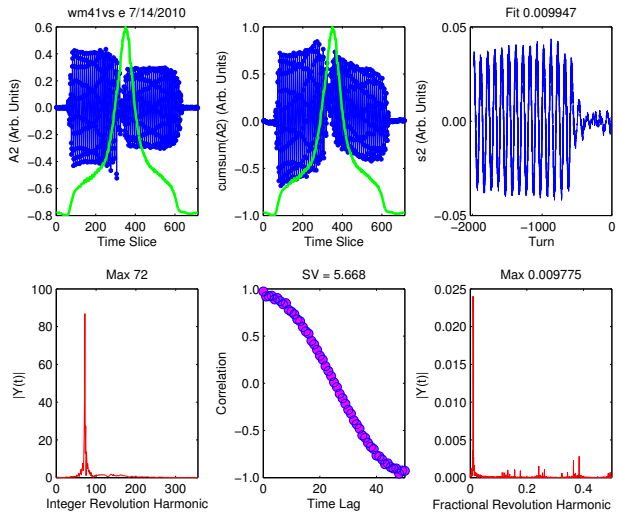


Figure 8: 201.25 MHz IC resulting from ICA of the sum signal when the revolution frequency is near an integer subharmonic of the linac frequency.

widths are 60 time slices, and the notch width between the two injection regions is 180 time slices. The spatial mode reproduces the notch injection scheme: inject for 30 ns, no injection for 90 ns (notch), and inject for 30 ns.

At the beginning of the 2010 LANSCE production run cycle, we observed an unusual amount of "hash" noise-like structure on the PSR beam profile. The hashy beam profile is an operational concern because it could excite longitudinal microwave instabilities and other longitudinal space charge effects[9, 10]. We wished to remove the hash from the beam profile, but did not know where to look for the cause, so we took data for ICA. ICA of the hashy beam signal yields the 201.25 MHz IC plotted in Fig. 8, which is reminiscent of Fig. 5 with obvious difference in the temporal mode (top right). A fractional revolution harmonic 0.009 describes an oscillation that repeats every 111 turns. The slow injection rastering causes the hashy beam profile by longitudinally stacking the beam upon itself more than in nominal operations where the fractional revolution harmonic is 0.07.

We calculate the PSR revolution frequency from the IC plotted in Fig. 8. The unchanging 201.25 MHz linac frequency is divided by the total revolution harmonic 72.009. This results in a revolution frequency of 2.7948 MHz. However, the PSR design revolution frequency is 2.7924 MHz. The revolution frequency differs from design by 2.4 kHz causing the hash on the current profile.

The PSR revolution frequency is created by a frequency generator into which a value is typed. For this run, the frequency generator value was inadvertently set to 2.7948 MHz; exactly the number predicted by ICA.

CONCLUSIONS

We adopt a new method applying ICA to slices along the beam and test with PSR sum and difference beam position signals digitized for an entire injection-extraction cycle.

We determine that PCA is inadequate for the BSS problem of slices along the bunch because PCA is unable to completely separate the source signals, yielding PCs describing a mixture of two or more source signals.

We discuss two classes of ICs identified by spatial and temporal modes. When viewed in concert, the ICs describing betatron motion along the pulse yielded the coherent space charge tune shift. Another IC described the longitudinal structure of newly injected beam and gave the longitudinal phase of injection and the revolution frequency.

REFERENCES

[1] A. Hyvärinen, et al., (John Wiley & Sons, New York, 2001).

[2] J. Irwin, et al., Phys. Rev. Lett. **82**, 1684 (1999); C.X. Wang, Ph.D. thesis, unpublished (Stanford University 1999).

[3] X. Huang, et al., PR-STAB **8**, 064001 (2005); X. Huang, Ph.D. thesis, unpublished (Indiana University 2005).

[4] F. Wang and S.Y. Lee, PR-STAB **11**, 050701, (2008); F. Wang, Ph.D. thesis, unpublished (Indiana University 2008).

[5] X. Pang and S.Y. Lee, Journal of Applied Physics, **106**, 074902 (2009); X. Pang, Ph.D. thesis, unpublished (Indiana University 2009).

[6] J. Kolski, et al., IPAC 2012 Conference Proceedings, WEPPR038, May 2012; J. Kolski, Ph.D. thesis, unpublished (Indiana University 2010).

[7] A. Belouchrani, et al., IEEE Trans. Signal Processing, **45**, 434-444, (1997).

[8] J.F. Cardoso and A. Souloumiac, SIAM J. Mat. Anal. Appl., **17**, 161, (1996).

[9] C. Beltran, et al., PAC 2003 Conference Proceedings, TOPD004, May 2003; C. Beltran, Ph.D. thesis, unpublished (Indiana University 2003).

[10] R. McCrady, et al., LANSCE-ABS 06-008, PSR 06-002 (unpublished).

BIG DATA ANALYSIS AND VISUALIZATION: WHAT DO LINACS AND TROPICAL STORMS HAVE IN COMMON?

E.W. Bethel*, S. Byna, J. Chou†, E. Cormier-Michel‡, C.G.R. Geddes,
M. Howison, F. Li, P. Prabhat, J. Qiang, O. Rübel, R.D. Ryne, M. Wehner, K. Wu
LBNL, Berkeley, CA 94720, USA

Abstract

While there is wisdom in the old adage "the two constants in life are death and taxes," there are unavoidable truths facing modern experimental and computational science. First is the growing "impedance mismatch" between our ability to collect and generate data, and our ability to store, manage, and gain understanding from it. The second is the fact that we cannot continue to rely on the same software technologies that have worked well for the past couple of decades for data management, analysis, and visualization. A third is that these complementary activities must be considered in a holistic, rather than Balkanized way. The inseparable interplay between data management, analysis, visualization, and high performance computational infrastructure, are best viewed through the lens of case studies from multiple scientific domains, where teams of computer and scientists combine forces to tackle challenging data understanding problems.

INTRODUCTION

Big data and its attendant challenges—managing it and gaining insight into its secrets—are the subject of a substantial amount of research and development in industry, academia, and the broader scientific community. These challenges are often cited as being among the greatest barriers facing scientific knowledge discovery [1].

Within the high performance visualization and analysis community, several different but complementary approaches contribute solutions to these challenges. One approach focuses on increasing the capacity of the data analysis and visualization processing pipelines [2], and recent results have shown such techniques capable of scaling to extreme concurrency for both production [3] and research codes [4].

An alternative approach is to focus instead on limiting visualization and analysis processing to the subset of data of interest where presumably the interesting subset is much smaller in size than the original data. Projects in this space have focused on blending index and query technologies from scientific data management with visualization and analysis tools to implement what is known as *query-driven visualization* [5]. This class of solution has proven applicable to diverse problems ranging from forensic cybersecurity [6] to accelerator modeling [7].

* Corresponding author ewbethel@lbl.gov
† jchou@cs.nthu.edu.tw
‡ ecormier@txcorp.com

The focus of this paper is on the nexus between those two approaches with an eye towards enabling scientific insight. Three case studies, one from climate modeling and two from accelerator modeling, show how high performance computing technology and advances in visualization and analysis software infrastructure combine to provide new scientific data understanding capabilities brought to bear on problems that can be characterized as "large data analysis and visualization."

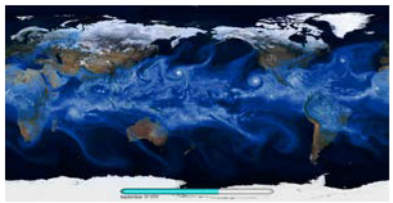
CASE STUDY: TROPICAL CYCLONES AND CLIMATE MODELING DATA ANALYSIS

As with many computational science domains, the study of climate and climate change, benefits from increasingly powerful computational platforms. Modern climate codes, which model processes in the atmosphere, ocean, ice caps, and more, produce massive amounts of data. As discussed below, when modeling the atmosphere at $0.25°$ resolution for a period of about two decades of simulation time, the CAM5.1 code [8] produces approximately 100TB of model output. These data sizes will only grow in size with time.

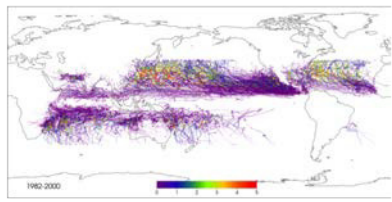
One challenge facing this computational science community is a rich legacy of visualization and analysis tools that are serial in implementation, yet such tools are not capable of processing modern-sized data sets due to memory constraints. Worse, such tools are often incapable of performing the type of analysis required to gain understanding in ever-larger and ever-richer collections of data.

Output from codes like CAM5—which sample and discretize both space and time in a regular fashion and are often generated by ensemble simulation experiments for varying physical input parameters and initial conditions—exposes data parallelism in many different ways to analysis and visualization tasks. Many types of visualization or analysis codes can be run independently, and in parallel, across individual ensemble members, timesteps, spatial regions, and individual grid points.

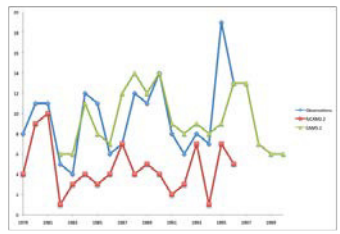
There are several pressing needs and challenges within the climate modeling community. First is the growing "impedance mismatch" between their ability to collect or generate data, and their ability to perform analysis and visualization on massive data sets. Second is the need to be able to quickly and easily create and test a variety of different types of quantitative analysis activities, such as spatiotemporal feature detection, tracking, and analysis. Third is the ability to execute such capabilities on large, parallel



(a) Visualization of 0.25° CAM5 output.



(b) Cyclone tracks computed from 100TB of CAM5 output.



(c) Comparing the annual count of observed vs. modeled cyclones.

Figure 1: In this climate science example, a high resolution atmospheric code produces massive amounts of data and the science objective is to study the number of cyclones that form over time. One timestep of model output is visualized (a). The TECA code is run in parallel to identify cyclones and their tracks over time (b). These results are compared to the counts of cyclones observed over the same time period, as well as to a third model's (fvCAM2.2) output (c). Images courtesy of Prabhat, Wehner, et al. (LBNL).

platforms, which have resources sufficient to process massive data sets in a reasonable amount of time.

Towards addressing these needs, Prabhat et al., 2012 [9], developed the Parallel Toolkit for Extreme Climate Analysis (TECA). Their objective was to enable rapid implementation of a variety of user-written and customizable spatiotemporal feature detection, tracking, and analysis algorithms in a single framework that accommodates different varieties of data parallelism. Within that framework, there are a number of capabilities that are common to all feature detection tasks, such as loading data files, accommodating different calendering systems, data scatter, etc.

To the user-developer who wants to implement a new feature detection and tracking algorithm, they write code that is handed a 1D/2D/3D sub-block of data and perform their computations on that block, storing results in a "local table." Later, the local tables are gathered and processed in a serial post-processing phase, which is typically much smaller in size compared to the original problem and easily accommodated in serial. The developer need not make any MPI calls, nor be concerned with the mechanics of data distribution or the gathering of results.

The authors present several case studies where TECA is applied to detecting, tracking, and analyzing different extreme-weather phenomena, including tropical cyclones, extra-tropical cyclones, and atmospheric rivers. Results from a similar, yet earlier tropical cyclones study, is shown below in Figure 1. One of the major accomplishments of their work is the ability to scale the tropical cyclone feature detection, tracking, and analysis code to 7,000 cores to complete analysis/processing of 100TB of CAM5 model output in approximately 2 *hours*, compared to an estimated serial processing time of approximately 583 *days*. More recent versions of this work, yet unpublished, have scaled to approximately 80,000 cores. In that example, the user-written code examines its local data block to determine if that block contains a cyclone, which is defined as a combination of locally high vorticity, low pressure within a defined distance from the high vortex, and a locally warm temperature combined with a temperature dropoff in nearby locations.

CASE STUDY: LINACS, LPAS, AND ACCELERATOR MODELING DATA ANALYSIS

While the previous section focuses on quantitative analysis, specifically feature detection and tracking in climate model output, this section focuses on visual data exploration and analysis of particle-based data produced by accelerator modeling codes. These codes output massive amounts of particle-based data at each timestep; the two examples below consist of hundreds of millions of particles per timestep. Generally speaking, the overall objective is to find and analyze "interesting particles," where the definition of "interesting" varies from study to study. Also, generally speaking, these so-called interesting particles can be characterized as satisfying some set of multivariate conditions in space and time.

As with the climate example, these communities have built up a rich legacy of serial tools over time for visualization and analysis, tools that cannot accommodate modern massive data sizes. The subsections below discuss work aimed at providing the infrastructure for next-generation visualization and analysis tools that are applicable to accelerator modeling output of massive size and on modern supercomputing platforms.

Laser-Plasma Accelerators

One central challenge in the analysis of large LPA simulation data arises from the fact that, while large numbers of particles are required for an accurate simulation, only a small fraction of the particles are accelerated to high energies and subsequently form particle features of interest. During the course of a simulation, multiple high-energy particle beams may form and additional particle bunches may appear in secondary periods of the plasma wave. Studying this acceleration phenomena requires gaining insight into which particles become accelerated, how are they trapped and accelerated by the plasma wave, and

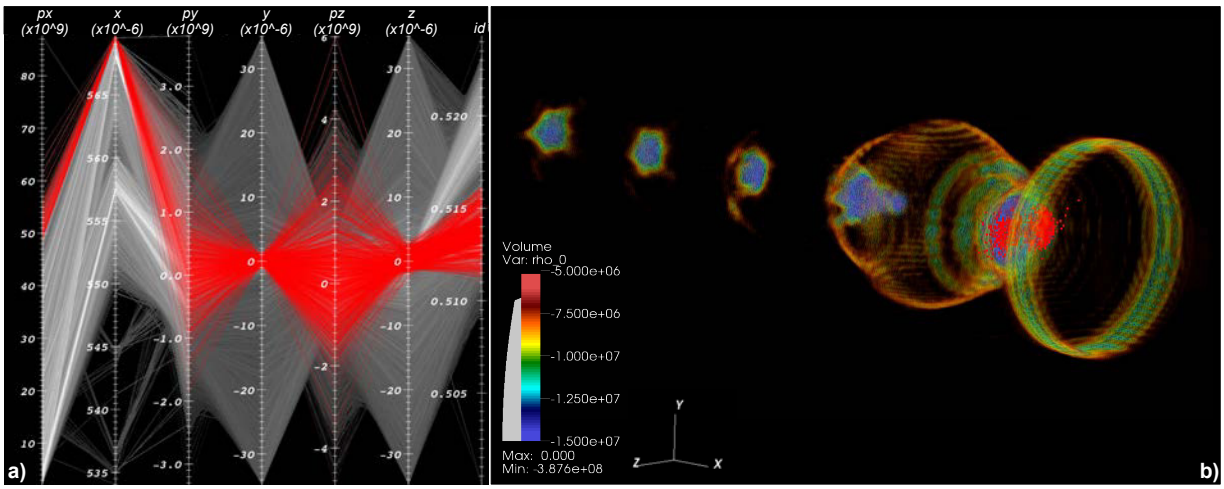


Figure 2: Query-driven visualization and analysis of a large 3D plasma-based particle acceleration data containing $\approx 90 \times 10^6$ particles per timestep. On the left (a), parallel coordinates of timestep $t = 12$ showing: (1) all particles above the base acceleration level of the plasma wave (Query: $px > 2 \times 10^9$) (gray) and (2) a set of particles that form a compact beam in the first wake period following the laser pulse (Query: $(px > 4.856 \times 10^{10})\text{AND}(x > 5.649 \times 10^{-4})$)(red). On the right (b), volume rendering of the plasma density illustrating the 3D structure of the plasma wave. The additional selected beam particles are shown in red. Image source: Rübel et al., 2008 [7].

how these beams evolve over time. These questions are the subject of work by Rübel et al, 2008 [7].

To identify those particles that were accelerated, the authors, who included two accelerator physicists, performed an initial threshold selection in px at a late timestep of the simulation. This initial selection restricts the analysis to a small set of particles with energies above the phase velocity of the plasma wave. Figure 2a shows an example of a parallel coordinates plot of such a selection (gray).

Based on the results of the first query, the selection was further refined to select the main particle beams of interest. The authors first increased the px threshold to extract the particles of highest energy. This initial refinement often results in the selection of multiple beam-like features trapped in different periods of the plasma wave.

As illustrated in Figure 2a (red lines), to separate the different particle beams and to extract the main particle beam, the selection is then often further refined through range queries in the longitudinal coordinate x and transverse coordinates y and z. In the selection process, parallel coordinates provide interactive feedback about the structure of the selection, allowing for fast identification of outliers and different substructures of the selection. High performance scientific visualization methods are then used to validate and further analyze the selected particles shown in Figure 2b.

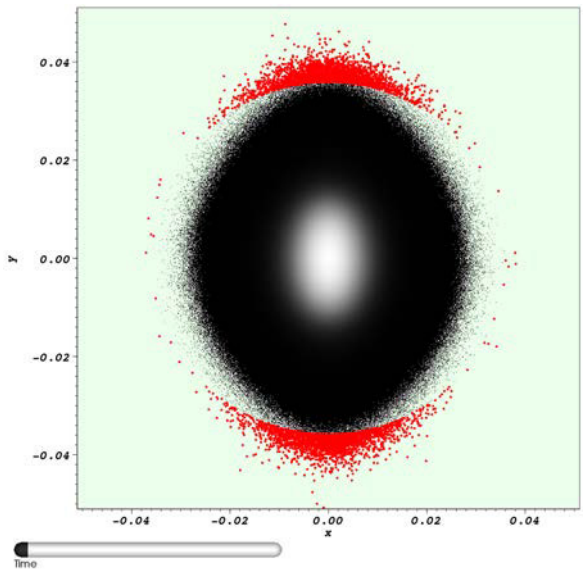
The next step was to trace, or follow, and analyze the high-energy particles through time. Particle tracing is used to detect the point in time when the beam reaches its peak energy and to assess the quality of the beam. Tracing the beam particles further back in time, to the point at which the particles enter the simulation window, supports analysis of the injection, beam formation, and beam evolution processes. Based on the information from different individual timesteps, a user may refine a query, to select, for example, beam substructures that are visible at different discrete timesteps.

This particular study by Rübel et al., 2008 [7], which included scalability tests out to 300-way parallel in 2008, reduced the time required to track and analyze 300 accelerated particles from hours, using legacy serial processing tools, to less than one second. Whereas the Rübel et al., 2008 study demonstrated tracking tens to millions of particles, due to the mechanics of its internal algorithm, the legacy serial processing tool would not be practical for use on increasingly larger problem sizes.

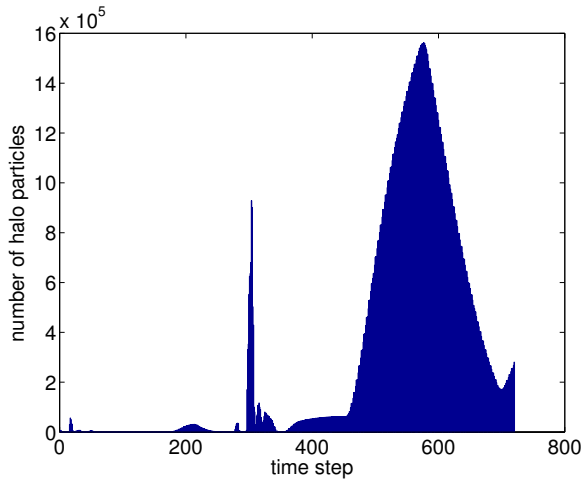
Linear Accelerators

This case study focuses on visual data exploration and analysis of output from large-scale, high resolution simulations of beam dynamics in electron LINACS for a proposed next-generation Xray free electron laser (FEL) at Lawrence Berkeley National Laboratory (LBNL) [10]. Particle-in-cell-based simulations of this type of accelerator require large numbers of macroparticles ($> 10^8$) to control the numerical macroparticle shot noise and to avoid the overestimation of microbunching instability, resulting in massive particle data sets [11].

Chou et al, 2011 [12] focus on studying the transverse halo and the longitudinal core. In both cases, the criteria a given particle satisfies to be classified as “halo” or “core” is a spatial one. Prior to this work, the typical workflow pattern would be to run a serial analysis tool on each timestep, where the serial code would examine each particle in turn, and evaluate whether or not it satisfied the given criteria. This process is $O(N)$ in computational complexity, given N particles per timestep. But, like the LPA example before,



(a) Particle density plot (gray) and particles selected (red) by the halo query for timestep 20 of the simulation.



(b) Plot showing the number of halo particles vs. timestep.

Figure 3: Particles of the transverse halo shown in red from a single simulation timestep identified using a spatial condition query (a). The number of halo particles (b) increases over time, indicating a potential problem with this particular accelerator design. Image source: Chou et al., 2011 [12].

this process is dramatically accelerated by using advanced index-query infrastructure: that infrastructure accelerates the processing time by avoiding the examine-every-particle approach and replacing it with one where only those particles that are likely to satisfy the condition are examined.

The authors identify halo particles using the query $r > 4\sigma_r$. The derived quantity $r = \sqrt[2]{x^2 + y^2}$, which is separately computed and indexed, describes the transverse radial particle location. The transverse halo threshold is given by $\sigma_r = \sqrt[2]{\sigma_x^2 + \sigma_y^2}$, where σ_x and σ_y denote the root mean square (RMS) beam sizes in x and y, respectively. Using r for the identification of halo particles is based on the assumption of an idealized circular beam cross section. To ensure that the query adapts more closely to the transverse shape of the beam, one may relax the assumption of a circular beam cross section to an elliptical beam cross section, by scaling the x and y coordinates independently by the corresponding RMS beam size, σ_x and σ_y using, for example, a query of the form $r_s^2(x, y) > 16$, with $r_s^2(x, y) = (\frac{x}{\sigma_x})^2 + (\frac{y}{\sigma_y})^2$. Halo particles identified with this methodology, for one timestep, are shown in Figure 3a. Details of the definition and analysis of the beam core are in the original study [12].

Figure 3b shows the number of halo particles per timestep identified by the halo query. It shows large variations in the number of halo particles, while in particular, the larger number of halo particles at later timesteps are indicative of a possible problem. In this case, the halo particles and observations of an increase in the maximum particle amplitude were found to be due to a mismatch in the beam as it traveled from one section of the accelerator to the next. These types of query-based diagnostics provide accelerator designers with evidence that further improvement of the design may be possible, and also provides quantitative information that is useful for optimizing the design to reduce halo formation and beam interception with the beam pipe, which will ultimately improve accelerator performance.

The authors used the FastQuery [12] system for the parallel processing of index and query operations on the NERSC Cray XE6 supercomputing system Hopper. The pre-processing time to build the indexes for all timesteps of the 50TB data set was about 2 hours. By using bitmap indexing, evaluating queries for finding the halo and core required only 12 seconds with 3000 cores, or around 20 seconds with 500 cores. The combination of visual data exploration and efficient data management in the query-driven visualization concept enables, in this way, repeated, complex, large-scale query-based analysis of massive data sets, which would otherwise not be practical, with respect to both time as well as computational cost.

LESSONS LEARNED

These case studies, each of which are field-leading in their own right, all have several themes in common.

Searching, querying. First, all studies involve some element of searching for data with specific features or characteristics. Over the years, this particular topic has been a subject of active research in the scientific data management community. By searching scientific data using the motifs presented here, which consist of high-dimensional, read-only query operations, the bitmap index has proven to be a useful and appropriate indexing structure [13]. The ac-

celerator modeling case studies here make use of FastBit, which is an Open Source, high performance compressed bitmap indexing implementation [14].

Data I/O. Increasingly, the cost of performing I/O from simulation codes dominates the analysis cycle. In the case of the LINAC case study, the accelerator scientists had never before seen the results of later simulation timesteps due to the inability to write out that much simulation data. This work was made possible through a partnership between computer and computational scientists whereby the IMPACT-T simulation code was fitted to use a high performance parallel I/O library. Looking further ahead, many in the field expect an increasing amount of visualization and analysis to be performed *in situ*, while particle and field data are still resident in-core inside the simulation.

Integrative technologies. In our work, we are finding that it is increasingly the case that disparate technologies—simulation codes, data management, analysis, and visualization infrastructure—must be designed and engineered to work together as a whole. It is increasingly unlikely that these technologies can be considered and used in isolation from one another as design choices for one can have a profound impact on others. This situation is exacerbated by increasingly large and complex data sets and more sophisticated lines of scientific inquiry.

ACKNOWLEDGMENT

This work was supported by the Director, Office of Science, Office and Advanced Scientific Computing Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, and used resources of the National Energy Research Scientific Computing Center (NERSC), which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

REFERENCES

[1] Richard Mount. The Office of Science Data-Management Challenge. Report from the DOE Office of Science Data-Management Workshops. Technical Report SLAC-R-782, Stanford Linear Accelerator Center, March-May 2004.

[2] Hank Childs and Mark Miller. Beyond Meat Grinders: An Analysis Framework Addressing the Scale and Complexity of Large Data Sets. In *SpringSim High Performance Computing Symposium (HPC 2006)*, pages 181–186, 2006.

[3] Hank Childs, David Pugmire, Sean Ahern, Brad Whitlock, Mark Howison, Prabhat, Gunther Weber, and E. Wes Bethel. Extreme Scaling of Production Visualization Software on Diverse Architectures. *IEEE Computer Graphics and Applications*, 30(3):22–31, May/June 2010.

[4] Mark Howison, E. Wes Bethel, and Hank Childs. Hybrid Parallelism for Volume Rendering on Large, Multi- and Many-core Systems. *IEEE Transactions on Visualization and Computer Graphics*, 99(PrePrints), 2011.

[5] Kurt Stockinger, John Shalf, Kesheng Wu, and E. Wes Bethel. Query-Driven Visualization of Large Data Sets. In *Proceedings of IEEE Visualization 2005*, pages 167–174. IEEE Computer Society Press, October 2005. LBNL-57511.

[6] E. Wes Bethel, Scott Campbell, Eli Dart, Kurt Stockinger, and Kesheng Wu. Accelerating Network Traffic Analysis Using Query-Driven Visualization. In *Proceedings of 2006 IEEE Symposium on Visual Analytics Science and Technology*, pages 115–122. IEEE Computer Society Press, October 2006. LBNL-59891.

[7] Oliver Rübel, Prabhat, Kesheng Wu, Hank Childs, Jeremy Meredith, Cameron G. R. Geddes, Estelle Cormier-Michel, Sean Ahern, Gunther H. Weber, Peter Messmer, Hans Hagen, Bernd Hamann, and E. Wes Bethel. High Performance Multivariate Visual Data Exploration for Extemely Large Data. In *Supercomputing 2008 (SC'08)*, Austin, TX, USA, November 2008. LBNL-716E.

[8] R. B. Neal et al. Description of the NCAR Community Atmosphere Model (CAM 5.0). Technical report, National Center for Atmospheric Research, Boulder, CO, USA, 2010. NCAR/TN-486+STR.

[9] Prabhat, Oliver Rübel, Surendra Byna, Kesheng Wu, Fuyu Li, Michael Wehner, and E. Wes Bethel. TECA: A Parallel Toolkit for Extreme Climate Analysis. In *Third Worskhop on Data Mining in Earth System Science (DMESS 2012) at the International Conference on Computational Science (ICCS 2012)*, Omaha, NE, June 2012. LBNL-5352E.

[10] J. N. Corlett, K. M. Baptiste, J. M. Byrd, P. Denes, R. J. Donahue, L. R. Doolittle, R. W. Falcone, D. Filippetto, J. Kirz, D. Li, H. A. Padmore, C. F. Papadopoulos, G. C. Pappas, G. Penn, M. Placidi, S. Prestemon, J. Qiang, A. Ratti, M. W. Reinsch, F. Sannibale, D. Schlueter, R. W. Schoenlein, J. W. Staples, T. Vecchione, M. Venturini, R. P. Wells, R. B. Wilcox, J. S. Wurtele, A. E. Charman, E. Kur, and A. Zholents. A Next Generation Light Source Facility at LBNL. In *Proceedings of PAC 2011*, New York, USA, April 2011.

[11] J. Qiang, R. D. Ryne, M. Venturini, and A. A. Zholents. High Resolution Simulation of Beam Dynamics in Electron Linacs for X-Ray Free Electron Lasers. *Physical Review*, 12(100702):1–11, 2009.

[12] Jerry Chou, Mark Howison, Brian Austin, Kesheng Wu, Ji Qiang, E. Wes Bethel, Arie Shoshani, Oliver Rübel, Prabhat, and Rob D. Ryne. Parallel Index and Query for Large Scale Data Analysis. In *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '11, pages 30:1–30:11, Seattle, WA, USA, November 2011. LBNL-5317E.

[13] Arie Shoshani and Doron Rotem, editors. *Scientific Data Management: Challenges, Technology, and Deployment*. Chapman & Hall/CRC Press, Boca Raton, FL, USA, 2010.

[14] Kesheng Wu, Sean Ahern, E. Wes Bethel, Jacqueline Chen, Hank Childs, Estelle Cormier-Michel, Cameron Geddes, Junmin Gu, Hans Hagen, Bernd Hamann, Wendy Koegler, Jerome Lauret, Jeremy Meredith, Peter Messmer, Ekow Otoo, Victor Perevoztchikov, Arthur Poskanzer, Prabhat, Oliver Rübel, Arie Shoshani, Alexander Sim, Kurt Stockinger, Gunther Weber, and Wei-Ming Zhang. FastBit: Interactively Searching Massive Data. In *SciDAC 2009*, 2009. LBNL-2164E.

List of Authors

Bold papercodes indicate primary authors

Institutes List

ANL
Argonne, USA
- Borland, M.
- Mametjanov, A.
- Min, M.
- Soliday, R.
- Wang, Y.

ANSL
Yerevan, Armenia
- Khojoyan, M.

BNL
Upton, Long Island, New York, USA
- Bassi, G.
- Fischer, W.
- Hao, Y.
- Litvinenko, V.
- Malitsky, N.
- Ranjbar, V.H.
- Wang, G.
- White, S.M.

CERN
Geneva, Switzerland
- Assmann, R.W.
- Bruce, R.
- De Michele, G.
- Deniau, L.
- Koukovini-Platia, E.
- Mounet, N.
- Métral, E.
- Redaelli, S.
- Renshall, H.
- Rijoff, T.L.
- Rumolo, G.
- Salvachua, B.
- Wollmann, D.
- Zannini, C.
- Zimmermann, F.

CIAE
Beijing, People's Republic of China
- Wang, C.
- Yin, Z.G.
- Zhang, T.J.

CIPS
Boulder, Colorado, USA
- Cary, J.R.

CLASSE
Ithaca, New York, USA
- Sagan, D.

CST
Darmstadt, Germany
- Balk, M.C.
- Wolfheimer, F.

Cyfronet
Kraków, Poland
- Szymocha, T.

DELTA
Dortmund, Germany
- Riemann, B.
- Weis, T.

DESY Zeuthen
Zeuthen, Germany
- Asova, G.
- Groß, M.
- Hakobyan, L.
- Isaev, I.I.
- Ivanisenko, Ye.
- Khojoyan, M.
- Klemz, G.
- Koss, G.
- Kourkafas, G.
- Krasilnikov, M.
- Kusoljariyakul, K.
- Li, J.
- Mahgoub, M.
- Malyutin, D.
- Marchetti, B.
- Oppelt, A.
- Petrosyan, B.
- Rimjaem, S.
- Shapovalov, A.
- Stephan, F.
- Vashchenko, G.

DESY
Hamburg, Germany
- Barber, D.P.
- Dohlus, M.
- Feng, G.
- Kot, Y.A.
- Zagorodnov, I.

Diamond
Oxfordshire, United Kingdom
- Apollonio, M.
- Bartolini, R.
- Martin, I.P.S.

ELETTRA
Basovizza, Italy
- Cornacchia, M.
- Di Mitri, S.
- Scafuri, C.

EPFL
Lausanne, Switzerland
- Zannini, C.

ESS
Lund, Sweden
- Danared, H.
- Eshraqi, M.
- Laface, E.
- Miyamoto, R.
- Molloy, S.
- Ponton, A.

ETH
Zurich, Switzerland
- Arbenz, P.

Fermilab
Batavia, USA
- Johnstone, C.
- Khabiboulline, T.N.
- Lunin, A.
- Yakovlev, V.P.

FZJ
Jülich, Germany
- Lehrach, A.
- Maier, R.
- Senichev, Y.
- Zyuzin, D.

GSI
Darmstadt, Germany
- Appel, S.
- Boine-Frankenheim, O.
- Fitzek, J.
- Franchetti, G.
- König, H.G.
- Kornilov, V.
- Schmidt, P.
- Sorge, S.
- Spädtke, P.

Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Institute of Radiation Physics
Dresden, Germany
- Bussmann, M.H.
- Debus, A.D.
- Pausch, R.
- Schramm, U.
- Widera, R.

HZB
Berlin, Germany
- Abo-Bakr, M.
- Anders, W.
- Knobloch, J.
- Kuske, B.C.
- Kuske, P.
- Meseck, A.
- Neumann, A.
- Richter, D.
- Scheer, M.

IBM Research - Zurich
Rueschlikon, Switzerland
- Bekas, C.
- Curioni, A.

IMP
Lanzhou, People's Republic of China
- Li, C.
- Qi, X.
- Yang, L.

IS
Beijing, People's Republic of China
- Xu, J.
- Zhao, R.
- Zhufu, X.

KEK
Ibaraki, Japan
- Forest, E.
- Ohmi, K.

KIRAMS
Seoul, Republic of Korea
- An, D.H.
- Cho, S.J.
- Hahn, G.
- Hong, H.B.
- Hong, S.S.
- Hwang, W.T.
- Jang, H.S.
- Jung, I.S.
- Kang, J.
- Kang, K.U.
- Kim, C.H.
- Kim, G.B.
- Lee, M.Y.
- Park, Y.S.
- Yang, T.K.
- Yim, H.

KIT
Karlsruhe, Germany
- Hiller, N.
- Huttel, E.
- Judin, V.
- Kehrer, B.
- Klein, M.
- Marsching, S.
- Meuter, C.A.J.
- Müller, A.-S.
- Naknaimueang, S.
- Nasse, M.J.
- Rossmanith, R.
- Schuh, M.
- Schwarz, M.
- Smale, N.J.
- Streichert, M.
- Wesolowski, P.

KU Leuven
Kortrijk, Belgium
- De Gersem, H.
- Deryckere, J.
- Masschaele, B.
- Roggen, T.

LANL
Los Alamos, New Mexico, USA
- Kolski, J.S.
- Macek, R.J.
- McCrady, R.C.
- Pang, X.

LBNL
Berkeley, California, USA
- Benedetti, C.
- Bethel, E.W.
- Byna, S.
- Chou, J.
- Cormier-Michel, E.
- Esarey, E.
- Geddes, C.G.R.
- Howison, M.
- Leemans, W.
- Li, F.
- Michael, M. F.
- Mitchell, C.E.
- Prabhat, P.
- Qiang, J.
- Ruebel, O.
- Ryne, R.D.
- Schroeder, C.B.
- Vay, J.-L.
- Wu, K.

LLNL
Livermore, California, USA
- Cohen, R.H.
- Friedman, A.
- Grote, D.P.

MAX-lab
Lund, Sweden
- Sjöström, M.

MSU
East Lansing, Michigan, USA
- Berz, M.
- Jagasia, R.
- Makino, K.
- Wittig, A.
- Zhang, H.

Naples University Federico II, Science and Technology Pole
Napoli, Italy
- Iadarola, G.

Northern Illinois University
DeKalb, Illinois, USA
- Erdelyi, B.

ORNL
Oak Ridge, Tennessee, USA
- Shishlo, A.P.

PSI
Villigen, Switzerland
- Adelmann, A.
- Bao, Y.
- Guo, H.
- Ineichen, Y.
- Oswald, B.S.C.

Pulsar Physics
Eindhoven, The Netherlands
- de Loos, M.J.
- van der Geer, S.B.

RIKEN Nishina Center
Wako, Japan
- Fukuda, N.
- Inabe, N.
- Kameda, D.
- Kubo, T.
- Kusaka, K.
- Suzuki, H.
- Takeda, H.

Rostock University, Faculty of Computer Science and Electrical Engineering
Rostock, Germany
- Bahls, C.R.
- Brackebusch, K.
- Flisgen, T.

- Galek, T.
- Heller, J.
- Markoviḱ, A.
- Papke, K.
- Pöplau, G.
- van Rienen, U.

RPI
Troy, New York, USA
- Fu, J.

Saint-Petersburg State University
Saint-Petersburg, Russia
- Alekhina, T.Yu.
- Tyukhtin, A.V.

SLAC
Menlo Park, California, USA
- Ge, L.
- Jensen, A.
- Ko, C.
- Lee, K.H.
- Li, Z.
- Ng, C.-K.
- Novokhatski, A.
- Xiao, L.

Solaris
Kraków, Poland
- Goryl, P.P.
- Wawrzyniak, A.I.

St. Petersburg State University
St. Petersburg, Russia
- Andrianov, S.N.
- Golovkina, A.G.
- Ivanov, A.N.
- Kotina, E.D.
- Kudinovich, I.V.
- Kulabukhova, N.V.
- Ovsyannikov, D.A.

Tech-X
Boulder, Colorado, USA
- Abell, D.T.
- Amyx, K.M.
- Balasalle, J.
- Bell, G.I.
- Bruhwiler, D.L.
- Cary, J.R.
- Cormier-Michel, E.
- Cowan, B.M.
- Hallman, E.J.
- Meiser, D.
- Pogorelov, I.V.
- Schwartz, B.T.
- Webb, S.D.

TEMF, TU Darmstadt
Darmstadt, Germany
- Ackermann, W.
- Boine-Frankenheim, O.
- Franke, S.
- Gjonaj, E.
- Görgen, P.A.
- Hartel, U.K.
- Klingbeil, H.
- Klopfer, K.
- Lecz, L.
- Niedermayer, U.
- Weiland, T.
- Yaman, F.

TRIUMF, Canada's National Laboratory for Particle and Nuclear Physics
Vancouver, Canada
- Chao, Y.-C.
- Gong, C.

TU Darmstadt
Darmstadt, Germany
- Bahlo, T.
- Burandt, C.
- Eichhorn, R.
- Enders, J.
- Konrad, M.
- Nonn, P.N.

TUD
Darmstadt, Germany
- Glesner, M.
- Spies, C.

TUE
Eindhoven, The Netherlands
- Luiten, O.J.
- van der Geer, S.B.
- Vredenbregt, E.J.D.

UMAN
Manchester, United Kingdom
- Appleby, R.
- Molson, J.
- Serluca, M.
- Toader, A.M.

Uni HH
Hamburg, Germany
- Rönsch-Schulenburg, J.

University of Huddersfield
Huddersfield, United Kingdom
- Barlow, R.J.

University of Malta, Information and Communication Technology
Msida, Malta
- Sammut, N.J.
- Valentino, G.

UNL
Lincoln, USA
- Kyle, K.
- Serguei, S.
- Shadwick, B.
- Umstadter, D.P.

UNM
Albuquerque, New Mexico, USA
- Bizzozero, D.
- Ellison, J.A.
- Heinemann, K.A.
- Lau, S.R.

USTC/NSRL
Hefei, Anhui, People's Republic of China
- Shang, L.

YSU
Yerevan, Armenia
- Kotanjyan, A.
- Saharian, A.A.