Global electromagnetic gyrokinetic particle-in-cell simulation

R. Hatzky

High Level Support Team

Outline

- **The High Level Support Team**
- **The Vlasov-Maxwell equations**
- **The Particle-In-Cell (PIC) method and its gyrokinetic discretization**
- **The cancellation problem**
- **Control variates as variance reduction method**
- **The adjustable control variate (ACV)**
- **Numerical results**
- **Summary**

The EFDA HPC project

High Performance Computer for Fusion Applications

HPC-FF

- 1080 nodes @ 8 cores, peak performance 101 TFlop/s
- Start of production: August 2009

High Level Support Team

HLST

core team (5 persons)

IPP

rzg
RECHENZENTRUM GARCHING

+ 4 ppy in other Associations

The High Level Support Team

HLST core team at IPP Garching

- Roman Hatzky – core team leader

- Tamás Fehér (plasma physicist with HPC experience)
- Matthieu Haefele (computer engineer – visualization specialist)
- Kab Seok Kang (numerical mathematician)
- Michele Martone (computer engineer)
- Tiago Ribeiro (plasma physicist with HPC experience)

≈50 applications were evaluated to select the members of the core team

Quite difficult to find candidates with extensive
High Performance Computing (HPC) & physics background

The High Level Support Team

HLST staff members

- Salomon Janhunen (TEKES, Finland)
- Huw Leggate (DCU, Ireland)
- Sebastian Petruczynik (PSNC, Poland)
- Trach-Minh Tran (CRPP, Switzerland)

The staff members are located at fusion relevant sites all over Europe

They dedicate 50% of their working time to HLST

The High Level Support Team

Management of HLST

- The work is coordinated by the HLST coordinator, Darren McDonald.
- The core team leader, Roman Hatzky, guides the daily work.
- The whole HLST meets personally twice a year in Garching.

The HLST has its own web site URL: www.efda-hlst.eu for dissemination purpose.

The High Level Support Team

Main tasks for HLST

The HLST team is a support unit to ensure optimal exploitation of HPC-FF, i.e. it is not focused on its own academic research.

Support for code development

- Single processor performance optimization
- Parallelization & optimization of codes for massively parallel computers
- Improvement of the parallel scalability of existing codes already ported to parallel platforms
- Implementation of algorithms and mathematical library routines to improve the efficiency of codes
- Visualization of large data sets

The High Level Support Team

Improvement of the efficiency of algorithms

Try to reduce as much as possible the number of FLOPs needed to solve a given problem → usage of highly efficient parallel algorithms

Example:
Using multigrid instead of the conjugated gradient method to solve a discretized PDE.

- The multigrid method is very efficient but complex.
- General purpose numerical libraries are usually not most efficient.
- A geometric multigrid solver has to be adapted by hand
  → know how is mandatory

Some algorithms become inefficient/efficient for massively parallel usage.

The High Level Support Team

Final goal of HLST evolution

The more insight into the physics/work flow of a code, the better are the chances to significantly improve its efficiency.

→ The HLST members have to further enhance their HPC skills to become finally specialists in algorithms used in certain fields of plasma physics simulations.

Without such experts we will not make efficient use of the IFERC-CSC petaflop computer for a significant number of plasma physics codes.

The IFERC-CSC computer

The system in operation:

- One node consists of 16 cores with 58 GB available memory.
- The total system is composed of 4410 nodes, i.e. 70,560 cores.
- The peak performance is 1.3 PFLOPS.
- The available memory is 256 TB.
- The network is an InfiniBand (IB) interconnect.

The HLST gives support for IFERC-CSC!

The High Level Support Team

HLST call

The call is launched once a year

- Addressed to scientists from the EFDA associates
- Maximal allocatable resources of 12 months (exceptions possible)

In the period 2009 – 2012:

- 46 proposals with a total request of 29.7 ppy received
- 42 projects have been approved
- The project coordinators were located in: Austria, Finland, France, Germany, Italy, Spain, Switzerland, and United Kingdom

The Vlasov equation

The usual basis for kinetic treatments of a collisionless plasma is the Vlasov equation:

$$\frac{\partial \hat{f}_s}{\partial t} + \mathbf{v} \cdot \frac{\partial \hat{f}_s}{\partial \mathbf{x}} + \frac{q_s}{m_s} \left(\mathbf{E} + \mathbf{v} \times \mathbf{B} \right) \cdot \frac{\partial \hat{f}_s}{\partial \mathbf{v}} = 0$$

Here, $\hat{f}_s(\mathbf{x}, \mathbf{v}, t)$ **is the distribution function of the** s**th species in six-dimensional phase space with the position** $\mathbf{x}$ **and the velocity** $\mathbf{v}$**.**

The Vlasov equation can be written in the following form:

$$\frac{\mathrm{D} \hat{f}_s}{\mathrm{D}t} \overset{\text{def}}{=} \frac{\partial \hat{f}_s}{\partial t} + \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} \cdot \frac{\partial \hat{f}_s}{\partial \mathbf{x}} + \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} \cdot \frac{\partial \hat{f}_s}{\partial \mathbf{v}} = 0$$

Its short form $\mathrm{D}\hat{f}_s/\mathrm{D}t = 0$ **means that the total derivative vanishes along the characteristics given by the integration of the equations of motion.**

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{v} \qquad , \qquad \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \frac{q_s}{m_s}\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right) \qquad \textbf{equations of motion}$$

In physical terms:
If we follow the particles along their trajectories by integrating the equations of motion in six-dimensional phase space, the initial value of $\hat{f}_s(\mathbf{x}(t_0), \mathbf{v}(t_0))$ will not change.

This method is well known by the name "method of characteristics" and can be used to evolve $\hat{f}_s$ in time (initial value problem).

The Vlasov-Maxwell equations

The self-consistent electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$ which appear in the force law are calculated from Poisson's equation and Ampère's law, two of the Maxwell equations:

$$\nabla \cdot \mathbf{E} = \frac{\varrho}{\epsilon_0} \qquad \nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \frac{1}{c^2}\frac{\partial \mathbf{E}}{\partial t}$$

Here, the charge density ϱ and current density $\mathbf{j}$ are to be obtained at each point in space from the appropriate moments of the distribution function itself:

$$\varrho(\mathbf{x}) = \sum_s q_s \int \hat{f}_s \, \mathrm{d}^3 v \qquad \mathbf{j}(\mathbf{x}) = \sum_s q_s \int \mathbf{v} \hat{f}_s \, \mathrm{d}^3 v$$

where the summation is over the species of particles present in the plasma.

The PIC method

The Particle-In-Cell method (PIC) is a numerical technique used to solve a certain class of partial differential equations:

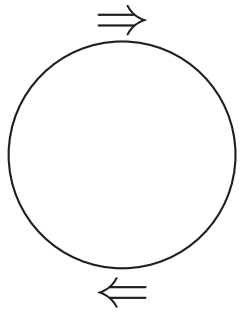
- **A total number of N (macro) particles in a Lagrangian frame are traced in continuous phase space**
- **Moments of the distribution function (ϱ and $\mathbf{j}$) are computed simultaneously on a Eulerian (stationary) mesh to solve the self-consistent field equations**

The PIC method is a so-called Particle-Mesh (PM) method which includes interactions of particles only through the average fields.

Schematic diagram of the PIC method

Integration of the equations of motion

Interpolation of the fields from the mesh to the particle locations



Interpolation of charge and current source terms onto the potential mesh

Computation of the potentials on the mesh

Monte Carlo evaluation of integrals

Of special interest is the evaluation of moments of the distribution function f, i.e. integrals over a volume element Ω_0 of the phase-space volume V

$$I(\Lambda) \stackrel{\text{def}}{=} \int_{\Omega_0} \Lambda(\mathbf{z}) f(\mathbf{z}) \, \mathrm{d}\mathbf{z}$$

where $\Lambda(\mathbf{z})$ is a general function of the phase-space coordinates $\mathbf{z}$.
For example, $I(\Lambda)$ would be the number density in configuration space if $\Lambda = 1$, and the integral is evaluated over the velocity space.

The sampling distribution of our Monte-Carlo sampling points (markers) is done by a continuous probability density function $g(\mathbf{z})$ such that

$$\int_V g(\mathbf{z}) \, \mathrm{d}\mathbf{z} = 1$$

Now, the integral for $I(\Lambda)$ **can be written in the following form**

$$\mathcal{E}[X] \stackrel{\text{def}}{=} \int_V X(\mathbf{z}) g(\mathbf{z})\,\mathrm{d}\mathbf{z} \qquad \textbf{where} \qquad X(\mathbf{z}) \stackrel{\text{def}}{=} \frac{\Lambda(\mathbf{z}) f(\mathbf{z})}{g(\mathbf{z})}$$

and $\mathcal{E}[X]$ **is the expected value of the random variable** X**. In addition, we define the variance of** X **by**

$$\sigma^2 \equiv \mathcal{V}[X] \stackrel{\text{def}}{=} \int_V (X(\mathbf{z}) - \mathcal{E}[X])^2 g(\mathbf{z})\,\mathrm{d}\mathbf{z}$$

The crude Monte-Carlo estimator for the integral $I(\Lambda)$ **is given by the sum over the marker weights** w_n

$$I(\Lambda) = \frac{1}{N} \sum_{n=1}^{N} \Lambda(\mathbf{z}_n)\, w_n \pm \epsilon \qquad \textbf{where} \qquad w_n \stackrel{\text{def}}{=} \frac{f(\mathbf{z}_n)}{g(\mathbf{z}_n)}$$

The statistical error is defined as:

$$\epsilon \stackrel{\text{def}}{=} \frac{\sigma}{\sqrt{N}}$$

- **The convergence rate of $1/\sqrt{N}$ is quite poor, i.e. to half the error one needs four times more particles**
- **If possible the standard deviation $\sigma \stackrel{\text{def}}{=} \sqrt{\mathcal{V}}$ should be reduced**

Variance reduction methods:

- **Importance sampling**
- **Stratified sampling**
- **Control variate method**

Liouville's theorem

A phase-space volume element $\Omega_p \stackrel{\text{def}}{=} \Omega(\mathbf{z}_p)$ **moving in phase space according to the equations of motion does not change its size, although in general it can change its shape (Liouville's theorem):**

$$\int_{\Omega_p} \mathrm{d}\mathbf{z} = \textbf{const.}$$

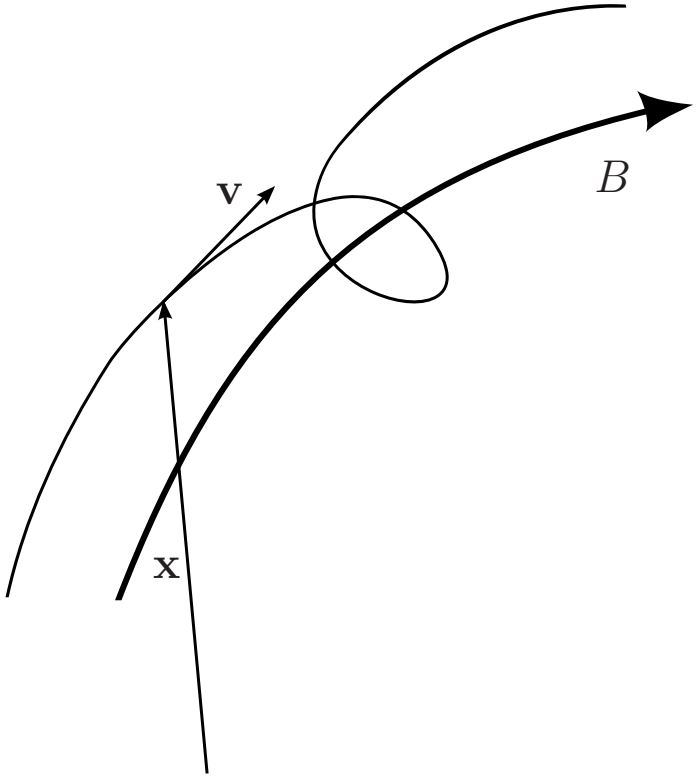
As the marker distribution and phase-space volumes Ω_p **are related by**

$$g_p = \frac{V}{\Omega_p} = \textbf{const.}$$

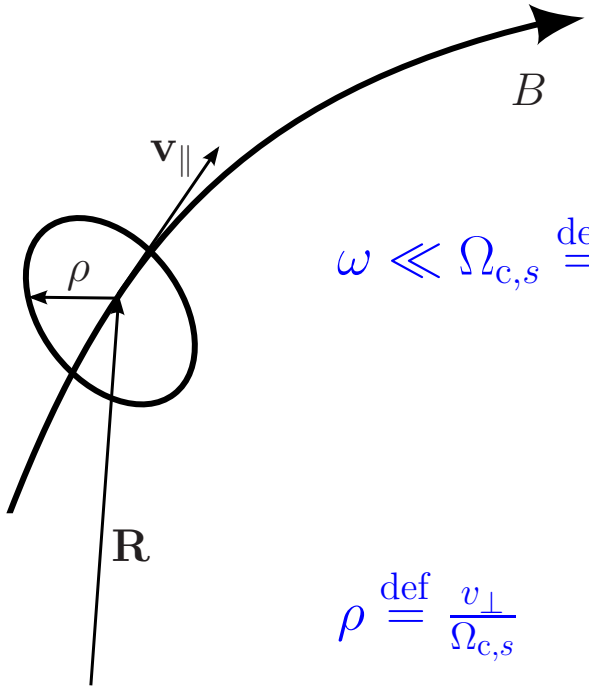
it follows for the marker weights from $f_p = \textbf{const.}$ **that**

$$w_p = \frac{f_p}{g_p} = \textbf{const.}$$

kinetics



gyrokinetics



$$\omega \ll \Omega_{\mathrm{c},s} \stackrel{\mathrm{def}}{=} \frac{q_s B}{m_s}$$

$$\rho \stackrel{\mathrm{def}}{=} \frac{v_\perp}{\Omega_{\mathrm{c},s}} \qquad \mu \stackrel{\mathrm{def}}{=} \frac{v_\perp^2}{2B}$$

$\hat{f}_s(\mathbf{x}, \mathbf{v}, t)$ **in 6-dim** $\Rightarrow$ $f_s(\mathbf{R}, v_\parallel, \mu, t)$ **in 5-dim**

The gyrokinetic Vlasov equation (W.W. Lee, 1983)

The average of the Vlasov equation over the fast gyro-motion leaves just the guiding center motion and thus reduces the dimensionality of the problem:

$$\frac{\partial f_s}{\partial t} + \frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t} \cdot \frac{\partial f_s}{\partial \mathbf{R}} + \frac{\mathrm{d}v_\parallel}{\mathrm{d}t} \cdot \frac{\partial f_s}{\partial v_\parallel} = 0 \quad , \quad \frac{\mathrm{d}\mu}{\mathrm{d}t} = 0$$

Here, $f_s(\mathbf{R}, v_\parallel, \mu, t)$ **is the guiding center distribution function of the** s**th species in the reduced five-dimensional phase space with the guiding center coordinate** $\mathbf{R}$ **and the parallel velocity coordinate** $v_\parallel$ **and the magnetic moment** μ**.**

The equations of motion depend on the gyro-averaged electrostatic and magnetic potentials $\langle\phi\rangle$ **and** $\langle\mathbf{A}\rangle$**.**

Equations of motion ($p_\parallel$-formulation)

$$\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t} = v_\parallel \mathbf{e}_\mathrm{B} + \frac{m_s}{q_s}\left[\frac{\mu B + v_\parallel^2}{B B_\parallel^\star}\mathbf{e}_\mathrm{B}\times\nabla B + \frac{v_\parallel^2}{B B_\parallel^\star}(\nabla\times\mathrm{B})_\perp\right] +$$

$$- \frac{q_s}{m_s}\left(\mathbf{e}_\mathrm{B} + \frac{m_s}{q_s}\frac{v_\parallel}{B B_\parallel^\star}\left[\mathbf{e}_\mathrm{B}\times\nabla B + (\nabla\times\mathrm{B})_\perp\right]\right)\langle A_\parallel\rangle + \frac{1}{B_\parallel^\star}\mathbf{e}_\mathrm{B}\times\nabla\langle\Psi_\mathrm{eff}\rangle$$

$$\frac{\mathrm{d}v_\parallel}{\mathrm{d}t} = -\mu\,\nabla B\cdot\left[\mathbf{e}_\mathrm{B} + \frac{m_s}{q_s}\frac{v_\parallel}{B B_\parallel^\star}(\nabla\times\mathrm{B})_\perp\right] +$$

$$- \frac{q_s}{m_s}\left(\mathbf{e}_\mathrm{B} + \frac{m_s}{q_s}\frac{v_\parallel}{B B_\parallel^\star}\left[\mathbf{e}_\mathrm{B}\times\nabla B + (\nabla\times\mathrm{B})_\perp\right]\right)\cdot\nabla\langle\Psi_\mathrm{eff}\rangle$$

where

$$B_\parallel^\star \overset{\mathrm{def}}{=} B + \frac{m_s}{q_s}v_\parallel \mathbf{e}_\mathrm{B}\cdot(\nabla\times\mathbf{e}_\mathrm{B}), \qquad \Psi_\mathrm{eff} \overset{\mathrm{def}}{=} \phi - v_\parallel A_\parallel$$

Potential equations

Quasineutrality equation (long wavelength approximation, valid up to $k_\perp \rho_\mathrm{i} \lesssim 1$):

$$\underbrace{-\nabla_\perp \cdot \left(\frac{n_\mathrm{i}}{B\Omega_{\mathrm{c,i}}} \nabla_\perp \phi \right)}_{\textbf{polarization density}} = \langle n_\mathrm{i} \rangle - n_\mathrm{e}$$

Parallel Ampère's law ($p_\parallel$-formulation):

$$-\nabla_\perp \cdot \left[(1-\beta_\mathrm{i}) \nabla_\perp A_\parallel \right] + \underbrace{\left(\frac{\beta_\mathrm{i}}{\rho_\mathrm{i}^2} + \frac{\beta_\mathrm{e}}{\rho_\mathrm{e}^2} \right) A_\parallel}_{\textbf{skin terms}} = \mu_0 \left(\langle j_{\parallel \mathrm{i}} \rangle + j_{\parallel \mathrm{e}} \right)$$

The discretized f

The full-f PIC approximation is given by a sum of δ functions:

$$f(\mathbf{R}, v_\parallel, \mu, t) = \sum_{p=1}^{N} \frac{w_p}{\mathcal{J}_{\text{red}}} \delta[\mathbf{R} - \mathbf{R}_p(t)]\, \delta[v_\parallel - v_{\parallel p}(t)]\, \delta[\mu - \mu_p(t_0)]$$

with the Jacobian in the reduced phase space $\mathcal{J}_{\text{red}} = 2\pi B_\parallel^\star$.

Each marker (macro particle, tracer, ...) p is defined by:

- **its position in the 5-dim phase space $(\mathrm{R}_p, v_{\parallel p}, \mu_p)$**
- **its phase-space volume Ω_p assigned by the initial marker distribution $g(t_0)$**
- **its constant weight w_p**

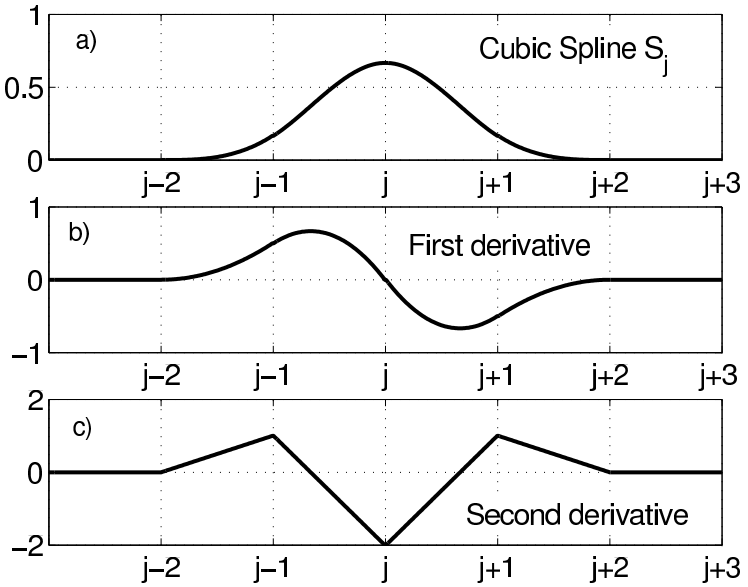
Discretization of potential equations

The potential equations are discretized with a finite element method, e.g.:

$$A_{\parallel}(\mathbf{x},t) = \sum_{\nu} c_{\nu}(t)\Lambda_{\nu}(\mathbf{x})$$

where $\Lambda_{\nu}(\mathbf{x})$ is a product of unidimensional B-splines S of order k:

$$\Lambda_{\nu}(\mathbf{x}) = S_l^k(x_1)S_m^k(x_2)S_n^k(x_3)$$



Advantages of finite elements:

- Conservation laws, e.g. particle number and energy conservation for PIC are consistently preserved
- Complicated geometries and non-equidistant meshes are easy to implement

Discretization of Ampère's law using Galerkin's method

LHS:

$$\sum_{\nu'} \int \left[(1-\beta_{\mathrm{i}}) \nabla_\perp \Lambda_\nu \cdot \nabla_\perp \Lambda_{\nu'} + \frac{\beta_{\mathrm{i}}}{\rho_{\mathrm{i}}^2} \Lambda_\nu \Lambda_{\nu'} + \frac{\beta_{\mathrm{e}}}{\rho_{\mathrm{e}}^2} \Lambda_\nu \Lambda_{\nu'} \right] \mathrm{d}\mathbf{x}\; c_{\nu'} = A\,\mathbf{c}$$

In matrix representation:

$$(L + S_{\mathrm{i}} + S_{\mathrm{e}})\,\mathbf{c} = A\,\mathbf{c} \qquad \textbf{where} \qquad A \stackrel{\mathrm{def}}{=} L + S_{\mathrm{i}} + S_{\mathrm{e}}$$

The finite support of the finite elements Λ_ν results in a sparse matrix A.

$\Rightarrow$ Use of parallel sparse matrix packages like e.g. IBM WSMP or PETSc

The interpolation of the current onto the grid

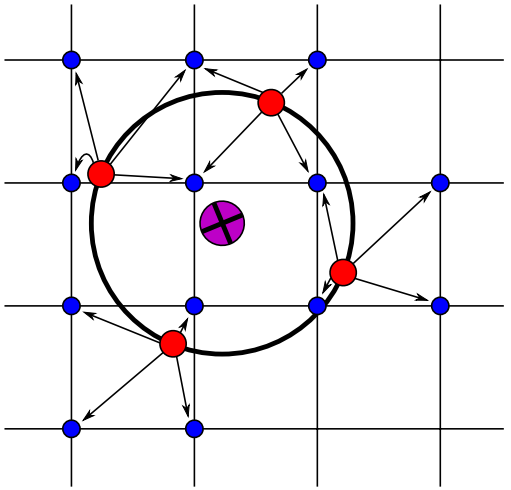
The construction of the RHS is the so-called current assignment:

$$j_{\parallel \mathrm{i},\nu} \overset{\text{def}}{=} \sum_{p=1}^{N} v_{\parallel p} w_{\mathrm{i},p} \frac{1}{2\pi} \int_0^{2\pi} \Lambda_\nu(\mathbf{R}_p + \boldsymbol{\rho}_{\mathrm{i},p}) \, \mathrm{d}\alpha$$

Projection of the weights $w_{\mathrm{i},p}$ in the form of gyro-rings onto the B-spline basis.

The charge assignment is a scatter operation, e.g. each sample point contributes to 64 grid points for cubic B-splines in 3-dim.

The figure shows linear interpolation for linear B-splines in 2-dim.



The interpolation of $\nabla\langle A_{\parallel}\rangle$ from the grid

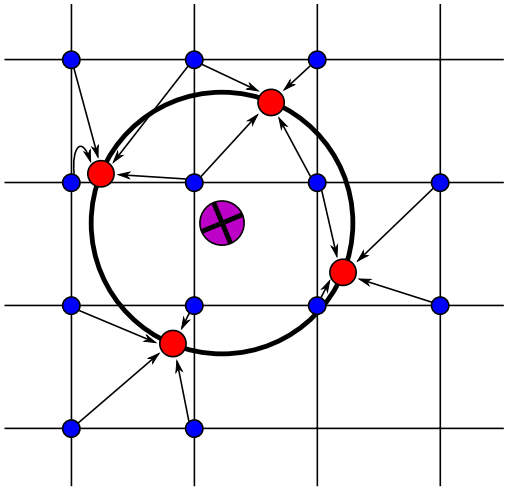
The gyro-averaged $\langle\nabla \mathbf{A}_{\parallel}\rangle$ is analytically represented within the B-spline basis:

$$\nabla\langle \mathbf{A}_{\parallel}\rangle \approx \langle\nabla \mathbf{A}_{\parallel}\rangle = \sum_{\nu} \frac{c_\nu}{2\pi} \int_0^{2\pi} \nabla\Lambda(\mathbf{R} + \boldsymbol{\rho}_{\mathrm{i}})\,\mathrm{d}\alpha$$

The N_{av} field vectors $\nabla \mathbf{A}_{\parallel n}$ on the gyro-ring are calculated from the B-spline representation of the potential and then averaged.

The $\nabla \mathbf{A}_{\parallel n}$ calculation is a gather operation, e.g. each sample point is assembled from 64 grid points for cubic B-splines in 3-dim.

The figure shows linear interpolation for linear B-splines in 2-dim.



The cancellation problem

The electron distribution function f_e includes an adiabatic (Boltzmann) part responding to $A_\parallel$:

$$\delta f_\mathrm{e} = \delta f_\mathrm{e}^\mathrm{ad} + \delta f_\mathrm{e}^\mathrm{nonad} \qquad \textbf{where} \qquad \delta f_\mathrm{e}^\mathrm{ad} = -\frac{e v_\parallel f_\mathrm{M,e}}{k_\mathrm{B} T_\mathrm{e}} A_\parallel$$

The electron skin term in Ampère's law coincides with the adiabatic current term:

$$\frac{\beta_\mathrm{e}}{\rho_\mathrm{e}^2} A_\parallel = \frac{\mu_0 n_0 e^2}{m_\mathrm{e}} A_\parallel = -\mu_0 e \int \mathrm{d}^6 Z \, \delta f_\mathrm{e}^\mathrm{ad} \, v_\parallel \, \delta(\mathbf{R} - \mathbf{x}) = \mu_0 j_{\parallel\mathrm{e}}^\mathrm{ad}$$

$$\Downarrow$$

In theory the skin terms cancel the adiabatic current terms.

$$-\nabla_\perp \cdot \left[(1 - \beta_\mathrm{i}) \nabla_\perp A_\parallel \right] = \mu_0 \left(\langle j_{\parallel\mathrm{i}} \rangle^\mathrm{nonad} + j_{\parallel\mathrm{e}}^\mathrm{nonad} \right)$$

The PIC method expresses inseparably the adiabatic and nonadiabatic current.

$$\Downarrow$$

Signal-to-noise problem when adiabatic electron current dominates.

$$\frac{j_{\parallel \mathrm{e}}^{\mathrm{ad}}}{\langle j_{\parallel \mathrm{i}} \rangle^{\mathrm{nonad}} + j_{\parallel \mathrm{e}}^{\mathrm{nonad}}} \approx \frac{\beta_{\mathrm{e}}/\rho_{\mathrm{e}}^2 A_\parallel}{\nabla_\perp^2 A_\parallel} = \frac{\mu_0 e^2 n_0 A_\parallel}{m_{\mathrm{e}} \nabla_\perp^2 A_\parallel} \overset{\text{cylinder}}{\approx} \frac{\mu_0 e^2 n_0}{m_{\mathrm{e}} k_\perp^2}$$

The nonadiabatic part is the smaller fraction of the electron distribution for

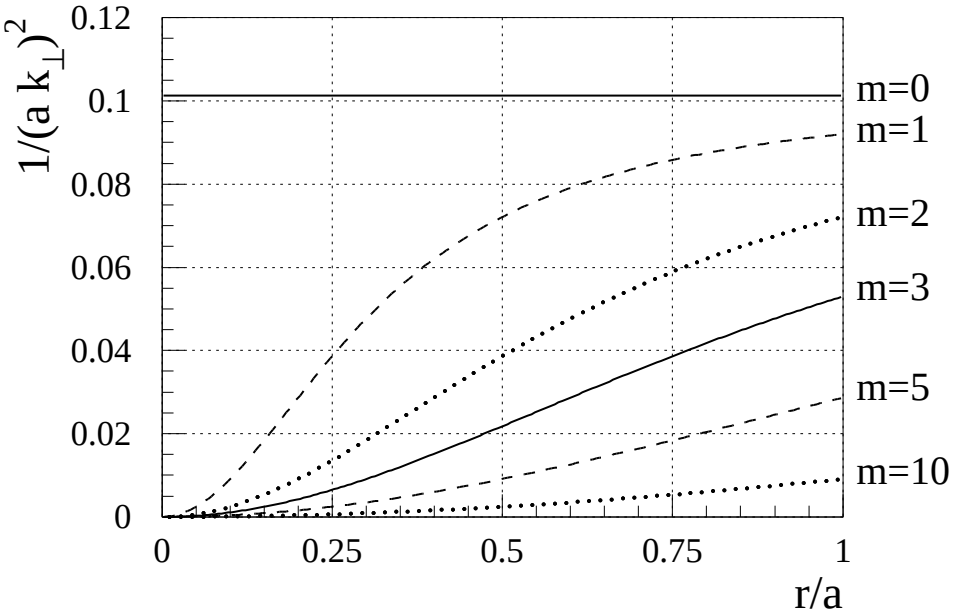
- **high beta cases,** $\beta_{\mathrm{e}} \gtrsim 1\ \%$
- **the MHD limit,** $k_\perp \to 0$

Due to the signal-to-noise problem the numerical cancellation of the skin terms is inaccurate which is called the "cancellation problem".

$1/k_\perp^2$ dependency of the cancellation problem for a cylindrical configuration

$$\frac{1}{k_\perp^2} = \frac{a^2}{\left(\frac{m}{r/a}\right)^2 + \pi^2}$$

where m is the poloidal mode number and the minimal k_r is assumed.



- **The cancellation problem is most pronounced for the $m = 0$ mode**
  $\Rightarrow$ Negative effect on the accuracy of the zonal flows
- **For $m \neq 0$ the cancellation problem is most pronounced at the edge**
- **The cancellation problem scales with a^2**

The cancellation problem in the quasineutrality equation

Due to the symmetry $f_{\mathrm{M,e}}(-v_\parallel) = f_{\mathrm{M,e}}(v_\parallel)$ **of the Maxwellian there is no contribution of the the adiabatic part** $\delta f_{\mathrm{e}}^{\mathrm{ad}}$ **to the electron number density** n_{e}**, i.e.**

$$-e^2 \int \frac{A_\parallel}{k_{\mathrm{B}} T_{\mathrm{e}}} \delta(\mathbf{R} - \mathbf{x}) \, \mathrm{d}\mathbf{R} \underbrace{\int v_\parallel f_{\mathrm{M,e}} \, \mathrm{d}\mathbf{v}}_{=0} = 0$$

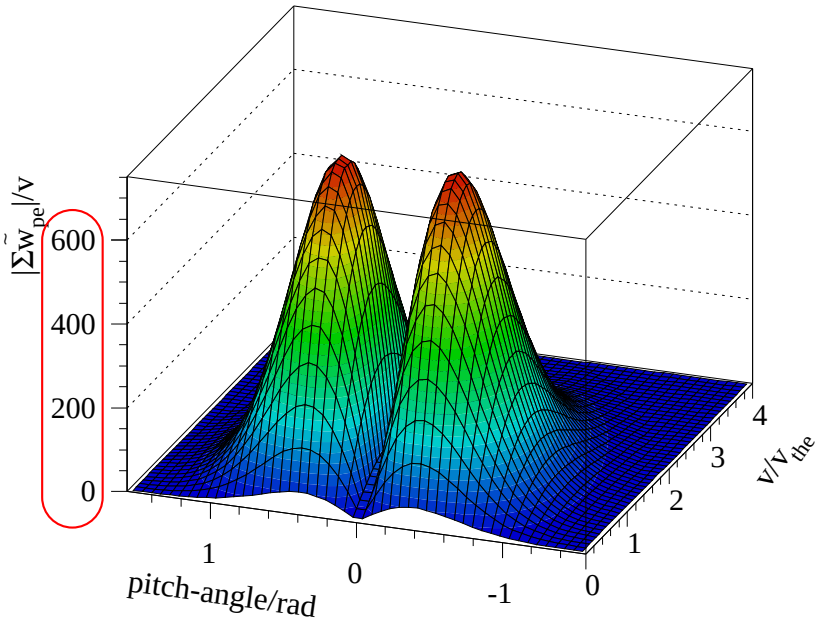
But, the statistical error ϵ **of the according Monte Carlo integration can be large!**

$\Rightarrow$ **The cancellation problem has its counterpart in the quasineutrality equation.**

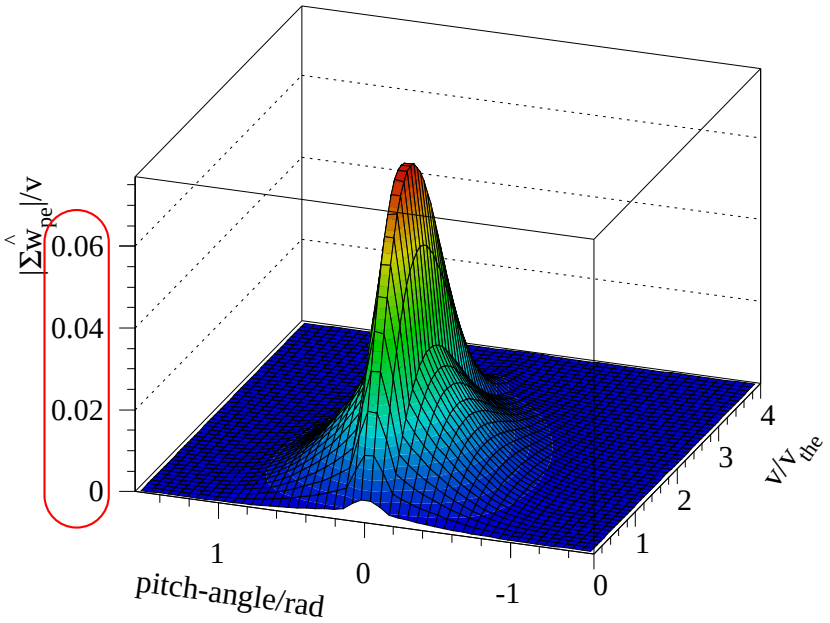
R. Hatzky, A. Könies, and A. Mishchenko, *Journal of Computational Physics*, **225: p. 568–590 (2007).**

The electron distribution function in velocity space at the MHD-limit

adiabatic and nonadiabatic part



nonadiabatic part



$\Rightarrow$ **The nonadiabatic part can be drastically smaller than the adiabatic part.**

Control variates as variance reduction method

One tries to utilize (strong) correlation between the observed variable X and some auxiliary variable Y, the so called control variable whose expected value $\mathcal{E}[Y] = \nu$ has to be known analytically.

The task is to estimate the expected value $\mathcal{E}[X] = \mu$ with a preferably smaller standard deviation than $\sqrt{\mathcal{V}[X]}$.

Hence, we define the variable Z which has the same expected value as $\mathcal{E}[X]$ by

$$Z \stackrel{\text{def}}{=} X - \alpha(Y - \nu) = \tilde{Z} + \alpha\nu \qquad \textbf{where} \qquad \tilde{Z} \stackrel{\text{def}}{=} X - \alpha Y$$

with

$$\mathcal{E}[Z] = \mathcal{E}[\tilde{Z}] + \alpha\nu = \mathcal{E}[X] - \alpha(\mathcal{E}[Y] - \nu) = \mathcal{E}[X].$$

Only the variable $\tilde{Z}$ will be discretized by our control variate schemes as the expected value ν is known analytically and can be added accordingly.

The variance of Z is

$$\mathcal{V}[Z] = \mathcal{V}[\tilde{Z}] = \mathcal{V}[X - \alpha Y] = \mathcal{V}[X] - 2\alpha \mathrm{Cov}[X, Y] + \alpha^2 \mathcal{V}[Y]$$

where the covariance is defined by

$$\mathrm{Cov}[X, Y] \stackrel{\mathrm{def}}{=} \mathcal{E}[(X - \mu)(Y - \nu)].$$

It follows the following condition:

$$\frac{\mathrm{Cov}[X, Y]}{\alpha \mathcal{V}[Y]} > \frac{1}{2} \qquad \Rightarrow \qquad \mathcal{V}[\tilde{Z}] < \mathcal{V}[X].$$

Effective control variates

The auxiliary variable Y is an effective control variate if $\mathcal{V}[\tilde{Z}] < \mathcal{V}[X]$.

For PIC simulations one can use the knowledge about the initial state $f(t_0)$ of the system to construct an effective control variate as long as the system does not evolve too far from its initial state.

For such situations the usage of a control variate is a valuable enhancement of the full-f PIC method which has naturally problems to resolve relatively small changes of the system.

The standard error $\epsilon \propto \sqrt{\mathcal{V}[\tilde{Z}]}$ (statistical noise) can be reduced in some cases drastically.

Construction of an effective control variates

A maximal reduction of the variance $\mathcal{V}[Z]$ **is given by the optimal parameter**

$$\alpha^\star \stackrel{\text{def}}{=} \frac{\text{Cov}[X,Y]}{\mathcal{V}[Y]}$$

which leads always to an optimal control variate. In practice, $\alpha^\star$ **has to be estimated with a sufficiently small statistical error.**

The corresponding optimal variance is given as

$$\mathcal{V}[Z] = \mathcal{V}[X] - \frac{\text{Cov}^2[X,Y]}{\mathcal{V}[Y]} = \mathcal{V}[X](1 - \text{Corr}^2[X,Y])$$

where the correlation coefficient is defined by

$$\text{Corr}[X,Y] \stackrel{\text{def}}{=} \frac{\text{Cov}[X,Y]}{\sqrt{\mathcal{V}[X]}\sqrt{\mathcal{V}[Y]}} \qquad \textbf{with} \qquad -1 \leq \text{Corr}[X,Y] \leq 1$$

- **The quality of the auxiliary variable Y as a control variate depends on the correlation between the variables X and Y.**
- **As long as X and Y are correlated the optimal parameter $\alpha^\star$ always reduces the variance of $\mathcal{V}[Z]$ even if X and Y are negatively correlated.**
- **In practice, the optimal parameter $\alpha^\star$ can be estimated from the Monte Carlo data.**
- **The performance does not depend strongly on α when α is close to $\alpha^\star$, where the derivative $\mathrm{d}\mathcal{V}[Z]/\mathrm{d}\alpha$ is zero.**
- **For very strong correlation between the variables X and Y the optimal parameter $\alpha^\star$ can be approximated by one.**

Example of variance reduction by a control variate

$$f(x) \stackrel{\text{def}}{=} a + b\sin(x), \qquad g(x) \stackrel{\text{def}}{=} \frac{1}{2\pi} \qquad \textbf{where} \qquad x \in [0, \ldots, 2\pi]$$

$$\mathcal{E}(f) \stackrel{\text{def}}{=} \int_0^{2\pi} f(x)\, g(x)\, \mathrm{d}x \qquad \Rightarrow \qquad \mathcal{E}(f) = a$$

$$\mathcal{V}(f) \stackrel{\text{def}}{=} \int_0^{2\pi} (f(x) - \mathcal{E}(f))^2\, g(x)\, \mathrm{d}x = \mathcal{E}(f^2) - [\mathcal{E}(f)]^2 \qquad \Rightarrow \qquad \mathcal{V}(f) = \frac{b^2}{2}$$

- **The parameter a influences only the expected value $\mathcal{E}$.**
- **The parameter b influences only the variance $\mathcal{V}$.**

Simple control variate

We choose the control variate:

$$f_0(x) \stackrel{\text{def}}{=} \begin{cases} c & : \quad 0 \le x \le \pi \\ -c & : \quad \pi < x \le 2\pi \end{cases} \qquad \textbf{and} \qquad \mathcal{E}(f_0) = 0, \qquad \mathcal{V}(f_0) = c^2$$

$$f - f_0 = a \mp c + \frac{1}{2\pi} b \sin(x) \qquad \Rightarrow \qquad \mathcal{E}(f - f_0) = a, \quad \mathcal{V}(f - f_0) = \frac{b^2}{2} + c^2 - \frac{2b}{\pi} c$$

We have an effective control variate under the assumption:

$$\mathcal{V}(f - f_0) \le \mathcal{V}(f): \qquad 0 \le \frac{2b}{\pi} c - c^2 \qquad \Rightarrow \qquad c \le \frac{2b}{\pi} \qquad \textbf{with} \qquad c_{\text{opt}} = \frac{b}{\pi}$$

Optimized control variate

We choose the optimized control variate:

$$\tilde{f}(x) \stackrel{\text{def}}{=} f - \alpha^{\star}[f_0 - \mathcal{E}(f_0)] \qquad \textbf{where} \qquad \alpha^{\star} \stackrel{\text{def}}{=} \frac{\mathrm{Cov}(f, f_0)}{\mathcal{V}(f_0)} = \frac{b}{\pi c}$$

The optimized control variate is better than the simple control variate:

$$\mathcal{V}(\tilde{f}) = \frac{b^2}{2} - \frac{b^2}{\pi^2} = \mathcal{V}(f - f_0(c_{\text{opt}})) \qquad \Rightarrow \qquad \mathcal{V}(\tilde{f}) \leq \mathcal{V}(f - f_0(c))$$

We always have $\mathcal{V}(\tilde{f}) \leq \mathcal{V}(f)$ **and for** $\alpha^{\star} = 0 \quad \Rightarrow \quad \mathcal{V}(\tilde{f}) = \mathcal{V}(f)$**.**

Second example[1] of variance reduction by a control variate

$$f(x) \stackrel{\text{def}}{=} ax + b\sin(x), \qquad g(x) \stackrel{\text{def}}{=} \frac{1}{2\pi} \qquad \textbf{where} \qquad x \in [0, \ldots, 2\pi]$$

$$\mathcal{E}(f) \stackrel{\text{def}}{=} \int_0^{2\pi} f(x)\, g(x)\, \mathrm{d}x, \qquad \mathcal{V}(f) \stackrel{\text{def}}{=} \int_0^{2\pi} (f(x) - \mathcal{E}(f))^2\, g(x)\, \mathrm{d}x$$

$$\Rightarrow \qquad \mathcal{E}(f) = \pi a \qquad \textbf{and} \qquad \mathcal{V}(f) = \frac{\pi^2}{3}a^2 + \frac{1}{2}b^2 - 2ab$$

[1] **private communication R. Kleiber**

Simple control variate

We choose the control variate:

$$f_0(x) \stackrel{\text{def}}{=} ax \qquad \textbf{and} \qquad \mathcal{E}(f_0) = \pi a, \qquad \mathcal{V}(f_0) = \frac{\pi^2}{3}a^2$$

$$f - f_0 = \frac{1}{2\pi}\, b \sin(x) \qquad \Rightarrow \qquad \mathcal{E}(f - f_0) = 0, \qquad \mathcal{V}(f - f_0) = \frac{1}{2}b^2$$

We have an effective control variate under the assumption:

$$\mathcal{V}(f - f_0) \leq \mathcal{V}(f): \qquad 0 \leq \frac{\pi^2}{3}a^2 - 2ab \qquad \Rightarrow \qquad b \leq \frac{\pi^2}{6}a$$

Optimized control variate

We choose the optimized control variate:

$$\tilde{f}(x) \stackrel{\text{def}}{=} f - \alpha^\star [f_0 - \mathcal{E}(f_0)] \qquad \textbf{where} \qquad \alpha^\star \stackrel{\text{def}}{=} \frac{\mathrm{Cov}(f, f_0)}{\mathcal{V}(f_0)} = 1 - \frac{ab}{\mathcal{V}(f_0)}$$

The optimized control variate is always better than the simple control variate:

$$\mathcal{V}(\tilde{f}) = \left(1 - \frac{6}{\pi^2}\right) \frac{1}{2} b^2 = \underbrace{\left(1 - \frac{6}{\pi^2}\right)}_{\approx 0.39} \mathcal{V}(f - f_0) \qquad \Rightarrow \qquad \mathcal{V}(\tilde{f}) < \mathcal{V}(f - f_0)$$

We always have $\mathcal{V}(\tilde{f}) \leq \mathcal{V}(f)$ and for $\alpha^\star = 0 \quad \Rightarrow \quad \mathcal{V}(\tilde{f}) = \mathcal{V}(f)$.

Charge assignment with a control variate

$$b(t) = \frac{1}{N}\sum_{p=1}^{N} w_p\,\Lambda(\mathbf{R}_p) \simeq \frac{1}{N}\sum_{p=1}^{N} \Omega_p \underbrace{\left[f_p - f_0(\mathbf{R}_p(t),\mathbf{v}_p(t))\right]}_{\tilde{Z}=X-Y}\Lambda(\mathbf{R}_p) + \underbrace{\int f_0\,\Lambda\,\mathrm{d}\mathbf{R}\,\mathrm{d}\mathbf{v}}_{\mathcal{E}[Y]=\nu}$$

$$= \frac{1}{N}\sum_{p=1}^{N} \underbrace{\left[w_p - \Omega_p f_{0,p}(t)\right]}_{\tilde{w}_p(t)}\Lambda(\mathbf{R}_p) + \hat{b} \qquad \textbf{where} \qquad \hat{b} \overset{\text{def}}{=} \int f_0\,\Lambda\,\mathrm{d}\mathbf{R}\,\mathrm{d}\mathbf{v}$$

- **The B-spline coefficient vector $\hat{\mathbf{b}}$ is the result of the analytic projection of the control variate f_0 onto the B-spline basis.**
- **As long as the control variate is time independent the B-spline coefficient vector $\hat{\mathbf{b}}$ has to be calculated only once at the initialization.**

The δf method

The popular δf method chooses the following ansatz:

$$\delta f = f - f_0$$

For PIC simulations it is only used as an efficient noise reduction method if

$$\delta f \ll f_0 = f_{\mathrm{M}}$$

It can be interpreted as a control variate method with $\alpha = 1$:

$$b = \frac{1}{N}\sum_{p=1}^{N} \Omega_p [f_p - f_{\mathrm{M},p}]\, \Lambda(\mathbf{R}_p) + \hat{b} = \frac{1}{N}\sum_{p=1}^{N} \Omega_p \delta f_p\, \Lambda(\mathbf{R}_p) + \hat{b} = \frac{1}{N}\sum_{p=1}^{N} \tilde{w}_p(t)\, \Lambda(\mathbf{R}_p) + \hat{b}$$

A.Y. Aydemir, *Physics of Plasmas*, **1, p. 822–831 (1994).**

The adjustable control variate

An effective control variate can be constructed under the assumption of the presence of a dominant part responding adiabatically to the magnetic potential in the perturbation of the distribution function of the electrons:

$$\delta f_{\mathrm{e}} \approx \delta f_{\mathrm{e}}^{\mathrm{ad}} = -\frac{e v_{\parallel} f_{\mathrm{M,e}}}{k_{\mathrm{B}} T_{\mathrm{e}}} A_{\parallel}$$

We define the stochastic variable $\tilde{Z}$ by introducing a "noise reduced" species of electron marker weights:

$$\bar{w}_{\mathrm{e},p} \stackrel{\mathrm{def}}{=} \Omega_{\mathrm{e},p}(\delta f_{\mathrm{e}} - \delta f_{\mathrm{e}}^{\mathrm{ad}}) = \tilde{w}_{\mathrm{e},p} + \Omega_{\mathrm{e},p} \frac{e v_{\parallel} f_{\mathrm{M,e}}}{k_{\mathrm{B}} T_{\mathrm{e}}} A_{\parallel} \Big|_{R_p, v_{\parallel p}, \mu_p}$$

Problem: We don't know $A_{\parallel}$ in advance!

Ampère's law discretized in a B-spline finite element basis

$$(L + S_{\text{i}} + S_{\text{e}})\,\mathbf{c} = \mu_0 \langle \mathbf{j}_{\parallel \text{i}} \rangle + \underbrace{\mu_0 \mathbf{j}_{\parallel \text{e}} - J^{\text{ad}}_{\parallel \text{e}}\,\mathbf{c}}_{\tilde{Z}=X-Y} + \underbrace{S_{\text{e}}\mathbf{c}}_{\mathcal{E}[Y]=\nu}$$

Capital letters depict matrices, $\mathbf{c}$ is the B-spline coefficient vector to be solved for and $\mathbf{j}_{\parallel}$'s are the coefficient vectors after current assignment of the weights.

$$\Downarrow$$

The cancellation problem can be resolved analytically by eliminating the electron skin term $S_{\text{e}}\mathbf{c}$ on both sides:

$$(L + S_{\text{i}} + J^{\text{ad}}_{\parallel \text{e}})\,\mathbf{c} = \mu_0 (\langle \mathbf{j}_{\parallel \text{i}} \rangle + \mathbf{j}_{\parallel \text{e}})$$

$$\Downarrow$$

This "noise reduced" matrix equation can be solved for $\mathbf{c}$ to achieve $A_{\parallel}(x)$.

Caveats of the scheme

The matrix operator $J^{\text{ad}}_{\|\text{e}}$ depends on the marker positions $(\mathrm{R}, v_\|, v_\perp)$:

- **The matrix has to be built up every time step**
- **In case of a direct solver a costly Cholesky decomposition is obligatory at every time step**

The iterative scheme

$$\mathbf{c} = (L+S)^{-1}[I-M]^{-1}\mathbf{b} = (L+S)^{-1}\left(\mathbf{b} + \sum_{i=1}^{\infty}\mathbf{b}^{(i)}\right)$$

where

$$S \overset{\text{def}}{=} (S_\text{i} + S_\text{e}), \quad M \overset{\text{def}}{=} (S_\text{e} - J^{\text{ad}}_{\|\text{e}})(L+S)^{-1}, \quad \mathbf{b} \overset{\text{def}}{=} \mu_0(\langle \mathbf{j}_{\|\text{i}} \rangle + \mathbf{j}_{\|\text{e}}), \quad \mathbf{b}^{(i)} \overset{\text{def}}{=} M^i\,\mathbf{b}$$

We use the Neumann series with the sufficient convergence condition $\|M\| < 1$.

The iterative scheme

The solution vector $\mathbf{c} = \lim_{i\to\infty} c^{(i)}$ **can be calculated iteratively by using the magnetic potential** $A_{\parallel}^{(n)}$ **from the previous iteration:**

$$\begin{aligned} \mathbf{c}^{(0)} &= (L + S_{\mathrm{i}} + S_{\mathrm{e}})^{-1}\,\mu_0(\langle \mathbf{j}_{\parallel \mathrm{i}} \rangle + \mathbf{j}_{\parallel \mathrm{e}}) \\ &\cdots \\ \mathbf{c}^{(n+1)} &= (L + S_{\mathrm{i}} + S_{\mathrm{e}})^{-1}[\mu_0(\langle \mathbf{j}_{\parallel \mathrm{i}} \rangle + \mathbf{j}_{\parallel \mathrm{e}}) + (S_{\mathrm{e}} - J_{\parallel \mathrm{e}}^{\mathrm{ad}})\mathbf{c}^{(n)}] \end{aligned}$$

The convergence condition $\|(S_{\mathrm{e}} - J_{\parallel \mathrm{e}}^{\mathrm{ad}})(L + S_{\mathrm{i}} + S_{\mathrm{e}})^{-1}\| < 1$ **seems to be benign.**

Usually just one or two iterations are sufficient!

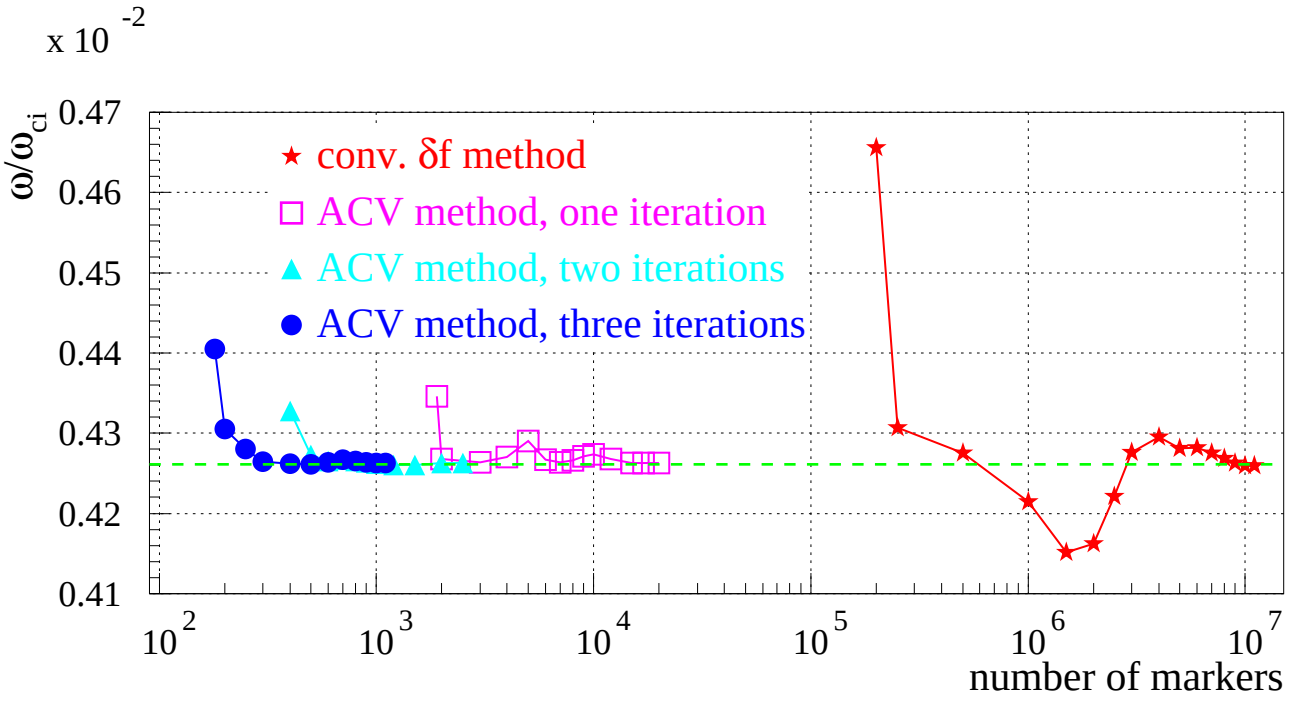
R. Hatzky, A. Könies, and A. Mishchenko, *Journal of Computational Physics*, **225: p. 568–590 (2007).**

Numerical results

Shear Alfvén wave in the MHD limit $k_\perp \to 0$

Slab model:

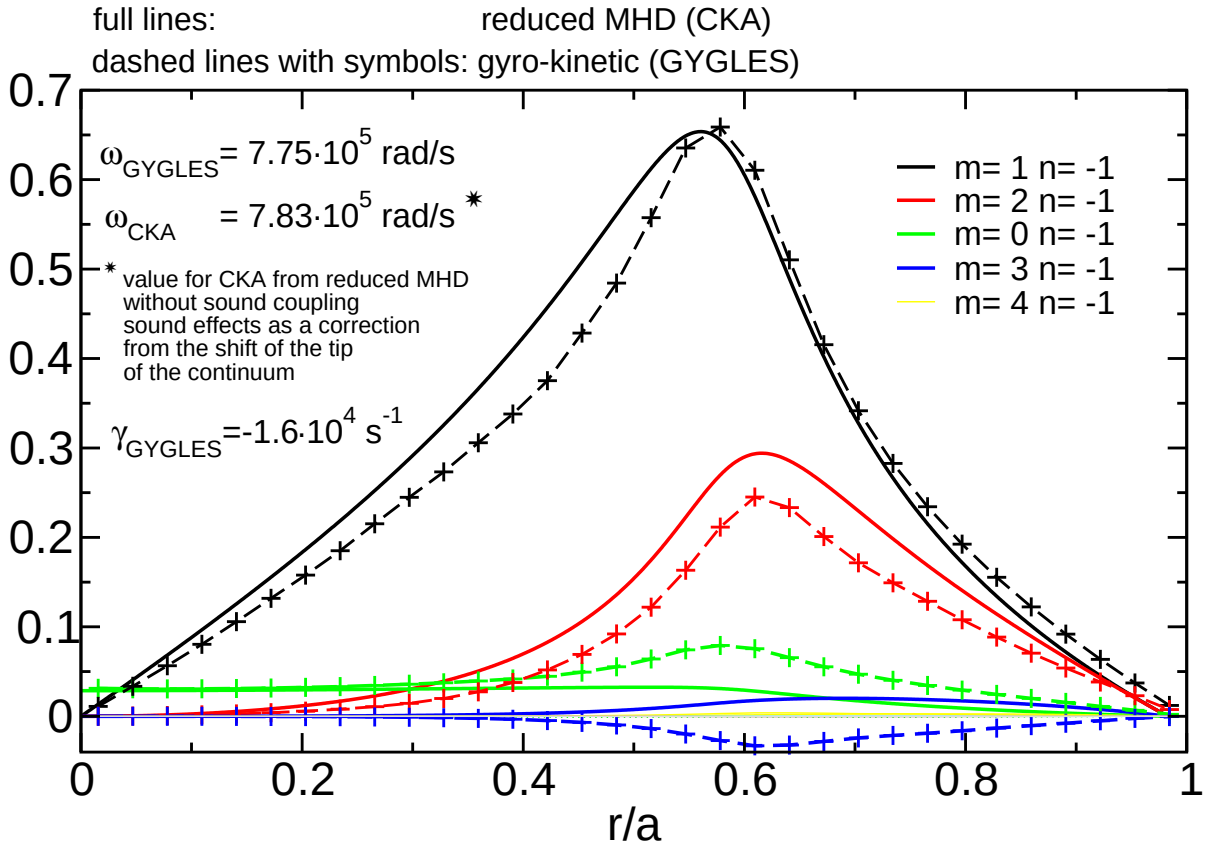
$k_x \rho_i = 0.023$
$k_y \rho_i = 0.015$
$k_z \rho_i = 7.43 \times 10^{-4}$
$\beta = 3.04\ \%$



Reduction of the number of markers by more than four orders of magnitude!

R. Hatzky, A. Könies, and A. Mishchenko, *Journal of Computational Physics*, **225: p. 568–590 (2007).**

Global linear tokamak simulation of a low n TAE with the GYGLES code



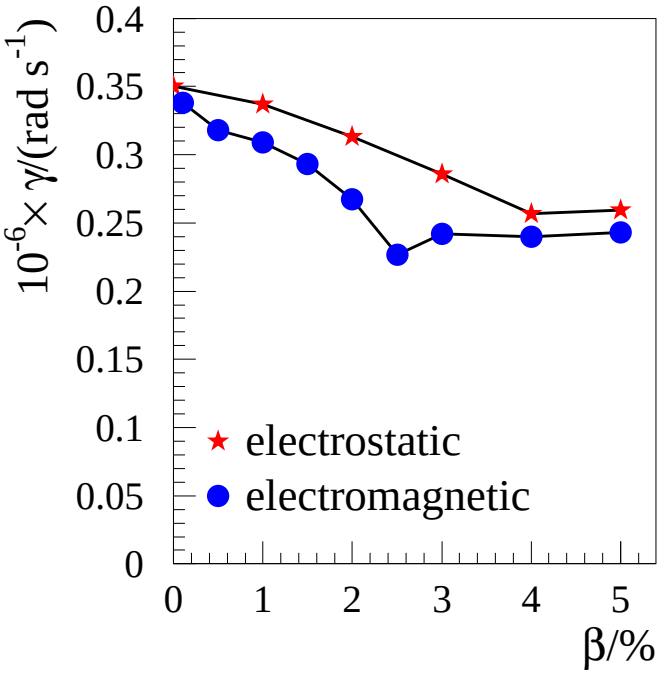
A. Könies et al., *12^{th} IAEA Technical Meeting on Energetic Particles in Magnetic Confinement Systems*, **2011.**

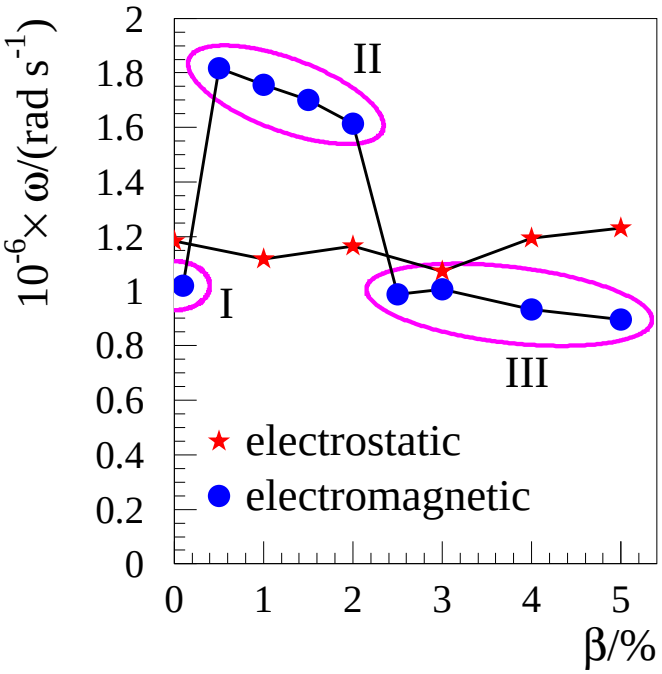
Global electromagnetic stellarator simulation with the EUTERPE code

- **Wendelstein 7-X equilibrium from VMEC code**
- **Linear simulation of electromagnetic ITG with** $\beta = 0, \ldots, 5\ \%$
- **Fully kinetic electrons** $\Rightarrow$ **small time step**
- **Marker numbers:** $N_{\mathrm{i}} = 32 \cdot 10^6$, $N_{\mathrm{e}} = 128 \cdot 10^6$
- **Grid size** $(s, \vartheta, \varphi) : 100 \times 128 \times 128$
- **Just one iteration for the adjustable control variate scheme needed**
- **Execution time on HPC-FF:** $\approx 100\,000$ **CPU hours on 128 cores**

EUTERPE originates from CRPP; now it is developed at IPP by R. Kleiber, R. Hatzky, and M. Borchardt

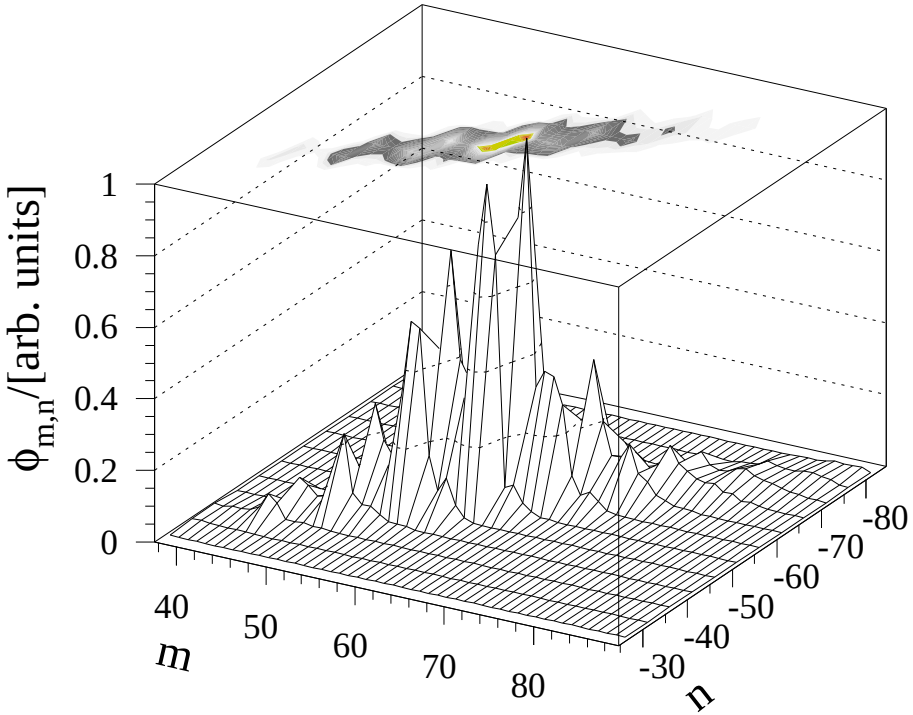
The electrostatic and electromagnetic simulation of ITG modes in W7-X

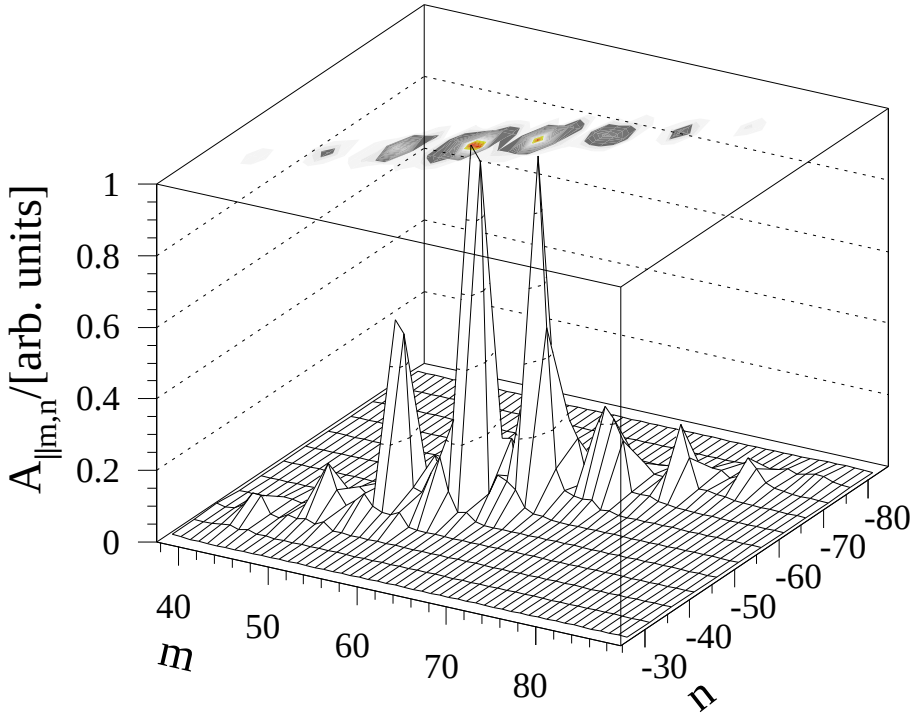




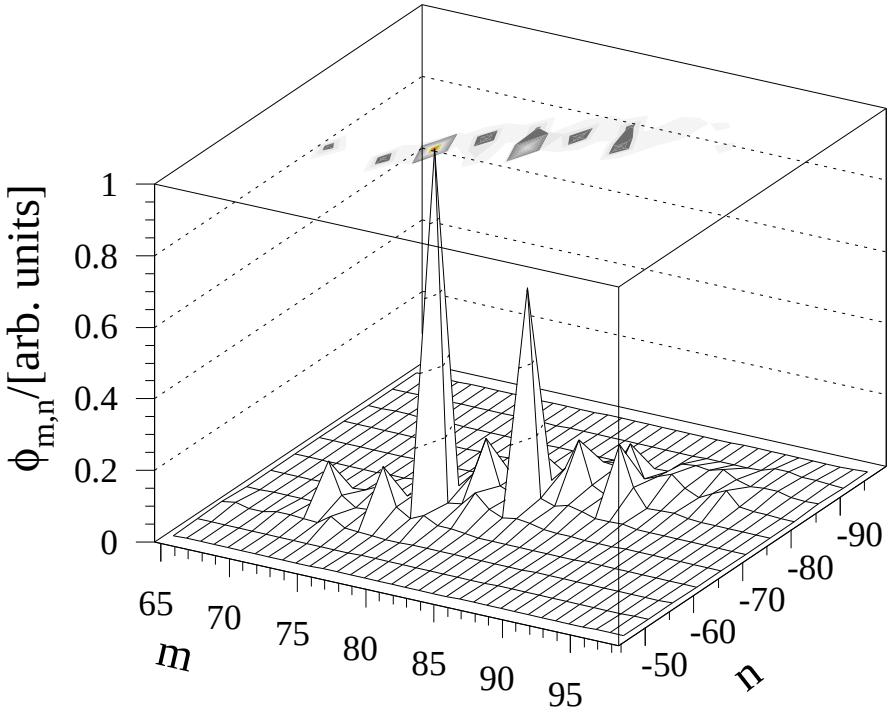
- **There is a significant difference between both models for $\beta > 0.1$ %**
- **It is possible to identify three regimes for the electromagnetic case**

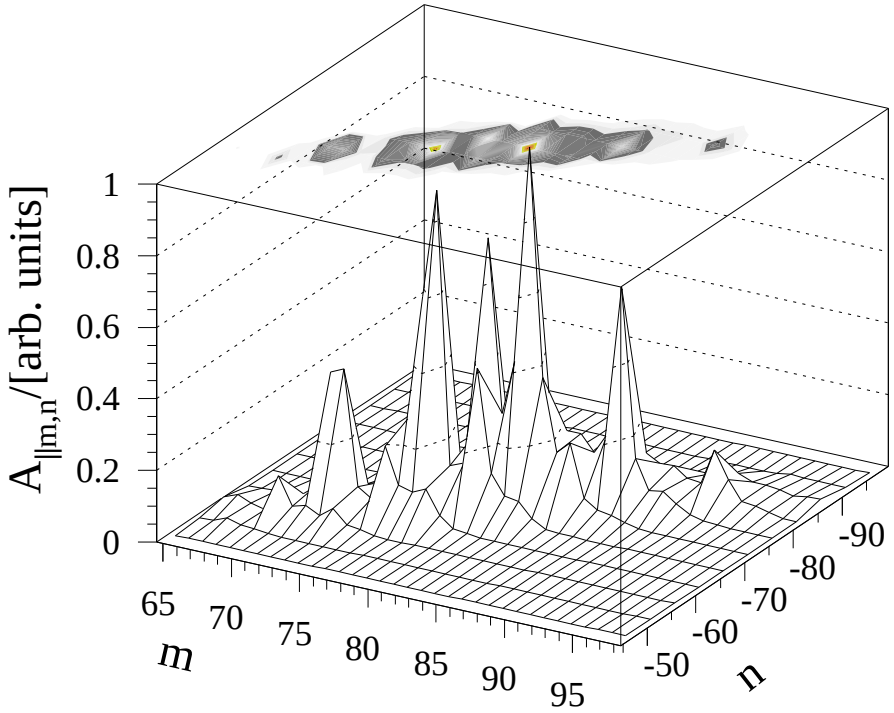
Electromagnetic Fourier spectra of an ITG mode in W7-X for $\beta = 0.1$ %



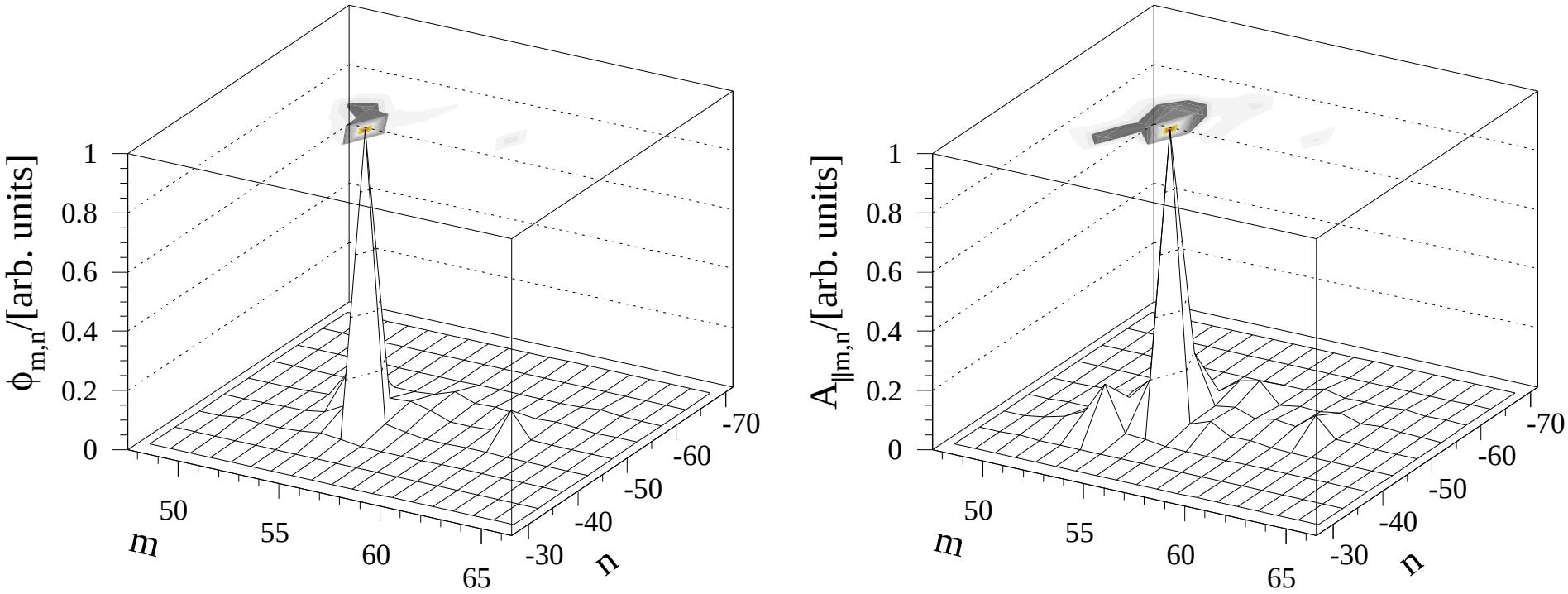


Electromagnetic Fourier spectra of an ITG mode in W7-X for $\beta = 0.5$ %



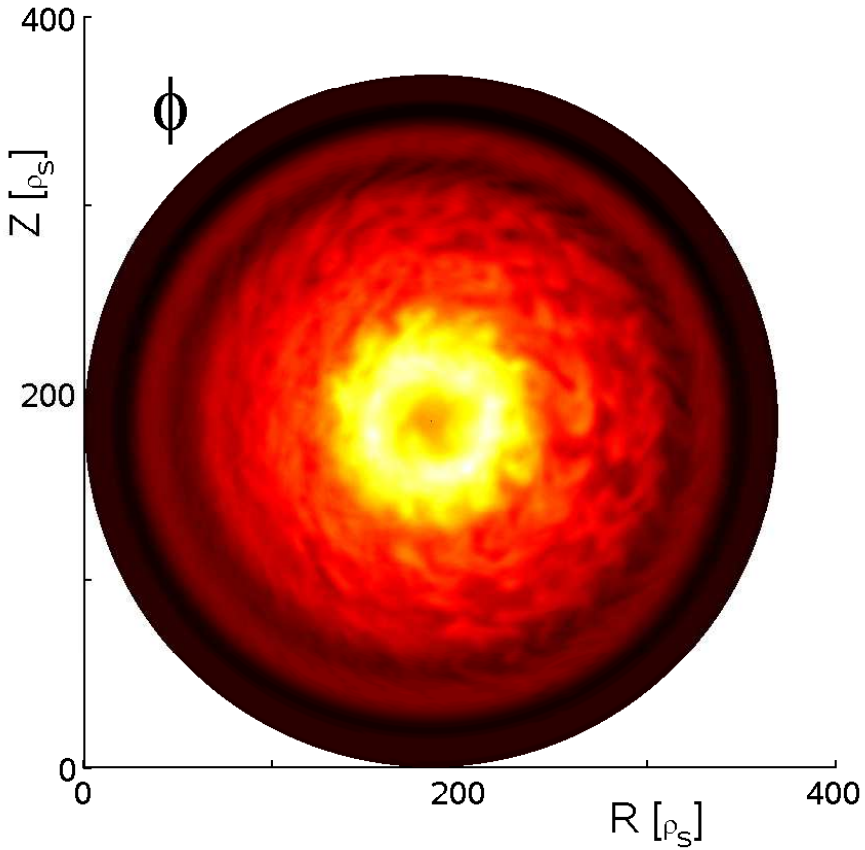


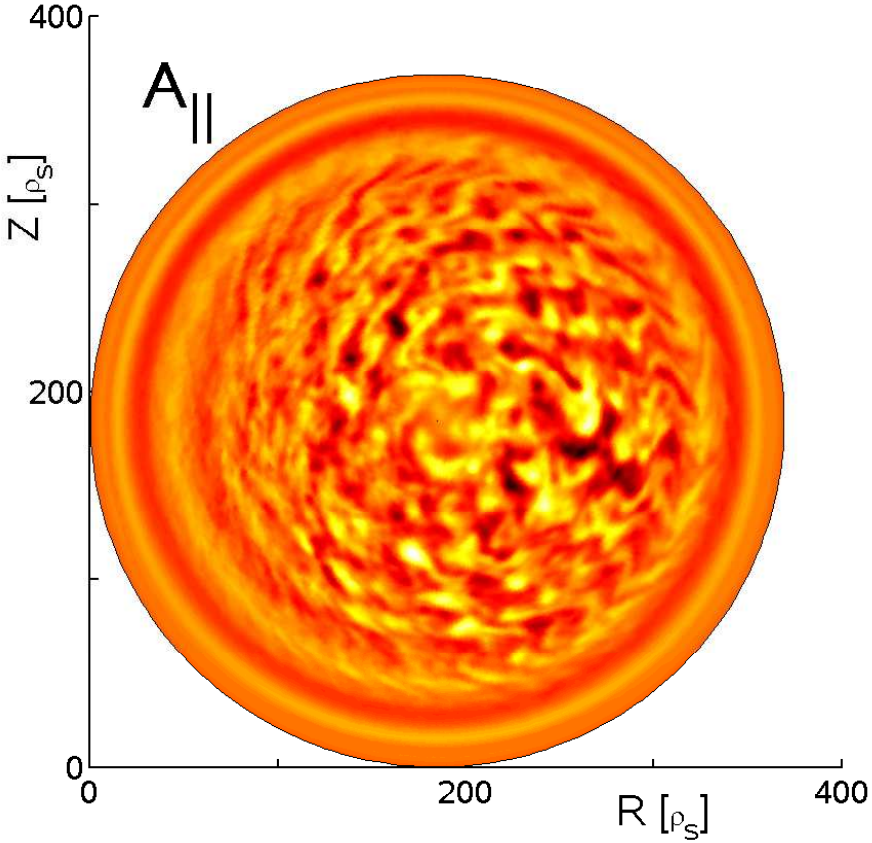
Electromagnetic Fourier spectra of an ITG mode in W7-X for $\beta = 5.0\ \%$



The mode has slab-like character!

Global nonlinear simulation in a tokamak with the NEMORB code





A. Bottino et al., *IEEE Transactions on Plasma Science*, **38, p. 2129–2135 (2010).**

Summary

- The established δf scheme for PIC is a control variate method belonging to a group of Monte Carlo methods used for variance reduction.
- There are no limitations concerning the geometry of the configuration.
- An iterative implementation of an adaptive control variate can be used to solve the electromagnetic "cancellation" problem very efficiently.
- The method works for linear and nonlinear simulations and has been implemented into the codes: EUTERPE, GYGLES and NEMORB.
- Gyrokinetic electromagnetic PIC simulations become more and more standard.

References

- **C.K. Birdsall and A.B. Langdon:** *Plasma physics via computer simulation*, **in Plasma Physics Series edited by E. W. Laing (Institute of Physics Publishing, Bristol, 1995).**
- **R.W. Hockney and J.W. Eastwood:** *Computer simulation using particles* **(Adam Hilger, Bristol, 1989).**
- **T.S. Hahm,** *Nonlinear gyrokinetic equations for tokamak microturbulence*, **Phys. Fluids 31, p. 2670–2673 (1988).**
- **A.Y. Aydemir,** *A unified Monte Carlo interpretation of particle simulations and applications to non-neutral plasmas*, **Physics of Plasmas 1, p. 822–831 (1994).**
- **R. Hatzky, A. Könies, and A. Mishchenko,** *Journal of Computational Physics*, **225: p. 568–590 (2007)**
- **URL: www.efda-hlst.eu**