DE LA RECHERCHE À L'INDUSTRIE

www.cea.fr

Exascale needs and bottlenecks for semi-lagrangian gyrokinetic simulations of turbulence in tokamak plasmas ➠ GYSELA code

V. Grandgirard[1]

Collaborations with physicists:

J. Abiteboul[2], G. Dif-pradalier[1], Y. Dong[3], D. Estève[1], X. Garbet[1], Ph. Ghendrih[1], J.B Girardo[1], C. Norscini[1], F. Palermo[1], Y. Sarazin[1], A. Strugarek[7], D. Zarzoso[2]

Collaborations with mathematicians:

A. Back[4], T. Cartier-Michaud[1], M. Mehrenberger[5], L. Mendoza[2], E. Sonnendrücker[2]

Collaborations with computer scientists:

G. Latu[1], J. Bigot[6], C. Passeron[1], F. Rozar[1,6]

[1] CEA, IRFM, Cadarache, France
[2] IPP Garching, Germany
[3] LPP, Paris, France
[4] CPT, Marseille, France
[5] IRMA, Strasbourg, France
[6] Maison de la Simulation, Saclay, France
[7] Montreal university, Canada

ANR GYPSI - ANR G8-Exascale Nufuse

ADT-INRIA SELALIB - AEN-INRIA Fusion

Turbulence governs Fusion plasma performance

- Scaling law in tokamaks: plasma volume $\times \tau_E \approx$ cte
  with τ_E = energy confinement time $\sim$ measure of thermal insulation.

➠ Two main possibilities to increase tokamak performances:

❶ increase the size of the machine or/and ❷ increase τ_E

- Turbulence governs τ_E
  - ➠ Generates loss of heat and particles
  - ➠ ↘ Confinement properties of the magnetic configuration

- Understanding, predicting and controlling turbulence for optimizing experiments like ITER and future reactors is a subject of utmost importance.

Outline

1. Gyrokinetic theory

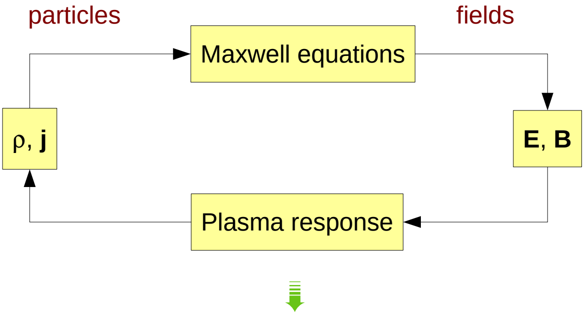
2. GYSELA code
   - Semi-lagrangian approach
   - MPI/OpenMP parallelisation
   - Global flux driven simulation

3. Exascale needs and associated challenges
   - Increase of core number : scalability, fault tolerance
   - Memory reduction and big data
   - Continuous integration

Complex interplay between fields and particles

- Charged particle motion governed by electromagnetic fields
- Electromagnetic fields governed by charge ρ and current $\mathbf{j}$ densities

➠ self-consistent treatment required



- Plasma response: The most accurate $\Rightarrow$ Kinetic

- **Fields ➠ Maxwell's equations**
  - Electrostatic ($\mathbf{B}$ = const): $\mathbf{E} = -\nabla\phi$ (ϕ electrostatic potential)
  - "large scale" ($> \lambda_{\text{Debye}} \sim 10^{-4} m$)

    ➠ Quasi-neutrality equation:

$$\rho(\mathbf{x}, t) = \sum_s n_s q_s = 0 \qquad \text{with} \qquad n_s = \int f_s d\mathbf{v}$$

- **Particles ➠ Kinetic approach mandatory**
  - Fusion plasmas weakly collisional $\Rightarrow$ fluid description not appropriate

    ➠ Boltzmann equation:

$$\frac{\partial f_s}{\partial t} + \mathbf{v} \cdot \frac{\partial f_s}{\partial \mathbf{x}} + \frac{d\mathbf{v}}{dt} \cdot \frac{\partial f_s}{\partial \mathbf{v}} = C(f_s) + S$$

6D function of s specie $f_s(\mathbf{x}, \mathbf{v})$ (3D in space and 3D in velocity)

Kinetic theory: ➟ 6D distribution function of particles
(3D in space and 3D in velocity) $F_s(r, \theta, \varphi, v_\parallel, v_\perp, \alpha)$

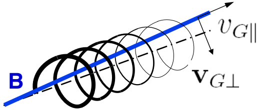
- Fusion plasma turbulence is low frequency:

$$\omega_{\text{turb}} \sim 10^5 s^{-1} \ll \omega_{ci} \sim 10^8 s^{-1}$$

- Phase space reduction: fast gyro-motion is averaged out
  - ➟ Adiabatic invariant: magnetic moment $\mu = m_s v_\perp^2/(2B)$
  - ➟ Velocity drifts of guiding centers

☺ Large reduction memory/CPU time

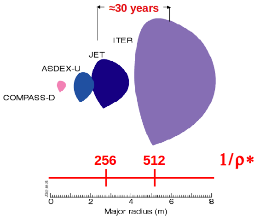
☹ Complexity of the system



Gyrokinetic theory: ➟ 5D distribution function of guiding-centers
$\bar{F}_s(r, \theta, \varphi, v_{G\parallel}, \mu)$ where μ parameter

Gyrokinetic codes require state-of-the-art HPC

- Gyrokinetic codes require state-of-the-art HPC techniques and must run efficiently on several thousands processors.
  - non-linear 5D simulations
  - multi-scale problem in space and time
    - time: $\Delta t \approx \gamma^{-1} \sim 10^{-6}s \rightarrow t_{\text{simul}} \approx \text{few } \tau_E \sim 10s$
    - space: $\rho_i \rightarrow$ machine size a $\quad \boxed{\rho_* \equiv \frac{\rho_i}{a} \ll 1}$



✓ $\rho_{*\,\text{ITER}} = 1/512$

✓ Number grid points $\sim (\rho_*)^{-3}$

⇓

Huge mesh for global simulations

ex: $1024^3 \times 128 v_\| \times 16\mu$

➠ several billiard of points

There are about ten 5D gyrokinetic codes for plasma fusion in the world.

- Various simplifications:
  - δf codes: scale separation between equilibrium and perturbation.
  - Flux-tube codes $\Rightarrow$ the domain considered is a vicinity of a magnetic field line.
  - Fixed gradient boundary conditions.
  - Collisionless.

- Various numerical schemes:
  - Lagrangian (PIC), Eulerian or Semi-Lagrangian

- A new generation of global full-f gyrokinetic codes is being developed with collisions and flux-driven boundary conditions.

GYSELA (GYrokinetic SEmi-LAgrangian code) is one of them

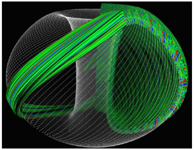
Global simulations required huge meshes

① "Flux-tube" approach (local)

- Simulate only a vicinity of magnetic field line

☺ drastic reduction of mesh size
+ periodic boundary conditions

☹ small scale structures only

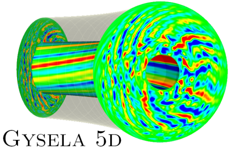


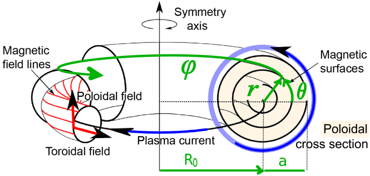
② Global approach

- Simulate the whole domain

☺ Capture large scale events

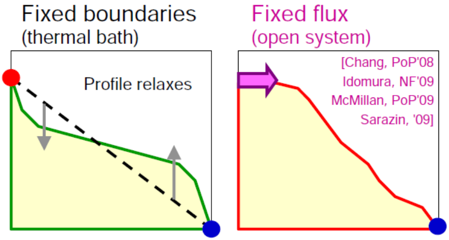
☹ Extremely large 3D meshes
+ boundary conditions





Flux-Driven boundary conditions
➠ Long-time simulations

- Vanishing gradient boundary conditions at inner boundary
  $\rightarrow$ temperature and flows evolve freely



☺ Source terms aims at maintaining the equilibrium profiles, which would otherwise relax towards marginal state

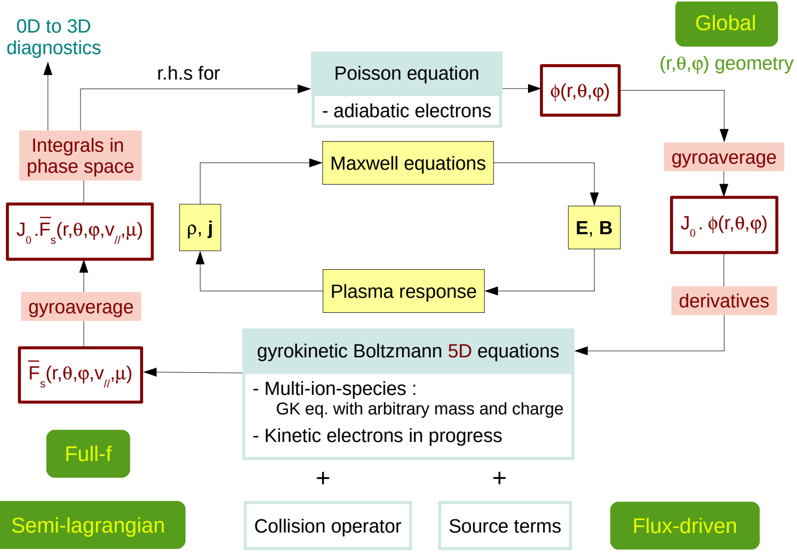
➠ Long-time simulations are available

$\Rightarrow$ Extremely expensive in terms of CPU time.

Outline

1. Gyrokinetic theory
2. GYSELA code
   - Semi-lagrangian approach
   - MPI/OpenMP parallelisation
   - Global flux driven simulation
3. Exascale needs and associated challenges
   - Increase of core number : scalability, fault tolerance
   - Memory reduction and big data
   - Continuous integration

GYSELA code - main characteristics



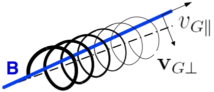
GYSELA - 5D Boltzmann equation

- Time evolution of the gyrocenter distribution function for s species $\bar{F}_s(r,\theta,\varphi,v_\parallel,\mu)$ governed by 5D gyrokinetic Fokker-Planck equation with an additional realistic heating source:

$$B^*_{\parallel s}\frac{\partial \bar{F}_s}{\partial t} + \nabla\cdot\left(\frac{d\mathbf{x_G}}{dt}B^*_{\parallel s}\bar{F}_s\right) + \frac{\partial}{\partial v_{G\parallel}}\left(\frac{dv_{G\parallel}}{dt}B^*_{\parallel s}\bar{F}_s\right) = \underbrace{C(\bar{F}_s)}_{\text{collision operator}} + \underbrace{S}_{\text{heating source}}$$

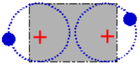
where $\frac{d\mathbf{x_G}}{dt} = \mathbf{v}_G = v_{G\parallel}\mathbf{b} + v_{G\perp}$

with $v_{G\perp} \approx \frac{\mathbf{E}\times\mathbf{B}}{B^2} + v_{d0}R\frac{\mathbf{B}\times\nabla B}{B^2}$



$\mathbf{E} = \nabla\left(J_0\cdot\phi\right)$ with $\phi(\mathbf{x})$ electrostatic potential and J_0 the gyroaverage operator

- Self-consistency ensured by a 3D quasi-neutrality equation:



$$\underbrace{\frac{e}{T_{e,\mathrm{eq}}}\left(\phi - \langle\phi\rangle_{\mathrm{FS}}\right)}_{\delta n_e \text{ for adiabatic electrons}} = \underbrace{\frac{1}{n_{e_0}}\sum_s Z_s\int J_0\cdot\left(\bar{F}_s - \bar{F}_{s,\mathrm{eq}}\right)d^3v}_{\sum_s \delta n_{GCs}} + \underbrace{\frac{1}{n_{e_0}}\sum_s Z_s\nabla_\perp\cdot\left(\frac{n_{s,\mathrm{eq}}}{B\Omega_s}\nabla_\perp\phi\right)}_{\delta n_{\text{polarization}} \text{ particles} \neq \text{guiding-centers}}$$

- Solving the 3D quasi-neutrality equation is equivalent to finding $\phi(r,\theta,\varphi)$ such that:

$$\frac{e}{T_{e,\mathrm{eq}}}\left(\phi - \langle\phi\rangle_{\mathrm{FS}}\right) - \frac{1}{n_{e_0}}\sum_s Z_s \nabla_\perp \cdot \left(\frac{n_{s,\mathrm{eq}}}{B\Omega_s}\nabla_\perp\phi\right) = \frac{1}{n_{e_0}}\sum_s Z_s \int J_0 \cdot \left(\bar{F}_s - \bar{F}_{s,\mathrm{eq}}\right) d^3v$$

- Numerical methods:
  - Fourier projection in periodic directions θ and φ
  - Finite differences in radial direction

- Difficulties:
  - ☹ R.H.S = integral over the velocity space
    $\Rightarrow$ *Parallel communications ++*
  - ☹ $\left\langle\phi\right\rangle_{\mathrm{FS}} = \int\int \phi\mathcal{J}_x \mathrm{d}\theta \mathrm{d}\varphi / \int\int \mathcal{J}_x \mathrm{d}\theta\mathrm{d}\varphi$ flux surface average of ϕ
    $\Rightarrow$ *Pb in Fourier due to coupling between θ and φ*

Time-splitting for Boltzmann equation

- A time-splitting of Strang is applied to the 5D non-linear Boltzmann equation:

$$B^*_{\parallel s}\frac{\partial \bar{F}_s}{\partial t} + \nabla\cdot\left(\frac{d\mathbf{x_G}}{dt}B^*_{\parallel s}\bar{F}_s\right) + \frac{\partial}{\partial v_{G\parallel}}\left(\frac{dv_{G\parallel}}{dt}B^*_{\parallel s}\bar{F}_s\right) = \mathcal{C}(\bar{F}_s) + S$$

- Let us define three advection operators (with $\mathcal{X}_G = (r,\theta)$)

$$B^*_{\parallel s}\frac{\partial \bar{F}_s}{\partial t} + \boldsymbol{\nabla}\cdot\left(B^*_{\parallel s}\frac{\mathrm{d}\mathcal{X}_G}{\mathrm{d}t}\bar{F}_s\right) = 0 \qquad : (\tilde{\mathcal{X}}_G)$$

$$B^*_{\parallel s}\frac{\partial \bar{F}_s}{\partial t} + \frac{\partial}{\partial\varphi}\left(B^*_{\parallel s}\frac{\mathrm{d}\varphi}{\mathrm{d}t}\bar{F}_s\right) = 0 \qquad : (\tilde{\varphi})$$

$$B^*_{\parallel s}\frac{\partial \bar{F}_s}{\partial t} + \frac{\partial}{\partial v_{G\parallel}}\left(B^*_{\parallel s}\frac{\mathrm{d}v_{G\parallel}}{\mathrm{d}t}\bar{F}_s\right) = 0 \qquad : (\tilde{v_{G\parallel}})$$

- And the collision operator $(\tilde{\mathcal{C}})$ on a Δt : $\partial_t\bar{F}_s = \mathcal{C}(\bar{F}_s)$
- And the source operator $(\tilde{\mathcal{S}})$ on a Δt : $\partial_t\bar{F}_s = \mathcal{S}$
- Then, a Boltzmann solving sequence $(\tilde{\mathcal{B}})$ is performed:

$$(\tilde{\mathcal{B}}) \equiv \left(\frac{\tilde{\mathcal{S}}}{2},\frac{\tilde{\mathcal{C}}}{2}\right)\left(\frac{\tilde{v_{G\parallel}}}{2},\frac{\tilde{\varphi}}{2},\tilde{\mathcal{X}}_G,\frac{\tilde{\varphi}}{2},\frac{\tilde{v_{G\parallel}}}{2}\right)\left(\frac{\tilde{\mathcal{C}}}{2},\frac{\tilde{\mathcal{S}}}{2}\right)$$

Time-splitting for Boltzmann equation

- A time-splitting of Strang is applied to the 5D non-linear Boltzmann equation:

$$B^*_{\parallel s}\frac{\partial \bar{F}_s}{\partial t} + \nabla \cdot \left(\frac{d\mathbf{x_G}}{dt} B^*_{\parallel s}\bar{F}_s\right) + \frac{\partial}{\partial v_{G\parallel}}\left(\frac{dv_{G\parallel}}{dt} B^*_{\parallel s}\bar{F}_s\right) = \mathcal{C}(\bar{F}_s) + S$$

- Let us define three advection operators (with $\mathcal{X}_G = (r,\theta)$)

$$B^*_{\parallel s}\frac{\partial \bar{F}_s}{\partial t} + \boldsymbol{\nabla} \cdot \left(B^*_{\parallel s}\frac{\mathrm{d}\mathcal{X}_G}{\mathrm{d}t}\bar{F}_s\right) = 0 \qquad : (\tilde{\mathcal{X}}_G)$$

$$B^*_{\parallel s}\frac{\partial \bar{F}_s}{\partial t} + \frac{\partial}{\partial \varphi}\left(B^*_{\parallel s}\frac{\mathrm{d}\varphi}{\mathrm{d}t}\bar{F}_s\right) = 0 \qquad : (\tilde{\varphi})$$

$$B^*_{\parallel s}\frac{\partial \bar{F}_s}{\partial t} + \frac{\partial}{\partial v_{G\parallel}}\left(B^*_{\parallel s}\frac{\mathrm{d}v_{G\parallel}}{\mathrm{d}t}\bar{F}_s\right) = 0 \qquad : (\tilde{v_{G\parallel}})$$

➠ Semi-Lagrangian scheme

- And the collision operator ($\tilde{\mathcal{C}}$) on a Δt : $\partial_t \bar{F}_s = \mathcal{C}(\bar{F}_s)$ ➠ Crank-Nicolson
- And the source operator ($\tilde{\mathcal{S}}$) on a Δt : $\partial_t \bar{F}_s = \mathcal{S}$ ➠ Crank-Nicolson
- Then, a Boltzmann solving sequence ($\tilde{\mathcal{B}}$) is performed:

$$(\tilde{\mathcal{B}}) \equiv \left(\frac{\tilde{\mathcal{S}}}{2}, \frac{\tilde{\mathcal{C}}}{2}\right)\left(\frac{\tilde{v_{G\parallel}}}{2}, \frac{\tilde{\varphi}}{2}, \tilde{\mathcal{X}}_G, \frac{\tilde{\varphi}}{2}, \frac{\tilde{v_{G\parallel}}}{2}\right)\left(\frac{\tilde{\mathcal{C}}}{2}, \frac{\tilde{\mathcal{S}}}{2}\right)$$

Example of Backward Semi-Lagrangian (BSL) approach for 1D advection operator

We consider the advection equation

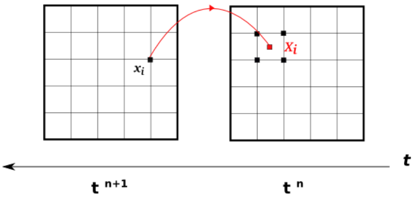
$$\frac{\partial f}{\partial t} + \mathbf{a}(x,t) \cdot \nabla_{\mathbf{x}} f = 0 \qquad (1)$$

The scheme: *(mix between PIC and Eulerian approach)*

- Fixed grid on phase-space *(Eulerian character)*
- Method of characteristics : ODE $\longrightarrow$ origin of characteristics *(PIC character)*
- Distribution function f is conserved along the characteristics

$$i.e. \qquad f^{n+1}(\mathbf{x}_i) = f^n(X(t_n; \mathbf{x}_i, t_{n+1})) \qquad (2)$$

- Interpolate on the origin using known values of previous step at mesh points (initial distribution f^0 known).



Backward Semi-Lagrangian parallelisation (1/2)

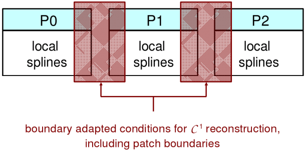
- ☺ Fixed grid in time ➠ perfect load balancing
- ☹ Complex parallelization due to **cubic spline interpolation**
  - Loss of locality (value of *f* on one grid point requires *f* over the whole grid)

➠ **Not possible to use a simple domain decomposition**

Two approaches are used in the GYSELA code

1. Work on the decomposed domain: A new numerical tool has been developed

➠ Hermite Spline interpolation on patches *[Latu-Crouseilles 2007]*



- Local splines on each subdomains with Hermite boundary conditions
- Derivatives defined to match as closely as possible those of global splines

☹ Some gradients can appear at the interfaces in the non-linear phase

2. Work on the global domain:

Data transposition

- Let us consider the transposition operation T_F and its inverse T_F^{-1}:

$$\bar{F}_s(r_{\text{block}}, \theta_{\text{block}}, \phi = *, v_{\|} = *, \mu = \mu_{\text{id}}) \quad \overset{T_F}{\underset{T_F^{-1}}{\rightleftarrows}} \quad \bar{F}_s(r = *, \theta = *, \phi_{\text{block}}, v_{\|\,\text{block}}, \mu = \mu_{\text{id}})$$

$\Downarrow$

Each processor has all information on ϕ and $v_{\|}$ directions so:

- ↪ 1D advection operator ($\tilde{\varphi}$) is possible
- ↪ as 1D advection operator ($\tilde{v_{G\|}}$)

$\Downarrow$

Each processor has all information on (r, θ) cross-section

- ↪ 2D advection operator $\tilde{\mathcal{X}}_G$

☹ Expensive in term of communication between processors

Outline

1. Gyrokinetic theory
2. GYSELA code
   - Semi-lagrangian approach
   - MPI/OpenMP parallelisation
   - Global flux driven simulation
3. Exascale needs and associated challenges
   - Increase of core number : scalability, fault tolerance
   - Memory reduction and big data
   - Continuous integration

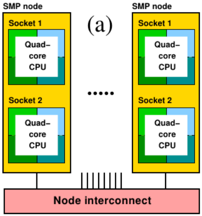
- GYSELA main characteristics:

☺ Complete knowledge at the institute.

- Written in Fortran90 + some routines in C (~ 50000 lines).
- Hybrid OpenMP/MPI parallelisation to use benefit of SMP cluster

For instance:

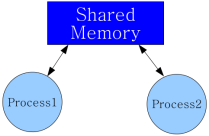
  - *MPI between nodes*
  - *OpenMP inside quad-core CPU*



SMP cluster scheme

- Message Passing Interface (MPI)
  - MPI is a library specification for message-passing, proposed as a standard by a broadly community.

- Open Multi Processing (OpenMP)
  - OpenMP is a specification for a set of compiler directives, library routines, and environment variables that can be used to specify shared memory parallelism in Fortran and C/C++ programs.



Global algorithm

- Let us consider the transposition operation T_F and its inverse T_F^{-1}:

$$\bar{F}_s(r_{\text{block}}, \theta_{\text{block}}, \phi = *, v_\parallel = *, \mu = \mu_{\text{id}}) \underset{T_F^{-1}}{\overset{T_F}{\rightleftarrows}} \bar{F}_s(r = *, \theta = *, \phi_{\text{block}}, v_{\parallel\,\text{block}}, \mu = \mu_{\text{id}})$$

- Input: Physics parameters + $\bar{F}_s^0(r_{\text{block}}, \theta_{\text{block}}, \phi = *, v_\parallel = *, \mu = \mu_{\text{id}})$
- For $k = 0$ to N:
  - Computation of r.h.s of quasi-neutrality: $\sum_s Z_s \int J_0 \cdot \bar{F}_s^k dv_\parallel d\mu$
  - Solve 3D QN equation: $\phi^k \to \phi^{k+1}$
  - For each species s and each value of $\mu = \mu_{\text{id}}$:
    - Gyroaverage computation: $J_0 \cdot \phi^{k+1}$
    - Solve 5D Boltzmann equation: $\bar{F}_s^k \to \bar{F}_s^{k+1}$

$$\left[\left(\frac{\tilde{\mathcal{S}}}{2}, \frac{\tilde{\mathcal{C}}}{2}\right)\left(\frac{\tilde{v_{G\parallel}}}{2}, \frac{\tilde{\varphi}}{2}\right)\right], T_F \left(\tilde{\mathcal{X}}_G\right) T_F^{-1}, \left[\left(\frac{\tilde{\varphi}}{2}, \frac{\tilde{v_{G\parallel}}}{2}\right)\left(\frac{\tilde{\mathcal{C}}}{2}, \frac{\tilde{\mathcal{S}}}{2}\right)\right]$$

    End for
  - Phase space reduction for 3D to 0D diagnostics at time t^{k+1}

  End for
- Output: Distribution function ($\bar{F}_s^N$) for restart + 0D to 3D diag. at several times

Parallel performance measurements

- Speed up = $T_{\mathrm{serial}}/T_{\mathrm{parallel}}(n)$
  - T_{serial} = 100 sec
  - $T_{\mathrm{parallel}}(2)$ = 80 secs
  - 25% speed up
- Efficiency = $T_{\mathrm{serial}}/(n \times T_{\mathrm{parallel}}(n))$
  - $100/(2\times 80)$ =
  - 62% efficiency

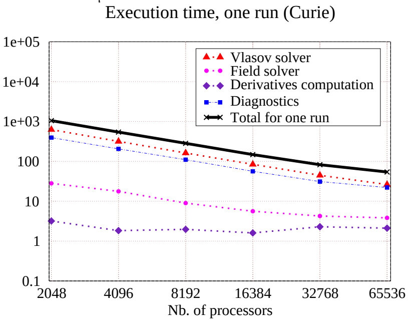
- Weak scaling The problem size (workload) assigned to each processing element stays constant and additional elements are used to solve a larger total problem
- Strong scaling The problem size stays fixed but the number of processing elements are increased

➡ In general, it is harder to achieve good strong-scaling at larger process counts since the communication overhead for many/most algorithms increases in proportion to the number of processes used.

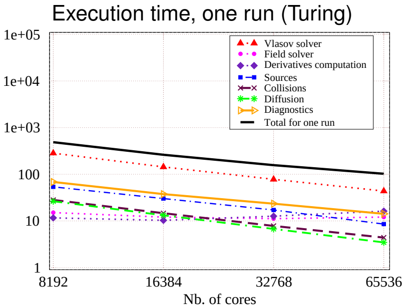
Scalability and bottlenecks

Strong scaling: $N_r = 512,\ N_\theta = 512,\ N_\varphi = 128,\ N_{v\|} = 128$

$M_\mu = 32$, main data=1 TiB



$M_\mu = 16$, main data=512 GiB

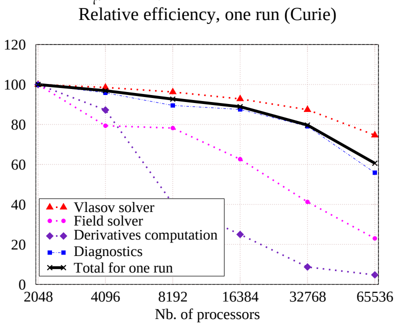


- Time dominated by Vlasov solver
- Scaling bottleneck: Poisson solver

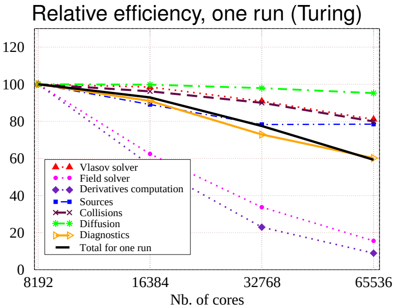
Scalability and bottlenecks

Strong scaling: $N_r = 512,\ N_\theta = 512,\ N_\varphi = 128,\ N_{v\|} = 128$

$M_\mu = 32$, main data=1 TiB



$M_\mu = 16$, main data=512 GiB



- Time dominated by Vlasov solver
- Scaling bottleneck: Poisson solver

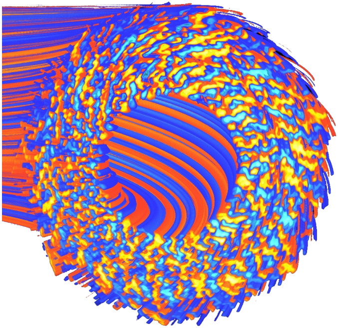
$\approx$ 60% efficiency at 64 k cores on both machines (Curie and Turing)

1. Gyrokinetic theory
2. GYSELA code
   - Semi-lagrangian approach
   - MPI/OpenMP parallelisation
   - Global flux driven simulation
3. Exascale needs and associated challenges
   - Increase of core number : scalability, fault tolerance
   - Memory reduction and big data
   - Continuous integration

GYSELA competitive in terms of physics (1/2)
➟ Biggest global simulation ever run

Grand Challenge CINES 2010: Biggest global simulation ever run

➟ A simulation close to ITER-size scenario ($\rho_* = 1/512$) performed on 1/4 torus with additional heating power of **60 MW during 1 ms**



ion temperature fluctuations in the turbulent saturated phase

- ✓ A 5D mesh of **272 10^9 points**

  $(r, \theta, \varphi, v_{\|}, \mu) = (1024 \times 1024 \times 128 \times 128 \times 16)$

- ✓ **$>$ 6.1 million hours monoproc.**
  - $\sim$ 31 days on 8192 processors

⇣

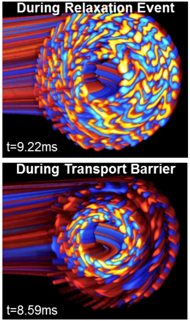
- ✓ **6.5 TBytes of data** to analyse
  - 1.5 TBytes for 2D and 3D savings
  - 5 TBytes for restart files

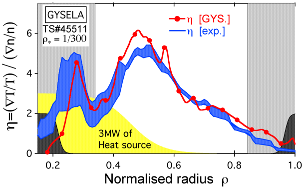
[J. Abiteboul EPS2010, Y. Sarazin IAEA2010]

GYSELA competitive in terms of physics (2/2)
➠ Flux-driven simulations

- Generation & transport of toroidal rotation / Role of turbulence & boundary conditions
  - *[J. Abiteboul et al., PPCF 2013]*
- Transport barrier relaxations with E_r shear
  - *[A. Strugarek et al., PPCF 2013]*
  - *[A. Strugarek et al., PRL 2013]*
  - *[Y. Sarazin, V. Grandgirard and A. Strugarek, La Recherche, nov. 2012]*
- Interaction energetic particles & turbulence via EGAMs
  - *[D. Zarzoso et al., PoP 2012, PRL 2013]*
- Comparison with experiments
  - *[invited G. Dif-Pradalier , TTF 2013]*
- Caracterisation of turbulent transport
  - *[C. Norscini, poster, Vlasovia 2013]*
  - *[T. Cartier-Michaud, poster, Vlasovia 2013]*

Snapshots of non-axisymmetric electric potential fluctuations

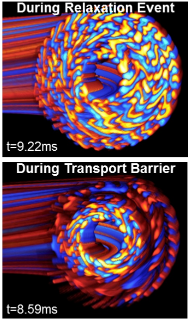


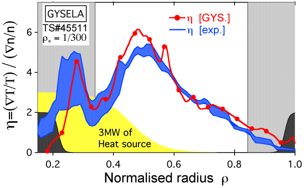


GYSELA competitive in terms of physics (2/2)
➠ Flux-driven simulations

- Generation & transport of toroidal rotation / Role of turbulence & boundary conditions
  - ☹ $N = 9$ instead $N = 18$ for ripple effects
- Transport barrier relaxations with E_r shear
  - ☹ Reduced $\rho_* = \rho_i/a$: $1/150$ instead of $1/500$
- Interaction energetic particles & turbulence via EGAMs
  - ☹ Not possible to treat very energetic particles
- Comparison with experiments
  - ☹ Several energy confinement times not accessible
- Caracterisation of turbulent transport
  - ☹ Not enough 3D data saved for good statistics

Snapshots of non-axisymmetric electric potential fluctuations





ITER-type simulation for an energy confinement time unreachable ➠ Exascale Needs

➠ GYSELA is already using currently Petascale machine (> 50 million hours/year)

☹ Compromise machine size & simul. up to energy confinement time must be found

- GYSELA simulation close to ITER-like parameters : 272 billions of points
- Longest time simulation: $2.10^6/\Omega_c \sim 1$ energy confinement time

	Number of Points ($\rho^*=\rho/a$)	Time / Ω_c	Number of cores	Number of days of simulation
Gd Challenge CINES 2010	272 billions ($\rho^*=1/512$)	147 840	8192	31
Gd Challenge CURIE 2012	33 billions ($\rho^*=1/150$)	678 510	16 384	15
	=> Adding of tritium		32768	6
Comparison with experiment (in progress)	87 billions ($\rho^*=1/300$)	2 000 000	5520	46

➠ GYSELA will require Exascale machine for realistic kinetic electrons

- With electrons: $\rho_{\text{ions}}/\rho_{\text{elec}} = 60$ ➠ mesh size $\times 60^3$ and time step$/60$!!!

Outline

1. Gyrokinetic theory

2. GYSELA code
   - Semi-lagrangian approach
   - MPI/OpenMP parallelisation
   - Global flux driven simulation

3. Exascale needs and associated challenges
   - Increase of core number : scalability, fault tolerance
   - Memory reduction and big data
   - Continuous integration

Petascale era ...
On the road towards Exascale

- At the moment, Petascale machines (in operation since 2008):
  ↪ more than 33 PetaFlops (1 PFlops= 10^{15} floating point operations per second)

TOP500® JUNE 2014

PRESENTED BY

ICL UT INNOVATIVE COMPUTING LABORATORY THE UNIVERSITY of TENNESSEE · BERKELEY LAB Lawrence Berkeley National Laboratory · PROMETEUS PROFESSOR MEUER TECHNOLOGIEBERATUNG

FIND OUT M
www.top5

	NAME	SPECS	SITE	COUNTRY	CORES	RMAX PFLOP/S
1	Tianhe-2 (Milkyway-2)	NUDT, Intel Ivy Bridge (12C, 2.2 GHz) & Xeon Phi (57C, 1.1 GHz), Custom interconnect	NSCC Guangzhou	China	3,120,000	33.9
2	Titan	Cray XK7, Operon 6274 (16C 2.2 GHz) + Nvidia Kepler GPU, Custom interconnect	DOE/SC/ORNL	USA	560,640	17.6
3	Sequoia	IBM BlueGene/Q, Power BQC (16C 1.60 GHz), Custom interconnect	DOE/NNSA/LLNL	USA	1,572,864	17.2
4	K computer	Fujitsu SPARC64 VIIIfx (8C, 2.0GHz), Custom interconnect	RIKEN AICS	Japan	705,024	10.5
5	Mira	IBM BlueGene/Q, Power BQC (16C, 1.60 GHz), Custom interconnect	DOE/SC/ANL	USA	786,432	8.59

- Nobody knows what will exactly be the future “Exascale machine” but:
  ↪ Exascale implementations projected by 2018
  ↪ Several millions of cores with small memory per core (< 1 GBytes)

Some exascale bottlenecks for GYSELA and others ... (1/2)

- Applications will need to be scalable on millions of cores
- Exascale machines could be close to BlueGene Architecture or ... ?

↪ Adapting the code for BlueGene architecture

[J. Bigot, F. Rozar et al., ESAIM proceedings 2013]

↪ Adapting the code to the new Intel-Xeon Phi technology

➠ Tests on IFERC machine with a prototype application

[G. Latu, M. Haefele, CEMRACS 2014 project]

- Increase of number of cores ⇒ Probability of crashes increases

➠ Post-Doc ANR-Nufuse G8@Exascale: *O. Thomine (oct 2011-oct 2013)*

↪ Non-blocking writing of restart files *[O. Thomine et al., ESAIM proceedings 2013]*

↪ Fault tolerance improvement *[J. Bigot, CEMRACS 2014 project]*

➠ Coupling with FTI library (developed by F. Capello)

Some exascale bottlenecks for GYSELA and others ... (2/2)

- Big data ~ Several hundreds TBytes: Issues of transfer, storage, visualisation

↪ HLST support (IPP Garching) for data compression and parallel writting

[S. Espinoza, HLST report 2013]

↪ How to improve data transfer ? ➠ Actually more than one week

↪ Where and how to archive ?

↪ CINES team (long time storage)

↪ Visualisation with SDvision (IRFU/DSM)

- Memory reduction per nodes:

➠ PhD Maison De la Simulation / IRFM: *F. Rozar (dec 2012-dec 2015)*

↪ Development of dedicated tools for memory scalability. (MTM C/Fortran library)

↪ First gain up to 50% of memory on a large simulation run.

[F. Rozar et al., submitted to PPAM2013]

- Big efforts of parallelisation since 2009
- Maximum of Gd Challenge opportunities taken to improve GYSELA efficiency

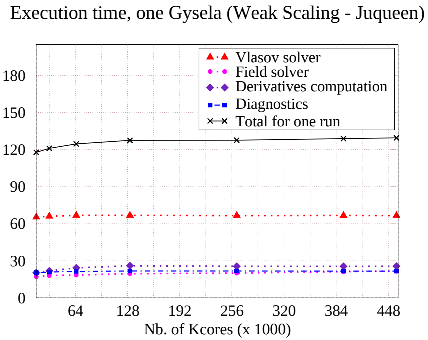
	Relative efficiency		Number of cores
	Weak scaling	Strong scaling	
Gd Challenge CINES (march 2010)	92 %	82 %	8192
Gd Challenge CURIE (march 2012)	91 %	61 %	65 536
Porting on Blue Gene Architecture => Communication schemes rewritten			
Gd Challenge TURING (january 2013)	92 %	61 %	65 536
Access to totality of JUQUEEN (may 2013)	91 %		458 752

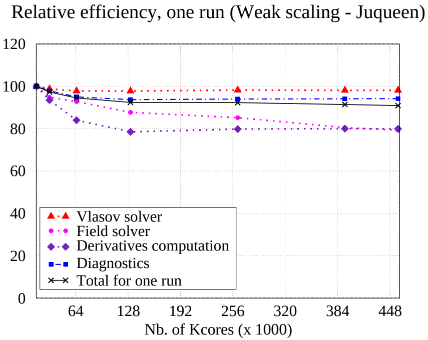
x56

↪ Weak scaling: Relative efficiency of 91% on 458 752 cores on the totality of the biggest european machine (Juqueen - 1.8 Mthreads)

WEAK scaling: (on JUQUEEN - Juelich) Relative efficiency of 91% on 458 752 cores

- Parallel communication schemes completely rewritten
- Tests performed on the totality of JUQUEEN/Blue Gene machine (Juelich)

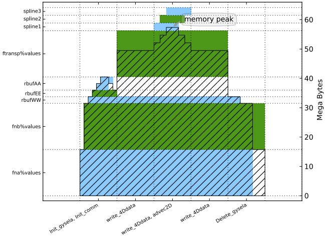




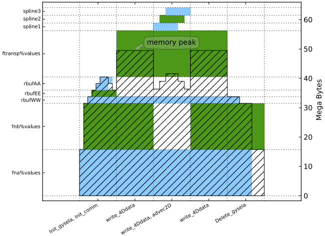
- Weak scaling: Relative efficiency of 91% on 458 752 cores.
  - PRACE preparatory access *(April 2012 - Nov 2012)*: 250 000 hours
  - ANR G8-Exascale via P. Gibbon.

Improvement of memory consumption

- GYSELA is global ➠ Huge meshes ➠ Constrained by memory per node
- Development of the MTM library in progress (Modelization & Tracing Memory consumption)
  - Identification of memory peak
  - Prediction of memory required before submit ➠ Avoid memory exhaust



Before optimisation



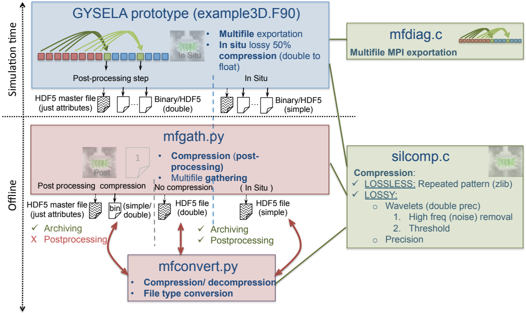
After optimisation

- Static to dynamic memory alloc. + improvement of algorithms

➠ Gain of factor 50% on 32*k* cores

[F. Rozar et al., accepted to PPAM2013]

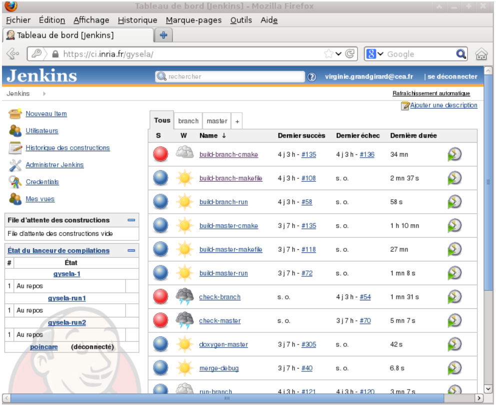
- Problem of memory and time scalability for GYSELA 3D diagnostics
- Development of the LCHD library performed by HLST-IPP Garching
  - 6 months project - S. Espinoza & M. Haefele *[S. Espinoza, HLST Report 2013]*
  - Fast multi-file multi-variable exportation
  - Lossless and lossy 3D data compression



- I/O bandwidth ×26 with parallel efficiency of 95% from 256 to 1280 cores
- Lossless: 8% compression;
- Lossy: from 50% to 70% achieved without altering physics

Continuous integration for non-regression tests

- Based on the Inria continuous integration platform
  - Jenkins + CloudStack
- Each time compilation in many modes (43) ➠ Error + warning analysis
- Non-regressing physical tests



Conclusion

- Each GYSELA simulation = a numerical experiments
  ↪ Several weeks on several thousands of core
  (ex: Grand Challenge Curie 2012: 15 days on 16384 cores)
  ↪ Several TBytes of data to store and to analyse

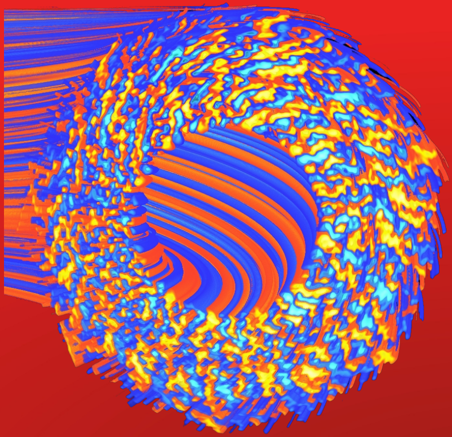
- Exascale HPC are required for realistic kinetic simulations with both ions and electrons
  ↪ Promising results: Weak scaling - relative efficiency of 91% on 458 752 cores

- Lots of bottlenecks need to be overcome for all gyrokinetic codes to be ready to run on exascale machines.

➠ High level collaboration with computer scientists is mandatory.

Collaborations:

- ANR GYPSI (2010-2014)
  ↪ Strasbourg, Nancy, Marseille
- ANR Nufuse G8@exascale (2012-2016)
  ↪ France, Germany, Japan, US, UK
- ADT INRIA Selalib (2011-2015)
  ↪ Strasbourg, Bordeaux
- Action C2S@Exa - IPL INRIA (march 2013-2017)
  ↪ Nice, Bordeaux
- New project following AEN INRIA Fusion (evaluation in progress)
  ↪ Strasbourg, Lyon, Nice
- Collaborations with IPP Garching (Germany) since 2012
- Collaborations with “Maison de la Simulation”- Saclay (Paris) since 2012



Commissariat à l'énergie atomique et aux énergies alternatives
Centre de Cadarache | 13108 Saint Paul Lez Durance Cedex
T. +33 (0)4 42 25 46 59 | F. +33 (0)4 42 25 64 21

DSM
IRFM
SCCP/GTTM

Etablissement public à caractère industriel et commercial | RCS Paris B 775 685 019