DE LA RECHERCHE À L'INDUSTRIE

cea

irfm

www.cea.fr

GYSELA: Exascale needs

V. Grandgirard[1], G. Latu[1]

Collaborations with physicists:

J. Abiteboul[2], Y. Dong[3], D. Estève[1], X. Garbet[1], J.B Girardo[1], Ph. Ghendrih[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:

J. Bigot[6], C. Passeron[1], F. Rozar[1,6], O. Thomine[1]

[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

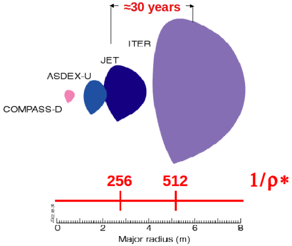
ANR GYPSI - ANR G8-Exascale Nufuse
ADT-INRIA SELALIB - AEN-INRIA Fusion

GYSELA is a 5D non-linear gyrokinetic code used to study turbulence (self-organisation & control) in Tokamak plasmas.

There exist ten 5D gyrokinetic codes for plasma fusion in the world: 5 (US), 3 (UE) and 1 (Japan).

- Various numerical schemes:
  - Lagrangian (PIC), Eulerian or Semi-Lagrangian
- Various simplifications:
  - δf codes: scale separation between equilibrium and perturbation.
  - Flux-tube codes ⇒ the domain considered is a vicinity of a magnetic field line.
  - Fixed gradient boundary conditions.
- A new generation of global full-f gyrokinetic codes is being developed with collisions and flux-driven boundary conditions.
  - GYSELA one of them

- GK codes require state-of-the-art HPC techniques and must run efficiently on more than 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 $\quad \rho_* \equiv \frac{\rho_i}{a} \ll 1$



✓ $\rho_{*\,\text{ITER}} = 1/512$

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

Huge mesh for global simulations

- Objectives of parallel optimisation of the code and development of new numerical schemes ➠ Always more physics,
  ➠ Closer and closer to experimental parameters

- Work in progress with physicists: (Not discussed here)
  - Energetic particles *[J.B Girardo (PhD)]*
  - Transport of impurities *[D. Esteve (PhD)]*
  - Spectral transfers *[Y. Dong (PhD-LPP)]*
  - Trapped electrons *[T. Cartier-Michaud (PhD)]*
    [F. Palermo (Post-Doc ANR GYPSI)]

- Work in progress with mathematicians: (Not discussed here)
  - Development of a numerical test platform for Vlasov solvers *ADT INRIA Selalib*
  - Treatment of realistic geometry (WEST) *[A. Back (Post-doc ANR GYPSI-CPT)]*
  - Aligned coordinates (kinetic electrons) *[L. Mendoza (PhD-IPP Garching)]*

- Presentation focused on Computer science developments

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

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

$^{(*)}$ *where $\mu = m v_\parallel^2/2B$ plays the role of a parameter*

with the equations of motion:

$$B_{\parallel s}^* d_t \mathbf{x}_G = v_{G\parallel} \mathbf{B}^* + \frac{1}{e} \mathbf{b} \times \nabla \Lambda$$

$$B_{\parallel s}^* m_s d_t v_{G\parallel} = -\mathbf{B}^* \cdot \nabla \Lambda$$

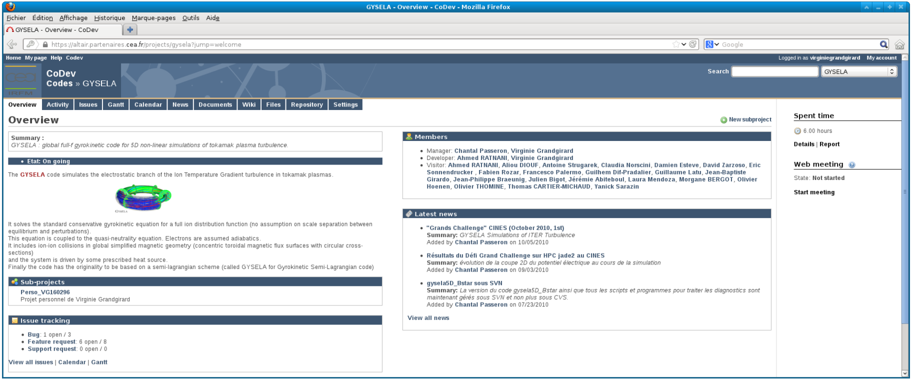
where $\mathbf{B}^* = \mathbf{B} + (m_s v_{G\parallel}/e) \nabla \times \mathbf{b}$ and $\Lambda = e J_0 \phi + \mu B$;

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

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

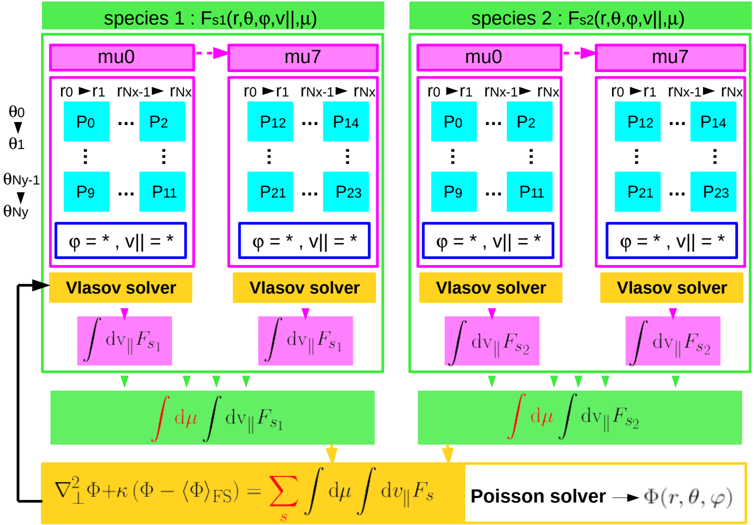
GYSELA: A collaborative environment for development

- GYSELA is mainly written in Fortran90 with some routines in C.
- Hybrid parallelisation MPI/OpenMP
- Collaborative development under SVN and GIT.
- Collaborative platform: CEA-CODEV (based on Redmine)



GYSELA: Parallelisation description (1/2)
➠ Vlasov solver + Poisson solver

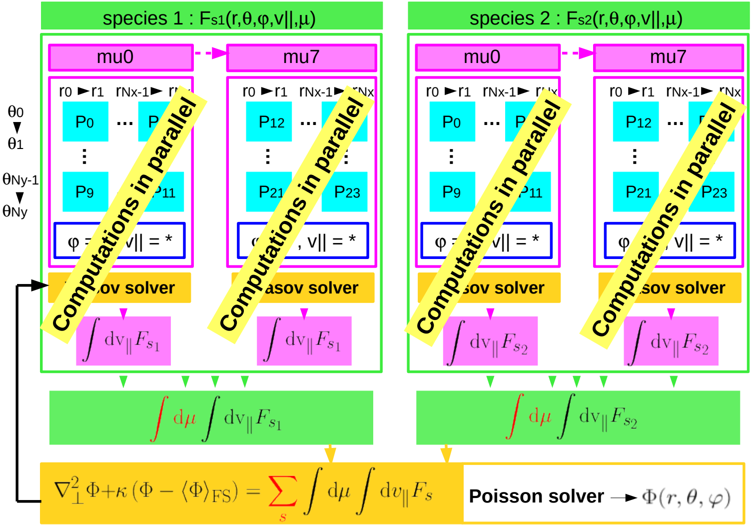
- Parallel decomposition example for $Nproc\_r = 3$, $Nproc\_\theta = 4$, $Nproc\_\mu = 8$



GYSELA: Parallelisation description (1/2)

➠ Vlasov solver + Poisson solver

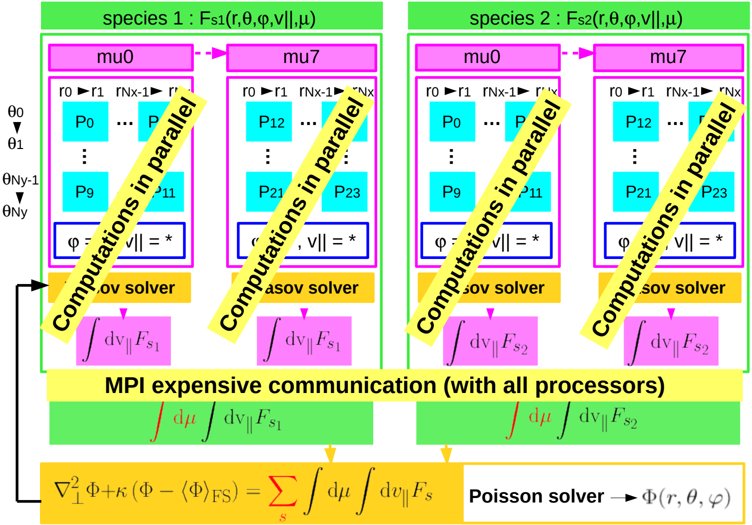
- Parallel decomposition example for $Nproc\_r = 3$, $Nproc\_\theta = 4$, $Nproc\_\mu = 8$



GYSELA: Parallelisation description (1/2)

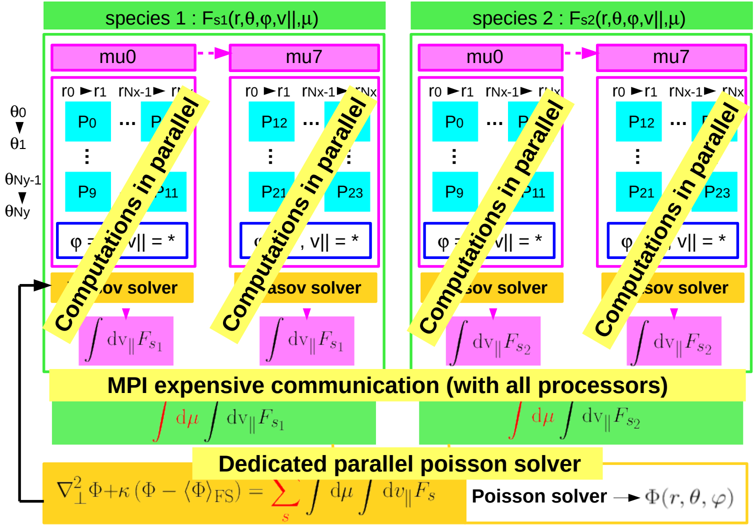
➠ Vlasov solver + Poisson solver

- Parallel decomposition example for $Nproc\_r = 3$, $Nproc\_\theta = 4$, $Nproc\_\mu = 8$

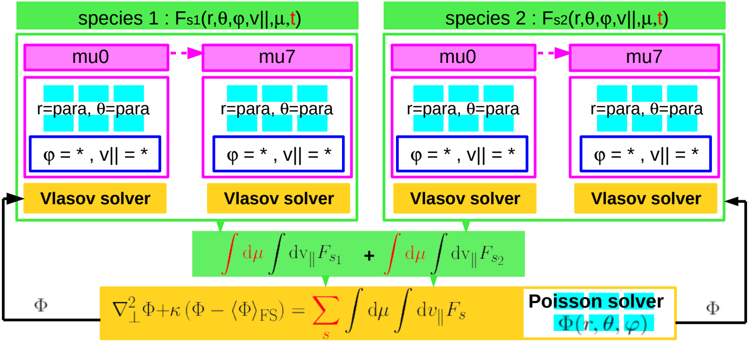


GYSELA: Parallelisation description (1/2)
➟ Vlasov solver + Poisson solver

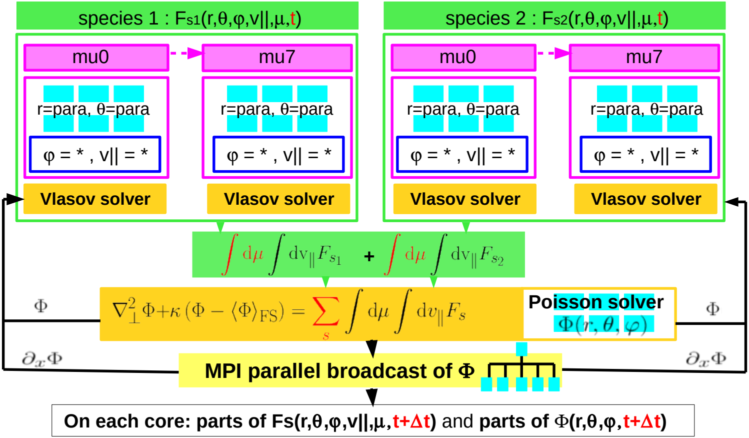
- Parallel decomposition example for $Nproc\_r = 3$, $Nproc\_\theta = 4$, $Nproc\_\mu = 8$



- One iteration of the code:



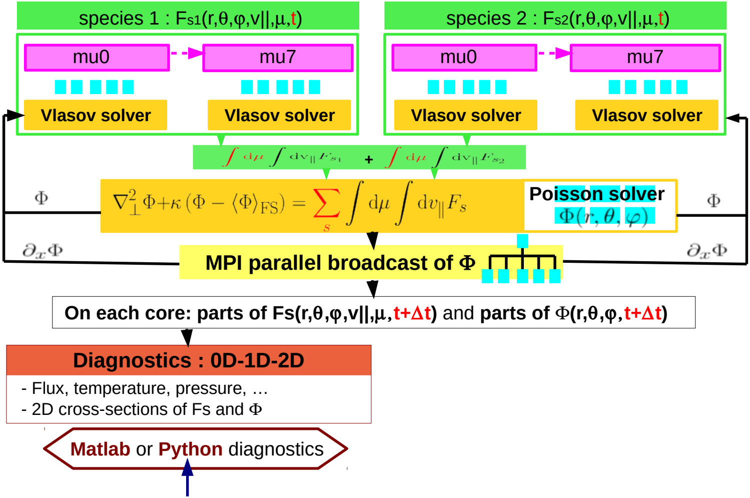
- One iteration of the code:



GYSELA: Parallelisation description (2/2)

Diagnostics

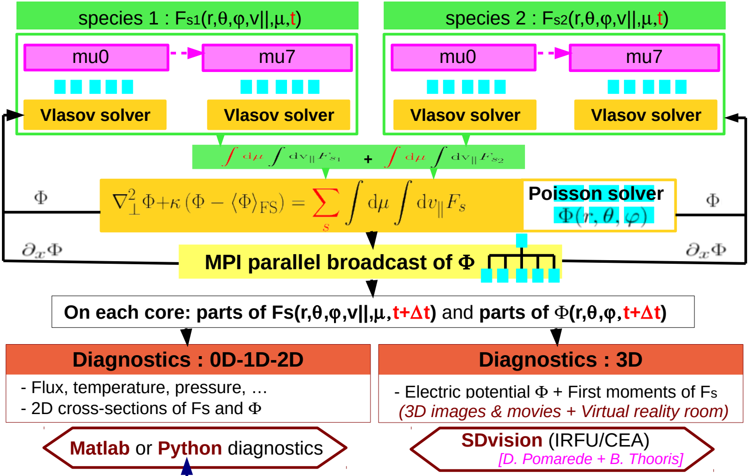
- One iteration of the code:



- Completely rewritten for several species

[Ch. Passeron, VG]

- **One iteration** of the code:



- Completely rewritten for several species

[Ch. Passeron, VG]

\+ **HLST 2013 project** :

(6 ppm : S. Espinosa + M. Haefele)

- parallel 3D writting to reduce memory problem,
- 3D compression

[G. Latu, Ch. Passeron, VG]

A typical run: several restarts

One typical GYSELA run

Nb iterations x one iteration *(described before)* → **Saving of restart files *(Fs 5D)*** → **Automatic** new submission on HPC

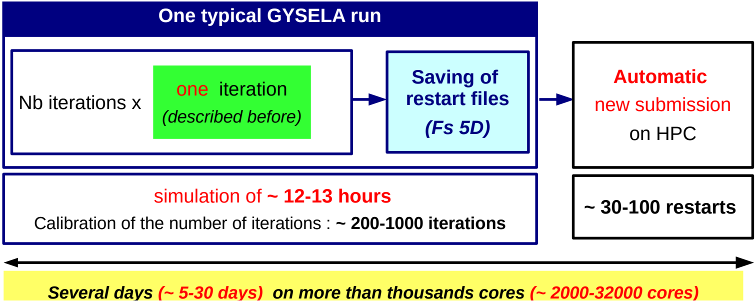
simulation of **~ 12-13 hours**

Calibration of the number of iterations : **~ 200-1000 iterations**

~ 30-100 restarts

Several days (~ 5-30 days) on more than thousands cores (~ 2000-32000 cores)

A typical run: several restarts



- Example of simulation in progress *[working group Comp. Simu/Exp. with LPP: O. Gurcan, P. Hennequin, P. Morel, L.Vermare]*
  - $\rho_* = 1/300$ for a quarter of torus,
  - mesh of 86 billion of points ➠ $(N_r, N_\theta, N_\varphi, N_\mu) = (512, 512, 128, 128, 20)$
  - restart files: $\sim$ 1.3 TBytes ➠ $2 \times$ (320 files of 2 GB)
  - run on 5520 cores ➠ $(\mathrm{Nproc}_r = 4, \mathrm{Nproc}_\theta = 4, \mathrm{Nproc}_\mu = 20, \mathrm{Nb}_{\mathrm{thread}} = 16)$
  - 33000 iterations already performed (1 million of Ω_c time)
  - performed on IFERC machine (Rokkasho-Japan) during $\sim$ 23 days

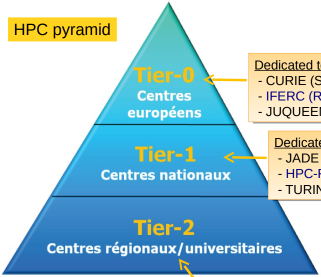
$=$ 3 millions of mono-processor hours *[G. Dif-Pradalier et al., TTF 2013]*

GYSELA deployed on 9 HPC machines

- GYSELA is actually present on 9 different HPC machines.
  - ↪ Deployments in 2012-2013: *CURIE, HELIOS, CCAMU, Poincaré*
  - ↪ Code adaptation for BlueGene architecture (2012-2013):*JUQUEEN, TURING*

[Ch. Passeron, G. Latu, J. Bigot (Post-Doc MDS)]

HPC pyramid



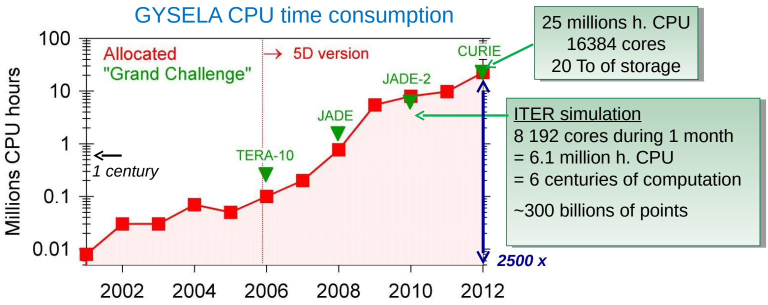
Dedicated to production :
- CURIE (Saclay/France) : ~ 80.000 cores
- IFERC (Rokkasho/Japon) : ~ 80.000 cores
- JUQUEEN (Juelich/Germany) : 458.752 cores

Dedicated to production :
- JADE (CINES Montpellier/France) : ~ 12.000 cores
- HPC-FF (Juelich/Germany) : ~ 10.000 cores
- TURING (IDRIS Orsay/France) : 65.536 cores

Dedicated to development :
- Local cluster : NORMA (256 cores) *(2007)*
- « Mésocentres »: CCAMU-Marseille (~1000 cores) *(2012)*
  Maison De la Simulation-Saclay (~1500 cores) *(2013)*
  ► *512 cores dedicated to GYSELA (paid by ANR GYPSI)*

- Development of non-regressive tests (common work with JOREK)

[Ch. Passeron, G. Latu]



- GYSELA CPU time needs increase exponentially
- 36 millions of hours for 2013

  ➟ More than 90% dedicated to production runs

+ Grand Challenge TURING: 13 millions of hours

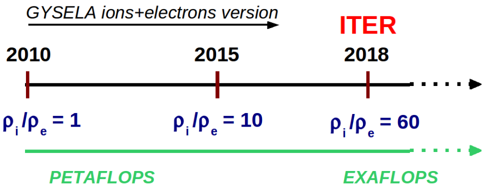
- GYSELA is already using currently Petascale machine

  ↪ Now ITER-like ion simulation: $272\ 10^9$ points ➠ $6\ 10^6$ CPU/hours

- 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 !!!



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

- Increase of number of cores ⇒ Increase of crashes
  ➠ Post-Doc ANR-Nufuse G8@Exascale: *O. Thomine (oct 2011-oct 2013)*
  ↪ Fault tolerance improvement
  ↪ Non-blocking writing of restart files
  [O. Thomine et al., ESAIM proceedings 2013]

- Memory reduction per nodes:
  ➠ PhD Maison De la Simulation / IRFM: *F. Rozar (dec 2012-dec 2015)*
  ↪ Development of dedicated tools for memory scalability
  ↪ First gain up to 50% of memory on a big case
  [F. Rozar et al., submitted to PPAM2013]

- Exascale machines will probably be close to BlueGene Architecture
  ➠ Post-Doc MDS/PRACE: *J. Bigot (july 2012-july 2014)*
  ↪ Adapting the code for BlueGene architecture
  [J. Bigot, F. Rozar al., ESAIM proceedings 2013]

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

- More modularity is required to improve future parallelisation capabilities
  ↪ Improve diagnostic parallelisation *(J. Bigot)*
  ↪ Facilitate interaction with SELALIB platform
  (in collaboration with E. Chacon-Golcher and P.Navaro)

- Open question: How to treat these huge amounts of data (> 100 TBytes) ?
  ↪ Data transfer, data analysis and data long-term storage...

Ultimate Goal

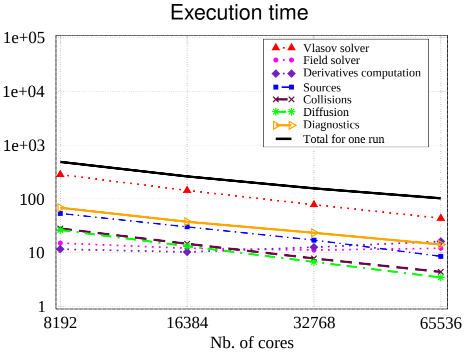
⇓

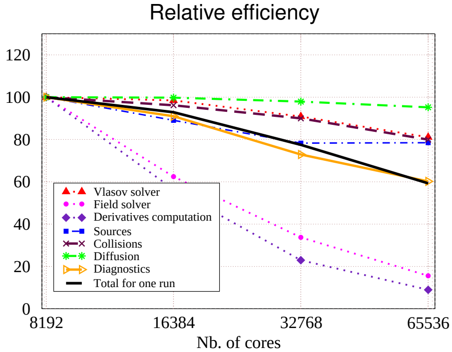
- Improve scalability of the code on more than 1 million of cores

  ↪ Recent results presented in the following

STRONG scaling (on TURING - Paris): Relative efficiency of 61% on 65 536 cores

- "Grand Challenge" Turing (IDRIS/Paris): *1st December 2012 - 15th January 2013*.
- Strong scaling performed on the totality of the machine: 65 536 cores.
- Relative efficiency of 61% on 65 536 cores.

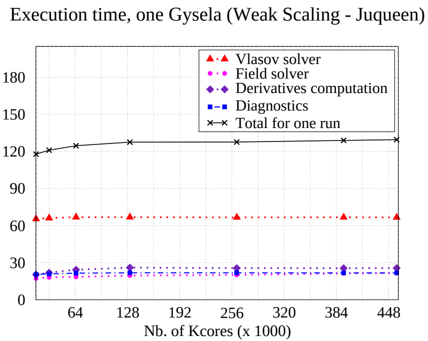


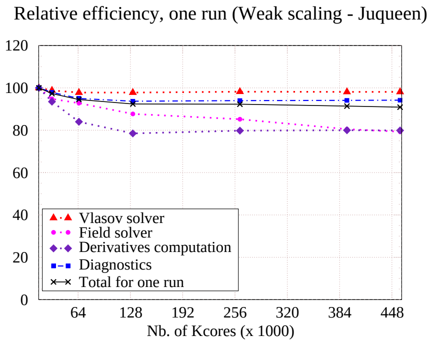


[G. Latu, J. Bigot & GYSELA team, "Grands Challenges IDRIS-GENCI 2012" publication]

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)





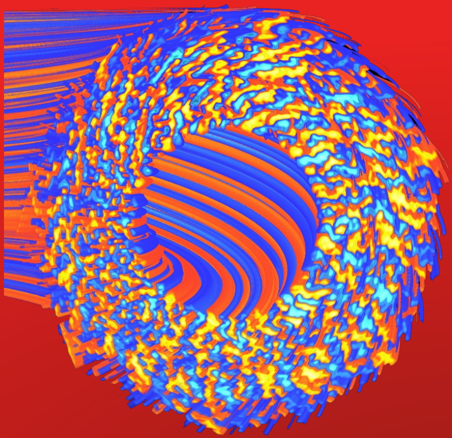
- Weak scaling: Relative efficiency of 91% on 458 752 cores.
  - PRACE preparatory access *(April 2012 - Nov 2012)*: 250 000 hours
  - ANR G8-Exascale via P. Gibbon.

Conclusion - Perspectives

- 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 will be required for realistic simulation with both ions and kinetic electrons
  - ↪ Development of lot of collaborations to prepare GYSELA for exascale
  - ↪ Promising results: Weak scaling - relative efficiency of 91% on 458 752 cores

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
- 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