

Les défis scientifiques pour mener les simulations numériques de demain

Colloque “Modélisation : succès et limites” organisé par
l’Académie des Technologies et le CNRS

Christophe DENIS

Ingénieur-Chercheur à EDF R&D
Chargé de Recherche au Centre de Mathématiques et de leurs Applications, ENS
Paris-Saclay

« Although we do not know the detail of the hardware that will be available, we can be certain that the level of parallelism will increase significantly, that machines will be more complex and heterogeneous [...]. Thus, in common with our colleagues in the US and Japan, we recognize that considerably more effort and manpower will be required to even begin to address this complexity ... »

— I. S. Duff, European Exascale Software Initiative: Numerical Libraries, Solvers and Algorithms

- 1 Remerciements
- 2 Introduction : enjeux et besoins autour du calcul haute performance
- 3 Free lunch is over : mutation architecturale en cours
- 4 Quelques défis scientifiques au niveau des codes pour utiliser pleinement les futures puissances de calcul
- 5 Conclusion

Remerciements



- Illustration de l'enjeu industriel de la simulation numérique fournie le groupe EDF
- Remerciements à :
 - Jean-Paul Chabard, Directeur Scientifique de EDF R&D
 - Ange Caruso, Responsable Programme Technologies de l'Information de EDF R&D

Les finalités de la simulation numérique pour EDF

Domaines d'application très variés

- Production électrique, Environnement, Réseau/Smarties, Commerce, Management de l'énergie, ..

3 grandes finalités

- 1 **Simuler pour comprendre**, meilleure compréhension de la complexité des phénomènes physiques
 - Evolution des contraintes réglementaires, nouvelles opportunités d'optimisation
- 2 **Simuler pour décider**, meilleure prédiction et fiabilité des simulations de systèmes complexes
 - Meilleure marge de fonctionnement, prise de décision sur des objectifs économiques
- 3 **Simuler pour innover**, obtenir des informations de plus en plus précises
 - Nouveaux domaines, nouveaux produits et nouveaux services
 - Améliorer les méthodes et les outils

Les usages du calcul haute performance à EDF

1 La simulation pour les études au quotidien

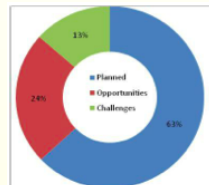
- ex : mécanique des fluides calcul à 100 millions de mailles d'une grille de mélange d'un assemblage combustible

2 La simulation pour préparer les études de demain

- ex : mécanique des fluides calcul à 1 milliard de mailles de mailles d'un assemblage combustible en entier

3 La simulation numérique pour explorer de nouvelles frontières

- R&D amont pour lever des verrous à terme
- ex : mécanique des fluides calcul à 100 milliard de mailles d'un assemblage combustible en entier
- ⇒ **Besoin d'une puissance de calcul toujours croissante**



Evolution des ressources et des besoins HPC

Illustration sur un cas d'étude CFD provenant d'EDF R&D

- Objectif : apporter une aide toujours plus efficace sur la prise de décision
 - Meilleure connaissance des phénomènes physiques multi-échelles et couplés
 - Calcul sur des géométries de plus en plus fines et complexes

Année	2003	2006	2007	2013	2018
Evolution de la finesse de la géométrie					
# de cellules	10^6	10^7	10^8	$3 \cdot 10^9$	$5 \cdot 10^{10}$
Evolution des ressources de calcul					
Machine	Fujitsu	IBM Power 5	IBM BG/L	Cray XE6	?
Mémoire	2Gb	25Gb	250Gb	3Tb	60Tb
# procs	1 p.	400 p.	8 000 p.	4 000 c	> 60k c.

Besoins toujours croissant de puissance de calcul !

L'utilisation de la simulation numérique pour face à la diversité des enjeux techniques, environnementaux, financiers et concurrentiels nécessite de :

- réduire l'écart entre le résultat de la simulation numérique et les phénomènes physiques étudiés
- de prendre en compte les incertitudes et le manque de connaissance des modèles utilisés
- vérifier la mise en place de nouvelles méthodologies alternatives (ex : réduction de modèles)
- combler le manque d'efficacité des codes de calcul : la majorité des codes n'utilisant qu'une faible partie des ressources allouées

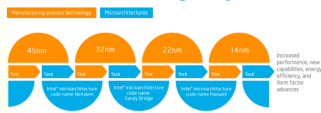
⇒ Nécessite une puissance de calcul toujours croissante

Free lunch ...

Pour répondre à la demande effrénée de puissance de calcul, deux leviers sont successivement utilisés :

- 1 augmentation de la performance intrinsèque des micro-processeurs
- 2 modification de l'architecture pour multiplier le nombre de micro-processeurs

The Tick-Tock model through the years



source Intel

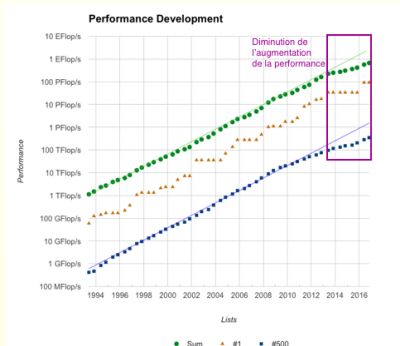
Course à la performance “bon marché” pour les codes de calcul durant les dernières decennies (Free Lunch)

- Grâce à la diminution constante de la vitesse de gravure
- Mais non optimisation fine des codes (économiquement non rentable)

Free lunch is over

En 2005, Herb Sutter sonne le glas de la course à la performance « bon marché » (Free Lunch is Over)

- Diminution plus complexe de la finesse de gravure des microprocesseurs
- Répercussion sur la puissance des architectures de calcul



Mutation architecturale en cours

Motivation : continuer à fournir une puissance de calcul croissante malgré la stagnation des micro-processeurs

- Calcul exaflopique : puissance crête d'un milliard de milliards d'opérations par seconde !
- Architecture fortement hétérogène et distribuée contenant plus d'un million de cœurs de calculs
- Problème crucial de la consommation électrique
- Le ratio Performance/\$/W devient incontournable

« The hardware costs for this forthcoming generation of high performance computers are estimated to be in the order of \$200 million with probably at least \$20 million a year in electricity costs to run them. »

— I. S. Duff, European Exascale Software Initiative : Numerical Libraries, Solvers and Algorithms

Quelques pistes pour utiliser efficacement les futures puissances de calcul

- 1 Limitation de la consommation électrique
- 2 Gestion efficace de la tolérance aux fautes
- 3 Evaluation et optimisation de la performance
- 4 Quantification des incertitudes
- 5 Validation et reproductibilité numériques

Ré-écrire ou adapter les codes de calcul ?

- Décision propre à chaque code de calcul en fonction de la complexité des modifications et de la rentabilité économique
- Adaptation facilitée par l'utilisation de bibliothèques de calcul dédiées (pérenité des développements)

Limitation de la consommation électrique

Le transfert de données est la part la plus importante de la consommation électrique d'une simulation

- Effort au niveau des algorithmes pour privilégier la localité des données
- Réduire la précision des nombres flottants, dans certaines parties du code et sans affecter la précision des résultats
- Algorithmes en précision mixte

Gestion efficace de la tolérance aux fautes

Le temps moyen de fonctionnement avant panne est inversement proportionnel au nombre de processeurs utilisés.

- Indicateur qui devient aussi important que l'accélération d'un code de calcul.
- Techniques efficaces de reprise de calcul hiérarchiques
- Méthodes de gestion de la tolérance aux fautes entraînant un surcoût limité en termes de temps de calcul, de consommation énergétique et de mémoire.

Evaluation fine et optimisation de la performance

Le paradigme de programmation parallèle traditionnellement utilisé vise à diminuer les volumes de calcul au détriment de la localité des données

- Remise en cause de ce paradigme dans des architectures massivement hétérogènes et distribuées (runtime)
- Besoin de méthodologies et d'outils d'évaluation fine de la performance
 - Thèse de Jean Pourroy au CMLA financé par HPE et encadré par C. Denis
- L'utilisation de bibliothèques de calcul dédiées pour certaines micro-architectures

Traitement des incertitudes

Deux principaux enjeux

- 1 Validation des résultats produits les codes de calcul complexes notamment pour les études de conception, de sureté ou de prevision.
 - Prise en compte des
 - incertitudes épistémiques (modélisation)
 - incertitudes stochastiques (paramètres physiques, conditions limites)
 - incertitudes numériques (approximation numérique lors de la résolution des modèles)
- 2 Optimisation de l'utilisation des codes pour faciliter la prise de décision
 - Analyse de sensibilité
 - Planification de plan expérience
 - Développement de modèles réduits

Validation et reproductibilité numériques

Prise en compte de l'usage intensif de l'arithmétique flottante (approximation sur ordinateur de l'arithmétique exacte)

- Plusieurs exécutions d'une même simulation numérique peuvent produire des résultats différents
- Développement de méthodologies et d'outils logiciels pour déterminer des statistiques sur la précision numérique
- Compromis entre la performance et la précision numérique tout en limitant la consommation électrique requise

Reproductibilité numérique \neq Précision numérique



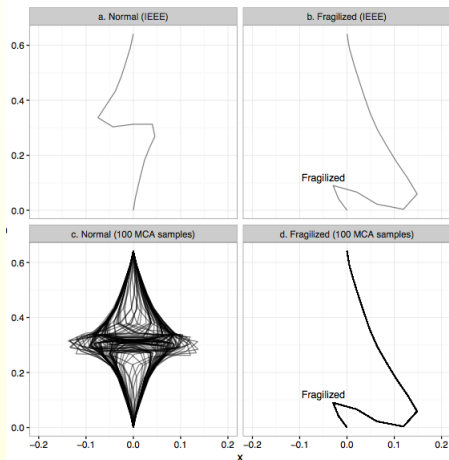
précision et reproductibilité
élevées



précision basse et
reproductibilité élevée

Illustration de la problématique de la reproductibilité numérique

Flambage d'une poutre



Conclusion

- La profonde modification architecturale des super-calculateurs pose de sérieux défis scientifiques au niveau du logiciel.
- Mise au point de méthodes alternatives (réduction de modèle) plus économe en temps de calcul et consommation électrique.
- La résolution de ces défis imposent des collaborations fortes entre académiques et industriels
- Un effort de formation/enseignement est également indispensable puisque ce type d'architecture bouleverse les paradigmes de programmation enseignés dans le passé.