

numéro spécial  
décembre 2017

GRANDS CHALLENGES GENCI/CINES 2017



DOSSIER spécial

## CALCUL INTENSIF

L'accès à des supercalculateurs de niveau mondial est un enjeu majeur pour les chercheurs et enseignants-chercheurs de l'Enseignement Supérieur et la Recherche (ESR) dont les recherches s'appuient de plus en plus sur des simulations numériques massives. Le calcul intensif irrigue maintenant tous les grands domaines scientifiques et la mise à disposition d'environnements HPC (*High Performance Computing*) de dernière génération est une condition sine qua non pour soutenir et accompagner au mieux les chercheurs dans la compétition mondiale.

La mise en production début 2017 de la seconde phase d'Occigen, soit l'ajout de 1260 nœuds Broadwell correspondant à un total de 35280 cœurs, répond précisément à cette exigence puisqu'elle a permis d'offrir 300 millions d'heures de calcul supplémentaires aux projets de recherche. Occigen, supercalculateur fourni par Bull-Atos, est maintenant dans sa configuration nominale de 3366 nœuds et 85824 cœurs, pour une puissance de calcul maximale de 3,5 Pflops qui le place au 54<sup>e</sup> rang mondial au dernier classement TOP500 de novembre 2017.

Cette réussite repose sur un partenariat stratégique fort entre GENCI (Grand Equipement National de Calcul Intensif) et le CINES (Centre Informatique National de l'Enseignement Supérieur), l'un qui finance l'achat du supercalculateur et l'autre qui l'exploite au quotidien après avoir défini conjointement à l'origine les spécifications attendues de la machine. Les équipes d'ingénieurs du CINES assurent en effet le bon fonctionnement d'Occigen qui passe par une infrastructure redondée de haut niveau (refroidissement par eau, électricité, climatisation etc...), une administration système et réseau des nœuds de calcul et bien sûr le support applicatif aux utilisateurs. L'efficacité de cette dernière mission, soit le support et l'expertise auprès des chercheurs qui veulent utiliser au mieux les capacités du supercalculateur, a été décisive dans l'envol rapide d'Occigen2 début 2017 et repose sur des échanges quasi quotidiens entre les équipes du CINES, de GENCI et de Bull-Atos.

Mais les chercheurs eux-mêmes ont contribué à ce succès via leur participation active aux Grands Challenges mis en place fin 2016 et début 2017. En effet, après la traditionnelle Vérification d'Aptitude contractuelle avec le constructeur, Occigen2 a été préalablement ouverte à une dizaine de gros projets ayant été sélectionnés par les Comités Thématiques GENCI. Chaque projet s'est ainsi vu offert la possibilité de faire « tourner » son code de simulation sur la quasi-totalité de la configuration dans sa phase d'installation finale. Ces premiers runs sont très intéressants à la fois pour les chercheurs concernés (levée de verrous scientifiques) et pour les équipes techniques du constructeur Bull-Atos et du CINES qui peuvent étudier la réponse du système de calcul fraîchement installé en condition de production maximale.

Cette gazette spéciale du CINES présente les résultats scientifiques majeurs obtenus lors de ces Grands Challenges sur Occigen2. Tous les grands domaines friands en calcul intensif sont représentés, soit les écoulements non réactifs, les Sciences de l'Univers, la physico-chimie des matériaux, la physique théorique, la modélisation moléculaire, la bio-santé ou les mathématiques appliquées. Chaque porteur de projet y présente le contexte scientifique de son Grand Challenge, sa méthodologie de recherche et les résultats importants qu'il a obtenus à cette occasion.

Le succès de cette campagne de Grands Challenges a ainsi reposé sur une collaboration étroite entre les équipes du CINES et de GENCI, le constructeur Bull-Atos et bien sûr les chercheurs qui portaient scientifiquement les projets sélectionnés après un important travail d'expertise mené par les Comités Thématiques GENCI. La qualité de tout ce travail collaboratif ressort bien dans cette gazette spéciale et nous remercions chaleureusement l'ensemble des acteurs de cette opération qui est emblématique de ce que l'ESR peut faire de mieux en matière de calcul intensif. Bonne lecture ! ■



**Boris DINTRANS**  
Directeur du CINES



**Philippe LAVOCAT**  
PDG de GENCI

# LE SUPERCALCULATEUR OCCIGEN EN CHIFFRES

## LE SUPERCALCULATEUR OCCIGEN

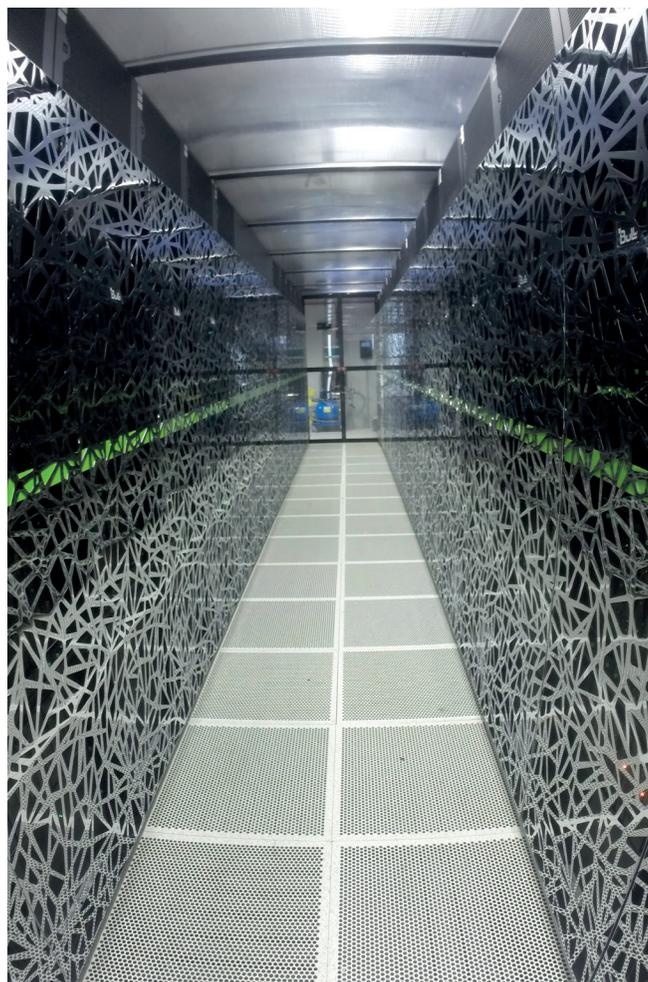
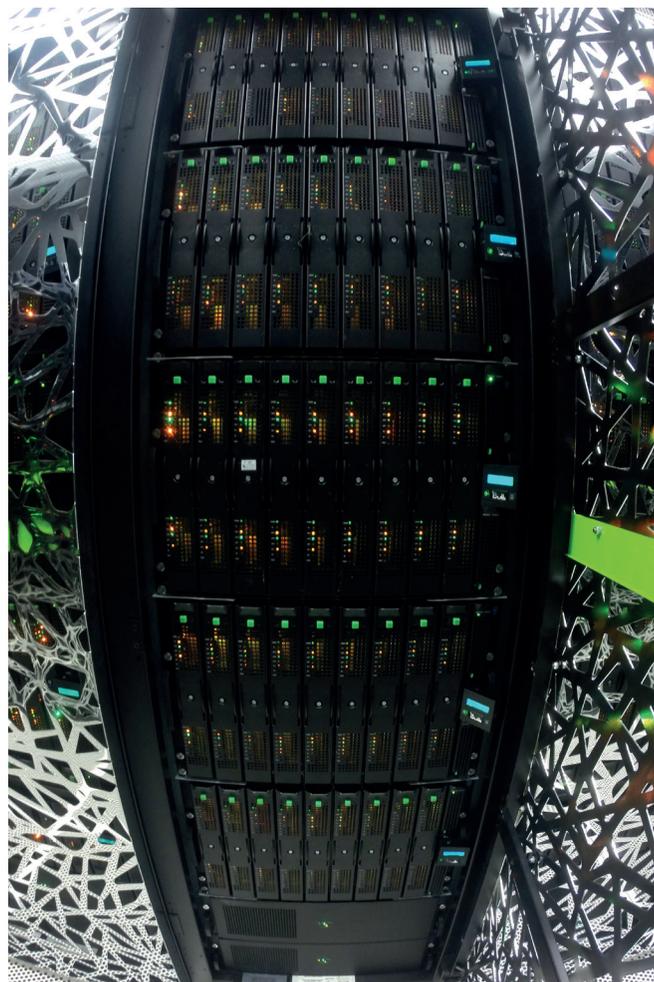
Le supercalculateur Occigen, financé par GENCI et exploité par le CINES depuis 2015, a accru de façon significative sa capacité de calcul à l'orée 2017, pour atteindre une puissance crête théorique de 3,5 PFlops (3,5 millions de milliards d'opérations flottantes par seconde), qui le situe au 54<sup>e</sup> rang mondial dans le classement de novembre 2017 des machines les plus performantes.

Les caractéristiques de cet environnement exceptionnel dédié à la simulation scientifique sont désormais les suivantes :

- **38 baies de calcul**, refroidies par un système haut rendement à eau tiède, directement sur les nœuds (technologie Direct Liquid Cooling – DLC) ;
- **7 baies** assurant l'accès à l'ensemble du cluster (via 5 nœuds de connexion et d'échange de données), son administration et la gestion de ses disques ;
- **85 824 cœurs** répartis sur **3336 nœuds** de calcul comprenant 2 processeurs chacun ; chaque

processeur comprend : 12 cœurs (Intel Haswell à 2.6 GHz) avec plus de 2,6 Go de mémoire par cœur (ou plus de 5,3 Go pour les nœuds à 128 Go), ou 14 cœurs (Intel Broadwell à 2.6 GHz) avec plus de 2,6 Go de mémoire par cœur ;

- Le réseau utilisé pour connecter les nœuds entre eux est un réseau **Mellanox Infiniband** (IB 4x FDR) dont la topologie est en forme d'arbre (Fat Tree) à trois étages ;
- **5 Pétaoctets** en ligne pour le stockage des résultats des calculs (système de fichiers parallèle Lustre) avec une bande passante maximum supérieure à 100 Go/s ;
- **260 Téraoctets** en ligne pour le stockage des fichiers utilisateurs et des codes à exécuter (système de fichiers Panasas), avec une bande passante de 10 Go/s ;
- **650 millions d'heures de calcul** disponibles sur l'année, **388 projets de recherche** et **900 chercheurs utilisateurs** (novembre 2017) ■



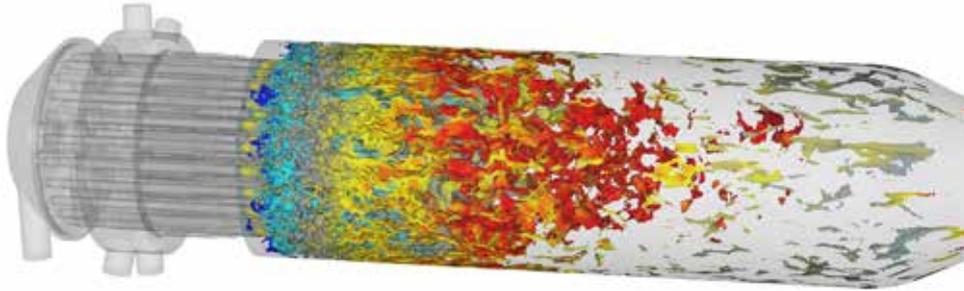
## Écoulements non réactifs

8

### BRUIT D'UN TURBOREACTEUR : EFFET DE TEMPERATURE CRYOTECHNIQUES

8

1	Contexte, enjeux et objectifs généraux	9	5	Évolution temporelle	11
2	Configuration	9	6	Influence de l'instabilité thermo-acoustique	11
3	Setup numérique	10	7	Conclusions & perspectives	12
4	Topologie des flammes (flammes stables)	10	8	Bibliographie	12



13

## Astronomie et géophysique

### LE PREMIER MILLIARD D'ANNÉES DE L'UNIVERS

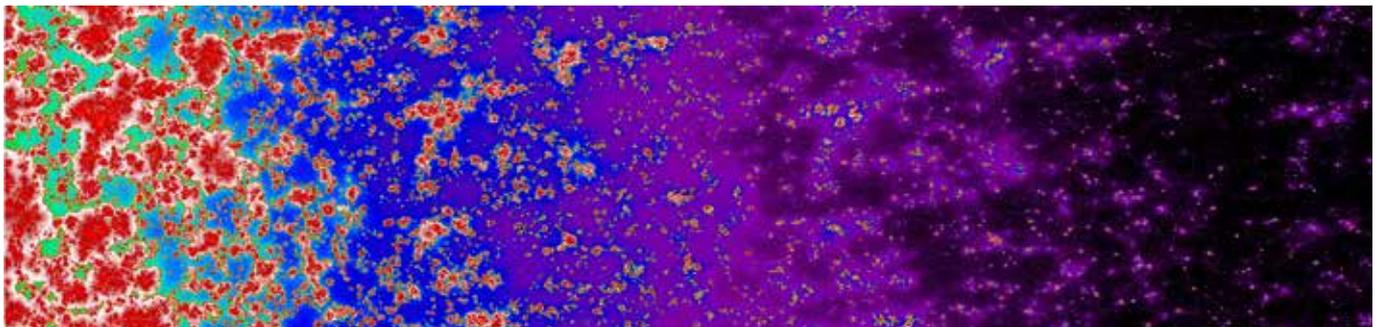
13

1	Contexte	14	4	Premiers résultats	15
2	Méthodes numériques	14	5	Perspectives	18
3	Mise en œuvre sur Occigen	15	6	Bibliographie	18

### ANALYSE DE SENSIBILITÉ EN MODÉLISATION SISMIQUE 3D

19

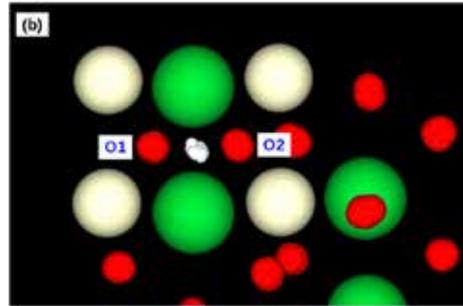
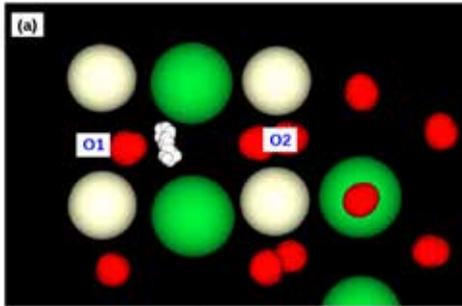
1	Contexte	20	6	Étude des performances parallèles	22
2	Description du code de calcul	20	7	Carte de sensibilité	24
3	Simulations réalisées	20	8	Conclusions, limites et perspectives	25
4	Cas test E2VP	20	9	Remerciements	25
5	Procédure de calcul et Mise en œuvre de l'application	21	10	Bibliographie	25



## Physique, chimie et propriétés de matériaux 26

### TRANSPORT DE L'HYDROGÈNE DANS UN ÉLECTROLYTE DE PILE À COMBUSTIBLE 26

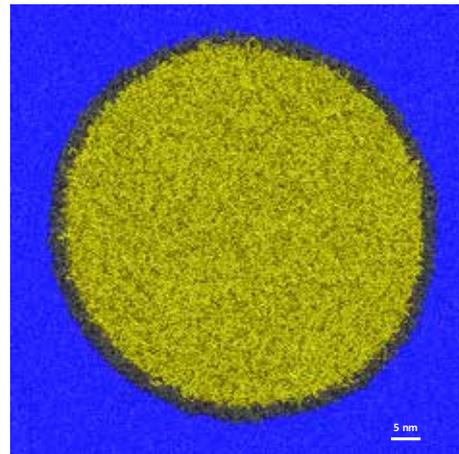
<b>1</b> Contexte, enjeux, objectif .....	27	<b>4</b> Observations et résultats scientifiques .....	29
<b>2</b> Code de calcul et moyens .....	27	<b>5</b> Conclusion .....	30
<b>3</b> Simulations réalisées .....	28	<b>6</b> Références .....	30



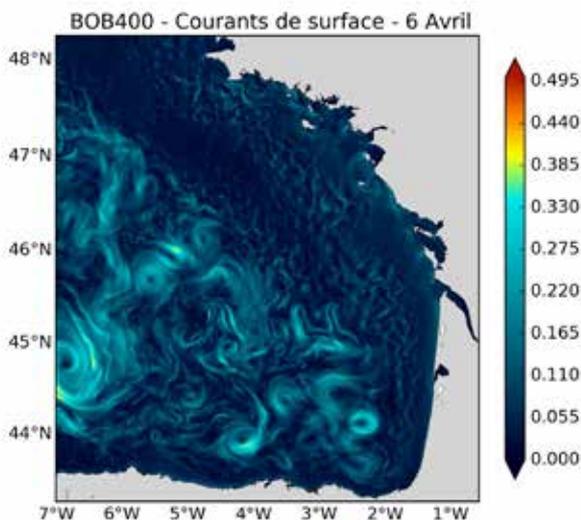
## 31 Dynamique moléculaire appliquée à la biologie

### LE STOCKAGE CELLULAIRE DES CORPS GRAS 31

<b>1</b> Contexte, enjeux, objectif .....	32
<b>2</b> Code de calcul et moyens .....	33
<b>3</b> Simulations réalisées .....	34
<b>4</b> Observations et résultats scientifiques .....	35
<b>5</b> Conclusion et remerciements .....	36
<b>6</b> Bibliographie et références .....	36



## Environnement 37



### INFLUENCE DES INSTABILITÉS SOUS-MÉSOÉCHELLES DANS LE GOLFE DE GASCOGNE 37

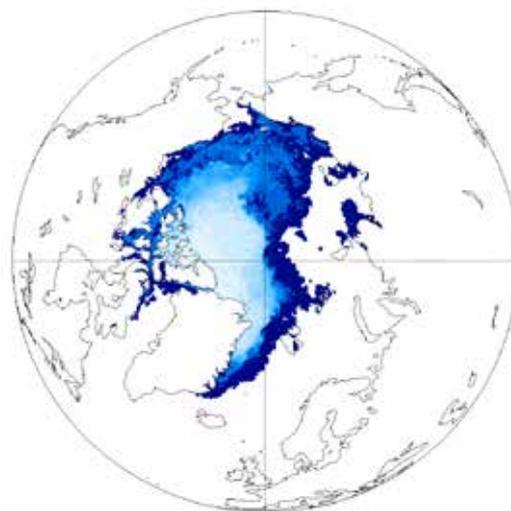
<b>1</b> Contexte scientifique .....	38
<b>2</b> Une transition pour la modélisation océanique côtière dans le golfe de Gascogne : de MARS3D à CROCO .....	39
<b>3</b> Premiers résultats : vers un océan instable .....	41
<b>4</b> Conclusions .....	43
<b>5</b> Références .....	44

## Environnement

45

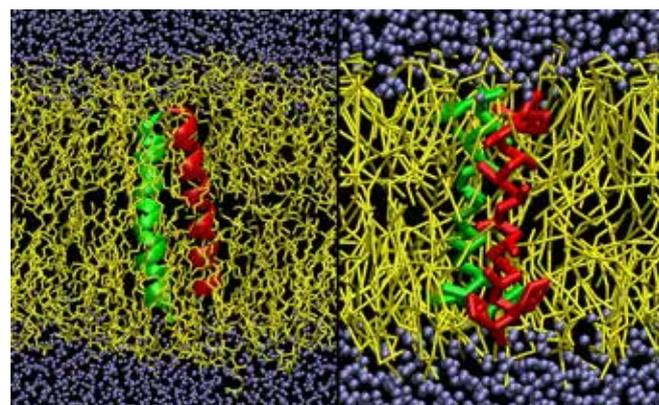
### SIMULATION NOUVELLE GÉNÉRATION DE L'OcéAN TURBULENT GLOBAL ..... 45

1	Contexte	46
2	Code, moyens de calcul, simulation	47
3	Résultats	48
4	Conclusions	52
5	Remerciements	52
6	Bibliographie	52



53

## Biologie et santé



### PHYSIQUE DU BOURGONNEMENT DES GOUTTELETTES LIPIDIQUES ..... 53

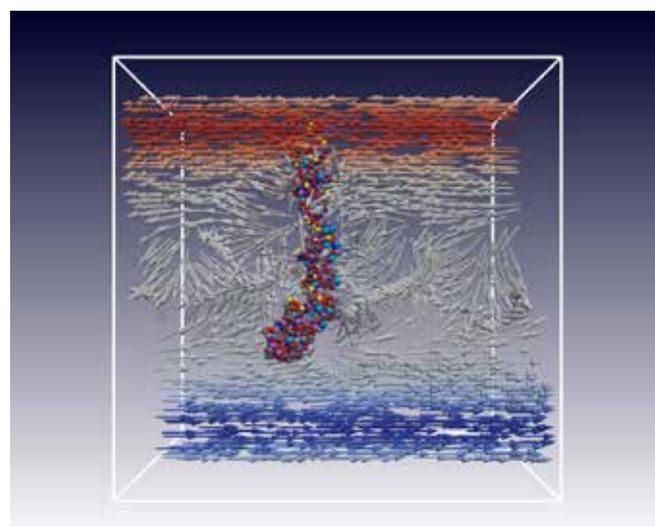
1	Résumé	54
2	Contexte	54
3	Objectifs	55
4	Méthodologie	55
5	Retour d'expérience sur l'utilisation de la machine	56
6	Résultats et discussion	57
7	Remerciements	58
8	Références	58

## Dynamique moléculaire appliquée à la biologie

59

### FIMH, UNE PROTÉINE CATCH-BOND QUI FONCTIONNE SOUS FLUX DE CISAILLEMENT ..... 59

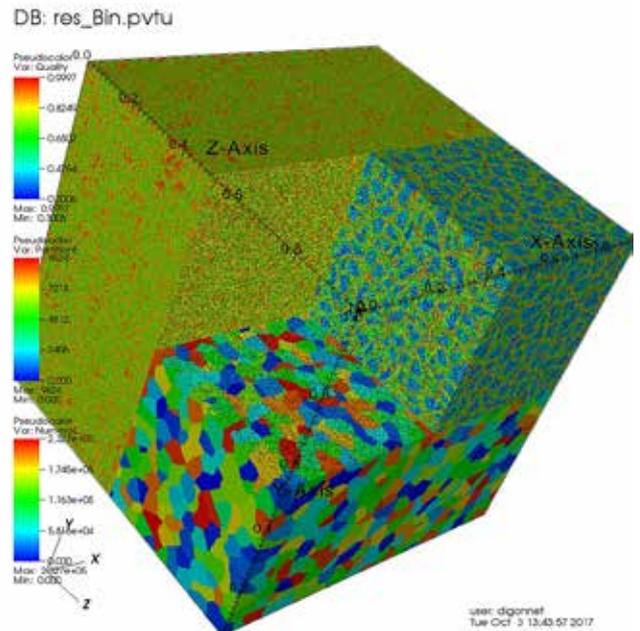
1	Contexte Scientifique et le grand défi.	60
2	La méthode	61
3	Le modèle	61
4	Le code Muphy	62
5	Résultats	63
6	Bibliographie	64



## 65

## Informatique, algorithmique et mathématiques

<b>TERA-MESH</b> .....	<b>65</b>
<b>1</b> Contexte .....	66
<b>2</b> Objectif du grand défi .....	67
<b>3</b> Code de calcul .....	68
<b>4</b> Résultats .....	68
<b>5</b> Perspectives .....	70
<b>6</b> Remerciements .....	71
<b>7</b> Références .....	71



## Physique théorique et physique des plasmas

## 72



## WATT - WEST ASPECT RATIO AND TURBULENT TRANSPORT .....

## 72

<b>1</b> Le contexte : ITER, le chemin vers la fusion thermonucléaire par confinement magnétique .....	73
<b>2</b> WEST et GYSELA en support à ITER .....	74
<b>3</b> Code de calcul GYSELA .....	75
<b>4</b> Préparation de GYSELA pour OCCIGEN2 .....	75
<b>5</b> GYSELA sur OCCIGEN2 .....	76
<b>6</b> Observations et résultats scientifiques .....	77
<b>7</b> Conclusion remerciement .....	79
<b>8</b> Bibliographie et références .....	79

# BRUIT D'UN TURBOREACTEUR : EFFET DE TEMPERATURE CRYOTECHNIQUES

Prédiction des instabilités de combustion  
haute fréquence dans les moteurs fusée.

Auteur

Gabriel STAFFLELBACH  
(CERFACS)



## Contributeurs

Thomas SCHMITT, Sebastien DUCRUIX (CNRS-EM2C)

Bertrand CIROU, Naima ALAOUI (CINES)

## SOMMAIRE

<b>1</b> Contexte, enjeux et objectifs généraux.....	9	<b>5</b> Évolution temporelle .....	11
<b>2</b> Configuration .....	9	<b>6</b> Influence de l'instabilité thermo-acoustique .....	11
<b>3</b> Setup numérique .....	10	<b>7</b> Conclusions & perspectives .....	12
<b>4</b> Topologie des flammes (flammes stables).....	10	<b>8</b> Bibliographie .....	12

## 1 Contexte, enjeux et objectifs généraux

L'industrie aérospatiale est confrontée à une concurrence internationale de plus en plus exacerbée. Les cycles de recherche et développement qui permettent d'innover sont de plus en plus contraints : il faut faire toujours plus vite et moins cher. Parce qu'appliquée avec succès dans d'autres domaines, la simulation numérique est un candidat de choix susceptible de révolutionner la recherche aérospatiale. Cependant, les technologies impliquées nécessitent des niveaux de précision qui vont bien au-delà des standards actuels et des ressources informatiques au plus haut niveau.

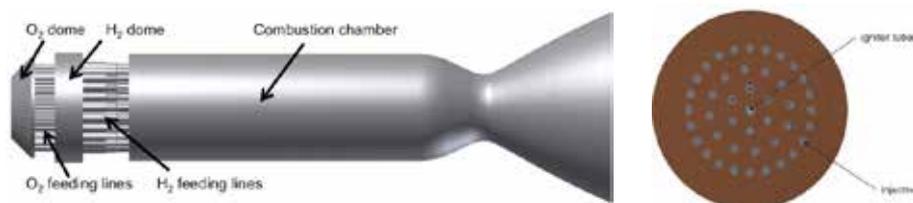
Les premiers développements des moteurs de fusées à ergols liquides au cours du XX<sup>e</sup> siècle ont souvent conduit à des instabilités de grande amplitude avec de graves conséquences et ce problème reste d'actualité. Malheureusement, la prédiction de ces instabilités demeure un des défis présents lors du design de moteurs fusées. Une instabilité de combustion (IC) provient du couplage résonnant entre le taux de dégagement de chaleur instationnaire des flammes et les modes propres acoustiques du système. Les très grandes densités de puissance dans ces moteurs (de l'ordre de 50 GW m<sup>-3</sup>) induisent des niveaux extrêmes de fluctuations de pression lors d'épisodes instables, qui augmentent les flux

de chaleur vers la paroi endommageant très rapidement la chambre de combustion et menant dans la plupart des cas à un échec du lancement. Généralement, ces instabilités sont à haute fréquence, correspondant à un mode acoustique transverse (noté nT avec n l'ordre du mode). Elles sont les plus dangereuses et destructives. À ce jour il n'existe pas de moyen de prédire avec précision, leur apparition et, le cas échéant, leur fréquence et leur amplitude.

Le CERFACS et le laboratoire CNRS EM2C, travaillent conjointement à l'extension du code AVBP (<http://www.cerfacs.fr/avbp>) pour prédire les instabilités de combustion dans les moteurs de fusée. Dans le cadre d'un grand challenge GENCI-CINES sur la machine OCCIGEN2 nous avons effectué des simulations haute-fidélité de la chambre de combustion d'un moteur de fusée à échelle réduite BKD (Gröning et al., 2014) étudiée expérimentalement par le laboratoire DLR.

L'expérience pouvant fonctionner à la fois dans des conditions stables et instables, elle donne l'occasion d'évaluer la capacité des outils numériques à prédire et distinguer entre les états stables et instable d'un moteur, le graal pour la conception de moteur de fusée. ■

## 2 Configuration



**Figure 1 :** Représentation du banc BKD du DLR. Gauche : vue de la configuration numérique. Droite : plan d'injection.

L'expérience du BKD (Gröning et al., 2014) est un modèle réduit d'un moteur de fusée, composé de 42 injecteurs coaxiaux. De nombreuses mesures ont été effectuées en variant les débits d'oxygène (O<sub>2</sub>) et de Di-Hydrogène (H<sub>2</sub>) ainsi que leurs températures. Pour ce grand challenge nous nous sommes focalisés sur l'influence de la température de l'hydrogène à l'injection vis-à-vis de l'apparition des instabilités thermo-acoustiques. Le

tableau 1 décrit en détail les 2 cas étudiés expérimentalement et simulés lors de ce grand challenge.

Il s'agit du cas T57 avec une température de H<sub>2</sub> de 57K et le cas T131 avec une température de H<sub>2</sub> de 131K. Le premier contient un mode 1T stable et un mode 2T faiblement perturbé alors que le second révèle un mode 2T stable et un mode 1T excité. ■

Cas	Débit O <sub>2</sub> [Kg/m <sup>3</sup> ]	Débit H <sub>2</sub> [Kg/m <sup>3</sup> ]	Température O <sub>2</sub> [K]	Température H <sub>2</sub> [K]
T57	5,8	0,97	107,6	57,6
T131	5,8	0,97	107,6	131,7

**Table 1 :** Description des cas simulés

### 3 Setup numérique

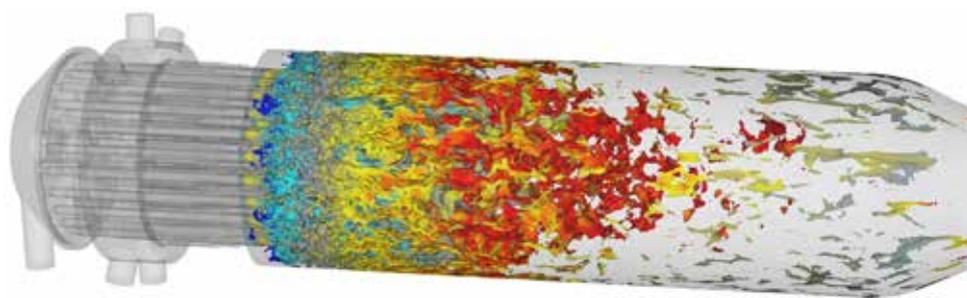
Les simulations ont été effectuées sur 143 et 286 nœuds de la machine OCCIGEN2 de GENCI au CINES avec le code AVBP-RG développé par le CERFACS et l'IFP Energies Nouvelles, étendu pour la physique gaz réel par EM2C et le CERFACS. On utilise le schéma numérique TTGC (3<sup>e</sup> ordre) et le modèle de sous-maille Wale.

Le passage à l'architecture Broadwell sur OCCIGEN2 offre une accélération d'un facteur 2 nœud à nœud comparée

à la machine CURIE du TGCC (équipée de processeurs Sandybridge).

Les simulations ont été effectuées sur un maillage de 300 millions de cellules raffiné dans les zones proches des injecteurs pour mieux capter la physique et les éléments géométriques de petite échelle (de l'ordre de 0.1mm). Les parois de la chambre sont isothermes (500K), les autres parois sont considérées comme adiabatiques. ■

### 4 Topologie des flammes (flammes stables)



**Figure 2 :** Visualisation des flammes (isosurface de température à 2100K coloriée par la vitesse axiale).

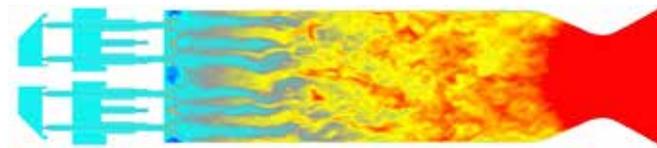
La figure 2 montre une visualisation instantanée des flammes avec une iso-surface de température à 2100 K coloriée par la vitesse axiale. On y observe des flammes très plissées et contenues dans la première partie de la chambre.

Des coupes longitudinales de masse volumique et de vitesse axiale sont montrées sur les figures 4 et 5. Ces images ont été prises avant l'apparition de l'instabilité pour le cas T57 discutée dans le paragraphe suivant.

On y observe que les jets denses pénètrent de manière similaire dans la chambre pour les deux cas simulés, jusqu'à un tiers de sa longueur. Leur topologie est très similaire dans les deux cas et ils sont entourés par des zones de haute vitesse d'hydrogène dont résulte les forts plissements des flammes. Les injections d'hydrogène produisent des zones de recirculation à grande vitesse au niveau du plan d'injection. ■



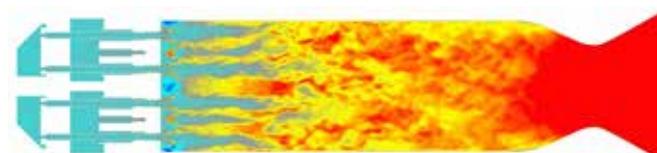
T57



T57



T131



T131

**Figure 3 :** Coupe longitudinale de masse volumique (rouge : 1150 kg/m<sup>3</sup>, bleu : 5 kg/m<sup>3</sup>).

**Figure 4 :** Coupe longitudinale de vitesse axiale (rouge : 500 m/s, bleu : -250 m/s).

## 5 Évolution temporelle

L'évolution temporelle de la pression dans la chambre est montrée sur la figure 5. Les fréquences inférieures à 5 kHz ont été filtrées. Ces figures montrent l'évolution de la pression depuis le début la simulation. Après quelques

millisecondes la pression se stabilise dans les deux cas mais après 10ms le cas T57 voit une augmentation soudaine et forte des fluctuations de pression rendant le système instable.

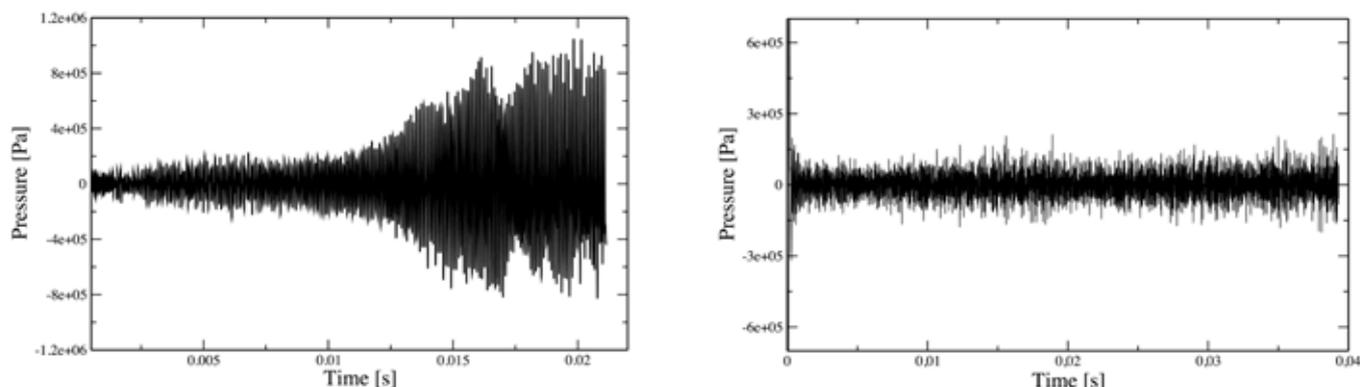


Figure 5 : Evolution de la pression. Droite : Cas T57, Gauche T131. Signal filtré [5khz-50khz]

Ces résultats vont à l'inverse des observations expérimentales où le T131 est fortement instable et le T57 plutôt stable. Cependant, le mode instable retrouvé est semblable à celui vu dans les expériences (en termes de fréquence et d'amplitude). Bien que la stabilité observée soit à l'inverse des expériences, il est important de noter qu'il

s'agit ici de la première simulation mondiale captant la transition naturelle de la stabilité d'un moteur de fusée en fonction des conditions d'opération. Ces résultats sont très encourageants et offrent l'opportunité d'observer les effets des instabilités sur les flammes. ■

## 6 Influence de l'instabilité thermo-acoustique

La figure 6 montre des champs de température, vitesse axiale, masse volumique et pression instantanées pour le cas T57 après la croissance de l'instabilité (22 ms sur la figure 5 à gauche). Par rapport aux illustrations présentées en Fig. 3 et 4, les flammes sont maintenant beaucoup plus courtes, notamment au centre de la chambre où la vitesse acoustique transversale est maximale. Ces

flammes, situées dans un nœud de pression et un ventre de vitesse, présentent une dynamique de type sinuoux. Les flammes extérieures, au contraire, essentiellement localisées dans un ventre de pression produisant des modulations de vitesse en sortie d'injecteur, présentent une forme variéuse. ■

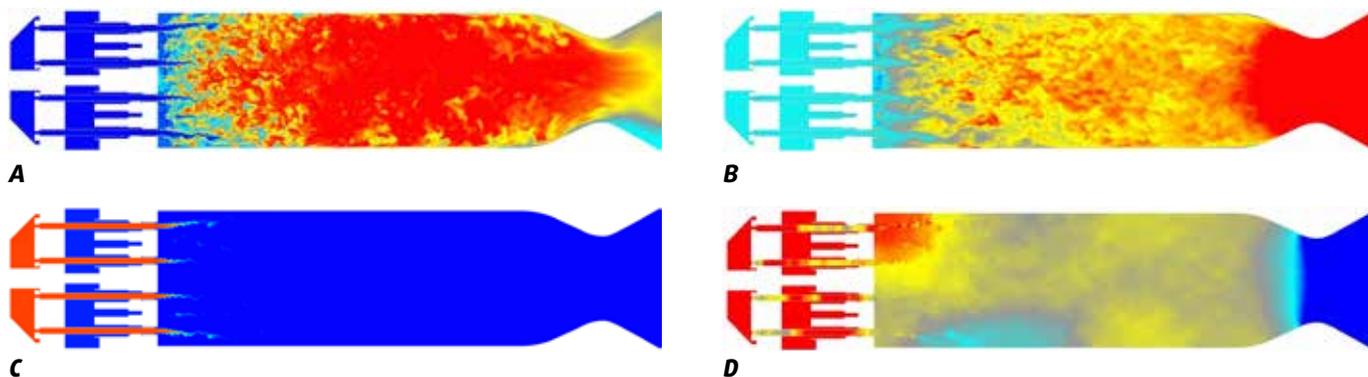


Figure 6 : Cas T57 (t=22 ms). Coupes longitudinales des champs instantanés de (a) température (rouge : 3600 K, bleu : 50 K), (b) vitesse axiale (rouge : 500 m/s, bleu : -250 m/s), (c) masse volumique (rouge : 1150 kg/m<sup>3</sup>, bleu : 5 kg/m<sup>3</sup>) et (d) pression (rouge : 90 bar, bleu : 60 bar).

## 7 Conclusions & perspectives

L'accès à la machine OCCIGEN2 durant ce challenge nous a permis de réaliser la première simulation au monde de transition naturelle entre un calcul stable et instable dans un moteur de fusée cryotechnique. Même si le comportement exact de l'expérience n'a pas été retrouvé, la structure et les fréquences des modes sont presque identiques entre la simulation et l'expérience ce qui constitue une avancée majeure pour la recherche dans ce domaine. Des travaux complémentaires sur les 3 Tera-octets de données générées sont toujours en

cours pour analyser les causes de cette transition et les incertitudes qui expliqueraient la divergence entre le numérique et l'expérience. OCCIGEN2 a permis d'effectuer ces travaux dans un temps record, avec une puissance de calcul deux fois supérieure nœud à nœud en comparaison avec l'architecture SandyBridge. Ces travaux ont fait l'objet d'une publication lors de la 7<sup>e</sup> conférence Européenne pour les sciences aéronautiques et spatiales (EUCCASS 2017) (Schmitt et al., 2017) et d'autres publications sont en préparation. ■

## 8 Bibliographie

■ [Colin] O. Colin, M. Rudgyard, "Development of High-Order Taylor-Galerkin Schemes for Unsteady Calculation", *J. Comp. Physics*, 162(2) :338-371, 2000.

■ [Gröning] Gröning D. Suslov, J. S. Hardi, and M. Oswald. Influence of hydrogen temperature on the acoustics of a rocket engine combustion chamber operated with lox/h<sub>2</sub> at representative conditions. *Proceedings of Space Propulsion*, 2014.

■ [Schmitt] T. Schmitt, G. Staffelbach, S. Ducruix, S. Gröning, J. Hardi, M. Oswald, Large-Eddy Simulations of a sub-scale liquid rocket combustor : influence of fuel injection temperature on thermo-acoustic stability. *7th European Conference for Aeronautics and Aerospace Sciences (EUCASS)*, 2017.

## LE PREMIER MILLIARD D'ANNÉES DE L'UNIVERS

Cosmologie, couplage gravitation et hydrodynamique

Auteur

Benoit SEMELIN  
(LERMA, Observatoire de Paris, UPMC)



### Collaborateurs

Florian BOLGAR, Evan EAMES (LERMA, Observatoire de Paris)

Victor CAMEO PONZ (CINES)

## SOMMAIRE

1	Contexte	14	4	Premiers résultats	15
2	Méthodes numériques	14	5	Perspectives	18
3	Mise en œuvre sur Occigen	15	6	Bibliographie	18

## 1 Contexte

Quelques centaines de milliers d'années après le Big Bang, l'univers est chaud, dense et presque homogène. L'hydrogène, l'hélium et des traces d'éléments plus lourds produits lors de la nucléosynthèse primordiale, se trouvent à l'état de plasma. L'expansion cosmologique provoque un refroidissement adiabatique de cette matière baryonique. Quelque 370 000 ans après le Big Bang, la densité d'énergie baisse suffisamment pour que les ions et les électrons se recombinent en atomes. Il s'en suit une période que l'on nomme « Age sombre », pendant laquelle la matière, neutre et froide, interagit très peu avec le fond de rayonnement cosmologique existant et produit très peu de nouveaux photons. Pendant cette période, les très faibles fluctuations de densité présentent dans l'univers primordial continuent de croître sous l'effet de leur propre gravité jusqu'à ce que se forment les premières étoiles, une centaine de millions d'années après le Big Bang. Ces premières étoiles sont très massives (jusqu'à cent fois la masse du soleil) et leur rayonnement, très énergétique, commence à ioniser le milieu qui les entoure. C'est le début de l'« époque de la réionisation ». La formation d'étoiles s'accélère, les galaxies primordiales apparaissent et génèrent dans le milieu intergalactique des bulles ionisées qui croissent jusqu'à percoler. Un milliard d'année après le Big Bang, l'univers est totalement réionisé.

Si comprendre les origines de l'univers tel que nous l'observons aujourd'hui constitue en soit un objectif fondamental, il y a une raison pour s'intéresser maintenant à l'époque de la réionisation. En effet, dans les quelques années à venir, deux très grands instruments astronomiques vont voir le jour et vont dévoiler directement cette époque en observant des objets très lointain dans l'espace et donc dans le temps. Le radiotélescope géant SKA (<https://skatelescope.org>) va nous permettre de dresser une carte tridimensionnelle du milieu intergalactique pendant l'époque de la réionisation en observant l'émission dans la raie 21-cm de l'hydrogène neutre. Cette carte dessinera (entre autres) en creux les bulles d'hydrogène ionisé autour des galaxies primordiales, tandis que le télescope spatial JWST (<https://www.jwst.nasa.gov>) observera directement l'émission optique (décalée dans l'infrarouge par l'expansion cosmologique) de ces galaxies primordiales.

## 2 Méthodes numériques

Le code LICORICE est un code multi-physique qui fonctionne dans un cadre cosmologique, c'est à dire qui prend en compte l'effet d'une expansion uniforme de l'univers sur les processus physiques. Le champ de matière est modélisé grâce à une approche lagrangienne basée sur

La simulation numérique est un outils indispensable, non seulement pour définir les spécifications des ces instruments en caractérisant le signal qu'ils vont observer, mais aussi, par la suite, pour interpréter les observations et remonter aux processus physiques qui régulent l'évolution de l'univers pendant cette période. Il s'agit néanmoins d'une tâche ardue. Pour modéliser un échantillon représentatif de l'univers, en particulier pour le signal à 21-cm du milieu intergalactique, il faudrait une boîte de simulation de 500 Mpc de côté (près de 1.8 milliard d'années lumière). Par ailleurs, pour résoudre toutes les galaxies dans lesquelles des étoiles peuvent se former facilement (sans même imaginer résoudre leur structure interne dans la même simulation), il faut que la masse d'un élément de résolution ne dépasse pas 107 masses solaires. La combinaison de ces deux contraintes se traduit par  $\sim 10^{12}$  éléments de résolution. De plus, pour pouvoir prédire les propriétés des galaxies ou l'émission à 21 cm, il est nécessaire de prendre en compte la gravité, l'hydrodynamique, le transfert radiatif multifréquence couplé à la dynamique, ainsi qu'un certain nombre de processus sous-grille liés en particulier à la formation stellaire.

Pendant longtemps, les simulations étudiant l'époque de la réionisation ont été limitées par la résolution du transfert radiative (à typiquement 108 éléments de résolution). Récemment des travaux s'intéressant autant à l'émission des galaxies qu'au signal 21-cm ont atteint et dépassé les 109 éléments de résolution, y compris pour le transfert radiatif (ex., Ocvirk et al. 2016, Semelin et al. 2017). Pour cette simulation dans le cadre d'un grand défi, nous avons opté pour une boîte de 300 Mpc de côté, un volume minimal vis à vis de notre objectif principal, modéliser le signal 21 cm, et une résolution en masse permettant d'identifier les galaxies de masses  $> 10^9$  masses solaires. En effet, sous ce seuil, différents effets tels que la rétroaction stellaire et le chauffage lié à la photo-ionisation limitent l'efficacité de la formation stellaire (Ocvirk et al. 2016), et donc l'impact sur le processus global de réionisation. En conséquence la simulation contient 20483 éléments de résolution, et les différents processus physiques sont tous traités au même niveau de résolution. ■

des particules. Le champ de gravitation est calculé grâce à un algorithme en arbre (Barnes & Hut, 1986) dont le coût est en  $Nx\ln(N)$ , où  $N$  est le nombre de particules. L'hydrodynamique (compressible) est prise en compte par la méthode du Smooth Particle Hydrodynamics

(SPH, voir par exemple Monaghan, 1992) qui permet d'interpoler les champs et leurs dérivées sur une grille irrégulière constituée par la positions des particules. Le transfert radiatif enfin, dans le continuum ionisant de l'hydrogène ( $E > 13.6$  eV) et dans le continuum X ( $100$  eV  $< E < 10$  keV), est réalisé par ray-tracing Monte Carlo (voir Baek et al, 2019). Le photo-chauffage résultant du transfert radiatif est pleinement couplé à l'hydrodynamique. D'autres processus physiques sont pris en compte sous forme de physique sous grille, comme la formation stellaire, ou en post-traitement, comme le transfert radiatif dans la raie Lyman- $\alpha$  qui est nécessaire pour calculer le signal 21-cm.

Le code bénéficie d'une parallélisation mixte OpenMP-MPI. La partie OpenMP, historiquement la première, a pu être optimisée pendant plusieurs années. On obtient typiquement une accélération d'un facteur 7 sur 8 threads,

et 13 sur 16 threads, hors situations pathologiques. La partie MPI date de 2013. Elle repose sur une décomposition en domaines spatiaux. Un choix radical a été fait sur la géométrie des domaines : une grille cartésienne de côté  $2^n$ . Ce choix a un avantage, il optimise les communications pour le calcul de la gravité. Le code utilise un mélange de communications collectives pour la gravité, et de communications non-bloquantes pour le transfert radiatif. Il y a néanmoins deux inconvénients au choix de la géométrie des domaines : un nombre de tâches MPI très rigide (8, 64, 512, 4096, etc...), et une impossibilité de faire de l'équilibrage de charge par déformation des domaines. L'équilibrage de charge n'avait jusqu'alors pas été un problème dans les applications du code (déséquilibre inférieur à 30 %). Avant le grand défi, le code avait été utilisé au maximum avec 512 tâches MPI et sur 4096 cœurs en production et 8192 cœurs en test. ■

## 3 Mise en œuvre sur Occigen

Le dimensionnement de la simulation a été guidé d'une part par la physique, comme expliqué ci-dessus, et d'autre part par l'ordre de grandeur du nombre d'heures de calcul disponibles dans le cadre d'un grand défi, résultant dans le choix de  $2048^3$  particules. Néanmoins, la mémoire effectivement disponible au total ( $\sim 40$  To) et par nœud (60 Go), ont été des contraintes importantes. Comme un fonctionnement à 512 tâches nécessite pour cette simulation au minimum 80 Go par tâche MPI, nous avons dû passer à 4096 tâches. Ceci correspond sur la tranche Grand Défis d'Occigen à 6 tâches par nœud (sur 683 nœuds) au minimum, soit 10 Go de mémoire disponible par tâche. En pratique nous avons rencontré des incidents quand nous avons dépassé 7 Go par tâche. Heureusement, le code permet d'ajuster le nombre de photons utilisés pour le ray-tracing Monte Carlo (et donc la mémoire occupée), au prix d'un niveau de bruit plus élevé du champ photo-ionisant, entraînant une instabilité accrue dans la résolution du système d'équations différentielles raide réglant l'évolution de l'état d'ionisation et d'énergie. En pratique, nous avons typiquement utilisé quelques dizaines de milliards de photons à chaque instant de la simulation. Néanmoins, alors que la simulation était à 60 % complète, le niveau

d'instabilité lié au nombre insuffisant de photons traversant chaque cellule à chaque pas de temps (quelques dizaines) a bloqué l'avancement. Nous n'avons pas pu résoudre ce problème avant la fin de la période des grands défis. Mais par la suite, en utilisant les nœuds d'Occigen à 128 Go de mémoires sur notre quota d'heures GENCI, et en ré-augmentant le nombre de photons émis (sans, bien sûr, modifier l'énergie émise par seconde), nous avons pu terminer la simulation sans difficulté supplémentaire. Cette situation illustre l'utilité d'un ratio RAM par cœur élevé pour certaines applications.

Au total, la simulation a nécessité de l'ordre de 5 Mh de calcul, et les sorties occupent 40 To de stockage. Nous en retenons un passage réussi au dessus de la barrière des  $2^{31}$  éléments de résolution, un passage réussi à 4096 tâche MPI et 16383 cœurs de calculs. Nous retenons également une dégradation substantielle de l'équilibre de charge dans la partie finale de la simulation, liée en toute probabilité à des résidus d'instabilité liée au bruit Monte Carlo des photons. Avant d'envisager d'introduire de l'équilibrage dynamique, nous devront travailler sur le solver du système raide (pour l'instant un simple solver semi-implicite avec pas de temps adaptatif). ■

## 4 Premiers résultats

Les sorties de la simulation sont de deux types. Tout d'abord, des instantanés de l'état du système (ici un cube d'univers) sont stockés à intervalles de temps prédéfinis. Les différentes grandeurs physiques associées à chacune des  $2043^3$  particules sont enregistrées. Si

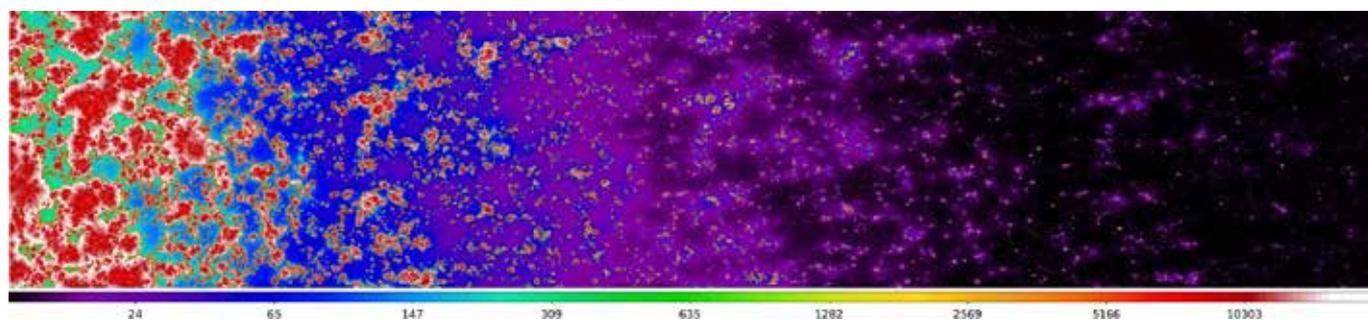
des telles sorties sont indispensables pour analyser la physique, elles ne correspondent pas à quelque chose de directement observable, même après avoir calculé l'intensité du signal 21-cm émis. En effet, la partie de l'univers que nous pouvons observer tel qu'il était il y

a un temps  $t$  dans le passé est une sphère de rayon  $cx t$  (où  $c$  est la vitesse de la lumière). Le champ d'un instrument comme SKA ayant une ouverture angulaire limitée (typiquement  $5^\circ \times 5^\circ$ ), nous devons construire ce que l'on appelle un cône de lumière, dans lequel chaque section perpendiculaire à l'axe correspond à un instant différent de la simulation. Malheureusement, entre le début et la fin de l'époque de la réionisation, la lumière parcourt une distance de l'ordre de 7 fois la taille de notre boîte de simulation. En prenant avantage de la périodicité de celle-ci, nous devons donc répliquer 7 fois notre boîte pour créer un cône de lumière (un cylindre à section carrée en pratique, en utilisant une approximation de type ciel plat). En calculant alors l'intensité du signal 21-cm, nous pouvons produire un signal directement observable. Lors de la simulation,

nous avons enregistré au vol, à chaque pas de temps, les particules dans la tranche contribuant au cône de lumière. Le post-traitement pour calculer le signal 21-cm sur le cône de lumière reste à faire. En effet l'intensité du signal s'exprime sous forme d'une température de brillance (Mellema et al., 2013) :

$$\delta T_B = 27. \dot{\delta} \frac{1-x}{\bar{\rho}} \frac{\rho}{\bar{\rho}} \frac{T_S - T_{CMB}}{T_S} \text{ mK}$$

Ici  $\rho$  est la densité locale du gaz,  $\bar{\rho}$  la densité moyenne de l'univers à cet instant,  $x_{\text{HII}}$  est la fraction d'ionisation du gaz,  $T_S$  la température de spin locale, et  $T_{\text{CMB}}$  la température du fond diffus cosmologique à cet instant.



**Figure 1 :** Cône de lumière (voir explication dans le texte) de la température du gaz dans la simulation. L'échelle de couleur donne la valeur de la température en Kelvin. L'univers est âgé de 360 millions d'année à droite de l'image, et de 760 millions d'année à gauche.

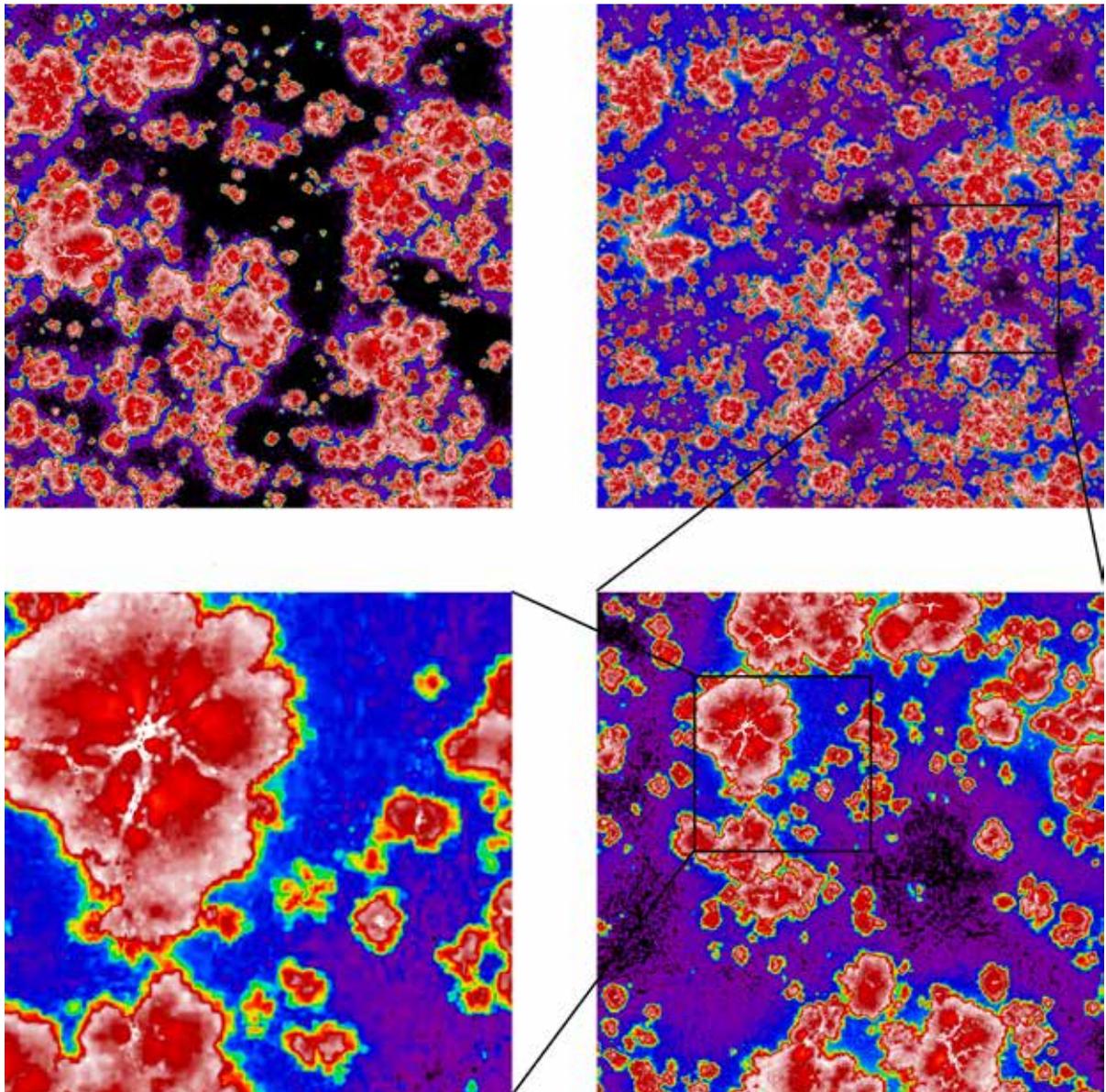
Or,  $T_S$  dépend entre autres de la température du gaz et du flux radiatif local à la longueur d'onde Lyman- $\alpha$ . C'est cette dernière quantité qu'il nous reste encore à calculer. La figure 1 présente donc une section du cône de lumière pour la température du gaz, qui à la fois détermine la température de spin et contient la trace de la fraction d'ionisation (le gaz ionisé est chaud, à  $T > 10^4$  K).

On peut observer les bulles ionisées chaudes (en rouge) qui apparaissent et croissent avec le temps (de la droite vers la gauche) jusqu'à percoler à la fin de la simulation.

L'objectif scientifique principal de la simulation est d'évaluer comment galaxies les moins massives impactent le signal 21-cm ainsi que les grandeurs dont il dépend comme la fraction d'ionisation et la température du milieu intergalactique. On s'attend en particulier à observer beaucoup plus de petites bulles ionisées que dans une simulation de moindre résolution. En effet le modèle cosmologique standard (sur lequel les conditions initiales et la physique de la simulation sont construites) prédit un beaucoup plus grand nombre de petites galaxies que de grosses. Pour vérifier cela, nous comparons la simulation Grand Défi, à une simulation à  $1024^3$  particules réalisée auparavant (Semelin et al., 2017) partageant les mêmes conditions initiales, à

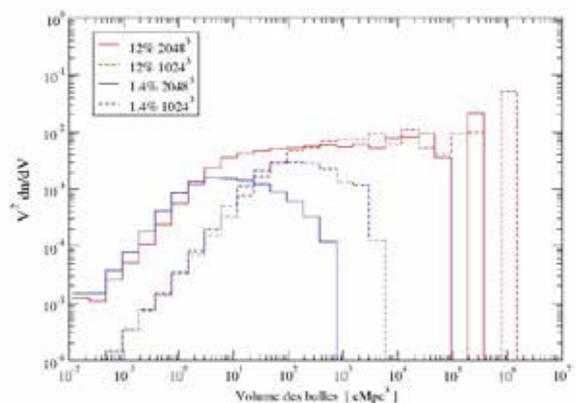
l'exception bien sûr des fluctuations aux plus petites échelles du champ de densité, seulement accessible avec la haute résolution. La différence de résolution en masse des deux simulations est donc d'un facteur 8. La figure 2 présente la même section d'un instantané (pas d'évolution temporelle dans l'image) du champ de température dans la boîte de simulation pour chacune des deux simulations. Les instants (différent pour les deux simulations) ont été choisis de manière à ce que le volume total occupé par le gaz ionisé soit le même. Comme attendu, on observe que les bulles ionisées (en rouge et blanc) sont plus grosses et moins nombreuses dans la simulation à plus basse résolution, hébergeant en leur centre des galaxies massive. Les galaxies de masse entre  $2 \times 10^9$  masses solaire et  $1.6 \times 10^{10}$  masses solaires ne sont pas résolue dans cette simulation, et les bulles ionisée correspondantes sont absentes. Pour quantifier l'impact de cette différence, nous calculeront le spectre de puissance des fluctuations de température et des fluctuations de l'intensité du signal 21-cm. Nous présentons ici un diagnostic plus simple et plus parlant : un histogramme du volume des bulles ionisées.

Si on note  $n$  le nombre de bulles dont le volume est inférieur à  $V$ , alors  $\frac{dn}{dV}$  est le nombre de bulles dans un intervalle de volume  $V$ , et  $V^2 \frac{dn}{dV}$  est le volume occupé par l'ensemble des bulles tombant dans un intervalle



**Figure 2 :** Cartes de température du gaz (échelle de couleur identique à Fig. 1) à un instant donné correspondant à 30 % du volume de la boîte ionisé. En haut à gauche, la simulation 10243, en haut à droite, la simulation 20483, en bas deux zooms.

logarithmique de volume. Cette grandeur est représentée sur la figure 3 sous forme d'histogramme, pour deux instants différents de l'histoire de réionisation et pour les deux niveaux de résolution. On a choisi deux époques assez précoces dans le processus de réionisation, 1.2 % et 12 % du volume réionisé, car par la suite le processus de percolation commence, et dès 30 % d'ionisation, une seule région constitue 90 % du volume d'ionisé et les histogrammes deviennent moins lisibles. On constate dans la simulation 1024<sup>3</sup> un fort déficit (un facteur > 10) de bulles d'une taille inférieure à 10 cMpc<sup>3</sup>. On peut attribuer cela à l'absence des galaxies de faible masse (et donc de faible luminosité ionisante) mais aussi au fait que la résolution étant plus faible, le gaz atteint des densités moins élevées sous l'effet de l'effondrement gravitationnel et a donc un temps de recombinaison plus long : une plus grande partie des photons atteint le front d'ionisation sans être absorbée et les bulles



**Figure 3 :** Histogramme du volume des bulles ionisées. La quantité  $v^2 \frac{dn}{dV}$  est égale au volume total occupé par les bulles d'un intervalle logarithmique de volume. Elle est ici normalisée au volume total de la boîte de simulation. Les volumes sont exprimés en Mpc<sup>3</sup> comobiles.

grossissent plus vite. ■

## 5 Perspectives

L'exploitation de la simulation ne fait que commencer. L'objectif principal est bien sûr de calculer l'intensité de l'émission 21-cm dans le cône de lumière et les grandeurs statistiques qui y sont associées comme le spectre de puissance des fluctuations. Nous comparerons alors aux mêmes grandeurs calculées pour la simulation 1024<sup>3</sup>, et établirons si la différence est supérieure au niveau de bruit attendu pour SKA. Dans le cas du spectre de puissance, la modification des fluctuations de la fraction ionisée pourrait induire un effet aux petites échelles, tandis que les fluctuations de température liée au chauffage par rayon X dans le gaz neutre (plus homogène pour des sources plus nombreuses) pourraient induire un effet aux grandes échelles.

Les possibilités d'analyse ouvertes par cette simulation sont cependant beaucoup plus vastes. Il sera en particulier intéressant d'étudier comment le taux de formation stellaire dépend de la masse de la galaxie hôte et comparer les résultats avec ceux de Ocvick et al. (2016). En effet notre résolution nous permet d'entrer (à peine) dans la gamme de masses des galaxies pour lesquelles le gaz chauffé par la photo-ionisation subit des forces de pression comparables à la gravité de

la galaxie. Il peut alors s'en échapper et ne plus être disponible pour la formation stellaire.

À plus long terme, rappelons que pour une simulation sans concession de l'époque de la réionisation et du signal 21-cm, il nous faudrait prendre en compte 8192<sup>3</sup> particules. Il faudrait par ailleurs pouvoir en réaliser suffisamment (plusieurs dizaines au moins) pour explorer l'espace des paramètres sous-grille qui influent sur le signal, une telle exploration servant alors d'échantillon d'apprentissage pour les méthodes d'inversion visant à remonter aux valeurs des paramètres à partir des observations (voir par exemple Shimabukuro et Semelin 2017, Kern et al. 2017 et Schmit et Pritchard 2017). Une simulation 4096<sup>3</sup> cube serait possible avec LICORICE dans le cadre de PRACE. Le programme d'observation de SKA s'étendra sans doute sur 5 ans (pour pouvoir accumuler 1000 h d'observation du même champ dans des conditions optimales), il est donc probable que nous avons encore près d'une décennie pour amener les simulations au niveau requis (une simulation nécessitant 64 fois plus de mémoire et de l'ordre de 1 milliard d'heures de calcul sur les processeurs actuels). L'objectif semble réalisable. ■

## 6 Bibliographie

- Baek et al., 2009, *A&A*, 495, 389
- Barnes J. and Hut P., 1986, *Nature*, 324, 446
- Kern et al., 2017, *ApJ*, 848, 24
- Mellema et al., 2013, *Experimental Astronomy*, 36, 235
- Monaghan J. J., 1992, *ARA&A*, 30, 543
- Ocvirk et al., 2017, *MNRAS*, 463, 1462
- Schmit, C. et Pritchard, J., 2017, arXiv :1708.00011
- Semelin B. et al., 2017, *MNRAS*, 472, 4508
- Shimabukuro H. et Semelin B., 2017, *MNRAS*, 468, 3869

## ANALYSE DE SENSIBILITÉ EN MODÉLISATION SISMIQUE 3D

Comment mieux quantifier les incertitudes en vue de mieux évaluer l'aléa lors de futurs séismes ?

### Auteurs

Fabrice DUPROS  
(BRGM)



Florent DE MARTIN  
(BRGM)



### Contributeurs

Philippe THIERRY (Intel)  
Abed BENAICHOUCHE (BRGM)  
Victor CAMEO PONZ (CINES)

## SOMMAIRE

<b>1</b> Contexte.....	20	<b>6</b> Étude des performances parallèles .....	22
<b>2</b> Description du code de calcul.....	20	Weak Scaling .....	22
<b>3</b> Simulations réalisées .....	20	Strong Scaling.....	23
<b>4</b> Cas test E2VP .....	20	I/O parallèle .....	23
<b>5</b> Procédure de calcul et Mise en œuvre de l'application ..	21	<b>7</b> Carte de sensibilité.....	24
Paramètres Incertains.....	21	<b>8</b> Conclusions, limites et perspectives .....	25
Maillages.....	21	<b>9</b> Remerciements .....	25
Filtres numériques et I/O parallèle .....	21	<b>10</b> Bibliographie .....	25

## 1 Contexte

Les simulations numériques tridimensionnelles deviennent de plus en plus prédictives et sont devenues essentielles pour améliorer la compréhension des phénomènes physiques complexes et souvent couplés en géosciences. Le calcul haute performance permet désormais une résolution spatiale plus fine que les données recueillies in situ pour construire les modèles utilisés par les simulations.

Dans le domaine du risque sismique, les méthodologies pour évaluer les niveaux réglementaires d'agressions sismiques sont en cours d'évolution. Autrefois basées sur un recueil d'informations telles que la géologie, la géotechnique, des données instrumentales, historiques,

les méthodologies en cours de développement ajoutent à ces dernières les résultats des simulations numériques 3D basées sur des interprétations physiques de la compréhension des séismes. Le BRGM anticipe de telles évolutions pour les futurs microzonages et les futures versions de carte d'aléa sismique. Le verrou pour intégrer les résultats des simulations 3D dans les futures cartes d'aléa repose sur la compréhension de l'origine de la variabilité du mouvement sismique et sur la quantification de cette variabilité. En conséquence, une meilleure compréhension des incertitudes épistémiques (liée à l'ignorance des paramètres du modèle) est maintenant un sujet crucial dans l'exploitation des prédictions numériques. ■

## 2 Description du code de calcul

En service depuis 2008 au BRGM, le code EFISPEC, «Element-FInis SPECtraux», (De Martin, 2011) permet de simuler des séismes, dans des milieux géologiques complexes. Il s'agit d'un code libre (CeCILL V2 et GNU GPL V3) qui résout les équations d'onde tridimensionnelles en

utilisant une méthode d'éléments finis spectraux (SFEM) pour calculer le champ d'onde sismique 3D (Komatitsch et Vilotte 1998). EFISPEC est programmé en FORTRAN95 et utilise la librairie MPI afin d'exploiter des architectures massivement parallèles. ■

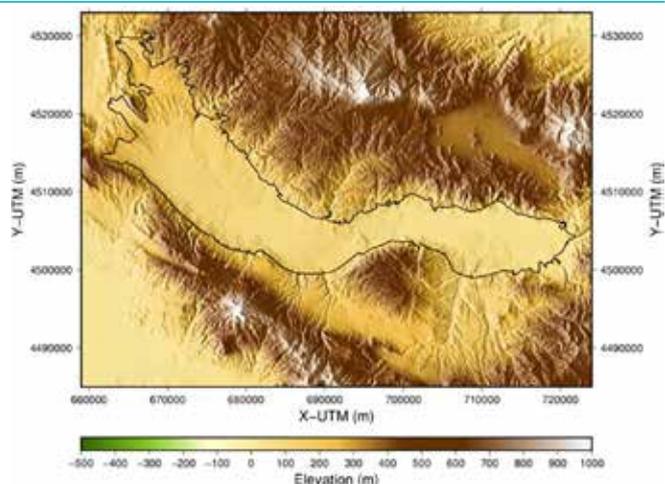
## 3 Simulations réalisées

L'effort à long terme de ce défi consiste à étudier la variabilité des mouvements du sol sismique associés aux incertitudes épistémiques du modèle E2VP. Afin d'éviter d'utiliser la méthode Monte Carlo standard, qui nécessiterait l'exécution d'un nombre trop important de modèles (par exemple, des centaines de milliers) pour explorer cette variabilité, notre étude repose sur des

méthodes de réduction de modèle. En plus des indicateurs stochastiques classiques tels que la moyenne et la variance des résultats préliminaires seront fournis concernant la sensibilité des grandeurs physiques (spectre de réponse) aux paramètres sismologiques. Nous avons également mené une étude approfondie des performances parallèles notamment des I/O. ■

## 4 Cas test E2VP

Le modèle de propagation des ondes, permettant de quantifier la variabilité du mouvement est issu du projet Euroseistest (E2VP). Il s'agit d'un projet collaboratif international notamment organisé conjointement par: l'Université Aristote de Thessalonique, Grèce; l'ITSAK (Institut de génie sismologique et de génie parasismique de Thessalonique), Grèce; le projet de recherche Cashima (soutenu par le CEA) et par ISTERre à l'Université de Grenoble Alpes, France. Le site cible de ce projet est le bassin de Mygdonian près de Thessalonique, en Grèce Il s'agit d'un site international de recherche et d'essai de nombreux projets internationaux de sismologie et d'ingénierie sismique. Il permet notamment une comparaison objective et quantitative entre les mouvements de sol sismiques enregistrés et leurs prédictions numériques (Chaljub et al. (2015). ■



**Figure 1 :** Modèle Numérique de Terrain du bassin E2VP (64 km x 42 km x 5km)

## 5 Procédure de calcul et Mise en œuvre de l'application

La figure ci-dessous montre le workflow de calcul incluant les phases de pré et post-traitement en vue de l'analyse de sensibilité à grande échelle. Tout d'abord, les paramètres incertains nous permettent de définir le plan d'expérience. Les modélisations tridimensionnelles sont réalisées en utilisant le logiciel EFISPEC. Enfin, une phase de filtre et de décimation des données est réalisée en parallèle afin de réduire la quantité de données sauvegardée.

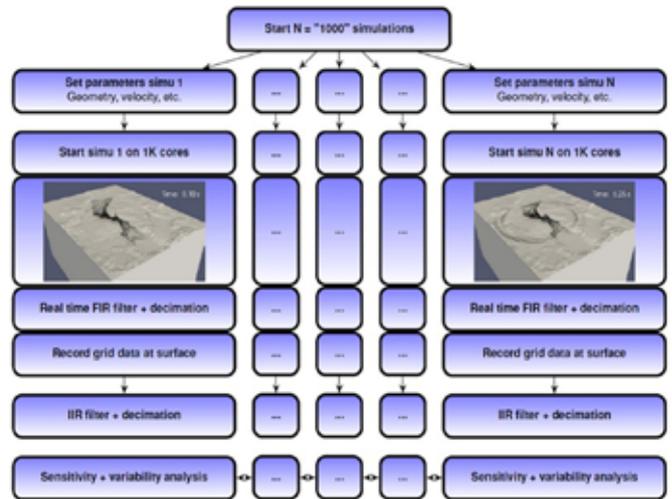
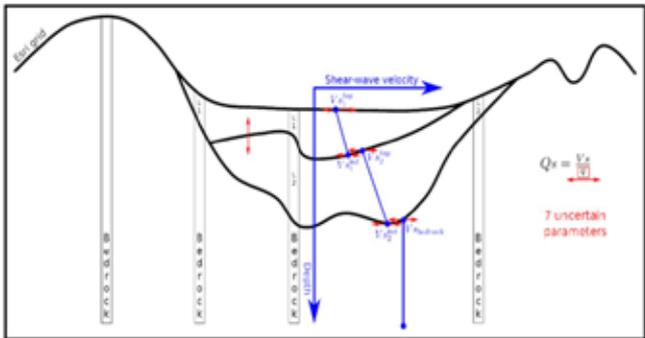


Figure 2 : Procédure de calcul pour l'analyse de sensibilité.

### Paramètres Incertains

Un maillage unique est généré sur la base du modèle le plus contraignant afin de respecter le critère de dispersion



numérique (5 nœuds GLL par longueur d'onde). Ce maillage est composé de 7 053 889 hexaèdres dont la taille varie de 312 mètres dans le rocher à 104 mètres à l'approche de la surface libre, et jusqu'à 35 mètres dans le bassin sédimentaire. Le raffinement du maillage est fait en utilisant une méthode de tripling proposée par le mailleur CUBIT. Nous utilisons une technique d'homogénéisation verticale permettant de garder un seul maillage.

Figure 3 : Paramètres incertains (7) utilisés pour l'étude de sensibilité cinq pour les vitesses des ondes de cisaillement, un pour l'amortissement et le dernier concerne la profondeur de l'interface des couches.

### Maillages

Un maillage unique est généré sur la base du modèle le plus contraignant afin de respecter le critère de dispersion numérique (5 nœuds GLL par longueur d'onde). Ce maillage est composé de 7 053 889 hexaèdres dont la taille varie de 312 mètres dans le rocher à 104 mètres à l'approche de la surface libre, et jusqu'à 35 mètres

dans le bassin sédimentaire. Le raffinement du maillage est fait en utilisant une méthode de tripling proposée par le mailleur CUBIT. Nous utilisons une technique d'homogénéisation verticale permettant de garder un seul maillage.

### Filtres numériques et I/O parallèle

Les sismogrammes à chaque Points Gauss-Lobatto-Legendre (GLL) (utilisés comme récepteurs) de la surface libre (c'est-à-dire 12 297 150 points GLL) sont sauvegardés pour chaque simulation. Afin de réduire le volume de données générées, la stratégie suivante est mise en œuvre : Les sismogrammes sont traités en cours d'exécution par un filtre à réponse finie-impulsionnelle (FIR). Les valeurs sont décimées d'un facteur 10 afin de conserver les informations pertinentes (cf. Figure). L'utilisation de ce filtre évite l'aliasing avant la décimation.

Après cette première phase de filtrage et décimation, la taille du fichier est réduite de 21 To à 2.1 To.

A la fin de chaque simulation, le fichier 2.1 To est relu en utilisant la librairie MPI I/O afin de transposer les données de la représentation initiale (Temps - Processeur - Quadrangle - Point GLL - Composante) vers une organisation optimale pour le post-traitement (Processeur - Quadrangle - Point GLL - Composante - Temps). Cette transposition vise à optimiser la continuité dans aux données lors de l'application des routines d'analyse

de quantification de l'incertitude. De plus, Les séries temporelles sont traitées par un filtre Butterworth et décimées par un facteur 15 pour atteindre une fréquence d'échantillonnage quatre fois supérieure à la

validité spectrale du maillage. Au total, l'espace disque est diminué d'un facteur 150 et la taille des données sauvegardées pour chaque simulation correspond à un fichier binaire de 137 Go. ■

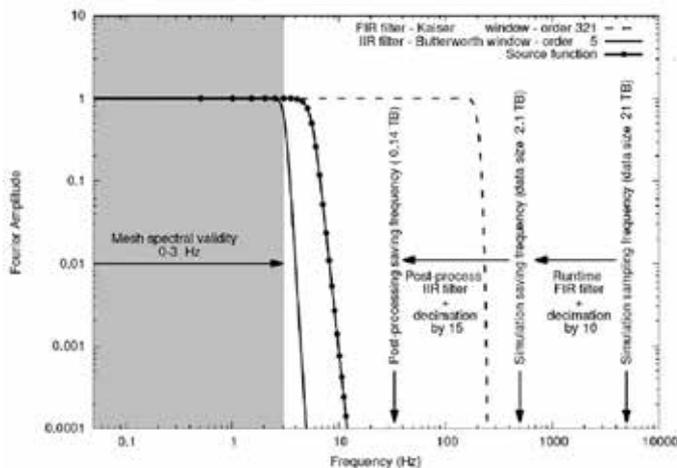


Figure 4 : Procédure de filtre à deux niveaux appliquée durant les phases de calcul et de post-traitement.

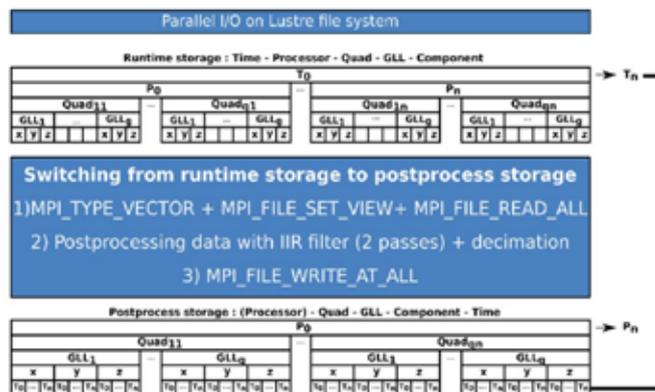


Figure 5 : Stratégie de transposition des données Durant la seconde phase de post-traitement.

## 6 Étude des performances parallèles

### Weak Scaling

L'étude du weak scaling est basée sur un domaine cubique initial discrétisé avec 25 000 hexaèdres par nœud de calcul. L'analyse est effectuée sur le supercalculateur Occigen 2. Nous comparons les résultats obtenus aux tests réalisés sur le supercalculateur ShaheenII localisé en Arabie Saoudite à l'université KAUST. Ce système est un Cray XC40 avec un peak théorique de 7.2 pétaflops et se compose de 6 174 nœuds de calcul (dual socket Intel Xeon E5-2698v3). Au total le système propose 197 568 cœurs de calcul.

Sur les deux systèmes, les résultats sont très similaires et le code montre un bon scaling jusqu'à 128 nœuds de calcul. Typiquement, on observe une efficacité de 94 % sur 3 584 coeurs sur Occigen2.

Les résultats sont obtenus sur un maximum de 128 nœuds de calcul sur Occigen2 et 2 048 nœuds sur Shaheen II.

Pour un plus grand nombre de cœurs une dégradation de l'efficacité est observée. Le volume croissant de communication et la concurrence sur le réseau expliquent en partie cette dégradation. En effet, les tests correspondent à une utilisation non-exclusive de la machine. Néanmoins le code maintient de bonnes performances avec une efficacité parallèle de 79% sur 65 536 cœurs sur la machine Shaheen II.

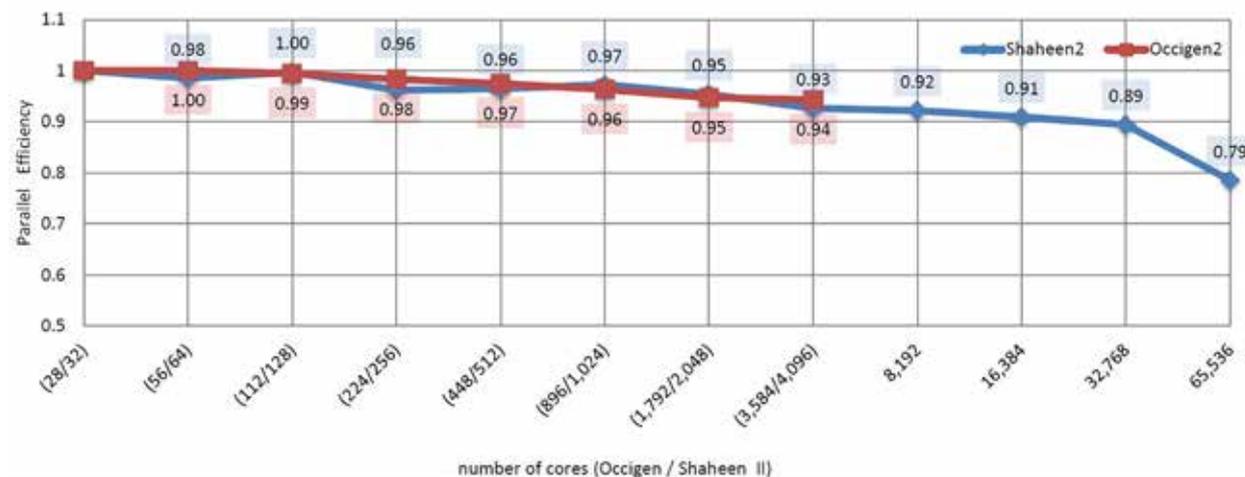


Figure 6 : Weak scaling sur les plateformes Occigen 2 et Shaheen2.

## Strong Scaling

Les tests de strong scaling sont basés sur un maillage composé de 1 906 624 hexaèdres. Ces expériences sont menées jusqu'à 7168 cœurs ce qui représente 256 nœuds dual-socket. Les implémentations MPI bloquantes et non bloquantes sont utilisées et les résultats obtenus sur deux nœuds de calcul sont utilisés comme référence.

La figure montre une performance super linéaire pour la version MPI non bloquante. Par exemple, sur 128 cœurs, nous observons une accélération de 135,3. Cette moyenne de 10% d'augmentation de la performance est cohérente avec les résultats rapportés par ailleurs dans Rietmann 2012 et vient d'une meilleure utilisation du cache lorsque la taille du problème diminue. Pour le nombre maximum de cœurs de calcul (7,168), la dégradation provient du ratio défavorable entre le volume de calcul et le coût des communications. En effet, chaque cœur ne calcule que 265 hexaèdres dans

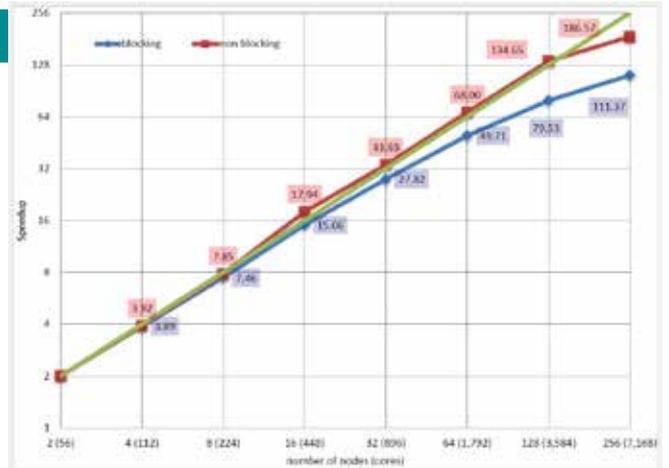


Figure 7 : Strong scaling sur la plateforme Occigen 2 pour les versions MPI bloquantes et non bloquantes.

ce dernier cas. Globalement, le ratio entre les versions MPI bloquantes et non bloquantes est de 1.5..

## I/O parallèle

### Phase de calcul dynamique

Comme expliqué dans la section précédente, nous sauvegardons les sismogrammes à chaque récepteur de la surface libre représentant 12 297 150 points. Ces séries chronologiques sont filtrées par un filtre à réponse impulsionnelle finie (FIR), puis décimées par un facteur 10 et finalement écrites sur un fichier partagé en utilisant la bibliothèque MPI-IO.

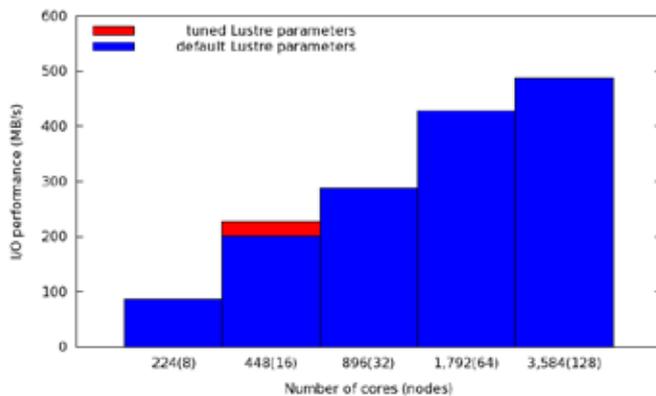


Figure 8 : Bande-passante applicative sur Occigen 2 obtenue avec le code EFISPEC. Les résultats correspondant aux paramètres Lustre par défaut sont en bleus et les améliorations mesurées en optimisant ces paramètres en rouge.

La figure ci-dessus montre la bande passante apparente de l'application. Cette métrique correspond à la différence de performance entre l'exécution de l'application avec et sans I/O (Paciucci2016). Nous effectuons deux types d'expériences : Premièrement, nous utilisons les paramètres Lustre par défaut pour réaliser un test de strong scaling. Deuxièmement, nous évaluons le gain potentiel en adaptant les paramètres lustre (stripe size et count size).

Les performances d'I/O augmentent avec le nombre de clients Lustre atteignant un maximum de 0.48 Go/s. En revanche, en utilisant un grand nombre de nœuds de calcul (au-delà de 64 nœuds), nous observons un peu de gain de performance.

De plus, les gains provenant de l'optimisation des paramètres Lustre sont limités. Les résultats montrent une amélioration maximale de 13 %. Dans la plupart des cas, les configurations Lustre par défaut fournissent déjà de très bons résultats. Globalement, le schéma d'I/O de notre application conduit à l'écriture de quelques kilooctets de données lors de chaque étape. Un mécanisme de bufferisation manuel des I/O seraient probablement bénéfiques à ce stade.

### Phase de post-traitement

A la fin de chaque simulation, nous appliquons une procédure de post-traitement en deux étapes. Dans un premier temps, nous relisons un fichier 2.1 To généré lors du calcul dynamique en utilisant tous les processus

	1MB	2MB	4MB	8MB	16MB	32MB
1	0.34	0.33	0.35	0.33	0.32	0.40
6	1.04	1.37	1.96	2.48	2.47	2.43
12	1.31	1.92	2.70	3.82	4.33	4.43
24	1.63	2.59	3.92	5.24	5.75	6.48
48	1.91	2.80	4.08	6.15	6.59	6.25
96	2.26	3.06	4.37	5.59	6.56	6.50

Figure 9 : Impact des paramètres du système de fichiers Lustre sur la bande passante mesurée pour la phase lecture des données (2.1 To). Le stripe size varie de 1 à 32 Mo horizontalement et le stripe count de 1 à 144 verticalement.

MPI disponibles (soit 2.048 cœurs pour nos expériences). Ensuite, les données sont filtrées et décimées. Enfin, la dernière opération d'écriture transpose les données afin d'optimiser les analyses statistiques futures.

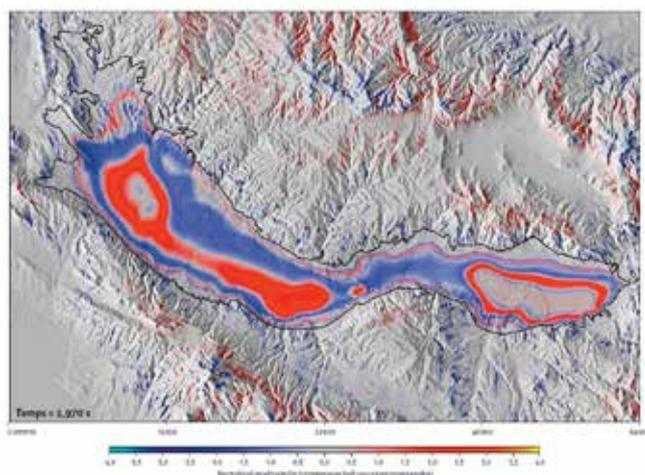
La figure ci-dessus montre les résultats pour cette phase de post-traitement. Deux paramètres Lustre sont étudiés (le stripe size représenté horizontalement et le stripe

count représenté verticalement). Nous pouvons remarquer l'importance de l'optimisation de ces paramètres, car les performances sont multipliées par un facteur de 11.8 avec des valeurs optimales (valeur maximale de 6,50 Go/s). Nous bénéficions probablement d'une meilleure distribution des blocs sur les OST du système Lustre mais également d'opérations d'I/O sur des données contigües. ■

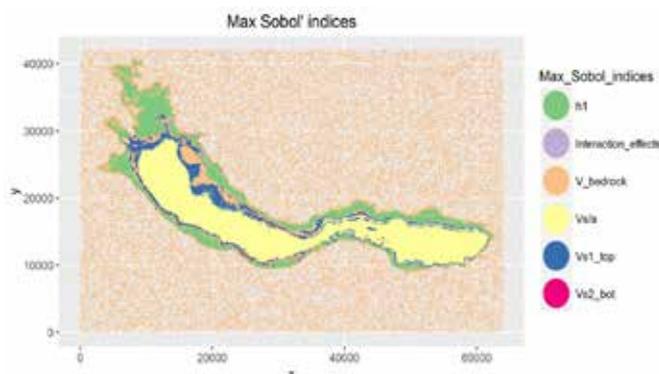
## 7 Carte de sensibilité

Nous avons effectué 500 simulations sur Occigen 2. Chaque simulation a été lancée sur 2.048 cœurs et a duré 4h30min générant un total de 70 To (137 Go de sismogrammes par simulation).

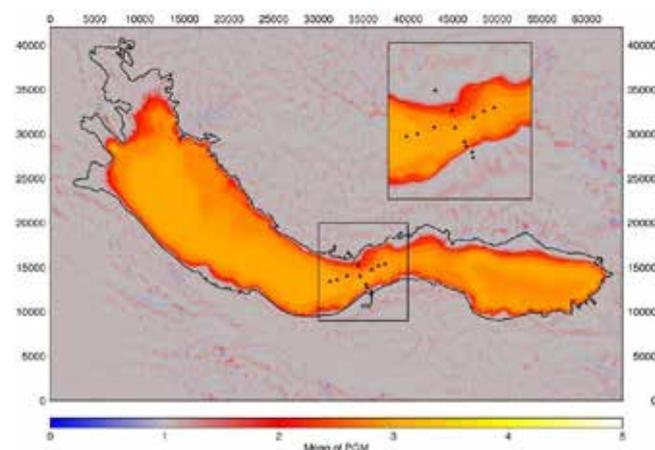
La Figure ci-dessous illustre, par un instantané en vue satellite, la propagation des ondes sismiques dans la bassin Mygdonien. L'amplitude des ondes est représentée par une échelle de couleur



**Figure 10 :** Vue satellite montrant un arrêt sur image de la propagation des ondes sismiques calculée en milieu géologique tridimensionnel complexe par le logiciel EFISPEC3D. L'amplitude des ondes est représentée par une échelle de couleur (rouge : positive, bleu : négative, gris : nulle).



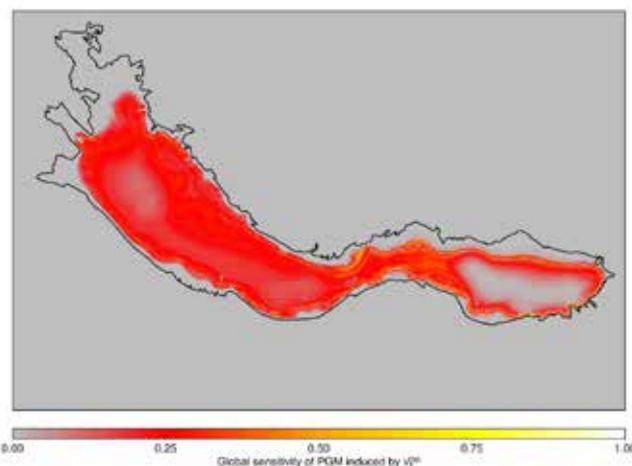
**Figure 12 :** Analyse des effets de premier ordre Sobol'.



**Figure 11 :** Peak Ground Motion (PGM) moyen.

A ce stade, l'analyse des 70 To de données constitue un effort à long termes. Les premiers résultats s'intéressent au Peak Ground Motion (PGM) normalisé comme quantité d'intérêt. Pour l'estimation des indices de Sobol avec 7 entrées aléatoires, un millier d'échantillons ont été générés en utilisant une approche de type Quasi-Monte-Carlo (Faure 2002). Les simulations montrent un PGM moyen entre 0.7 et 5 et un écart type entre 0 et 1.2..

L'analyse des effets de premier ordre Sobol sur le PGM montre qu'au centre du bassin, le facteur de qualité des ondes S ( $V_s/a$ ) est le paramètre le plus influent.



**Figure 13 :** Valeur de l'indice Sobol pour le paramètre  $V1\_top$

En dehors du bassin, la réponse est principalement contrôlée par le paramètre  $V_{\text{bedrock}}$  car le milieu de propagation est homogène (et dont la vitesse des ondes S est égale à  $V_{\text{bedrock}}$ ).

La figure 12 illustre la valeur de l'indice Sobol pour le paramètre  $V1_{\text{top}}$ . Il varie entre 0 (à la frontière et à l'extérieur du bassin) et 0.5. L'interprétation de ces résultats est intéressante d'un point de vue de la

sismologie car une telle cartographie n'est pas intuitive. Ces résultats montrent que les paramètres à privilégier pour réduire les incertitudes sur la prédiction du PGM dépendent fortement de la localisation spatiale. Par exemple, en milieu de bassin sédimentaire, les efforts devraient être concentrés sur la détermination du facteur de qualité de l'onde de cisaillement ( $Q_s = V_s / a$ ) afin de réduire l'incertitude sur le PGM. ■

## 8 Conclusions, limites et perspectives

Les études de sensibilité dans le domaine de la modélisation sismique tridimensionnelles constituent un enjeu majeur pour la prévention des risques. Les architectures pétaflopiques telles que Occigen 2 permettent d'envisager la réalisation de ces modélisations très gourmandes en temps de calcul.

L'optimisation des phases d'I/O demeure un point critique et des améliorations sont encore à envisager à ce niveau tant au niveau des implémentations applicatives, des supports d'exécution que des configurations

matérielles. Typiquement, un des verrous importants rencontrés durant ces expériences est venu du système de fichiers Lustre et de la difficulté à mener de manière concurrente l'ensemble des simulations. Ces dernières ont dû être découpées par lots, exploitant 2 048 cœurs pour chaque run.

Ces premières cartes de mouvement du sol sont prometteuses et ouvrent la voie à de nombreuses applications dans le domaine du littoral, du climat ou des ressources en eau. ■

## 9 Remerciements

Nous remercions le Supercomputing Laboratory at King Abdullah University of Science and Technology (KAUST)

in Thuwal, Saudi Arabia pour l'accès au supercalculateur ShaheenII ■

## 10 Bibliographie

■ Komatitsch, D., & Vilotte, J. P. (1998). The spectral element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures. *Bulletin of the seismological society of America*, 88(2), 368-392.

■ Le Maître, O., Knio, O.M. (2010) *Spectral methods for uncertainty quantification*, Springer.

■ Maufroy, E., E. Chaljub, F. Hollender, J. Kristek, P. Moczo, P. Klin, E. Priolo, A. Iwaki, T. Iwata, V. Etienne, F. De Martin, N. Theodoulidis, M. Manakou, C. Guyonnet-Benaize, K. Pitilakis, and P.-Y. Bard (accepted) Earthquake ground motion in the Mygdonian basin, Greece: the E2VP verification and validation of 3D numerical simulation up to 4 Hz. *BSSA*. vol. 105 [2015] p.1398-1418

■ Sochala, P., De Martin, F. : Surrogate combining harmonic decomposition and polynomial chaos for seismic shear waves in uncertain media. *Computational Geosciences*, 2017

■ G. Paciucci, F. De Martin, and P. Thierry, "Enhancing scalability and performance of parallel file systems." Available: <https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/enhancing-scalability-and-performance-white-paper.pdf>

■ F. De Martin, "Verification of a spectral-element method code for the southern California earthquake center loh.3 viscoelastic case," *Bull. Seism. Soc. Am.*, vol. 101, no. 6, pp. 2855-2865, 2011.

■ M. Rietmann, P. Messmer, T. Nissen-Meyer, D. Peter, P. Basini, D. Komatitsch, O. Schenk, J. Tromp, L. Boschi and D. Giardini, Forward and Adjoint Simulations of Seismic Wave Propagation on Emerging Large-Scale GPU Architectures, *Proceedings of the ACM / IEEE Supercomputing SC'2012 conference*, article #38, p. 38:1-38:11 (2012).

■ H. Faure, S. Tezuka, K. Fang, F. Hickernell, and H. Niederreiter, "Another random scrambling of digital (t, s)-sequences," *Monte Carlo and Quasi-Monte Carlo Methods 2000*, pp. 242-256, 2002.

# TRANSPORT DE L'HYDROGÈNE DANS UN ÉLECTROLYTE DE PILE À COMBUSTIBLE

Grand Défi OCCIGEN - CINES

Auteur

Marc TORRENT  
(CEA, DAM, DIF)



## Contributeurs

Grégory GENESTE (CEA, DAM, DIF)

Bertrand CIROU, Naima ALAOUI (CINES)

## SOMMAIRE

<b>1</b> Contexte, enjeux, objectif ..27	<b>5</b> Conclusion..... 30
<b>2</b> Code de calcul et moyens...27	<b>6</b> Références..... 30
<b>3</b> Simulations réalisées .....28	
<b>4</b> Observations et résultats scientifiques.....29	

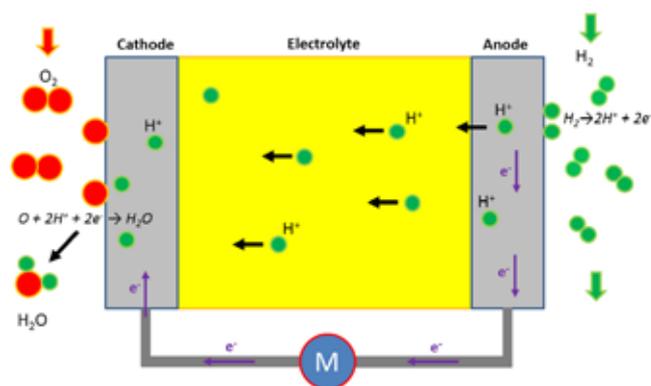
## 1 Contexte, enjeux, objectif

Le développement des énergies renouvelables, dans un contexte de recherche d'alternatives aux énergies carbonées, nécessite la mise au point de dispositifs de stockage de l'énergie efficaces. En effet, le caractère intermittent des énergies éoliennes, solaires, etc., impose de pouvoir stocker l'énergie non utilisée au moment de production, de façon à pouvoir la réutiliser ultérieurement. La filière hydrogène est une voie possible de stockage de l'énergie, dans laquelle l'énergie produite est emmagasinée dans la liaison chimique de la molécule  $H_2$ . L'hydrogène peut, par exemple, être produit par électrolyse de l'eau à haute température, et l'énergie ainsi stockée restituée sous forme d'électricité par le biais d'un dispositif électrochimique appelé pile à combustibles.

Une pile à combustibles transforme l'énergie d'une réaction de combustion (typiquement  $H_2 + \frac{1}{2}O_2 \rightarrow H_2O$ ) en électricité. Elle comporte deux électrodes et un électrolyte. L'électrolyte doit être un matériau conducteur ionique (ou protonique), mais pas conducteur électronique. Il existe de nombreux types de piles à combustibles, et la recherche actuelle vise à en améliorer les performances. Cela passe notamment par une recherche fondamentale pour comprendre les propriétés des matériaux mis en jeu. Dans ce travail, nous nous focalisons sur une famille particulière de composés, qui sont des oxydes de structure pérovskite, potentiels candidats pour servir de matériaux d'électrolytes dans les piles à combustibles de type PCFC (« Protonic Ceramic Fuel Cells »), dont le schéma de principe est montré sur la Figure 1. Dans ces dispositifs en fonctionnement, l'électrolyte est le siège d'un transport des protons. Notre travail donc vise à mieux comprendre les mécanismes physiques qui régissent les transferts des protons dans ces matériaux.

Nous avons choisi d'étudier le cérate de baryum ( $BaCeO_3$ ), matériau qui a déjà démontré ses capacités en termes

de bonne conduction protonique. Nous réalisons des simulations numériques de type *ab initio*, pour étudier le cas particulier du transfert d'un proton entre deux atomes d'oxygène du matériau. Même si les piles à combustibles de type PCFC fonctionnent à haute température, nous effectuons ces calculs à basse température (200 K), dans le but également de comprendre l'intrication entre les fluctuations quantiques affectant le mouvement du proton, et l'agitation thermique que subissent les atomes (plus lourds) du réseau. En effet, la technique que nous utilisons, basée sur les intégrales de chemin de Feynman (PIMD = « Path-Integral Molecular Dynamics »), permet de traiter à la fois les fluctuations thermiques et les fluctuations quantiques associées au mouvement des noyaux atomiques. ■



**Figure 1** : Schéma de principe simplifié d'une pile à combustibles de type PCFC. L'hydrogène  $H_2$  arrive à l'anode, se dissocie en  $2H^+$  et deux électrons. Les deux protons  $H^+$  franchissent l'électrolyte, par diffusion, et les deux électrons alimentent le circuit extérieur à la pile. À la cathode, les protons et électrons réagissent avec l'oxygène pour former une molécule d'eau.

## 2 Code de calcul et moyens

Le code de calcul utilisé dans le cadre de ce Grand Défi est le code ABINIT [1,2].

ABINIT est une suite logicielle open-source (licence GNU-GPL) --- d'environ 800 000 lignes de code Fortran --- développée dans le cadre d'un projet collaboratif international à grande visibilité, incluant des contributions de plusieurs laboratoires, essentiellement européens et nord-américains. ABINIT est utilisé intensivement, tant pour les travaux de recherche et de mise au point de méthodes que pour la production de données pour la Physique de

la Matière Condensée, à l'échelle microscopique. Il s'agit d'un code basé sur les « premiers principes » (modèles *ab initio*); les calculs sont effectués sans aucun ajustement sur des données expérimentales. Les systèmes, constitués de noyaux atomiques et d'électrons, sont étudiés à l'échelle microscopique. Le formalisme utilisé est celui de la Théorie de la Fonctionnelle de la Densité (DFT), aujourd'hui utilisé de manière routinière. Plusieurs types de calculs peuvent être effectués, notamment rechercher les propriétés de l'état fondamental de la matière, les états excités mais aussi les propriétés de

réponse à une sollicitation. Il est possible d'effectuer des calculs statiques mais aussi de la Dynamique Moléculaire ou de relaxer des structures atomiques.

Dans le cadre de l'étude présentée ici, nous avons utilisé la version 8.2 de ABINIT qui intègre plusieurs améliorations en vue de son utilisation sur les supercalculateurs à architecture *many-core*. Les nœuds *Broadwell* de *OCCIGEN* peuvent tirer profit d'une partie de ces nouveaux développements.

Le code utilise un parallélisme hybride *MPI-openMP* ; dans le cadre du *Grand Défi*, nous avons essentiellement utilisé le parallélisme distribué *MPI* car les simulations réalisées permettaient d'accéder à 4 niveaux de parallélisme relativement efficace sur quelques milliers de cœurs.

Le cœur d'un code de calcul dans le formalisme DFT est un algorithme itératif de diagonalisation. Il s'agit de déterminer les vecteurs propres d'un opérateur (dont nous n'avons pas une représentation matricielle) associées aux plus petites valeurs propres. Nous avons pu mettre en œuvre un algorithme récemment implanté --- algorithme de « filtrage de Chebyshev » ---, spécifiquement adapté aux architectures *petaflopiques* [3]. L'introduction de cet algorithme dans ABINIT a été récompensée par le second prix Bull-Fourier 2016, remis par ATOS et GENCI.

Notre étude nécessitait une grande quantité d'heures de calcul CPU --- plusieurs millions --- réparties en plusieurs simulations. Il s'agissait de réaliser le calcul de l'énergie totale de 10 systèmes indépendants. Pour chacun de ces systèmes, la simulation consistait en le calcul parallèle de plusieurs répliques légèrement différentes. Sur chacune de ces répliques, un calcul d'énergie totale dans le formalisme de la DFT, nécessitant un grand nombre d'itérations, a été effectué. Ce type de calcul se parallélise sur plusieurs centaines de

processus sans difficulté, notamment l'algorithme itératif de diagonalisation. Au total, il était possible d'effectuer des calculs parallèles sur plusieurs milliers de cœurs. A ceci, nous aurions pu ajouter un parallélisme multitâche de bas niveau, mais cela n'a pas été nécessaire. Pour consommer les heures de calcul de manière régulière et optimale, nous avons opté pour des jobs parallèles de 800 à 1000 processus.

L'installation de code ABINIT --- qui est disponible dans la plupart des centres de calcul --- s'est effectuée sans difficulté majeure.

Cependant, au début de la période d'essai suivant la mise en service du supercalculateur, nous avons constaté de fortes instabilités numériques, produisant des divergences dans des calculs simples de test. Après analyse du problème, nous avons diagnostiqué deux sources au problème : 1-une instabilité de la bibliothèque *MPI*, 2-des imprécisions numériques liées à la vectorisation, activée automatiquement. Le premier point a été rapidement réglé par des mises à jour successives de la bibliothèque ; quant à la vectorisation, nous avons dû compiler le code en désactivant (temporairement) certaines instructions.

Les problématiques liées à la vectorisation sur les processeurs modernes ne sont pas propres au supercalculateur *OCCIGEN*. Nous avons fait le même constat pour ABINIT dans d'autres centres de calcul. En effet, la vectorisation étant une opération de parallélisation non nécessairement déterministe, il peut se produire des cas de figure pour lesquels le code devient instable, notamment lorsque chaque *processus MPI* produit un résultat légèrement différent des autres. Ces petites différences ne sont pas supportées par l'algorithme itératif de diagonalisation. Dans le cadre d'une collaboration avec l'université de St-Quentin-en-Yvelines (UVSQ), nous avons entamé une étude de stabilité avec l'outil *verificarlo* [4], qui sera utilisé sur le calculateur *OCCIGEN*. ■

### 3 Simulations réalisées

Dans le cadre du projet présenté ici, nous souhaitons obtenir une estimation du coefficient de diffusion de l'hydrogène dans l'électrolyte de pile. Ce coefficient est directement dépendant de l'énergie nécessaire à un ion  $H^+$  (un proton) pour passer d'un site à l'autre dans le matériau. Plus la barrière d'énergie à franchir sera faible, plus le matériau sera conducteur de l'hydrogène.

Pour calculer la barrière d'énergie, il suffit théoriquement de déterminer le parcours emprunté par le proton et de calculer quelle est l'énergie du système (électrolyte et proton) pour différents points répartis le long de ce parcours. Il existe des méthodes standards pour ce type de calcul --- dites méthodes de recherche de chemin

d'énergie minimale --- mais elles ne permettent pas de prendre en compte tous les effets agissant dans le cas d'un atome d'hydrogène.

En effet, le noyau d'hydrogène est un élément si léger que son énergie n'est pas uniquement d'origine thermique (agitation liée à la température) ; elle est aussi d'origine quantique, et ce pour une part très importante. Pour déterminer qu'elle est l'amplitude de l'énergie du proton d'origine *quantique*, nous utilisons une méthode dite « dynamique moléculaire par intégrales de chemin ». Son principe repose sur une équivalence entre un *système quantique* et une chaîne de systèmes fictifs traités en mécanique classique. Cette chaîne de

systèmes *classique* doit être simulée dynamiquement, le long d'une trajectoire.

Il faut calculer une telle trajectoire pour chaque point le long du chemin emprunté par le proton dans l'électrolyte. La difficulté réside dans le fait que le proton, traité dynamiquement, va naturellement quitter le chemin que nous souhaitons lui faire emprunter. Pour l'y contraindre, nous avons implémenté dans le code ABINIT une technique spécifique de dynamique sous contrainte, appelée méthode « Blue moon ensemble ». Grâce à cette méthode, il nous est donc possible de réaliser, pour chacun des points le long du parcours du proton dans l'électrolyte, une trajectoire de dynamique sous contrainte pour une chaîne de systèmes fictifs basés sur le système initial.

La quantité de calculs à réaliser est donc importante, ce qui justifie que ce projet ait été sélectionné comme *Grand Défi*.

Les calculs à réaliser ont été dimensionnés de la manière suivante :

Le parcours du proton entre deux sites dans l'électrolyte de pile a été échantillonné sur 10 points.

Les chaînes de systèmes fictifs (utilisés pour simuler le système classique) étaient constituées de 16 répliques du système initial.

Les trajectoires de dynamique sous contrainte par la méthode « Blue moon ensemble » se sont déroulées pendant la totalité du temps imparti pour le projet ; nous avons effectué environ 1000 pas de dynamique.

La cellule de simulation était constituée de 20 atomes de baryum, cérium et oxygène, au milieu desquels le proton a été ajouté en position interstitielle.

Nous avons choisi de faire tourner 10 jobs simultanément, un par point échantillonnant le parcours. Chacun de ces jobs utilisait entre 800 et 1000 *processus* MPI, le parallélisme ayant été activé à tous les niveaux, notamment une distribution du calcul et des données sur tous les maillons de la chaîne de systèmes fictifs (un niveau de parallélisme très efficace). Dans ce régime, le code ABINIT bénéficie d'une *extensibilité* très bonne et d'une efficacité parallèle de l'ordre de 90%.

Les simulations de dynamique moléculaire sous contrainte se sont déroulées en continu par tranche de 24H, grâce à un système de reprise et de gestion d'historique de trajectoire, spécialement implémenté pour ce projet.

Au total, environ 2,5 millions d'heures de calcul ont été consommées à la fin de la période réservée au *Grand Défi*.

Les trajectoires de dynamique moléculaires obtenues, encore un peu courtes, sont tout fait exploitables pour une décrire qualitativement l'électrolyte étudié ici. ■

## 4 Observations et résultats scientifiques

Les calculs que nous avons effectués nous donnent une première image du chemin associé au transfert du proton d'un atome d'oxygène vers un autre. Bien que la statistique obtenue soit encore insuffisante pour tracer un chemin d'énergie libre le long de la coordonnée de réaction choisie pour étudier le transfert (cette coordonnée de réaction est la quantité  $O_1H-O_2H$  projetée le long de la direction du transfert), les calculs effectués permettent de comprendre comment ce transfert se déroule dans ses grandes lignes.

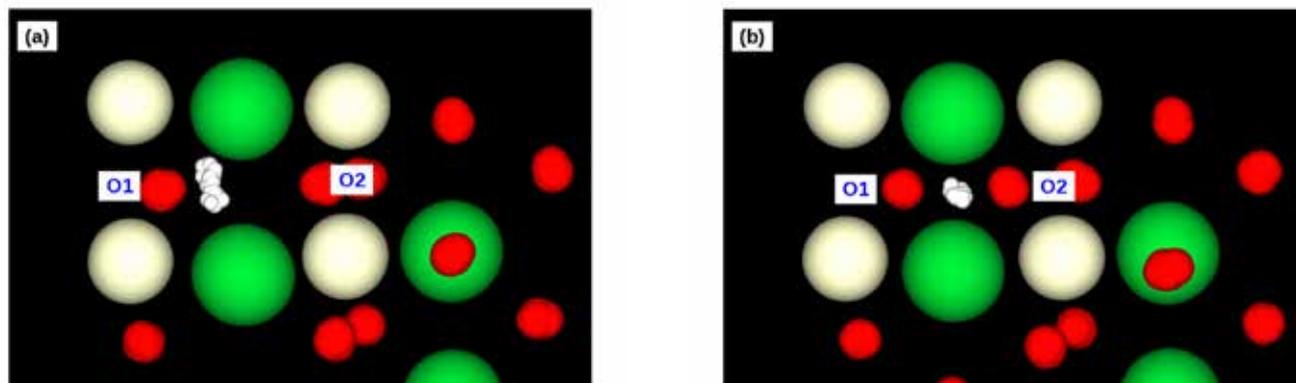
La Figure 2 montre deux images instantanées de la dynamique quantique. La première image correspond à une valeur de la coordonnée de réaction proche du point de départ (les « réactifs »). La seconde image correspond à une valeur de la coordonnée de réaction proche de « l'état de transition », c'est-à-dire au point où le proton effectue réellement son transfert d'un atome d'oxygène vers l'autre. Dans l'état initial, le proton fluctue autour de sa position stable. Ces fluctuations sont essentiellement d'origine quantique, car à basse température, le proton est complètement gelé dans son état fondamental vibrationnel (la température de

la simulation est ici de 200 K, fixée grâce à l'utilisation d'un « thermostat » de Langevin). On peut remarquer en particulier que la distance oxygène-oxygène est beaucoup plus faible dans l'état de transition que dans l'état de départ. Cela suggère que le transfert du proton se déroule dans des configurations particulières dans lesquelles cette distance est considérablement raccourcie. Ces configurations sont créées par le biais des fluctuations thermiques, car les atomes du réseau, plus lourds que le proton, ont un comportement beaucoup plus classique : ils subissent l'agitation thermique quand le proton, lui, ne la subit pas (on parle de « régime quantique thermiquement assisté »).

On peut voir également que l'extension du paquet d'images associées à l'atome d'hydrogène est relativement importante (typiquement de l'ordre de 0.15-0.20 Å à basse température), ce qui traduit le caractère « quantique » de son mouvement. Si cela est évident pour H, nous remarquons également que cette extension est plutôt grande pour les atomes d'oxygène, typiquement ici de l'ordre de 0.05 Å (en rouge sur la figure). En effet, bien que plus lourds que H, O possède une masse qui peut

occasionner des effets quantiques non négligeables jusqu'à des températures de l'ordre de la température ambiante dans les oxydes. Les autres atomes, plus

lourds (Ba, Ce), présentent des fluctuations quantiques beaucoup plus faibles, donnant lieu à une extension du paquet d'images relativement réduite. ■



**Figure 2 :** Images instantanées tirées de deux simulations de PIMD sous contrainte. (a) la coordonnée de réaction est fixée à une valeur proche de l'état initial (proton lié à l'atome d'oxygène O1). (b) la coordonnée de réaction est fixée à une valeur proche de l'état de transition. On voit que les deux atomes d'oxygène O1 et O2 se sont rapprochés. Les simulations emploient une cellule de 20 atomes correspondant à la maille primitive de l'oxyde BaCeO3 en structure orthorhombique. La température est  $T=200$  K.

## 5 Conclusion

La méthode de « Blue-Moon Ensemble » associée à la dynamique moléculaire par les intégrales de chemins (PIMD) a été implémentée dans le code ABINIT, et appliquée au cas du transfert d'un proton dans un oxyde à basse température. Cette méthode permet d'étudier de façon très fine des chemins de réaction ou de transition dans la matière, et de calculer les barrières d'énergie libre associées à ces processus élémentaires. Nous avons pu obtenir des informations sur la façon dont le transfert protonique se déroule à basse température dans ce

système. La méthode est désormais disponible dans le code ABINIT. Grâce à la mise à disposition du calculateur OCCIGEN dans le cadre d'un *Grand Défi*, le code a pu être testé sur une architecture à base de processeurs *Intel Broadwell*, en particulier le nouvel algorithme de diagonalisation à forte *extensibilité*. Même s'il subsiste des difficultés liées à la stabilité numérique, les performances obtenues sont satisfaisantes pour des calculs utilisant de l'ordre de 1000 cœurs CPU. ■

## 6 Références

- [1] [www.abinit.org](http://www.abinit.org).
- [2] Recent developments in the ABINIT software package, X. Gonze et coll. *Computer Physics Communications* 205, 106 (2016)
- [3] Parallel eigensolvers in plane-wave Density Functional Theory, A. Levitt, M. Torrent, M., *Computer Physics Communications* 187, 98 (2015)
- [4] Verificarlo: Checking Floating Point Accuracy through Monte Carlo Arithmetic, C. Denis, P. de Oliveira Castro, E. Petit, 2016 IEEE 23rd Symposium on Computer Arithmetic (ARITH), 55 (2016)

## LE STOCKAGE CELLULAIRE DES CORPS GRAS

### Une vue structurale de l'accumulation de lipides dans les cellules

**Auteur**

**Stefano VANNI**  
(Institut de Pharmacologie Moléculaire et Cellulaire)



#### Contributeurs

**Romain GAUTIER**

(Institut de Pharmacologie Moléculaire et Cellulaire)

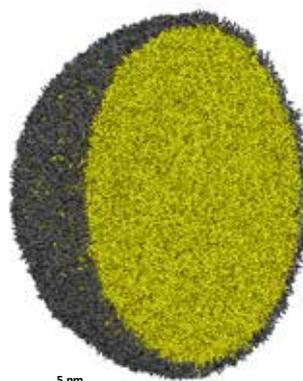
**Patrick FUCHS, Amelie BACLE**

(Institut Jacques Monod UMR 7592, CNRS Université Paris-Diderot, Sorbonne)

#### Abréviations

[GL = gouttelette lipidique, HA = hélice amphipathique, LN= lipides neutres, PL= phospholipides, TS= tension de surface, RE= réticulum endoplasmique, MD= dynamique moléculaire, GG= « gros-grain » ]

[interdigitation : ce terme anglais désigne le frottement des chaînes alcanes des phospholipides et des triglycérides entre le cœur et la membrane de la GL]



Graphical Abstract

## SOMMAIRE

- |   |  |
|---|--|
| <b>1</b> Contexte, enjeux, objectif ...32               | <b>5</b> Conclusion et remerciements .....36 |
| <b>2</b> Code de calcul et moyens ...33                 | <b>6</b> Bibliographie et références .....36 |
| <b>3</b> Simulations réalisées .....34                  |  |
| <b>4</b> Observations et résultats scientifiques.....35 |  |

## 1 Contexte, enjeux, objectif

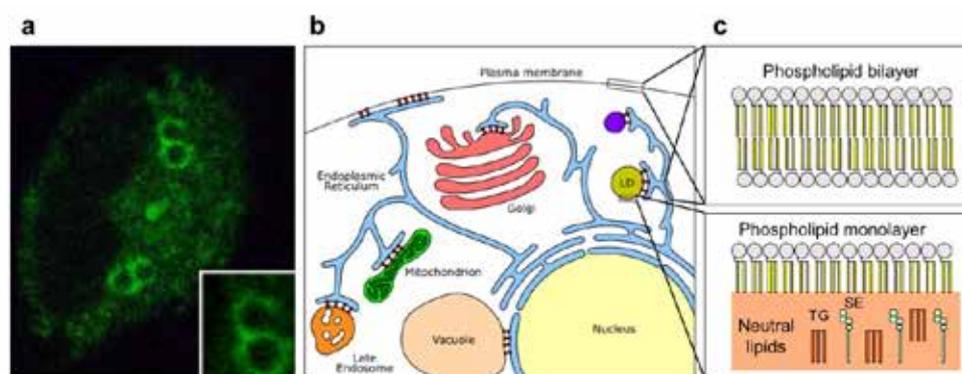
Les cellules eucaryotes sont caractérisées par une grande variété de membranes qui diffèrent par leur morphologie et leur composition<sup>1</sup>. Ces membranes entourent les différents organites cellulaires et sont de plus en plus considérées comme des plateformes d'activité protéique. Elles jouent ainsi un rôle essentiel dans la biochimie des interfaces, particulièrement dans les voies de signalisation ainsi que dans le trafic cellulaire.

De nombreuses techniques de biologie cellulaire et de biochimie ont récemment montré que les paramètres physico-chimiques des membranes régulaient les processus cellulaires à travers leurs interactions avec des protéines (incluant la courbure, l'élasticité, les interactions électrostatiques, la composition lipidique, la formation de domaines lipidiques, l'épaisseur hydrophobe ou la tension de surface)<sup>2</sup>. Ces méthodologies sont en général incapables d'atteindre une résolution à l'échelle du nanomètre, à cause de nombreuses limitations des méthodes biophysiques et biochimiques d'étude de systèmes mimant les membranes *in vitro*. Dans ce contexte, la modélisation moléculaire apparaît comme une méthode puissante et complémentaire permettant d'étudier le comportement des membranes et des systèmes interagissant avec celles-ci au niveau microscopique. Notamment, les simulations de dynamique moléculaire (MD) de membranes deviennent une méthode de choix pour caractériser ces phénomènes au niveau structural<sup>3</sup>.

La membrane des gouttelettes lipidiques (GLs) est particulièrement fascinante. Elle a été toutefois très peu caractérisée au niveau structural principalement à cause de limitations techniques au niveau expérimental et de modélisation. Originellement considérées comme inertes, les GLs ont vu un regain d'intérêt ces dernières années, principalement parce qu'elles sont impliquées dans les maladies métaboliques (incluant la lipodistrophie et l'obésité)<sup>4</sup>. L'obésité représente aujourd'hui

un problème de santé publique majeur, menant à des prédispositions pour de graves maladies telles que le syndrome métabolique, le diabète, les maladies cardiaques ou encore le cancer. Les GLs sont par ailleurs impliquées dans le cycle reproductif de pathogènes comme le virus de l'hépatite C (HCV), celui de la dengue ou de la chlamydia<sup>4</sup>. Ils ciblent pour cela les cellules métabolisant les corps gras ou les voies métaboliques des lipides pour se propager.

La fonction principale des GLs est de stocker l'énergie métabolique, la plupart du temps sous la forme de lipides neutres (LN) comme les triglycérides, afin que la cellule survive à de longue période de carence nutritive. Ainsi, elles participent également à la protection contre l'excès d'acides gras qui peut se révéler toxique. Bien qu'elles jouent un rôle particulièrement important dans les adipocytes (cellules spécialisées dans le stockage de corps gras), les GLs sont ubiquitaires, on les trouve chez les plantes, les levures et la plupart des cellules animales. Au point de vue biologie cellulaire, les GLs ont vu un intérêt croissant pour leur rôle déterminant dans le métabolisme des lipides<sup>5</sup>. Toutefois, bien que notre compréhension de la biologie des GLs augmente sans cesse, les mécanismes fins régissant ces processus restent peu compris, principalement à cause des difficultés d'étude de ces systèmes *in vitro* : contrairement à tous les autres organites qui sont entourés d'une membrane et dont la lumière est une phase aqueuse, les GLs sont uniques car leur intérieur est constitué d'une phase huileuse (Fig. 1). Leur surface est constituée d'une monocouche de phospholipides (PL) principalement zwitterioniques, comme les lipides phosphatidyl-choline (PC) ou phosphatidyl-ethanolamine (PE), entourant un cœur de LN (Fig. 1), principalement des triglycérides (TGs) et des esters de stérol (ESs), servant au stockage d'énergie et comme précurseurs de la synthèse des membranes lipidiques.



**Figure 1.** : Les Gouttelettes Lipidiques en un coup d'oeil. **a.** Image de microscopie de fluorescence de GLs, révélées avec des protéines se liant aux GLs. Image adaptée de 6. **b.** Représentation schématique des organites intracellulaires. **c.** Comparaison schématique entre une bicouche lipidique et une monocouche lipidique d'une GL.

Malgré leur structure unique, il a été montré que de nombreuses protéines se lient aux bicouches lipidiques pouvaient également se lier sur les GLs in vivo et in vitro<sup>7</sup>. D'un point de vue moléculaire, ces observations posent la question de la manière dont ces protéines sont capables de se lier à des membranes de structure si différente (une bicouche lipidique pour le réticulum endoplasmique (RE) ou le Golgi, vs, une monocouche entourant un cœur de lipides neutres pour les GLs).

Dans ce contexte, les simulations de MD représentent une approche alternative puissante pour étudier les propriétés d'interface des GLs. Toutefois, les méthodes classiques souffrent aussi de limitation, notamment pour reproduire certaines propriétés fondamentales des GLs. On ne compte plus les travaux de simulations de MD portant sur les bicouches lipidiques dans la littérature. En revanche, très peu d'études ont porté sur l'interface complexe huile/PL/eau. Un défi majeur dans l'investigation des GLs avec les simulations de MD provient de la complexité d'équilibrer le système non seulement dans le plan de la membrane, mais aussi de manière verticale le long de la normale au plan de la membrane. Ainsi, les simulations tout-atome sont limitées lorsqu'on étudie des systèmes type GL avec des interfaces huile/PL/eau et les simulations de MD de ces structures complexes doivent être accomplies en utilisant des simulations dites « gros-grain » (GG). Par ailleurs, il existe de nombreuses limitations des modèles GG courants, particulièrement sur les systèmes

étudiant les émulsions, justifiant ainsi le nombre limité d'études dans le domaine<sup>8</sup>. Pourtant ces approches sont nécessaires, notamment justifiées par des études récentes qui ont montré que la taille et la tension de surface avaient des conséquences sur la structure des émulsions « huile-dans-l'eau »<sup>7,9</sup>.

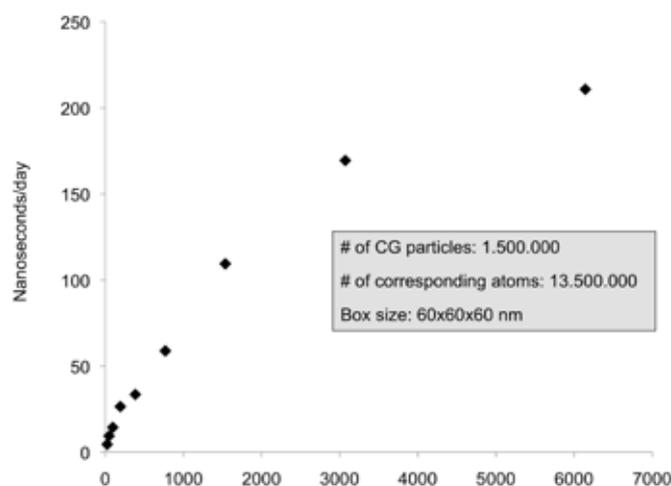
Afin de s'affranchir de ces limitations, nous avons récemment développé de nouveaux paramètres GG pour les triglycérides, basés sur le modèle GG nommé Shinoda-Devane-Klein<sup>8</sup>, afin d'étudier les GLs au niveau structural<sup>10</sup>. Ce modèle a été spécifiquement développé pour étudier les tensioactifs, et nos résultats préliminaires suggèrent que nos simulations reproduisent les propriétés des GLs avec une exactitude comparable aux simulations atomistiques<sup>10</sup>.

Le but de ce projet Grand-Défi a été de modéliser des GLs nanoscopiques de taille comparable à celles que l'on peut obtenir expérimentalement dans les études in vitro, et qui ont été montrées comme existant dans la cellule. En utilisant cette approche, nous avons étudié l'influence du type d'huile (triglycérides vs alcanes purs) et de la tension de surface sur les propriétés de surface des GL. Notre but était de comprendre comment ces propriétés pourraient réguler de manière sélective le ciblage de protéines et comme elles pourraient rendre la surface des GLs uniques par rapport aux bicouches « plus classiques ». ■

## 2 Code de calcul et moyens

Nous avons utilisé le logiciel LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) des Sandia National Laboratories<sup>11</sup>. Ce code est libre et utilisé par de nombreux laboratoires dans le monde. Sa meilleure implémentation est avec l'accélération MPI/OpenMPI ([http://lammps.sandia.gov/doc/accelerate\\_omp.html](http://lammps.sandia.gov/doc/accelerate_omp.html)) ainsi que la bibliothèque FFTW, même si d'autres accélérations sont possibles. Un patch spécifique pour faire des simulations GG avec le modèle Shinoda-Devane-Klein a déjà été développé (proposant diverses manières de gérer les paires et les angles spécifiques à ce modèle). Pour le projet, nous avons réalisé des benchmarks pour LAMMPS et les simulations GG sur Occigen (Fig. 2). Ce benchmark a été réalisé avec l'accélération MPI, mais sans l'accélération OMP, celle-ci se révélant très complexe à installer sur Occigen.

Pour faire tourner nos simulations, nous avons obtenu un total de 2.600.000 h CPU. Nous les avons utilisées pour lancer des simulations individuelles avec 4480 cœurs pour chaque. Avec le nouveau système sur Occigen, nous avons atteint jusqu'à 100 ns/jour sur de gros systèmes de plus de 30 millions d'atomes. ■



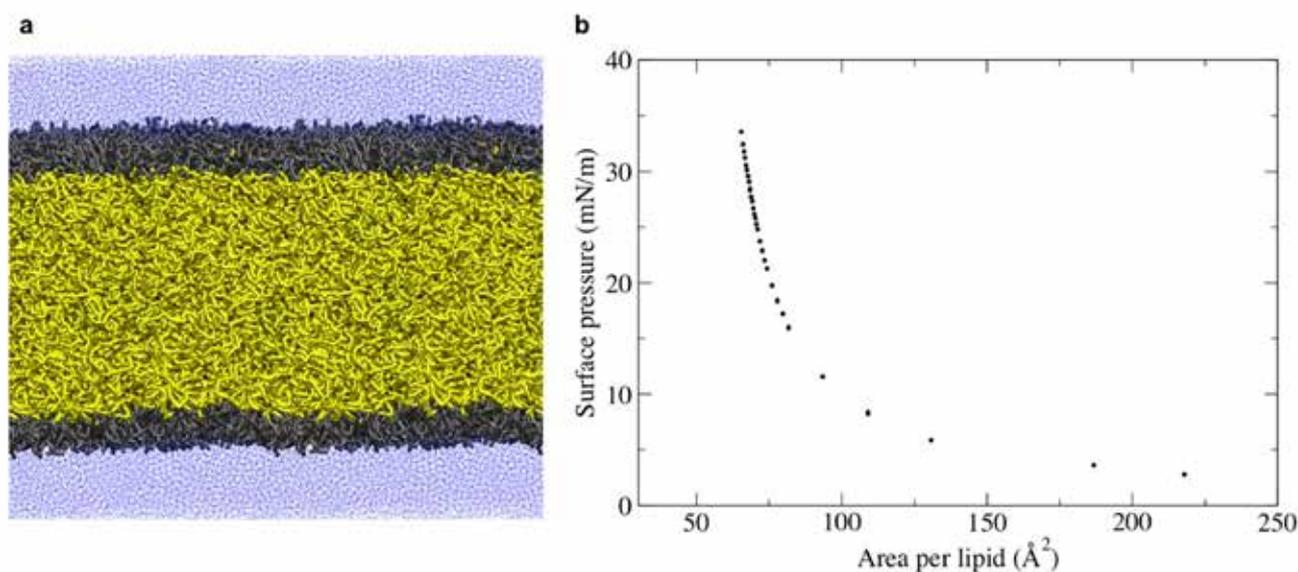
**Figure 2.** : Benchmark LAMMPS sur Occigen pour un système GG d'une « petite » nanogouttelette ( $R=30$  nm). Le scaling est spectaculaire jusqu'à 2000 cœurs. Celui-ci ainsi que la performance seront sans doute améliorés avec l'accélération OMP.

### 3 Simulations réalisées

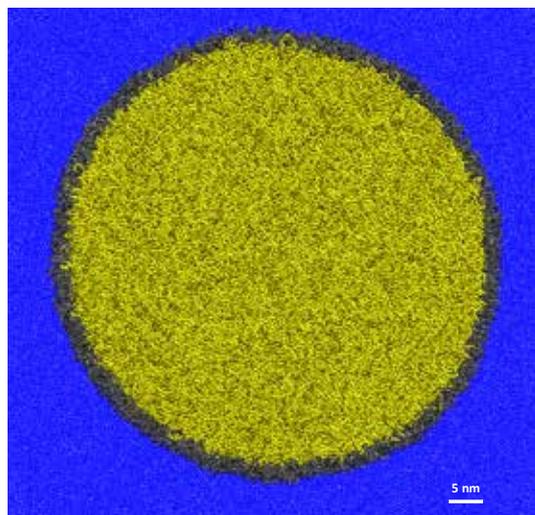
Grâce aux ressources généreuses allouées à notre projet, nous avons pu étudier en détail de nombreux paramètres des GLs qui ont été montrés comme extrêmement importants en biologie cellulaire. Premièrement, nous avons étudié la structure des GLs à différentes tensions de surface, et ce sur toute la gamme de 0 à 32 mN/m (la valeur d'une émulsion pure triglycérides / eau). Pour cela, nous avons utilisé un système plat plus petit nous permettant ainsi de tester de nombreuses simulations indépendantes et de calculer un isotherme de compression de GLs modèles (Fig. 3).

Ensuite nous avons préparé des nanogouttelettes individuelles avec des rayons de 15 à 30 nm. Ces systèmes sont par les plus grandes simulations GG tout en gardant

une résolution « chimique », avec au total un nombre de particules GG de 3,5 millions, correspondant à environ 35 millions d'atomes (Fig. 4). En plus de la taille des gouttelettes, nous avons également varié la tension de surface ainsi que la composition lipidique avec une attention particulière sur les différences entre les lipides neutres physiologiques (triglycérides), et les chaînes alcanes synthétiques comme le dodécane ou l'hexadécane. Ces simulations montrent qu'il existe de réelles différences dans le comportement moléculaire de ces différentes huiles. Toutefois, nos résultats sont pour le moment préliminaires, ainsi nous ne les discuterons pas en détail dans le prochain paragraphe, mais nous nous focaliserons plutôt sur les implications biologiques de nos simulations. ■



**Figure 3.** : Modèle plat de GL. a. Cliché d'un modèle plat de GL dans les simulations de MD GG. Jaune : triglycérides, gris : phospholipides, bleu : molécules d'eau. b. Isotherme de compression correspondant au système plat de GL. Image adaptée de 10.



**Figure 4.** : Modèle de nanogouttelette lipidique. Cliché d'une section de GL modèle de rayon 25-nm simulée dans cette étude. La nanogouttelette entière est sphérique et entièrement entourée de phospholipides. Jaune: triglycérides, gris: phospholipides, bleu: molécules d'eau.

## 4 Observations et résultats scientifiques

En développant de nouvelles approches pour l'étude des interactions triglycérides / eau et en réalisant des simulations à très large échelle au CINES, nous avons pu déduire la relation structure-fonction possible entre la structure unique des GLs et de nombreux phénomènes biologiques cruciaux ayant lieu à leur surface, comme par exemple le ciblage de protéines périphériques, la dégradation des triglycérides et la biogénèse des GLs.

Notre travail a été initialement motivé par l'observation déroutante que de nombreuses protéines périphériques qui sont traditionnellement adressées au RE et/ou à l'appareil de Golgi peuvent se lier aussi aux GLs. Toutes ces protéines possèdent une hélice amphipathique (HA). Nous avons récemment proposé un modèle suggérant que les défauts de packing mesurés en simulation de MD peuvent être utilisés pour prédire la liaison de protéines périphériques contenant des HAs, et ce in vitro et in vivo<sup>12,13</sup>. En simulant des modèles de GL, nous avons pu mesurer les défauts de packing des lipides à l'interface des GLs. Nous avons trouvé que la plupart des propriétés moléculaires de la surface de GLs modèles, dont les défauts de packing, ressemblent à celles des membranes modèles de RE et de Golgi. Cela nous a fourni une explication structurale sur la capacité de certaines protéines contenant des AHs à se lier sur des GLs ainsi que d'autres organites semblables.

Nous avons pu aussi quantifier l'importance de la tension de surface (TS) sur la liaison de protéines sur les GLs. Bien que celle-ci soit généralement très basse sur les bicouches lipidiques (valeurs typiques sous les 1 mN/m), cela n'est pas forcément le cas pour les émulsions huileuses. Comme attendu, les défauts de packing augmentent lorsque la TS augmente. Toutefois, cette augmentation est douce pour des TS basses (sous les 10 mN/m), puis augmente significativement lorsque la TS est forte (plus de 10 mN/m). Ce comportement suggère un scénario nouveau et intrigant : lorsque la TS est basse, seules les protéines qui possèdent une hélice bien adaptée aux GLs et à la liaison de membrane peuvent cibler de manière sélective les GLs; lorsque la TS est forte, de nombreuses protéines, dont celles qui ne doivent pas se lier aux GLs dans des conditions physiologiques normales, pourraient être mobilisées sur les GLs. Ce dernier scénario survient sûrement dans des conditions non physiologiques comme lorsqu'on observe la liaison de protéine aux GLs en microscopie de fluorescence (quand on les charge en oléate pour induire la formation de GLs).

Une conséquence directe des observations précédentes est que nous avons pu prédire que la TS des GLs intracellulaires ne devait pas dépasser ~ 10 mN/m de manière non régulée, sinon les conséquences sur le trafic intracellulaire pourraient être potentiellement

dévastatrices. De telles hautes TSs pourraient mener à des protéines périphériques contenant des HAs piégées sur les GLs alors qu'elles étaient destinées à d'autres organites. Cela pourrait mener à un trafic aberrant et à des fusions de GL spontanées de GL. En collaboration avec le laboratoire de R. Thiam (École Normale Supérieure, Paris), nous avons récemment montré que la TS de GLs purifiées était dans la gamme 2-4 mN/m, confirmant ainsi notre prédiction basée uniquement sur la modélisation moléculaire<sup>9</sup>.

De plus, nous avons également observé que l'« interdigitation » entre les triglycérides et les phospholipides était constante pour des TS < 10 mN/m, alors qu'elle augmentait de manière spectaculaire pour des TS > 10 mN/m. Ce résultat surprenant pourrait avoir des implications importantes sur le métabolisme des triglycérides. Les enzymes du métabolisme des lipides, et spécialement celles qui dégradent les triglycérides (comme par exemple l'adipose triglyceride lipase, ATGL), sont en général situées à la surface des GLs. Même si leur mécanisme d'action n'est pas connu, l'augmentation de substrat accessible, comme c'est le cas lorsque l'« interdigitation » augmente, pourrait augmenter leur activité et mener à une dégradation massive des triglycérides. Ainsi, nos simulations ne suggèrent qu'une augmentation marginale de l'activité d'ATGL pour des TS < 10 mN/m, et une augmentation importante de celle-ci pour des valeurs > 10 mN/m. La dégradation de triglycérides pourrait non seulement baisser le nombre de molécules d'huile dans le cœur, mais aussi augmenter le nombre de molécules tensioactives (la dégradation d'une molécule de triglycéride mène à la formation d'un acide gras et d'un diacylglycérol). Etant donné que ces molécules réduisent la tension de surface, nos observations suggèrent la présence possible d'un mécanisme de rétrocontrôle : des TS d'environ 10 mN/m augmenteraient de manière importante l'activité des lipases, ou inversement une augmentation de TS au-dessus des 10 mN/m serait fortement empêchée en présence de lipases. ■

## 5 Conclusion et remerciements

En résumé, bien que nos simulations ne représentent qu'une première étape vers la compréhension structurale de la biologie des GLs, il apparaît clairement que ces techniques computationnelles peuvent apporter des détails décisifs dans la compréhension des mécanismes complexes médiés par la membrane. Grâce à notre travail, ceci inclut maintenant les interfaces non conventionnelles, mais fascinantes, comme celles impliquant des GLs. Pour aller plus loin, plusieurs choses devront être améliorées : (i) de meilleurs champs de forces, (ii) des développements plus poussés de logiciels scientifiques et (iii) un investissement continu dans les ressources de calcul.

Comme les GLs jouent un rôle majeur dans les maladies métaboliques, nous sommes confiants qu'une compréhension plus poussée de leurs propriétés moléculaires aidera à concevoir des stratégies thérapeutiques contre ces maladies très répandues. Par exemple, il a été montré que les enzymes responsables de la synthèse de triglycéride dans les GLs étaient directement impliquées dans les maladies faisant intervenir les GLs, soit par un

manque (par exemple, la lipodystrophie), soit par un excès (obésité) de GLs. Ainsi les GLs ont été proposées comme des cibles possibles dans le développement de médicament<sup>4</sup>. Par conséquent, une compréhension structurale des propriétés de surface des membranes de GLs devraient aider à comprendre le rôle majeur des protéines interagissant avec les GLs. Cela constituera sans nul doute un premier pas vers le développement de médicaments qui interfèrent spécifiquement avec ces mécanismes.

Au final, nous profitons de cette occasion pour remercier pour nous avoir offert le support et les meilleures conditions de travail pour bien mener ce projet: l'Institut de Pharmacologie Moléculaire et Cellulaire (IPMC) du CNRS, son directeur Pascal Barbry et Bruno Antonny de l'équipe Dynamique des membranes et manteaux protéiques; l'Institut Jacques Monod (IJM), et notamment Alenka Copic et Catherine Jackson, de l'équipe Dynamique des membranes et trafic intracellulaire, pour des discussions stimulantes. ■

## 6 Bibliographie et références

1. Meer, G. Van, Voelker, D. R. & Feigenson, G. W. Membrane lipids : where they are and how they behave. *Nat. Cell Biol.* 9, 112–124 (2008).
2. Bigay, J. & Antonny, B. Curvature, Lipid Packing, and Electrostatics of Membrane Organelles : Defining Cellular Territories in Determining Specificity. *Dev. Cell* 23, 886–895 (2012).
3. Marrink, S. J. & Tieleman, D. P. Perspective on the Martini model. *Chem. Soc. Rev.* 42, 6801–6822 (2013).
4. Walther, T. C. & Farese Jr, R. V. Lipid Droplets and Cellular Lipid Metabolism. 81, 687–714 (2012).
5. Penno, A., Hackenbroich, G. & Thiele, C. Phospholipids and lipid droplets ☆. *BBA - Mol. Cell Biol. Lipids* 1831, 589–594 (2013).
6. Mishra, S. et al. Mature lipid droplets are accessible to ER luminal proteins. *J. Cell Sci.* 129, 3803–3815 (2016).
7. Thiam, A. R. et al. COPI buds 60-nm lipid droplets from reconstituted water – phospholipid – triacylglyceride interfaces , suggesting a tension clamp function. 110, 1–6 (2013).
8. Shinoda, W., DeVane, R. & Klein, M. L. Computer simulation studies of self-assembling macromolecules. *Current Opinion in Structural Biology* 22, 175–186 (2012).
9. Kalthoum, B. M. et al. ER Membrane Phospholipids and Surface Tension Control Cellular Lipid Droplet Formation. 41, 591–604 (2017).
10. Bacle, A., Gautier, R., Jackson, C. L., Fuchs, P. F. J. & Vanni, S. Interdigitation between Triglycerides and Lipids Modulates Surface Properties of Lipid Droplets. 1417–1430 (2017). doi:10.1016/j.bpj.2017.02.032
11. Plimpton, S. Fast Parallel Algorithms for Short – Range Molecular Dynamics. *J. Comput. Phys.* 117, 1–42 (1995).
12. Vanni, S., Hirose, H., Barelli, H., Antonny, B. & Gautier, R. A sub-nanometre view of how membrane curvature and composition modulate lipid packing and protein recruitment. 5, 4916 (2014).
13. Magdeleine, M., Gautier, R., Gounon, P., Vanni, S. & Antonny, B. A filter at the entrance of the Golgi that selects vesicles according to size and bulk lipid composition. 5, e16988 (2016).

## INFLUENCE DES INSTABILITÉS SOUS-MÉSOÉCHELLES DANS LE GOLFE DE GASCOGNE

Océanographie Physique Côtière - CT1: Environnement

Auteur

Sébastien THEETTEN  
(IFREMER, centre de Brest)



### Contributeurs

Guillaume CHARRIA, Frédéric VANDERMEIRSCH (IFREMER, centre de Brest)

Nicole AUDRIFFREN (CINES)

## SOMMAIRE

<b>1</b>	<b>Contexte scientifique</b> .....	<b>38</b>			
	Une région océanique côtière ..	38			
	Des processus océaniques à fines échelles : un défi scientifique .....	38			
	Un modèle numérique réaliste et performant.....	38			
	Un supercalculateur de nouvelle génération : vers un défi technologique !....	39			
<b>2</b>	<b>Une transition pour la modélisation océanique côtière dans le golfe de Gascogne : de MARS3D à CROCO</b> .....	<b>39</b>			
	Une nouvelle configuration pour un golfe de Gascogne réaliste .....	39			
				Un ensemble de simulations produites .....	39
				Nouveau code et nouveau calculateur: quid des performances ? .....	40
<b>3</b>	<b>Premiers résultats : vers un océan instable</b> .....	<b>41</b>			
<b>4</b>	<b>Conclusions</b> .....	<b>43</b>			
<b>5</b>	<b>Références</b> .....	<b>44</b>			

## 1 Contexte scientifique

Le défi réalisé peut se résumer par l'équation suivante:

Une région océanique côtière + des interrogations sur les processus océaniques à fines échelles + un modèle numérique réaliste et performant + un supercalculateur

de nouvelle génération = un défi scientifique et technologique !

Afin de résoudre cette équation, il convient d'explicitier chacun de ces termes:

### Une région océanique côtière

Ce défi a été conçu sur la région du golfe de Gascogne. Cette région, située le long des côtes française est un concentré de processus hydrodynamiques complexes de part sa morphologie (un plateau continental peu profond qui s'élargit en allant vers le Nord du golfe), ses rivières (trois rivières majeures - Loire, Gironde et

Adour qui apportent une quantité importante d'eau douce dans l'océan), ses marées (région où les courants de marée sont intenses dans la partie Nord du golfe), et le tout soumis à des forçages atmosphériques et une circulation grande échelle qui vont redistribuer l'ensemble des masses d'eau.

### Des processus océaniques à fines échelles: un défi scientifique

Dans ce système dynamique complexe, soumis aux changements globaux, la circulation se développe à différentes échelles spatiales et temporelles. Dans ce large spectre d'échelles, le rôle des fines échelles (~1-10 km - aussi appelées sous-mésoéchelles), incluant, par exemple, les tourbillons, les fronts et les instabilités associées, sur la dynamique générale de la zone étudiée aux échelles pluriannuelles reste peu connu. En milieu côtier, tel que le golfe de Gascogne, l'hétérogénéité du milieu (e.g. petits fonds / grands fonds, eaux salées / eaux dessalées à proximité des panaches de rivières) complexifie cette exploration, mais constitue un exemple type d'interactions en zone côtière.

Aujourd'hui, les observations avec une couverture synoptique de ces fines échelles restent rares et constituent un des enjeux de la future mission d'altimétrie satellite SWOT (Surface Water and Ocean Topography). Par

ailleurs, la modélisation numérique océanique côtière à la résolution nécessaire pour reproduire complètement les processus à fines échelles de manière réaliste reste peu accessible. De plus, les hypothèses classiquement utilisées dans ces modèles (approximation hydrostatique) arrivent en limite de validité lorsque les résolutions spatiales descendent largement en dessous du kilomètre.

L'enjeu scientifique de ce défi réside donc en un saut supplémentaire vers la reproduction et la compréhension des processus à fines échelles en atteignant des résolutions (400 m de résolution spatiale - comparé à 1 km de résolution dans les études précédentes) qui devraient permettre de résoudre une grande partie du spectre des processus à fines échelles. Cette nouvelle étape représentera un pas significatif dans la compréhension des processus à fines échelles et de leur influence sur la dynamique à grande échelle de l'océan côtier.

### Un modèle numérique réaliste et performant

La modélisation de la dynamique océanique réaliste en milieu côtier nécessite des outils capables de résoudre les équations primitives (version simplifiée et adaptée à l'océan ou l'atmosphère des équations de la mécanique des fluides de Navier-Stokes). Ces outils existent et sont largement utilisés pour la compréhension de l'environnement côtier. Par exemple, la configuration du code MARS 3D utilisée lors du précédent défi sur Occigen (Theetten et al., 2017) a permis d'étudier la variabilité interannuelle de la dynamique à fines échelles dans la région profonde du golfe de Gascogne (Charria et al., 2017).

Cependant, les outils existants peuvent présenter des limites en terme de performances et/ou de validité des hypothèses. Afin d'accéder aux limites de ce type de modélisation, le défi réalisé s'inscrit dans une dynamique nationale visant à développer une future plateforme de

modélisation communautaire inter-organismes (CNRS, INRIA, IFREMER, IRD, SHOM, Université Paul Sabatier, Université de Bretagne Occidentale), appelée CROCO (Coastal and Regional Ocean COmmunity model - <http://www.croco-ocean.org/>) en cours de structuration.

Pour aller explorer les fines échelles, l'étude a donc nécessité d'aller vers cette plateforme de modélisation CROCO et de l'adapter à notre cas d'étude réaliste le golfe de Gascogne. Ce développement de configuration a constitué la première composante du défi.

## Un supercalculateur de nouvelle génération: vers un défi technologique !

Enfin, la dernière composante de l'équation de ce défi réside dans l'utilisation de ce code CROCO à partir du calculateur Occigen dans sa version étendue. Ce portage

avec des performances acceptables pour envisager d'utiliser la grande capacité de calcul d'Occigen a constitué la partie technologique de ce défi. ■

## 2 Une transition pour la modélisation océanique côtière dans le golfe de Gascogne : de MARS3D à CROCO

### Une nouvelle configuration pour un golfe de Gascogne réaliste

Sur la base de ce nouveau code CROCO (<http://www.croco-ocean.org/>), deux configurations ont été développées: BOB400 (correspondant à une maille de 400m) et BOB2400 (correspondant à une maille de 2400m). Le code CROCO est basé sur le modèle ROMS-AGRIF (Debreu et al., 2012; Penven et al., 2006) largement utilisé pour différents types d'applications dans le domaine côtier (de la dynamique à l'échelle de la plage jusqu'à une partie de bassins océaniques). Ce code, écrit en Fortran présente des coordonnées verticales de type sigma généralisées. Le schéma de couche limite KPP (K-profile parameterization - Large et al., 1994) permet de paramétrer les processus de mélange vertical aux échelles inférieures à la grille de calcul. Le code utilise les techniques de parallélisation MPI pour l'optimisation

des calculs et peut ainsi être déployé sur différentes plateformes de calcul.

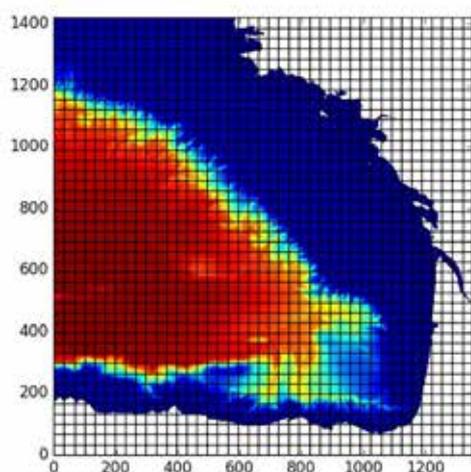
Le domaine des configurations BOB s'étend de 43.2°N à 48.5°N et de 8°W à 0.6°W pour des profondeurs pouvant atteindre 4849 m. La configuration présente deux frontières ouvertes à l'Ouest et au Nord au niveau desquelles sont utilisés des situations climatologiques. Les champs simulés à l'aide du modèle MARS3D lors du précédent défi (configuration BACH1000 à 1km de résolution - Charria et al., 2017) ont été aussi utilisés pour d'autres simulations réalisées lors de ce défi. Ces mêmes produits sont utilisés en condition initiale des simulations avec les configurations BOB.

### Un ensemble de simulations produites

La parallélisation du modèle est faite par une méthode de décomposition de domaine classique (i.e. le découpage est géographique). La grille de calcul est découpée en sous-maillages bidimensionnels sur l'horizontale (il n'y a pas de parallélisation sur la verticale). Les calculs entre les différents sous-domaines sont réalisés par l'échange

de message avec la bibliothèque MPI. La Figure II.1 représente le découpage de 1684 sous-domaines retenu pour la réalisation du Grand Défi sur Occigen.

Différentes expériences ont été réalisées dans le cadre de cette étude (e.g. différents forçages atmosphériques réalistes ou climatologiques, avec/sans apport des rivières). Pour donner un aperçu des progrès réalisés dans le cadre de ce défi, deux configurations sont présentées. Les deux configurations présentent 40 niveaux verticaux. La première simulation, nommée BOB2400, présente une résolution horizontale de 2400m soit  $226 \times 230 \times 40 = 2\,079\,200$  points de grille. La seconde simulation, nommée BOB400, présente une résolution horizontale de 400m soit  $1340 \times 1360 \times 40 = 72\,896\,000$  points de grille. Les deux simulations présentées utilisent les forçages atmosphériques climatologiques (d'autres simulations non présentées ici sont basées sur des forçages réalistes). La dynamique de marée n'a pas été implémentée dans ces simulations. Ces deux simulations couvrent une période commune de 16 mois (la simulation à basse résolution BOB2400 a simulé 10 années).



**Figure II.1:** Découpage du domaine de calcul de la configuration BOB400 (400m). Le nombre total de sous-domaines est de 1680 (40x42). La bathymétrie est représentée par le gradient de couleurs.

## Nouveau code et nouveau calculateur: quid des performances ?

La première étape d'analyse du défi a porté sur l'analyse des performances du code CROCO sur Occigen. Pour évaluer ces performances, plusieurs découpages MPI ont été testés sur la configuration BOB400 pour une simulation de 4320 itérations temporelles sans écriture de résultat afin d'évaluer uniquement les performances de calcul.

La figure II.2 montre ainsi le résultat des analyses pour des simulations allant jusqu'à des découpages impliquant plus de 8000 cœurs (expérimentations uniquement rendues possibles par la participation à ce défi). Le code CROCO présente des performances excellentes (avec une accélération dans certains cas supérieure aux attentes théoriques) pour des découpages inférieurs à 3000 cœurs. Ces accélérations simulées restent exceptionnelles pour ce type de modèles numériques utilisés en océanographie côtière. Au delà de 3000 cœurs, les

performances déclinent avec une accélération réduite en comparaison de l'accélération théorique.

Dans la gamme des expériences où le taux d'accélération est efficace, différents découpages géographiques des domaines ont été testés. Ainsi, par exemple, pour 2520 cœurs, une liste de découpage a été définie (24x105, 35x72, 40x63, 42x60, 45x56, 56x45, 72x35) et montre une variation du temps de simulation entre 290s et 325s avec un temps minimum pour un découpage avec un nombre de domaine équivalent dans les deux directions (Figure II.2).

L'exploration de ces différents découpages permet de tirer deux conclusions principales:

- il n'apparaît pas de règle claire pouvant permettre d'optimiser le découpage selon le nombre de cœurs

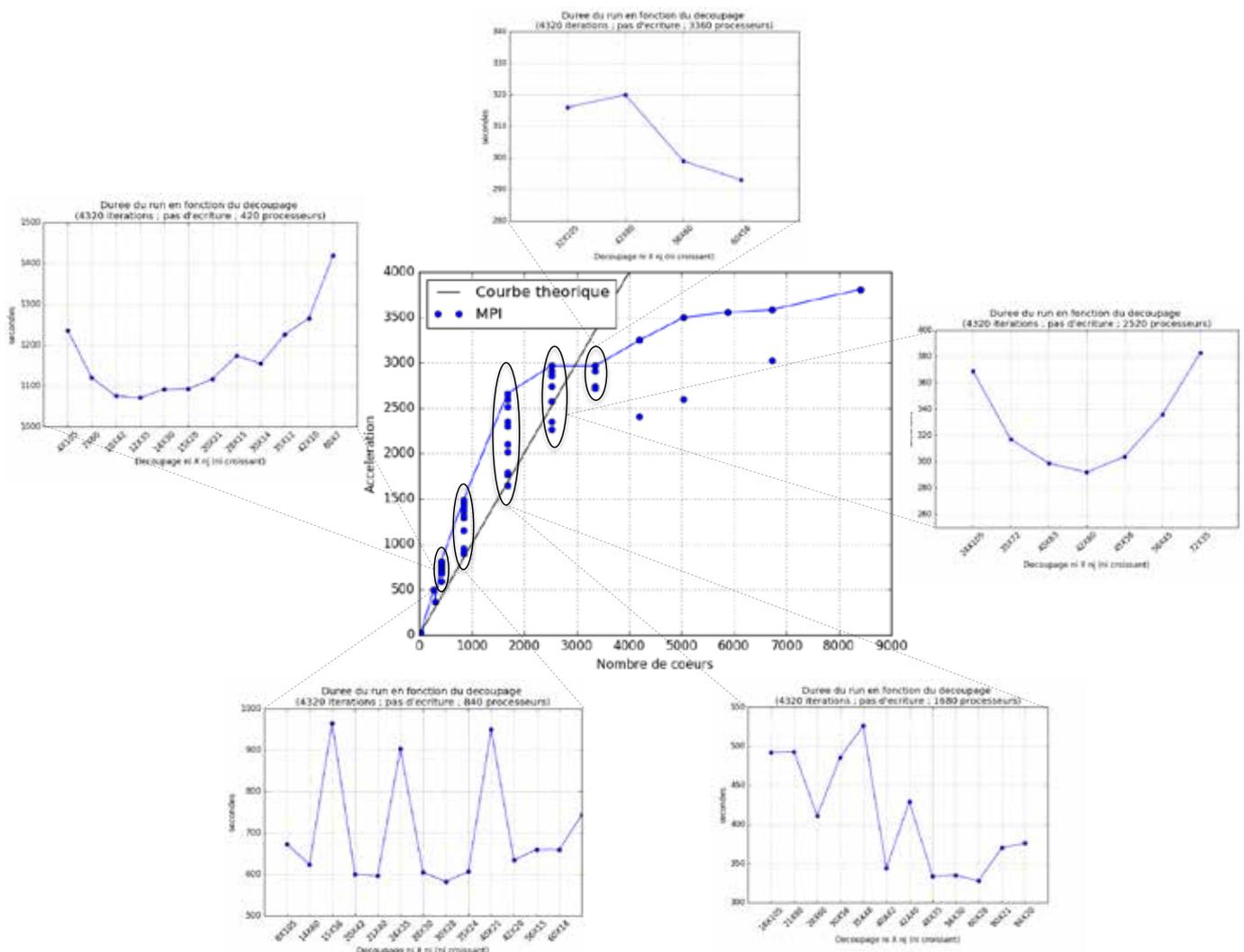


Figure II.2: Performances du modèle CROCO selon le nombre de cœurs utilisés et différents découpages en sous-domaines.

utilisés dans ce cadre d'expérience. Notons que ces expériences incluent aussi une mise en production de la nouvelle machine Occigen et donc une phase de réglage du système.

- les performances présentent une grande variabilité de temps d'exécution selon les différents découpages. Ces fluctuations n'ont pas pu être complètement expliquées

(e.g. performance des processeurs, optimisation des échanges MPI, autres réglages d'Occigen) et pourront faire l'objet d'explorations futures.

Sur la base de cette analyse et pour la production de simulations devant être analysées, une configuration à 1680 sous-domaines (40 x 42) a été retenue (Figure II.1). ■

## 3 Premiers résultats : vers un océan instable

La modélisation numérique de l'océan côtier reste un défi car les échelles de temps des processus physiques peuvent être dominées par des échelles de temps courtes. Ces situations transitoires ne permettent pas le développement de structures grandes échelles stables et cohérentes dans le temps. Nous sommes donc face à une dynamique très transitoire dans laquelle les structures cohérentes se dissipent rapidement, entre autre par frottement sur le fond au-dessus du plateau continental.

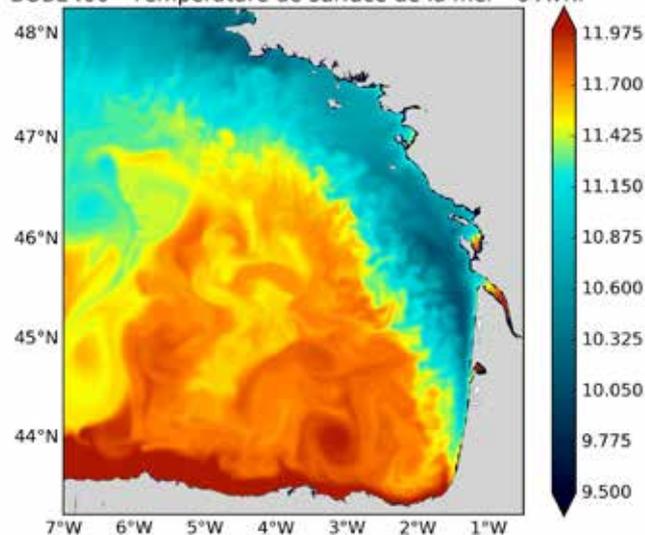
Dans ce contexte, les simulations BOB2400 et BOB400 réalisées montrent explicitement le développement de ces fines échelles lorsque les résolutions spatiales tendent vers des échelles beaucoup plus petites. Sur la figure III.1, les températures de surface de la mer simulées aux deux résolutions spatiales présentent une distribution générale similaire avec des températures plus basses dans la partie Nord du domaine et sur le plateau continental le long des côtes françaises. Cependant, il apparaît clairement que les échelles spatiales résolues sont plus fines dans BOB400. Ainsi, de fines structures apparaissent sur le plateau continental ainsi que dans la partie profonde avec de nombreux fronts et filaments

qui n'apparaissent pas dans la simulation à plus basse résolution.

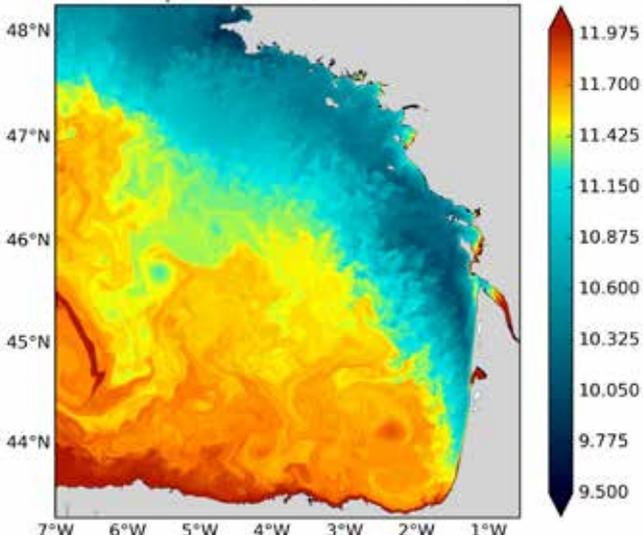
Ce contraste entre les deux simulations est d'autant plus marqué lorsque sont analysés les courants de surface. Sur la Figure III.2, il est intéressant de noter que la résolution de structures à plus fines échelles est associée à des courants de surfaces plus intenses. En effet, les fines échelles sont associées à des structures en rotation (e.g. tourbillons) qui dans le cas de la simulation à 400m de résolution sont plus nombreuses. Cette activité se traduit donc par un plus grand nombre de régions présentant des courants sensiblement plus intenses.

La figure III.3 montre un autre indicateur de l'intensification de l'activité à sous-mésoséchelle (i.e. fines échelles) lorsque la résolution spatiale est augmentée. En particulier, sur le plateau du golfe de Gascogne, la composante verticale de la vorticité relative montre une situation beaucoup plus turbulente à 400m de résolution à l'exception d'une frange proche des côtes françaises sous l'influence des panaches des rivières qui présente une activité réduite. L'activité frontale dans cette région et les instabilités

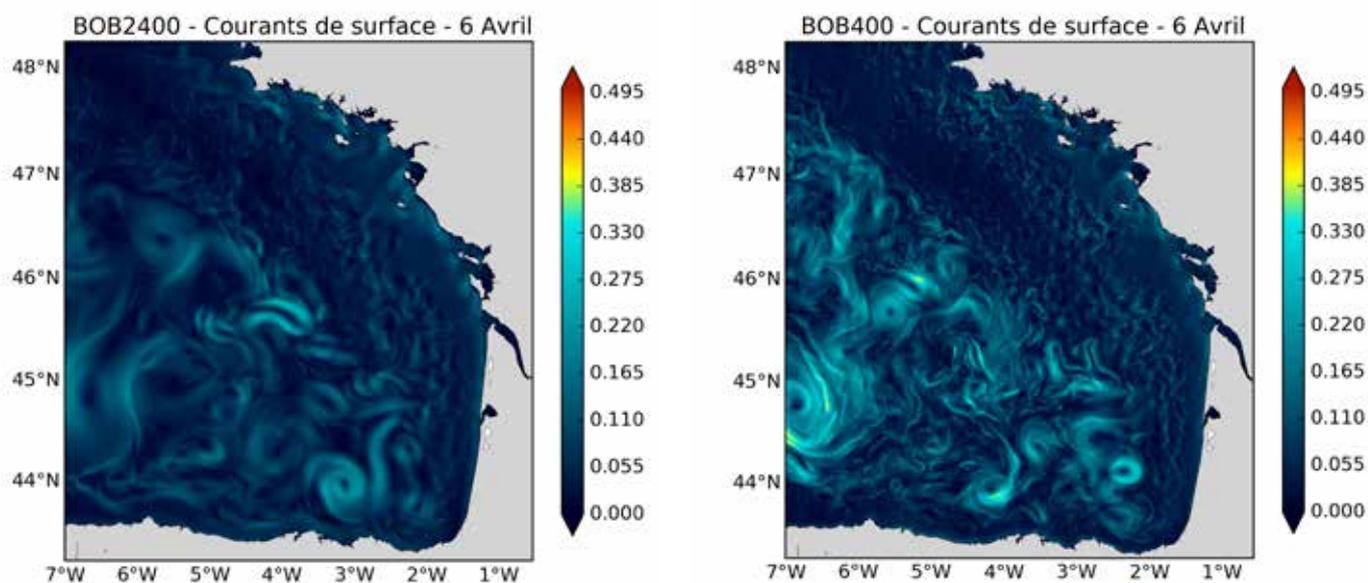
BOB2400 - Température de surface de la mer - 6 Avril



BOB400 - Température de surface de la mer - 6 Avril



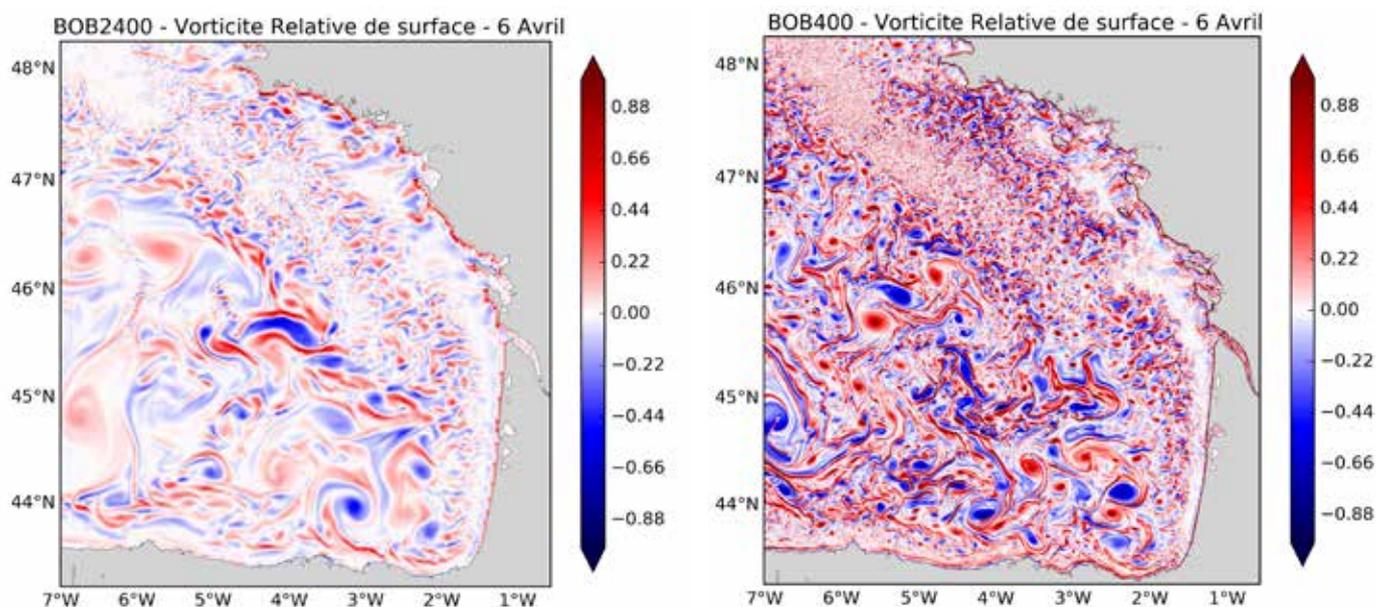
**Figure III.1:** Température de surface de la mer (°C) le 6 avril de la deuxième année de simulation pour la simulation BOB2400 (gauche) et la simulation BOB400 (droite).



**Figure III.2:** Intensité des courants de surface ( $m\ s^{-1}$ ) le 6 avril de la deuxième année de simulation pour la simulation BOB2400 (gauche) et la simulation BOB400 (droite).

associées ont été explorées par Yelekçi et al. (2017). La zone à proximité des côtes présentant une vorticité plus faible correspond aux zones d'influence des eaux dessalées à la lisière desquelles se développent des instabilités baroclines expliquant une intensification de la turbulence au milieu du plateau continental.

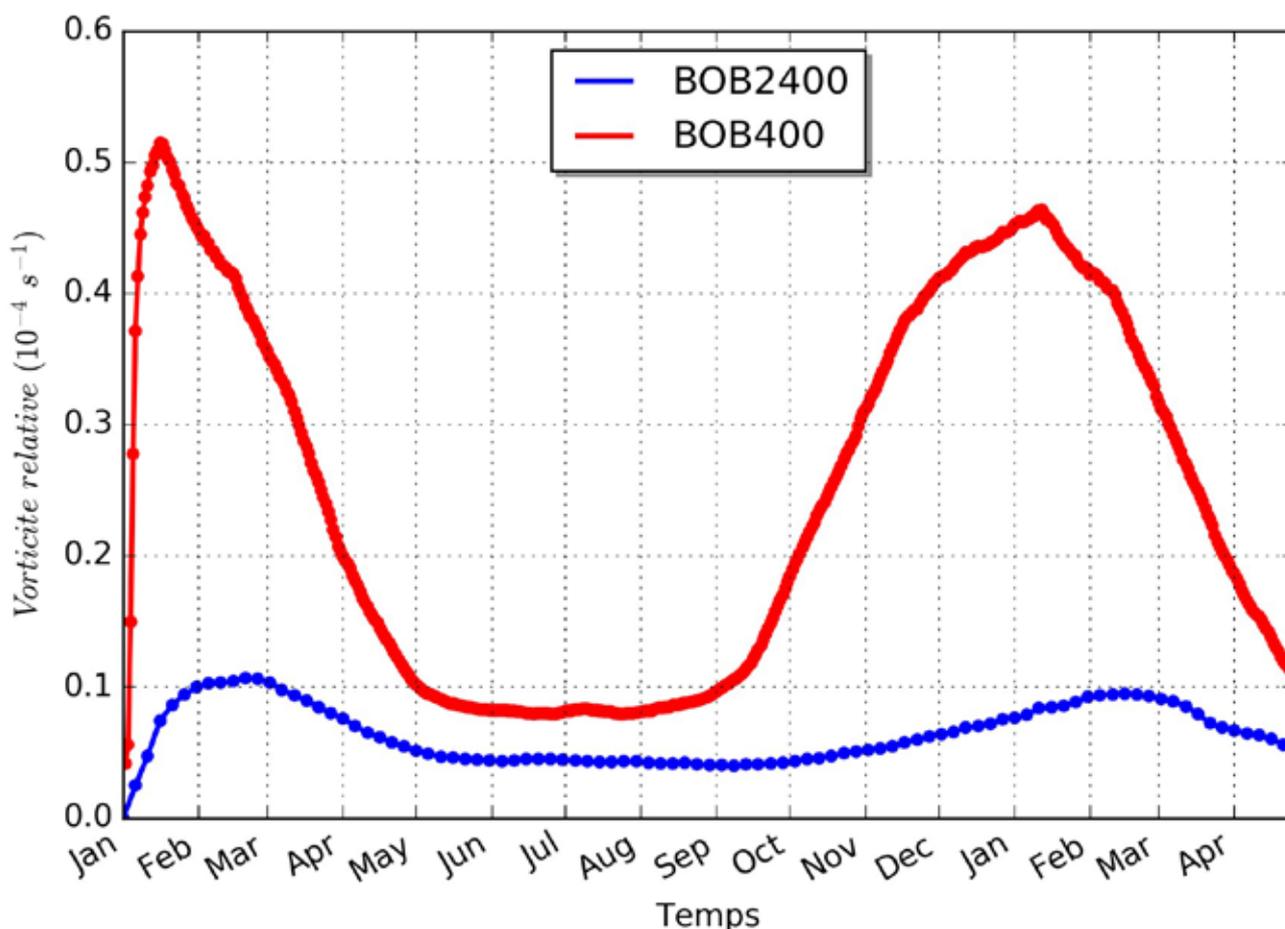
Lorsque cette vorticité relative de surface est moyennée sur l'ensemble du domaine (Figure III.4), la simulation avec la configuration BOB400 présente des valeurs jusqu'à 4 fois supérieures aux résultats obtenus à partir de la configuration BOB2400. Cette intensification de la vorticité relative et donc des structures tourbillonnaires



**Figure III.3:** Composante verticale de la vorticité relative de surface ( $10^{-4}\ s^{-1}$ ) pour la configuration BOB2400 (gauche) et la configuration BOB400 (droite) pour le 6 avril de la deuxième année de simulation (Y+1).

à fines échelles est attendue. L'analyse des résultats devra permettre de comprendre comment ces structures à fines échelles et les instabilités associées vont

impacter la circulation de la région et en particulier le mélange vertical. ■



**Figure III.4:** Moyenne spatiale sur tout le domaine simulé de la composante verticale de la vorticité relative de surface ( $10^{-4} s^{-1}$ ) pour la configuration BOB2400 (bleu) et la configuration BOB400 (rouge).

## 4 Conclusions

Ce nouveau défi sur Occigen aura permis de résoudre une partie significative de l'équation à l'origine de ce projet. En particulier, l'accès à ces ressources a rendu possible:

- l'évaluation des performances du code CROCO dans un environnement de calcul massivement parallèle appliqué à une région côtière, le golfe de Gascogne.
- la réalisation de différentes simulations dans deux configurations de résolution intermédiaire (2400m) et pour la première fois à l'échelle de cette région à très haute résolution (400m).

L'analyse de ces simulations ouvre de nombreuses perspectives dans la compréhension des processus à sous-mésoéchelles dans le golfe de Gascogne et l'influence de cette dynamique sur la circulation océanique dans cette région. ■

## 5 Références

- Charria, G., Theetten, S., Vandermeirsch, F., Yelekçi, Ö., and Audiffren, N. (2017), Interannual evolution of (sub) mesoscale dynamics in the Bay of Biscay, *Ocean Sci.*, 13, 777-797, <https://doi.org/10.5194/os-13-777-2017>.
- Debreu, L., P. Marchesiello, P. Penven, and G. Cambon, 2012: Two-way nesting in split-explicit ocean models: algorithms, implementation and validation. *Ocean Modelling*, 49-50, 1-21.
- Large, W.G., McWilliams, J.C. Doney, 1994. Oceanic vertical mixing: a review and a model with a nonlocal boundary layer paramétrisation. *Rev. Geophys.* 32, 363- 403.
- Penven P., L. Debreu, P. Marchesiello, and J.C. McWilliams, 2006 : Evaluation and application of the ROMS 1-way embedding procedure to the central california upwelling system. *Ocean Modelling*, 12, 157-187.
- Theetten S., Vandermeirsch F., Charria G. (2017). BACH1000\_100lev-51 : a MARS3D model configuration for the Bay of Biscay. SEANOE. <http://doi.org/10.17882/43017>.
- Yelekçi, O., G. Charria, X. Capet, G. Reverdin, J. Sudre, H. Yahia, Spatial and seasonal distributions of frontal activity over the French continental shelf in the Bay of Biscay, *Continental Shelf Research*, Volume 144, Pages 65-79, ISSN 0278-4343, <http://dx.doi.org/10.1016/j.csr.2017.06.015>, 2017.

## SIMULATION NOUVELLE GENERATION DE L'OCEAN TURBULENT GLOBAL

### Influence des instabilités sous-mésoéchelles dans le golfe de Gascogne

#### Auteurs

Jean-Marc MOLINES



Bernard BARNIER



(Équipe de Modélisation des Écoulements Océaniques Multi-échelles  
Institut des Géosciences de l'Environnement (IGE))

#### Contributeurs

Nicolas JOURDAIN, Julien LE SOMMER,  
Ignacio MERINO, Thierry PENDUFF  
(IGE)

Pierre MATHIOT (UKMO)

Nicole AUDIFFREN (CINES)  
Gérard GIL (CINES)

Camille LIQUE, Claude TALANDIER,  
Anne-Marie TREGUIER (LOPS)

Julie DESHAYES, Gurvan MADEC  
(LOCEAN)

Romain BOURDALLE-BADIE  
(MERCATOR Océan)

## SOMMAIRE

<b>1 Contexte</b> .....	<b>46</b>	<b>Icebergs et plateformes glaciaires</b> .....	<b>47</b>
Interactions d'échelles dans l'océan.....	46	<b>Moyens de calcul et simulation</b>	<b>48</b>
Simulations numériques de l'océan global .....	46	<b>3 Résultats</b> .....	<b>48</b>
Objectifs scientifiques.....	47	Circulation globale turbulente .	49
<b>2 Code, moyens de calcul, simulation</b> .....	<b>47</b>	Fonte des icebergs .....	49
NEMO .....	47	Banquise dans les régions polaires .....	50
Configuration Océan Global eORCA12 .....	47	<b>4 Conclusions</b> .....	<b>52</b>
		<b>5 Remerciements</b> .....	<b>52</b>
		<b>6 Bibliographie</b> .....	<b>52</b>

# 1 Contexte

## Interactions d'échelles dans l'océan

La variabilité de la circulation océanique et son évolution induite par les activités humaines ont potentiellement d'énormes conséquences. En effet, la circulation océanique globale contrôle en grande partie la biomasse et les cycles biogéochimiques marins, l'absorption et la séquestration du CO<sub>2</sub> de l'océan et son acidification, ainsi que les structures régionales du niveau de la mer [1].

Les courants océaniques s'organisent sur des échelles spatiales bien inférieures à la taille des bassins. La turbulence de mésoéchelle (méandres et tourbillons de taille variant de 10 km à 100 km) et les courants de bord s'écoulant le long des talus et sur les plateaux continentaux contribuent pour une part importante aux échanges de chaleur et d'eau douce entre les latitudes subtropicales et les latitudes polaires [2]. Ces dynamiques de fine échelle et leurs contributions aux circulations de plus grande échelle par le biais d'interactions non-linéaires ne sont pas résolues par les modèles de projection climatique [3] utilisés dans le contexte du GIEC. Pour le climat comme pour les applications opérationnelles [4], de grandes questions restent largement ouvertes

pour le traitement desquelles ces interactions d'échelles ont un rôle crucial.

L'avancée des sciences océaniques nécessite donc un cadre d'étude cohérent dans lequel les interactions entre les fines échelles critiques et la variabilité multi-décennale de la circulation de grande échelle puissent être quantifiées, permettant une évaluation et une amélioration véritables de leurs paramétrisations dans les modèles de prévision climatique. Les observations des fines échelles de la turbulence océanique sont trop rares et trop éparpillées dans le domaine de l'*in-situ* et limitées à la surface dans le domaine du spatial, et ne couvrent pas des périodes suffisamment longues, pour permettre d'aborder ces questions d'interactions multi-échelles de façon pertinente.

Ainsi, la simulation de l'océan par des modèles numériques à très haute résolution apparaît comme une approche pertinente et indispensable pour progresser dans la connaissance de ces interactions.

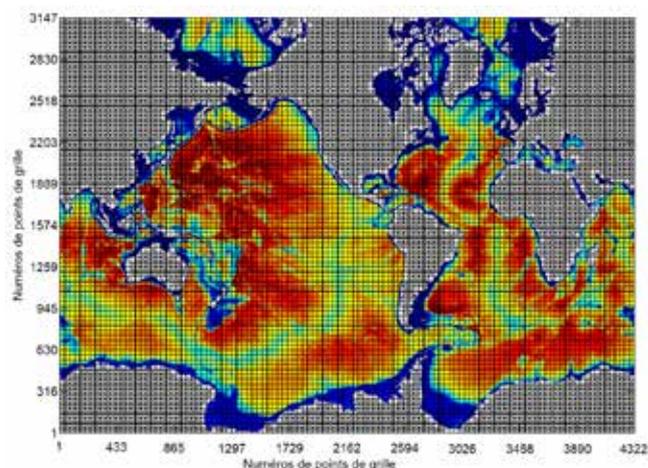
## Simulations numériques de l'océan global

Nos équipes développent et exploitent un cadre de simulation numérique «réaliste» de l'océan global qui permet de traiter les questions d'interactions d'échelles (processus, paramétrisations, impacts) pertinentes pour les études de prévision climatique et les applications opérationnelles. Ce cadre, intitulé DRAKKAR [16], est commun aux équipes utilisatrices du code communautaire NEMO [5] en Europe et est continuellement adapté aux supercalculateurs auxquels cette communauté a accès.

Une des configurations phares de DRAKKAR est la configuration ORCA12 [6] au 1/12° (Fig. 1). Utilisée opérationnellement par les services maritimes de COPERNICUS [4], elle préfigure les configurations de modèle d'océan qui seront utilisées à moyen terme par les systèmes de prévision saisonnière (c'est déjà le cas au UKMO) et à plus long terme par les *Earth System Models* (ESMs) pour les projections climatiques. Son utilisation se généralise peu à peu en Europe, malgré ses coûts de calcul encore importants qui nécessitent souvent des attributions spéciales.

Notre projet pour ces Grands Challenges concerne cette configuration phare dont une toute nouvelle version vient d'être mise en place, intégrant nombre d'innovations récentes dans le code NEMO. Nous avons proposé la réalisation d'une nouvelle simulation hindcast des ~40 dernières années (1979-2015). Les innovations que nous avons implémentées sont coûteuses en temps

de calcul et ont motivé notre participation aux Grands Challenges GENCI 2016.



**Figure 1 :** Décomposition du domaine global ORCA12 en 6523 sous-domaines. Chaque boîte de la grille représente un sous-domaine de calcul. Les boîtes marquées d'une croix sont les domaines « terre » et sont éliminés pour le calcul. La résolution de grille de 1/12° varie de ~9 km à l'équateur à ~3 km près des côtes de l'Antarctique. La couleur indique la profondeur de l'océan, discrétisée par 75 niveaux verticaux.

## Objectifs scientifiques

L'objectif scientifique général de notre projet est l'étude des impacts que les changements survenus dans l'atmosphère aux hautes latitudes lors des ~30 dernières années ont induits sur la circulation et la variabilité de l'océan global et de la banquise. Nous cherchons en particulier à déterminer les contributions de la mésoéchelle océanique d'une part et des flux d'eau douce (fonte des calottes, évolution de la banquise) d'autre part à la formation des masses d'eaux profondes et des eaux de fond formées dans les océans des hautes latitudes. Ces masses d'eau représentent plus de 80% du total de la masse de l'océan et contribuent à la régulation du climat sur les échelles décennales et plus longues.

Nous avons réalisé pour ces Grands Challenges une simulation globale à haute résolution ( $1/12^\circ$ ) de la circulation océanique et de la banquise longue de 37 ans (1979-2015) correspondant à la période couverte par la réanalyse atmosphérique ERAinterim [13], avec la toute nouvelle version de la configuration phare de DRAKKAR : eORCA12. Elle sera exploitée dans des études visant à quantifier les contributions de cette turbulence et des flux d'eau douce à la formation des masses d'eaux profondes dans l'Atlantique Nord (NADW), et des eaux intermédiaires (AAIW) et de fond (AABW) autour du continent Antarctique. ■

## 2 Code, moyens de calcul, simulation

La configuration eORCA12 que nous avons utilisée pour ces Grand Challenges représente une évolution sensible de la configuration ORCA12 utilisée par DRAKKAR ces

dernières années. Cette évolution concerne aussi bien le code numérique NEMO [5] que sa géométrie et les données de forçage.

### NEMO

Le code utilisé est le modèle de circulation océanique NEMO (Nucleus for European Modelling of the Ocean). Il combine différents modules permettant en particulier de simuler l'évolution de l'océan et de la banquise. Nous avons utilisé une version de développement (NEMO 4-alpha, rev 7046), afin de bénéficier des derniers schémas numériques disponibles. Pour la gestion des sorties du modèle, nous avons utilisé le serveur XIOS\_2.0. La parallélisation se fait par décomposition de domaines. C'est un code Fortran90 et pur MPI qui tourne régulièrement sur tout type de plateformes, dont les machines Tier-0 comme OCCIGEN (CINES) ou CURIE (TGCC). Les évolutions majeures concernant le code ont été les suivantes :

- Modèle de banquise LIM3 : Cette nouvelle version [7] représente une avancée majeure par rapport au modèle

précédent : multi-catégories, il utilise une rhéologie Elasto-Visco-Plastique. Autre innovation remarquée, le couplage avec le modèle d'océan se fait par échange de masse entre la banquise et l'océan.

- Nouvelle surface libre non-linéaire associée à une coordonnée verticale en volume variable (vvl) qui assure une conservation exacte du contenu en sel du modèle.

- Nouveaux schémas d'advection pour la quantité de mouvements (dont on attend une réduction du bruit sur les vitesses verticales ce qui réduira la diffusion diapycnale numérique).

- Paramétrisation du mélange vertical (nouveaux paramètres de fermeture turbulente TKE).

### Configuration Océan Global eORCA12

La géométrie de la configuration globale utilisée est baptisée eORCA12.L75 car sa géométrie est différente de la configuration ORCA12.L46 dont elle est issue.

- Le domaine modélisé a été étendu vers le sud pour prendre en compte les parties de l'océan situées sous les plateformes glaciaires Antarctiques, en prévision du futur couplage de NEMO aux modèles de calottes.

- La résolution verticale a été augmentée de 46 à 75 niveaux verticaux avec une résolution de 1m sur les dix premiers mètres de l'océan puis une augmentation progressive de la distance entre les niveaux vers le

fond pour une épaisseur de niveau maximale de 200 m dans les plaines abyssales. On en attend une meilleure représentation des flux verticaux dans les couches de surface.

## Icebergs et plateformes glaciaires

Le forçage en eau douce du modèle a été très largement modifié, surtout en ce qui concerne les flux d'eau douce continentale des régions polaires :

- NEMO est couplé à un module lagrangien de dérive et fonte des Icebergs, initialement développé au NOCS [8] et modifié et implémenté à l'IGE [9]. Chaque iceberg est traité comme une particule lagrangienne dont la dérive dépend de l'action du courant des couches de surface, et le taux de fonte dépend des températures océaniques ambiantes [9]. Ce module permet ainsi de calculer à chaque instant le flux d'eau douce lié à la présence d'icebergs libres d'évoluer. Le coût de calcul est élevé car des millions d'icebergs sont traités. Les icebergs sont émis à partir de plusieurs points sources en Antarctique et en Arctique, chaque point source ayant son propre taux de production calibré sur les

observations. Seuls les « petits » icebergs (de taille inférieure ou équivalente à la maille du modèle) sont traités. Ils représentent la plus grande part du flux.

- La paramétrisation des flux de fonte à la base des plateformes glaciaires Arctiques et Antarctiques développée à UKMO [10]. Cette paramétrisation permet de prendre en compte la distribution réaliste, tant géographique qu'en fonction de la profondeur, du flux d'eau douce associé aux interactions entre l'océan et les plateformes.

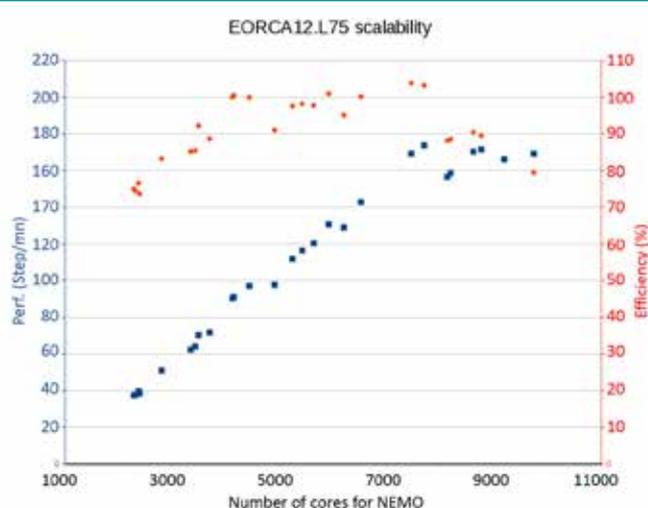
- La dernière version du Drakkar Forcing Set, le DFS5.2 [11], issu des réanalyses du ECMWF et couvrant la période 1979-2015 avec une fréquence à 3h a été utilisé pour les conditions limites de surface (vents, flux de chaleur et d'eau douce).

## Moyens de calcul et simulation

Dans le contexte de ces Grands Challenges, nous avons retenu un découpage en 6523 sous-domaines (Fig. 1) et utilisé 113 serveurs d'I/O. La configuration a donc tourné sur 237 nœuds Broadwell. Ce choix a été guidé par l'étude de la scalabilité de cette configuration sur Occigen2 qui a montré de bonnes performances jusque vers 7500 cœurs de calcul (Fig. 2).

Après un *spin-up* du modèle de 7 ans, nous réalisons la simulation des années 1979 à 2015, soit un total de 44 ans correspondant à 6.4 millions d'heures de calcul. Les travaux, réalisés dans la deuxième phase des Grands Challenges (à partir de Janvier 2017), se sont enchaînés d'une façon très fluide et avec très peu d'erreurs.

Le serveur d'I/O a été paramétré de telle façon que nous avons sauvegardé les sorties du modèle en moyenne journalière pour les champs bidimensionnel (flux, température et salinité de surface, élévation de la surface, champs de vitesses en surface et à 10, 30 et 40 m de profondeur), et en moyenne à 5 jours [14] pour les champs représentant l'état de l'océan tridimensionnel. Tous les fichiers produits sont au format netcdf4/hdf5 avec compression de donnée activée. L'ensemble des résultats de cette simulation correspond à un jeu de données primaire de 75 Téraoctets. ■



**Figure 2 :** Etude de scalabilité pour la configuration eORCA12.L75. En abscisse, le nombre de cœurs utilisés pour NEMO. En ordonnée à gauche (bleu) les performances du code en pas-de-temps/minutes. En ordonnée à droite l'efficacité en % (rapport entre speedup réel et speedup théorique). À noter que le speedup est calculé par rapport au cas 4125 cœurs. L'élimination des domaines correspondants aux continents permet de diminuer significativement le nombre de points de grille à calculer quand on augmente de nombre de cœurs.

## 3 Résultats

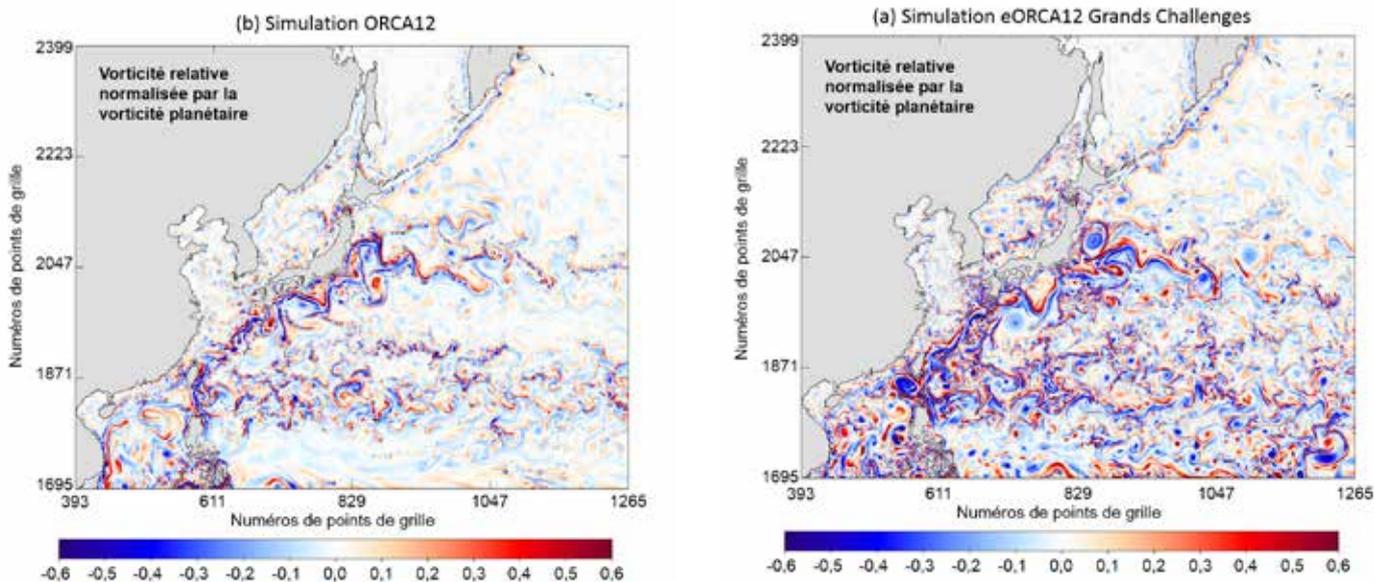
Nous présentons ici les résultats qui sont les plus pertinents pour les évolutions apportées au modèle, c'est-à-dire la représentation de la turbulence de moyenne échelle, des flux d'eau douce associés à la fonte des icebergs au cours de leur dérive, la couverture de la banquise et les propriétés des masses d'eaux au voisinage des

plateformes glaciaire antarctiques. Afin de caractériser les impacts des évolutions apportées à eORCA12, la simulation réalisée pour les Grands challenges est comparée avec une simulation réalisée avec la version précédente de la configuration ORCA12.

## Circulation globale turbulente

L'activité turbulente simulée par le modèle est notablement supérieure à celle des simulations précédentes. Ceci est illustré en figure 3 par des cartes d'instantané de vorticité relative, variable caractérisant bien l'activité tourbillonnaire. L'amplitude des tourbillons est supérieure dans la simulation des Grands Challenges et une plus grande gamme d'échelles spatiales est visible, indicatrices d'une turbulence mieux résolue. Ceci est remarqué dans toutes les régions du globe, même les régions

de faible activité où le modèle tend à sous-estimer la turbulence. Ceci est une conséquence de la moindre diffusion numérique liée aux nouveaux schémas et de l'augmentation de la résolution verticale. On s'attend à ce que cette meilleure représentation de la turbulence de mésoéchelle ait un impact sur les transports méridiens et les flux verticaux.

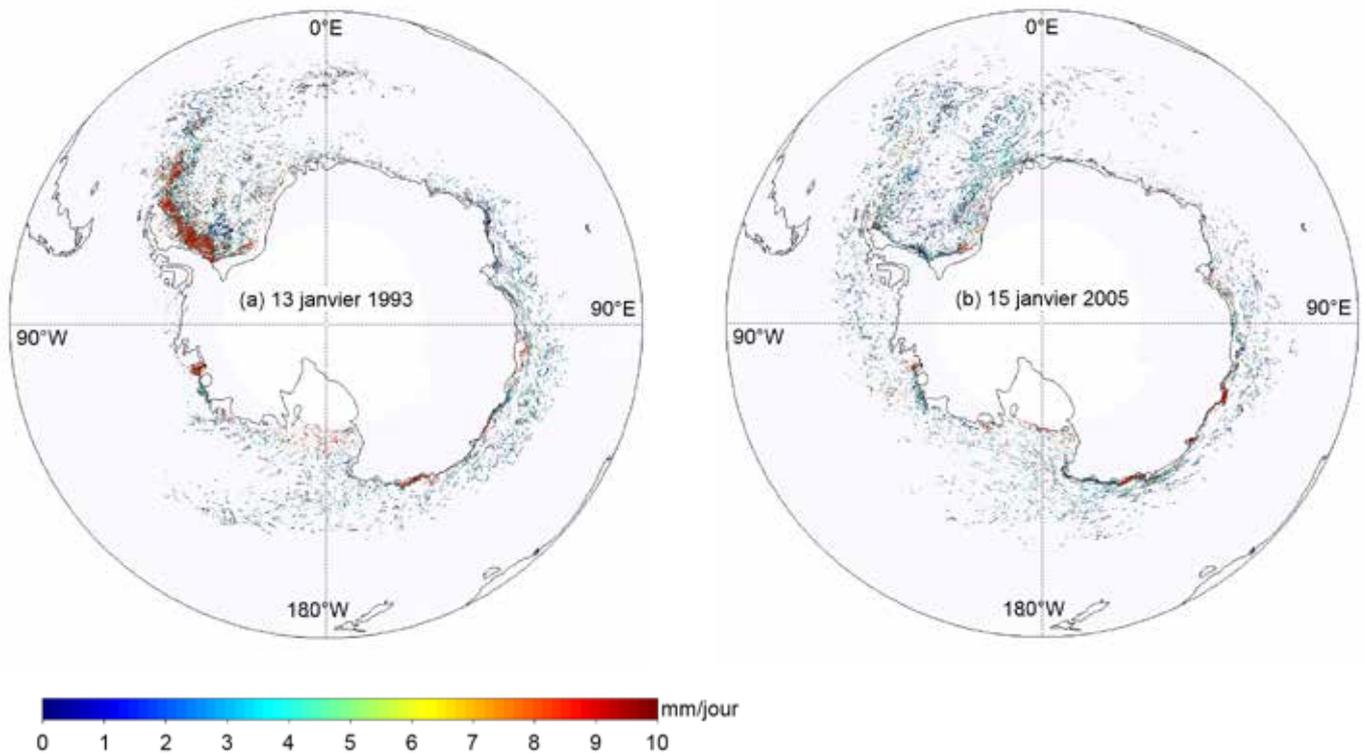


**Figure 3 :** Instantanés (date du 14 mars 1996) de vorticité relative (normalisée, sans unité) dans le Pacifique Nord-Ouest, siège du Kuroshio pour : (a) la simulation eORCA12 des Grands Challenges, et (b) pour la simulation réalisée avec une version antérieure de la configuration ORCA12. Les anticyclones (cyclones) sont représentés en bleu (rouge).

## Fonte des icebergs

Plusieurs millions d'icebergs ont été relâchés des calottes glaciaires antarctiques et arctiques pendant la simulation. Dérivants de façon lagrangienne au gré des courants de surface, ces icebergs relâchent de l'eau douce au cours de leur fonte, modifiant la salinité de surface de l'océan des hautes latitudes. Le flux d'eau douce instantané de cette fonte est présenté en figure 4 pendant l'été austral pour deux années différentes. On constate une forte variabilité interannuelle dans la distribution géographique des icebergs et dans l'amplitude de la fonte, particulièrement marquée le long de la péninsule Antarctique à l'ouest de la mer de Weddell. La fonte peut atteindre 10 mm/jour sur la surface d'une maille du modèle. La fonte est également marquée par un cycle saisonnier important (non montré).

Le flux d'eau douce moyen venant de la fonte sur la période d'intégration (non montré) est généralement inférieur à 0.5 mm/jour sauf près des sources où il atteint des valeurs de 1 à 3 mm/jour. La distribution géographique est sensiblement différente de celle proposée par les modèles de résolution plus faible [9]. La turbulence, en donnant aux icebergs une trajectoire plus chaotique qui leur permet de s'échapper du courant moyen, favorise une distribution plus homogène spatialement.



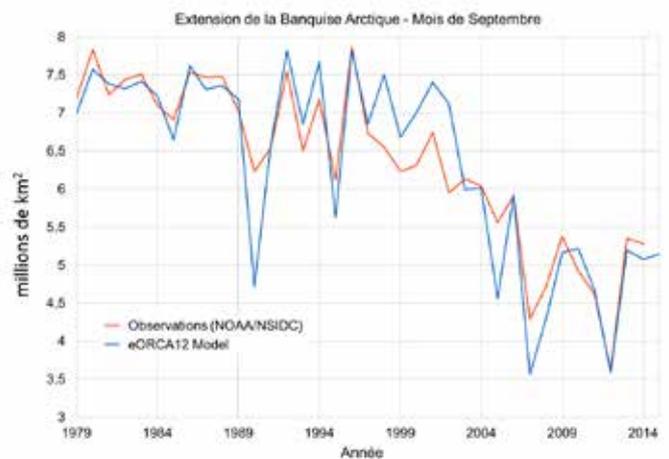
**Figure 4 :** Flux d'eau douce dû à la fonte des icebergs dans l'océan austral à deux instants différents. Chaque point représente la position d'un iceberg, la couleur indiquant son taux de fonte (mm/jour).

## Banquise dans les régions polaires

Une évolution majeure de NEMO concerne le modèle de banquise LIM3 [7]. Un résultat notable, illustré en figure 5, est le fait que dans l'Arctique le modèle montre une bonne capacité à reproduire les variations interannuelles et la décroissance persistante des surfaces couvertes par la banquise en été depuis le milieu des années 1990. Le modèle manque parfois de précision dans la représentation des valeurs extrêmes, mais simule fidèlement le maximum de 1996, et le minimum de 2012, ce dernier indiquant une réduction de moitié de la couverture de banquise depuis les années 1990. Si modèle et observations s'accordent relativement bien sur la surface globale couverte de banquise, ils ont encore des divergences notables sur sa répartition géographique (Fig. 6). Les régions libres de glace (concentration < 15%) simulées sont en effet trop grandes dans le bassin Eurasien et encore trop faibles dans le bassin Canadien. Ceci indique que l'accord sur la couverture globale est en partie le résultat de compensations régionales.

Dans l'océan Austral la couverture de banquise en hiver ne montre pas de tendance notable, en accord avec les observations. On remarque que le modèle génère des polynies côtières (zones libres de glace) pendant l'hiver austral (Fig. 7a), avec des conséquences sensibles sur les propriétés des masses d'eaux (Fig. 7b,c). La comparaison avec la simulation ORCA12, qui ne prends pas en compte la fonte à la base des plateformes glaciaires produit des ouvertures de surface plus limitée (non montré), montre que la présence de la polynie s'accompagne

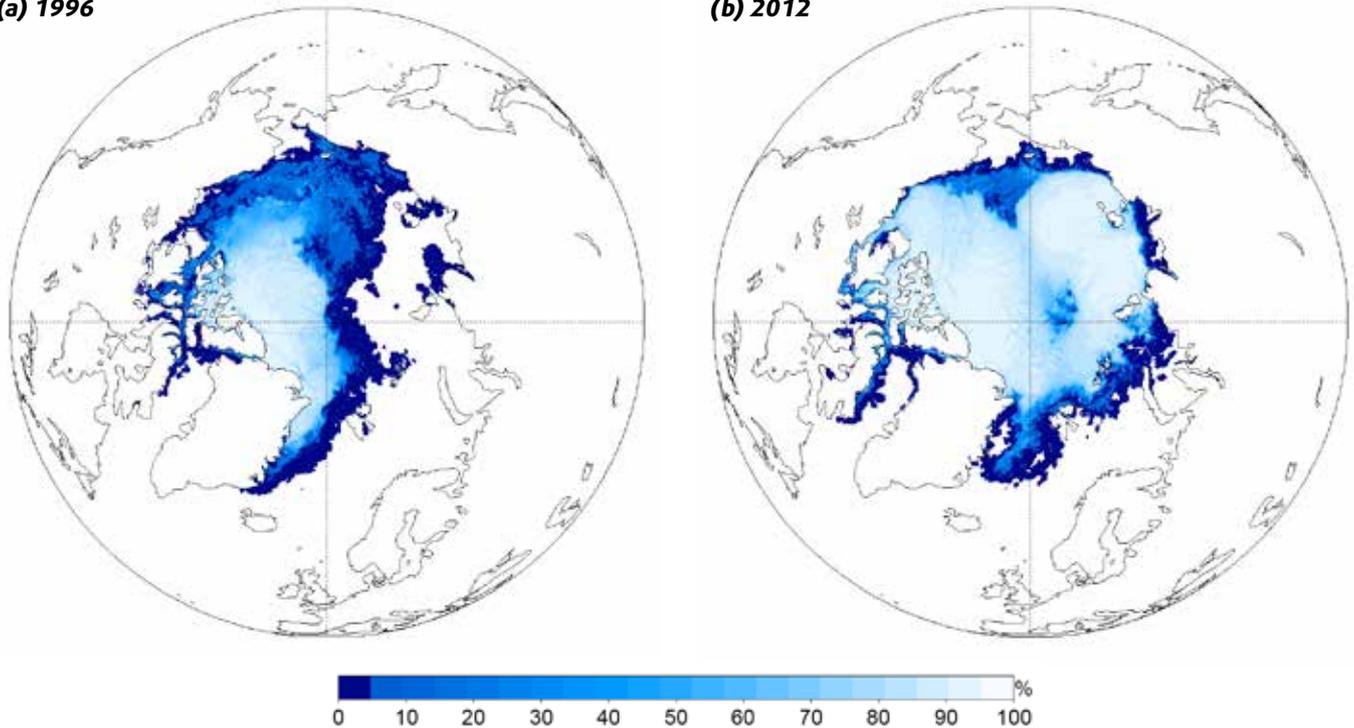
d'un fort mélange vertical entre les eaux de surface et de fond (line de mélange sur la figure 7b). Ce mélange est causé par la fonte en profondeur de la plateforme glaciaire et s'accompagne d'une remontée de chaleur vers la surface faisant fondre la glace et provoquant l'ouverture d'une zone d'eau libre (polynie).



**Figure 5 :** Variations interannuelles de la surface couverte par la banquise (plus de 15% de glace) en Arctique au mois de septembre simulées par eORCA12 (bleu) et estimées par l'observation spatiale (rouge, NOAA/NSIDC [14]).

(a) 1996

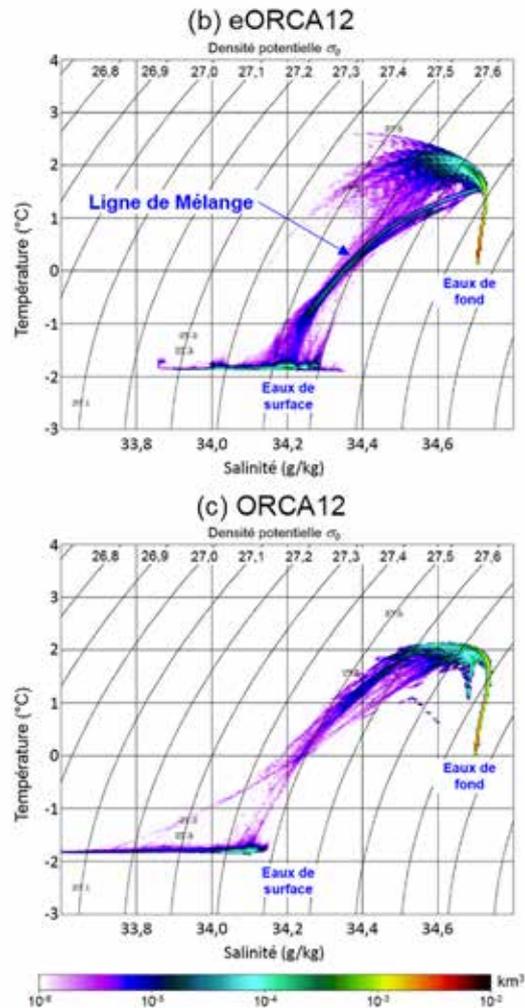
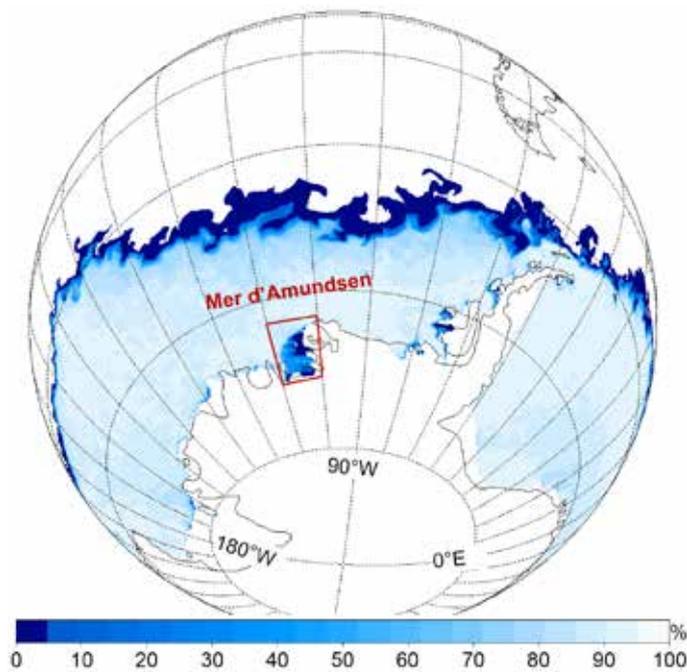
(b) 2012



**Figure 6 :** Concentration (en % de glace) de la banquise au début du mois de septembre, correspondant pour l'océan Arctique à (a) un maximum de couverture en 1996 et (b) un minimum en 2012.

(a) Concentration de glace

(b) Diagramme T,S volumique



**Figure 7 :** (a) Polynie (ouverture dans la banquise) en Mer d'Amundsen pendant l'hiver Austral (14 septembre 1996). La partie sud de cette mer (cadre rouge) montre une large région de très faible concentration de glace (<10%). (b et c) Diagramme T,S volumique indiquant les propriétés de température (T), salinité (S) et densité potentielle ( $\sigma_0$ ) des masses d'eaux occupant le plus de volume dans la polynie de la Mer d'Amundsen (cadre rouge) pour (a) la simulation eORCA12 Grands Challenges et (b) la simulation de référence ORCA12.

## 4 Conclusions

La simulation réalisée pour ces Grands Challenges GENCI est la première simulation sur la période 1979-2015 des émissions, dérives et fontes des icebergs dans un modèle de si haute résolution. Elle a donc permis de prendre en compte la turbulence océanique dans l'estimation de l'eau douce continentale apporté à l'océan par ce processus. Seul les Grands Challenges nous ont donné la puissance de calcul nécessaire au suivi lagrangien de millions d'icebergs. Cet estimé pourra être utilisé directement dans d'autres simulations, ce qui évitera

le coût de calcul associé à ce suivi. Le traitement des cavités sous les plateformes glaciaires doit être amélioré. L'analyse de nos simulations montre une dépendance encore sensible de la solution du modèle aux schémas numériques et aux paramétrisations sous-maille utilisés, même à cette haute résolution. Les travaux futurs de la coordination DRAKKAR se focaliseront donc sur le traitement de ces questions, en particulier dans les couches limites ou les courants interagissent avec le relief bathymétrique. ■

## 5 Remerciements

Nous remercions les organismes qui ont apporté leur soutien au GDRI DRAKKAR (INSU-CNRS, GENCI, LEFE, IFREMER, CNES). ■

## 6 Bibliographie

■ [1] IPCC (2013), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

■ [2] Volkov, D. L., T. Lee, and L.-L. Fu (2008), Eddy-induced meridional heat transport in the ocean, *Geophys. Res. Lett.*, 35, L20601, doi:10.1029/2008GL035490.

■ [3] Barnier B., Penduff T., and Langlais C., 2011: Eddying vs. laminar ocean circulation models and their applications. *Operational Oceanography in the 21st Century*. Schiller, Andreas; Brassington, Gary B. (Eds.), 1st Edition., 2011, X, 450 p., ISBN: 978-94-007-0331-5

■ [4] COPERNICUS Marine Environment Monitoring Services. <http://marine.copernicus.eu/>

■ [5] Madec, G., 2014. NEMO ocean engine. Note du Ple de modélisation, Number 27. Institut Pierre-Simon Laplace (IPSL).

■ [6] DRAKKAR Group (A.M. Treguier, B. Barnier, A.T. Blaker, A. Biastoch, C.W. Böning, A. Coward, J. Deshayes, A. Duchez, J. Hirschi, J. Le Sommer, G. Madec, G. Maze, J. M. Molines, A. New, T. Penduff, M. Scheinert, and C. Talandier), 2014: DRAKKAR: Developing high-resolution ocean components for European Earth system models. *CLIVAR Exchanges Newsletter*, No.64 (Vol 19, No.1).

■ [7] Rousset C., M. Vancoppenolle, G. Madec, T. Fichefet, S. Flavoni, A. Barthélemy, R. Benshila, J. Chanut, C. Levy, S. Masson, and F. Vivier, 2015: The Louvain-La-Neuve sea ice model LIM3.6: global and regional capabilities.

## PHYSIQUE DU BOURGEONNEMENT DES GOUTTELETES LIPIDIQUES

Auteur



Luca MONTICELLI  
(Molecular Microbiology and Structural  
Biochemistry (MMSB)  
UMR 5086, CNRS)

### Contributeurs

Abdou Rachid THIAM

(Laboratoire de Physique Statistique - École Normale Supérieure)

Mathieu CLOIREC (CINES)

### Abréviations

GL = gouttelette lipidique ; RE = réticulum endoplasmique ;  
MD = dynamique moléculaire ; CG = gros-grain.

## SOMMAIRE

<b>1</b>	<b>Résumé</b> .....	<b>54</b>	<b>5</b>	<b>Retour d'expérience sur l'utilisation de la machine</b> .....	<b>56</b>
<b>2</b>	<b>Contexte</b> .....	<b>54</b>		Comparaison de benchmarks : vieux nœuds vs. nouveaux nœuds .....	56
<b>3</b>	<b>Objectifs</b> .....	<b>54</b>	<b>6</b>	<b>Résultats et discussion</b> .....	<b>57</b>
	1. Bourgeonnement des gouttelettes lipidiques naissantes: rôle des phospholipides .....	55		Forme et composition de gouttelettes lipidiques .....	57
	2. Directionnalité du bourgeonnement de LD: rôle de la distribution des protéines .....	55		Bourgeonnement des gouttelettes .....	57
<b>4</b>	<b>Méthodologie</b> .....	<b>55</b>	<b>7</b>	<b>Remerciements</b> .....	<b>58</b>
	Modélisation à gros-grain .....	55	<b>8</b>	<b>Références</b> .....	<b>58</b>
	Logiciel .....	56			

## 1 Résumé

Les gouttelettes lipidiques (GL) sont les organelles qui régulent le métabolisme énergétique dans les cellules. Le mécanisme de formation des GL implique la formation d'une lentille d'huile entre les monocouches d'une bicouche lipidique. Une déformation asymétrique de cette lentille et de la membrane forme une protubérance presque sphérique. Ce processus est appelé «bourgeonnement». Le mécanisme et les forces motrices du bourgeonnement sont inconnus. Sur la base d'études expérimentales, deux hypothèses ont été avancées. Elles sont liées à (a) la nature et la répartition des lipides entre les deux monocouches de la membrane, et (b) la répartition asymétrique des protéines des deux côtés de la membrane. Dans ce projet, nous avons testé la première hypothèse en utilisant des simulations moléculaires dites "gros-grains". Lors de la première phase du

projet, nous avons caractérisés les déformations de la LD naissante qui accompagnent les changements dans la composition des phospholipides de la bicouche. Dans la deuxième partie, nous avons exploré l'effet de la distribution lipidique asymétrique dans la bicouche sur la direction du bourgeonnement. En raison de la taille typique des GL naissantes, la réalisation de simulations sur ces systèmes (atteignant 80 nm en taille latérale) n'est possible, couplé à l'utilisation d'une méthodologie "gros-grain", que dans le cadre d'un projet "Grand Challenge". Les simulations permettent d'observer des transformations dans les agrégats lipidiques avec une résolution nanométrique ; ce qui offre une occasion unique d'interpréter les résultats expérimentaux obtenus sur ces systèmes fascinants. ■

## 2 Contexte

Les gouttelettes lipidiques (LD) sont les organelles intracellulaires régulant le bilan énergétique de la cellule. À l'instar des gouttelettes d'émulsion huile-dans-eau, elles sont recouvertes d'une seule couche de phospholipides. Elles sont, ainsi, différentes de tous les autres organites cellulaires, entourés d'une double couche de phospholipides. En dépit de leur rôle fondamental dans le métabolisme cellulaire, leur implication dans les maladies majeures et leur nombre croissant de fonctions nouvellement identifiées, le mécanisme de formation de GL à partir du réticulum endoplasmique reste mal compris [2]. Pour former un GL, les molécules d'huile synthétisées dans le réticulum endoplasmique (RE, un autre organite cellulaire) s'accumulent entre les monocouches de la membrane et induisent une déformation de la membrane. Ce processus de formation se compose de trois étapes: nucléation, croissance et bourgeonnement [2]. Au cours de la nucléation, les molécules d'huile commencent à s'agréger et forment une pré-gouttelette, qui est thermodynamiquement instable. Si la pré-gouttelette atteint une taille critique, elle augmentera (par exemple, par fusion avec d'autres pré-gouttelettes) pour former une «lentille» dans les deux tracts de la membrane bicouche. Afin de produire un GL, une telle lentille devra ensuite bourrer, c'est-à-dire former une gouttelette presque sphérique enfermée dans une saillie entre les deux monocouches du RE. Par conséquent, le bourgeonnement implique un changement majeur dans la forme de l'inclusion de l'huile, le passage d'une forme de lentille à une forme presque sphérique. Quelles sont les forces motrices qui déterminent ce changement de forme? Les études expérimentales et théoriques ont avancé plusieurs hypothèses, mais elles sont difficiles à explorer expérimentalement, que ce soit in vitro ou

sur des systèmes modèles synthétiques. Une hypothèse met en évidence le rôle important de la nature chimique des phospholipides dans la membrane du RE: différents phospholipides favoriseront différentes formes de la lentille, favorisant ou défavorisant le bourgeonnement. Pourtant, les différences de forme de la lentille n'ont pas encore pu être observées expérimentalement. Les simulations moléculaires sont un excellent outil pour aborder cette question, car elles donnent un aperçu de la structure des agrégats lipidiques sur des échelles de longueur de quelques dizaines de nanomètres et ne sont pas accessibles aux techniques de microscopie.

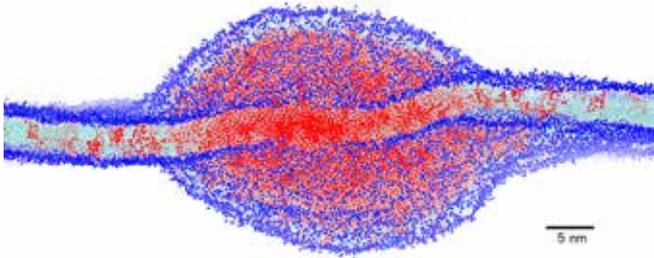
Une autre caractéristique importante du mécanisme de bourgeonnement de GL est sa «directionnalité»: les LD sont formées à partir de la membrane du RE qui est une bicouche lipidique. Dans la plupart des cas, elles se dirigent vers le côté cytosolique du RE, et non le côté luminal. Les forces motrices moléculaires responsables de ce phénomène sont inconnues. On suppose qu'elles seraient liées à la composition lipidique asymétrique de la bicouche ER et/ou à la répartition asymétrique des protéines liées à la surface du RE. Les simulations moléculaires seraient ici aussi un outil idéal pour aborder la question. ■

## 3 Objectifs

### 1. Bourgeonnement des gouttelettes lipidiques naissantes: rôle des phospholipides

Il a été suggéré que la composition des phospholipides jouerait un rôle de premier plan dans le processus de bourgeonnement des GL. Pour tester cette hypothèse, nous avons généré des modèles de GL naissantes (c'est-à-dire

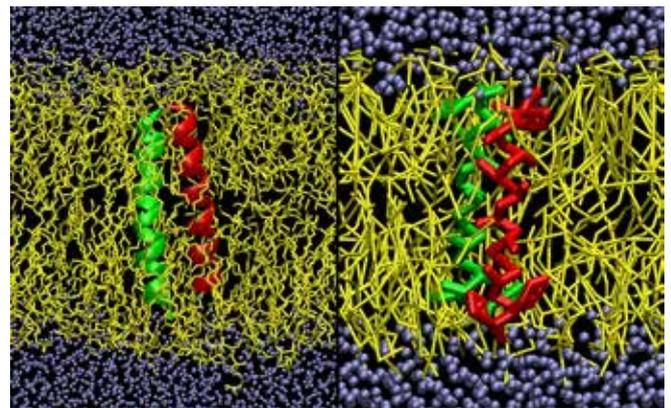
des lentilles d'huile incluses dans deux feuillettes d'une bicouche de phospholipides) et nous prévoyons d'examiner l'effet de la composition de phospholipides (PL) sur la forme de la GL naissante.



**Figure 1.** : Gouttelette lipidique naissante, consistant en une bicouche lipidique de DOPC (tête polaire des lipides en bleu, le reste du lipide en cyan) et une lentille de trioléin (TO). Le système contient 7500 molécules de TO, 18144 molécules de DOPC, et environ 1,5 M de particules d'eau (non représentées), pour un nombre total de particules d'environ 2M (correspondant à approximativement 23M atomes). La taille du système est de 78,5 x 78,5 x 36 nm.

### 2. Directionnalité du bourgeonnement de LD: rôle de la distribution des protéines

Les GL naissantes avec la même concentration et le même type de lipides et de protéines dans les brochures cytosoliques et lumenale devraient présenter une même probabilité de bourgeonnement sur chaque face. Dans les cellules, le bourgeonnement se produit de manière préférentielle vers le côté cytosolique - ce qui indique une asymétrie dans la composition lipidique ou protéique (ou les deux). Nous testons si un déséquilibre de la tension superficielle ou de la concentration de lipides seul est capable d'induire des déformations asymétriques dans la GL naissante.



**Figure 2.** : Instantanés à partir de simulations atomiques (gauche) et CG (droite) de peptides (rouge et vert) dans une membrane lipidique DOPC (jaune), avec une eau explicite (bleu).

## 4 Méthodologie

### Modélisation à gros-grain

La modélisation gros-grain (CG, « coarse-grain » en anglais) représente une simplification utile des systèmes atomiques, permettant des simulations de plus grandes portions de matière sur des échelles de temps plus longues. Dans les modèles CG, plusieurs atomes sont regroupés dans une particule « virtuelle » qui interagit à travers un potentiel moyen. Compte tenu de la taille typique des liposomes et des GL, et de l'échelle de temps de leur dynamique, l'utilisation de modèles gros-grain est essentielle pour notre projet.

Pour le projet actuel, nous avons choisi d'utiliser le champ de force MARTINI [1,3] (figure 2). Co-développé par le Dr Monticelli, MARTINI permet une reproduction semi-quantitative de nombreuses propriétés des agrégats lipidiques et est actuellement le modèle CG le plus utilisé. Il a été utilisé avec succès pour simuler divers systèmes membranaires et protéiques, (révisé dans la référence [4]).

## Logiciel

Toutes les simulations ont été réalisées avec le logiciel GROMACS, version 5.0.7 et 5.1.4 [5,6]. GROMACS utilise les bibliothèques FFTW, le multithreading OpenMP pour la parallélisation multicore intra-nœud et MPI pour la parallélisation inter-nœud (voir [http://www.gromacs.org/Documentation/Acceleration\\_and\\_parallelization](http://www.gromacs.org/Documentation/Acceleration_and_parallelization)). Actuellement, les options d'accélération prises en charge par GROMACS sont: SSE2, SSE4.1, AVX-128-FMA (AMD Bulldozer + Piledriver), AVX-256 (Intel Sandy + Ivy Bridge) et AVX2 (Intel Haswell / Haswell-E, Skylake). GROMACS peut également utiliser l'accélération GPU

en même temps, avec un équilibrage de charge CPU-GPU automatisé. Dans l'ensemble, GROMACS est l'un des logiciels MD parallèles les plus performants, et il atteint une excellente mise à niveau jusqu'à 2000 CPU dans le cas des systèmes typiques à utiliser dans l'étude proposée (environ 2M particules). La mise à l'échelle avec la taille du système est approximativement linéaire dans GROMACS, principalement en raison de l'utilisation de la décomposition du domaine et de l'équilibrage de charge dynamique (pour plus d'informations, voir le site Web de GROMACS: [www.gromacs.org](http://www.gromacs.org)). ■

## 5 Retour d'expérience sur l'utilisation de la machine

À l'ouverture de la nouvelle machine, nous avons dû résoudre un certain nombre de problèmes pour que des simulations de production puissent commencer. Nous avons notamment eu des problèmes avec la procédure de connexion, puis avec les versions manquantes du logiciel Gromacs, des difficultés à compiler des versions plus anciennes et plus récentes de Gromacs, l'absence d'« output » lors des simulations et des problèmes avec le script exécutant les simulations, en raison des

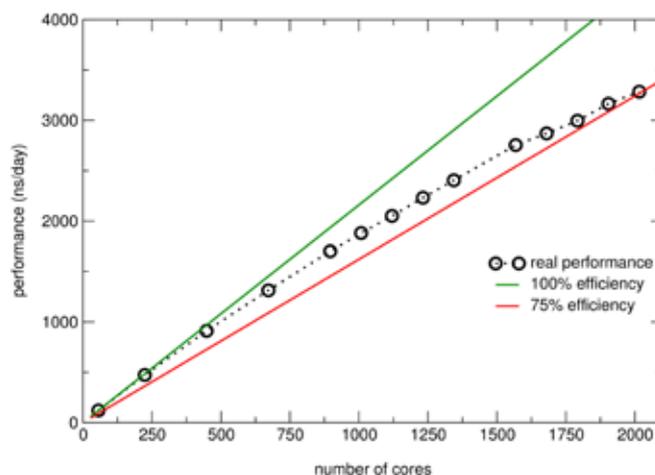
différences dans l'installation du logiciel. Il a fallu environ un mois pour résoudre ces problèmes. A ce moment-là, il était impossible de connaître la quantité de temps CPU utilisé - qui n'a été résolue qu'après la fin de la période Grand Challenge.

Une fois que les simulations de production ont commencé, aucun problème significatif n'a été rencontré à n'importe quel stade et le système s'est avéré très stable.

### Comparaison de benchmarks : vieux nœuds vs. nouveaux nœuds

En 2015, nous avons mesuré l'efficacité du logiciel GROMACS dans nos simulations MD sur Occigen (vieux nœuds) pour un certain nombre de systèmes de taille différente, entre 20 000 particules et particules 1,2 M. Nous avons observé que, pour les plus petits systèmes, la performance maximale est déjà atteinte avec 96 cœurs. D'autre part, la plus grande échelle de système était excellente (> 70% d'efficacité) jusqu'à 864 cœurs (c'est-à-dire 36 nœuds), et la performance maximale a été atteinte à plus de 1000 cœurs. Avec une efficacité de 75%, la performance était d'environ 1,7  $\mu$ s / jour.

Les systèmes à utiliser dans le projet Grand Challenge se composaient d'environ 2M de particules, donc nous nous attendions à ce que les performances maximales atteignent environ 2000 cœurs. Cela s'est vérifié dans la phase initiale du projet. Une référence est présentée ci-dessous pour un système typique utilisé dans le présent projet. La taille du système est d'environ 80x80x36 nm, le nombre de particules gros-grain d'environ 2M (correspondant à environ 23M atomes). La fonction d'énergie potentielle provient de MARTINI v2.2, sans interactions électrostatiques à longue distance. L'efficacité des 2016 cœurs (72 nœuds) était supérieure à 75%, ce qui a donné près de 3,3  $\mu$ s par jour, environ 2 fois plus que les précédents nœuds Occigen. ■



**Figure 3.** Performance absolue en ns / jour obtenue avec GROMACS (v5.1.4) sur Occigen (nouveaux nœuds). Le système simulé était semblable à celui représenté à la figure 1 (contenant environ 2M de particules). Les calculs ont été effectués avec le champ de force gros-grain MARTINI.

## 6 Résultats et discussion

### Forme et composition de gouttelettes lipidiques

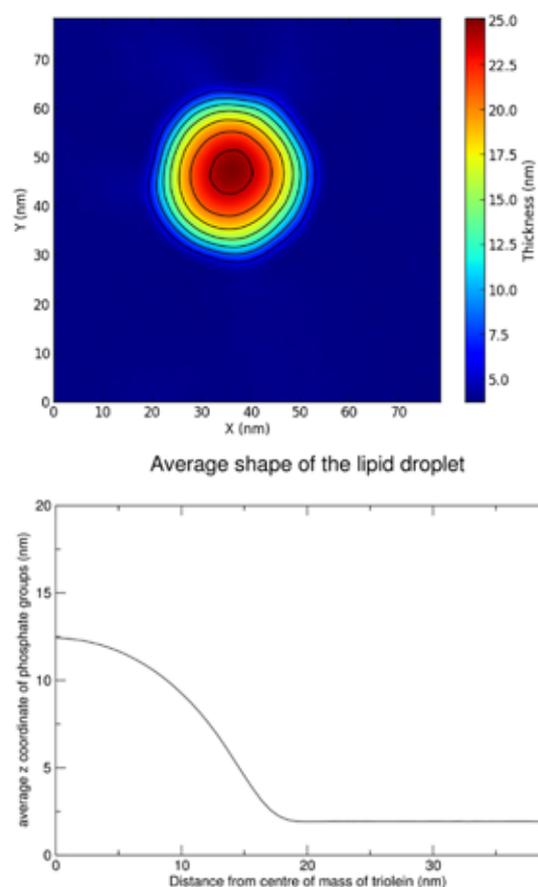
Afin de pouvoir observer les variations de la forme des LD en fonction de leurs compositions en phospholipides, nous devons faire face à deux obstacles majeurs liés à l'échantillonnage: l'échelle de longueur et l'échelle de temps des simulations. En termes d'échelle de longueur, il a été nécessaire d'utiliser de grandes tailles de système. En effet, la taille de la GL naissante doit avoir un diamètre de plus de 30 nm, ce qui implique une taille latérale de la bicouche lipidique de plus de 70 nm. En gardant la composition de la phase huileuse constante (100% de trioléin), nous avons préparé des systèmes avec différentes bicouches phospholipides, augmentant et diminuant la rigidité de la membrane. Nous avons remplacé le DOPC par le DPPC et le DLiPC (respectivement lipides saturés et polyinsaturés) et DOPE (chaînes d'acyle identique au DOPC mais groupe de tête différent). Nous avons procédé par étapes afin d'éviter les problèmes d'instabilité. En ce qui concerne les échelles de temps, l'analyse des simulations préliminaires du système DOPC (10 microsecondes), indique qu'un échantillonnage plus long est nécessaire pour obtenir un ensemble proche de l'équilibre. Nous estimons donc que 30  $\mu$ s de simulations devraient permettre un échantillonnage suffisamment proche de l'équilibre.

Le tableau suivant rapporte la composition phospholipidique dans chacune des 8 simulations réalisées. La taille des systèmes est d'environ 2M particules (correspondant à approximativement 23M atomes).

**Tableau 1. :** Liste des simulations effectuées, avec la composition lipidique pour chacune d'elles.

Lipid	S1	S2	S3	S4	S5	S6	S7	S8
DOPC	100	80	80	80	60	60	60	40
DPPC	0	20	0	0	40	0	0	0
DLiPC	0	0	20	0	0	40	0	60
DOPE	0	0	0	20	0	0	40	0
Simulation time (ms)	30	30	15	23	30	40	30	30

Le problème suivant est l'analyse quantitative de la forme des gouttelettes lipidiques. Pour cela, nous avons conçu un algorithme fonctionnant sur une grille, calculant la distance moyenne entre les groupes de tête lipidique dans chaque point de la grille. Implémentée dans l'outil GROMACS, l'algorithme permet également de calculer l'épaisseur de la membrane en fonction de la distance du centre de masse de n'importe quel groupe. Des exemples des résultats obtenus sont présentés dans la figure 4. Un manuscrit est en préparation, nous empêchant de présenter ici des résultats plus détaillés. ■



**Figure 4. :** (Panneau de gauche) Distance moyenne entre les groupes phosphate dans les deux faces de la membrane. (Panneau de droit) La forme moyenne de la membrane, y compris la lentille d'huile.

### Bourgeonnement des gouttelettes

Théoriquement une induction d'une asymétrie dans la tension superficielle entre les lamelles de membrane déterminerait le bourgeonnement de la LD vers la monocouche de moindre tension. Nous avons simulé le processus expérimental en générant des GL naissantes avec une asymétrie de tension. Nous avons effectivement observé qu'un déséquilibre de tension

entraînait le bourgeonnement de la goutte. Nos résultats préliminaires suggèrent que le bourgeonnement ne nécessite pas forcément des manteaux protéiques ; des protéines spécifiques pourraient tout simplement moduler les tensions de monocouches. Un manuscrit est en préparation, nous empêchant de présenter ici des résultats plus détaillés.

## 7 Remerciements

---

LM remercie Guillaume Launay (Molecular Microbiology and Structural Biochemistry, CNRS/Université de Lyon) pour sa lecture attentive du manuscrit. ■

## 8 Références

---

■ 1 Monticelli, L. et al. *J. Chem. Theory Comput.* 4, 819-834 (2008).

■ 2 Thiam, A. R. and Forêt, L. *BBA - Molecular and Cell Biology of Lipids* 1861, 715-722 (2016).

■ 3 Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., and de Vries, A. H. *J. Phys. Chem. B* 111, 7812-7824 (2007).

■ 4 Marrink, S. J. and Tieleman, D. P. *Chem. Soc. Rev.* 42, 6801-6822 (2013).

■ 5 Abraham, M. J. et al. *SoftwareX* 1-2, 19-25 (2015).

■ 6 Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. *J. Chem. Theory Comput.* 4, 435-447 (2008).

## FIMH, UNE PROTÉINE CATCH-BOND QUI FONCTIONNE SOUS FLUX DE CISAILLEMENT

Une modélisation multi-échelle

**Auteur**

Fabio STERPONE  
(LBT, CNRS)



### Contributeurs

Philippe DERREUMAUX (LBT, CNRS)

Simone MELCHIONNA (ISC, CNR)

Nicole AUDIFFREN (CINES)

## SOMMAIRE

<b>1</b> Contexte Scientifique et le grand défi.....	60	<b>4</b> Le code Muphy.....	62
<b>2</b> La méthode.....	61	<b>5</b> Résultats.....	63
<b>3</b> Le modèle.....	61	<b>6</b> Bibliographie.....	64

## 1 Contexte Scientifique et le grand défi

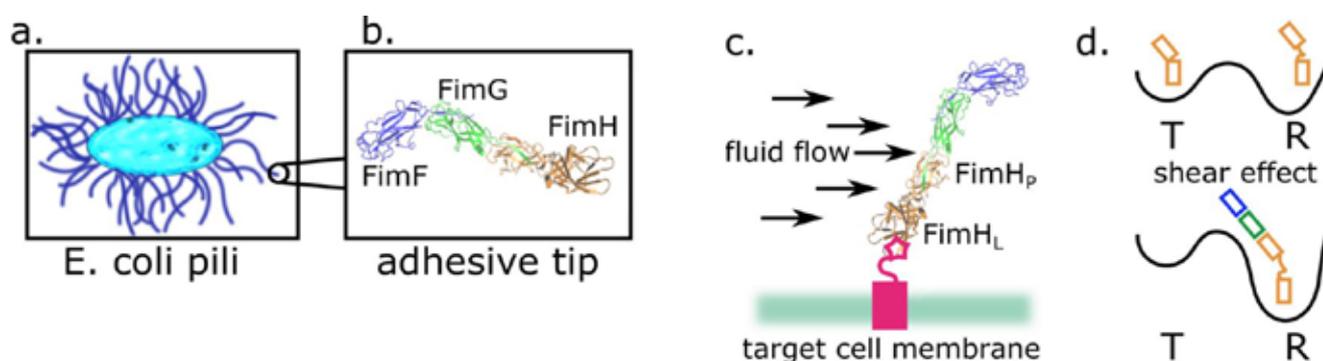
De nombreux phénomènes biophysiques, importants, impliquent la réponse des protéines au stress mécanique généré par le fluide environnant, les forces résultantes étant appelées forces de « tension ». Par exemple, dans le système vasculaire ou dans les voies urinaires, des protéines dites 'catch-bonds' permettent l'adhésion cellulaire, et sont activées par le flux de cisaillement qui se génère physiologiquement<sup>1</sup>. De même, les blessures des vaisseaux sanguins, altérant le flux du sang, activent la réponse du *facteur von Willebrand* (vWF), une protéine multimérique, qui favorise l'ancrage des plaquettes aux endroits touchés<sup>2</sup>.

Dans ce « Grand Défi », nous avons étudié une protéine de type « catch-bond » impliquée dans les infections des voies urinaires en favorisant l'adhésion des cellules bactériennes (*E. coli*) à l'hôte. Les cellules *E. coli* sont attachées aux cellules hôtes par de longs poils (pili). Ces longs poils sont composés de plus de 3 000 copies de la protéine FimA et leurs extrémités sont composées de trois unités nommées, FimF, FimG et FimH. La protéine terminale FimH, constituée de deux domaines appelés FimHL et FimHP, se lie au récepteur situé sur la cellule hôte, ce qui lui permet d'ancrer *E. coli*. Il a été prouvé que la liaison avec le récepteur était plus forte sous l'action d'une force de tension. Il a été proposé que l'allongement de la protéine multi-domaine, causé par le stress généré par le flux du fluide, réduit les interactions compétitives inter-domaines, donc rend plus forte la liaison avec le récepteur<sup>3,4,5</sup>. Une représentation schématique du système et de son *modus operandi* est donnée dans la Figure 1.

Nous avons effectué des simulations moléculaires basées sur la technique Lattice Boltzmann Molecular Dynamics (LBMD)<sup>6</sup> afin d'étudier l'effet du flux sur l'étirement conformationnel du complexe formé par les trois protéines FimF, FimG et FimH, qui constituent l'extrémité adhésive du pilus de la bactérie (voir la Figure 1, partie a).

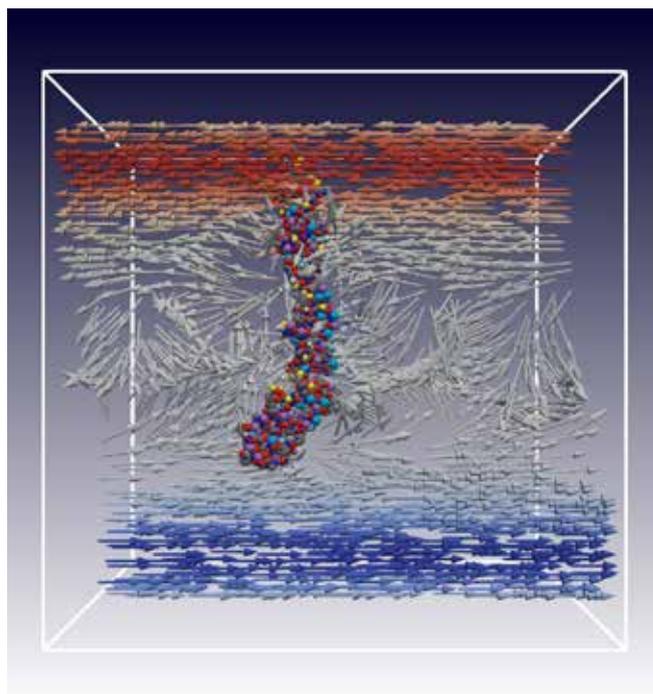
Les simulations cherchent à modéliser l'effet des forces hydrodynamiques qui agissent sur le système dans les conditions *in vivo*. Les trois sous-unités protéiques sont de tailles moyennes et composées respectivement de 154, 144, 279 acides aminés. Chaque sous-unité a été modélisée par le champ de forces gros-grain OPEP<sup>7,8</sup>. Dans cette étude, nous nous sommes seulement intéressés à l'orientation du complexe par rapport au flux externe. Nous avons donc restreint les mouvements internes des protéines par un réseau élastique et les trois sous-unités interagissent entre elles par des interactions standard non liées du champ de forces OPEP. Cette stratégie a permis d'augmenter le temps d'intégration pour la dynamique moléculaire,  $\tau = 10$  fs<sup>9</sup>, tout en conservant la flexibilité conformationnelle relative des sous-unités. Le système a été simulé dans une grande boîte de dimensions  $L_x = L_y = L_z = 240$  Å. La dynamique du fluide a été intégrée de manière synchrone avec la dynamique moléculaire avec un paramètre de couplage (frottement)  $\gamma = 0.1$  fs<sup>-1</sup>, une valeur pour la viscosité cinématique égale à celle de l'eau pure, et en utilisant une grille régulière avec un espacement  $dx = dy = dz = 3$  Å<sup>6,10</sup>.

Le système a été simulé à la fois en absence et en présence de flux de fluide. Pour ce dernier scénario, nous avons considéré deux types de perturbations d'écoulement, le flux de cisaillement (Couette) et le flux de Poiseuille, et diverses amplitudes du stress généré. Afin d'imiter l'effet du flux du fluide sur la fixation, nous avons lié le site de liaison du domaine adhésif FimHL au bord de la boîte de simulation représentant ici la membrane cellulaire de la cible. La fixation est réalisée en gelant le mouvement du premier résidu de la sous-unité FimH. Chaque simulation a été étendue jusqu'à 1  $\mu$ s dans le cas du flux de cisaillement et 0,5  $\mu$ s pour le flux de Poiseuille. Chaque simulation a été allouée sur 512 coeurs.



**Figure 1.** : Les pili de *E. coli* (panneau a) sont essentiels pour l'adhésion bactérienne via l'interaction pointe-récepteur. La pointe d'un pilus est formée par une série de domaines Fim (panneau b). L'interaction avec le récepteur de la membrane se produit sous un écoulement de fluide physiologique tel que représenté dans le panneau c, dont l'intensité favorise fortement la conformation à haute affinité (ou état décontracté, Relaxed, dans la terminologie de la description allostérique) contre la conformation à faible affinité (état « Tensed ») telle que représentée en panneau d.

Il convient de noter que, si nous transposons notre modèle à une représentation atomique en solvant explicite, notre système correspondrait à environ 1,3 million de particules. Un système de cette taille rend nécessaire une approche multi-échelle telle que celle encodée dans la technique LB (Lattice Boltzmann) couplée à un modèle gros-grain en solvant implicite pour les protéines. Outre l'infaisabilité d'une approche atomistique, il faut souligner aussi que la procédure hors équilibre pour générer les flux du fluide est naturelle dans la technique LBMD, une caractéristique qui en fait un outil approprié pour modéliser ces types de perturbations sur les macromolécules en solution. Une image du système est donnée en Figure 2 où les atomes de la composante moléculaire sont représentés par des sphères, et les vitesses du fluide par de flèches. D'autres groupes de recherche ont traité ce type de problème par exemple en essayant de quantifier le taux de cisaillement minimal provoquant des changements de conformation fonctionnels dans les protéines multi-domaines vWf<sup>11</sup>, ou le déploiement de protéines<sup>12</sup>. Toutes ces études étaient basées sur une représentation simplifiée du composant moléculaire et utilisaient la dynamique brownienne, cette dernière étant inefficace d'un point de vue de la scalabilité pour de grands systèmes. ■



**Figure 2.** : Représentation du système moléculaire dans la boîte de simulation et des lignes du champ de vitesses générées par le flux de cisaillement. Le gradient de couleur représente le gradient des vitesses du fluide le long l'axe X.

## 2 La méthode

Dans LBMD<sup>6</sup>, un modèle moléculaire à solvant implicite est couplé à la technique Lattice Boltzmann afin d'inclure des interactions hydrodynamiques dans la dynamique moléculaire. L'action du fluide sur la dynamique d'une particule est modélisée selon le formalisme de Stokes:

$$F_i^D = -\gamma (v_i - \tilde{u}_i)$$

où  $v_i$  est la vitesse de la particule  $i$ ,  $\tilde{u}_i$  est l'extrapolation de la vitesse de réseau du fluide sur la particule

$i$ , et  $\gamma$  est la friction, un paramètre réglable dans la méthodologie. La force de traction s'ajoute aux forces conservatrices habituelles, dérivées du Hamiltonien du système,  $F_i^C$ , et un terme aléatoire utilisé pour imiter le bruit blanc dans la dynamique type Langevin,  $F_i^R$ . En respectant le principe d'action/réaction, le mouvement d'une particule affecte l'évolution temporaire du fluide par un terme de perturbation. La dynamique LB est résolue numériquement à l'aide de l'opérateur collision BGK (Bhatnagar-Gross-Krook)<sup>13</sup> en ajoutant la perturbation moléculaire. ■

## 3 Le modèle

OPEP<sup>14</sup> est un modèle gros-grain en solvant implicite optimisé afin de d'échantillonner la stabilité, la flexibilité, le repliement et l'agrégation de protéines solubles. Dans le modèle OPEP, le squelette des acide-aminés est décrit à un niveau atomique tandis que chaque chaîne latérale est décrite par un seul centre de force. L'Hamiltonien est composé de deux classes d'interaction, les termes liés et les termes non-liés. Dans la première classe, on trouve des potentiels standard pour décrire les liaisons, les angles, et les angles dièdres. Des potentiels ad hoc sont introduits afin de échantillonner correctement les diagrammes de Ramachandran et donc pour reproduire

les structures secondaires ( $\alpha$ -hélices, feuillets  $\beta$ ). Dans la deuxième classe, on trouve des potentiels entre particules non liées, Lennard-Jones et des termes à deux et quatre corps pour la formation des liaisons hydrogène au sein du squelette, et des termes plus complexes pour les ponts ioniques entre les chaînes latérales des acide-aminés chargés. Une présentation des applications du modèle OPEP est fournie dans un travail récemment paru en littérature<sup>7</sup>. Contrairement à d'autres modèles qui contraignent la structure secondaire de la protéine (par exemple Martini), la résolution détaillée du squelette avec OPEP permet de reproduire la stabilité et la flexibilité

conformationnelle de la protéine sans utiliser de biais externes vers l'état natif. Cela impacte le temps de calcul en raison des petites valeurs du temps d'intégration utilisé (comparable aux simulations MD atomistique). Cependant, des astuces comme la redistribution de masse le long du squelette permet d'augmenter le pas

de temps d'intégration à 4 fs. De plus, quand le détail de la flexibilité interne n'est pas important, OPEP peut être transformé en un modèle de réseau élastique qui contraint la structure de la protéine, permettant ainsi d'atteindre un temps d'intégration de 10 fs. ■

## 4 Le code Muphy

Les simulations ont été effectuées avec le logiciel MUPHY<sup>15</sup> développé par le groupe de S. Melchionna (CNR, Rome) et en partie par le groupe de F. Sterpone. MUPHY est un code conçu pour les simulations multi-échelles qui intègre la technique du LB avec la dynamique de particules afin de modéliser de très nombreux systèmes bio-fluidiques, tels que des solutions de protéines et d'ADN, les fluides de globules rouges, etc. La méthode LB est particulièrement adaptée pour gérer des géométries complexes comme par exemple la structure des vaisseaux sanguins et la diffusion des globules rouges, car l'utilisation d'une grille cartésienne, régulière, facilite la gestion de configurations irrégulières.

MUPHY emploie un modèle de programmation hybride et a été conçu pour exploiter des ressources hardware de pointe avec des communications rapides entre processeurs, autant que des superordinateurs équipés de GPU. Le logiciel maintient des performances hautement compétitives, malgré la conception générale sous-jacente des composants logiciels. Du point de vue du parallélisme, pour la partie LB, MUPHY offre des performances quasi-idéales en fonction du nombre de cœurs utilisés. Dans le concours Gordon Bell de 2013, MUPHY a atteint les 20 PFlops de performance soutenue tout en simulant une suspension de 18 000 protéines (70 millions de particules) sur le supercalculateur de Titan. Afin d'avoir une communication locale entre les parties de fluide et de dynamique moléculaire, le code exploite la stratégie dite de décomposition de domaines, où les domaines de la MD et les domaines LB coïncident. De cette façon, chaque tâche parallèle effectuée à la fois les calculs LB et MD, et les interactions entre les particules avec le fluide sont quasi-locales. Le maillage LB sous-jacent sert à identifier les particules qui appartiennent au domaine via un test d'adhésion: une particule avec la position  $R$  appartient au domaine si le vecteur des nombres entiers les plus proches coïncide avec un point de maillage du domaine.

Le transfert d'information entre la grille et les particules nécessite l'échange des vitesses des particules à la frontière entre domaines, de manière similaire à ce qui est fait pour le calcul des forces entre un couple de particules qui appartiennent à des domaines adjacents et qui nécessite l'échange de positions. Dans le cas d'un couple de particules, une fois les positions échangées entre les domaines adjacents, chaque cœur calcule l'interaction mais accumule la force sur les particules qui lui appartiennent seulement. Quant au calcul du transfert des vitesses des particules vers le fluide, une fois effectué l'échange de vitesses des particules qui résident dans le domaine adjacent, chaque cœur calcule le terme de couplage qui modifie la dynamique du fluide, et transfère par communication parallèle la force  $F_i^p$  qui agit sur la particule  $i$  localisée dans le domaine adjacent.

Cette approche est particulièrement efficace par rapport à d'autres stratégies. Ainsi en échangeant les points de maillage situés à l'intérieur des cellules frontières à la place des particules, on pourrait réduire le nombre de données échangées à une seule étape de communication. Toutefois, une telle alternative imposerait une grande surcharge de communication car tous les noeuds de maillage situés à proximité des interfaces entre les domaines devraient être échangés.

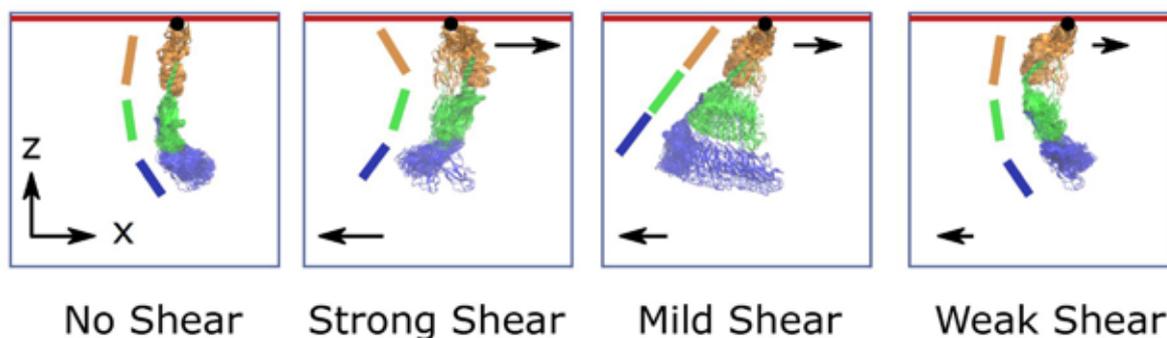
Enfin, en ce qui concerne la scalabilité, la composante de DM en gros grain est problématique car les particules ne remplissent pas nécessairement l'espace de manière uniforme, et peut se poser le problème de charge non équilibrée (load-unbalance). Récemment, nous avons implémenté la méthode dite du point moyen (Midpoint) pour calculer les voisins en interaction permettant ainsi de réduire la charge de transfert d'environ un ordre de grandeur. Des tests sur des systèmes complexes sont en cours. ■

## 5 Résultats

Nous discutons d'abord l'effet du flux de Couette. Dans nos simulations, nous avons considéré une gamme de valeurs pour le taux de cisaillement  $\dot{\gamma} \in [10^6, 10^9] \text{ s}^{-1}$ . Pour des valeurs trop élevées,  $\dot{\gamma} \geq 10^9 \text{ s}^{-1}$ , les trois sous-unités FimF, FimG et FimH se désassemblent assez rapidement, la séparation se produisant dans l'échelle de nanosecondes. Au contraire pour  $\dot{\gamma} \leq 10^8 \text{ s}^{-1}$ , les unités restent assemblées et répondent au gradient de vitesse du fluide. La réponse dépend de l'amplitude du stress. Pour la plus haute valeur tolérable  $\dot{\gamma} = 10^8 \text{ s}^{-1}$ , les unités

FimH et FimF, extrémités de la protéine, se courbent suite à l'action du gradient de vitesse du fluide, alors que pour des valeurs intermédiaires,  $\dot{\gamma} \sim 10^7 \text{ s}^{-1}$ , les trois sous-unités s'alignent le long du gradient de vitesse.

Ce régime semble correspondre à ce qui est attendu, la force de traction écartant les deux domaines de l'unité FimH, FimHL et FimHP. Fait intéressant, pour le taux de cisaillement échantillonné le plus bas,  $\dot{\gamma} \sim 10^6 \text{ s}^{-1}$ , l'orientation des sous unités dans la cellule de simulation

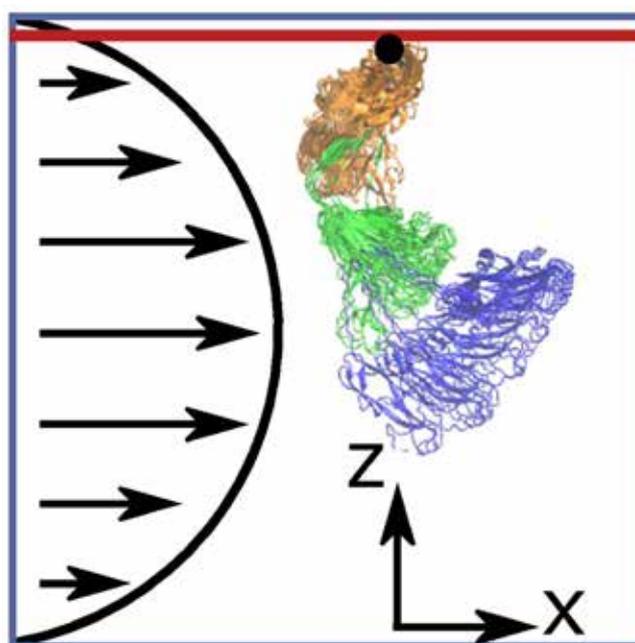


**Figure 3.** : Conformations échantillonnées le long de la trajectoire de  $1 \mu\text{s}$ . Le panneau de gauche se réfère à la simulation en l'absence de flux de fluide, puis les trajectoires représentatives aux taux de cisaillement élevés, moyens et faibles, voir le texte. Le codage des couleurs pour les domaines protéiques est comme dans la Figure 1, l'ancrage du domaine adhésif FimHL à la paroi de la membrane est représenté par un cercle noir.

ne suit pas le gradient de vitesse, semble insensible au flux de fluide et les conformations acquises ressemblent à celles échantillonnées en l'absence de flux. Un aperçu des résultats est représenté dans le panneau central de la Figure 3 où l'évolution de la conformation protéique pendant la trajectoire d'une microseconde est représentée par des états représentatifs successifs.

Nous avons aussi exploré l'effet du flux dit de Poiseuille. Encore une fois, nous avons effectué une analyse du débit associé et mesuré comme la dérivée du profil de la vitesse du fluide à la paroi de la boîte de simulation,  $Z = L_z$ . Pour la valeur de  $\geq 10^6 \text{ s}^{-1}$ , les trois sous-unités se désassemblent très rapidement, comme pour les taux de cisaillement élevés avec le flux de Couette. À la valeur maximale tolérable,  $\sim 10^5 \text{ s}^{-1}$ , la protéine se courbe selon le profil parabolique de la vitesse, et l'unité FimF subit la traction maximale de la vitesse du fluide, voir la Figure 4. Ce résultat montre encore que, sous l'écoulement du fluide, la conformation des trois sous-unités est considérablement modifiée et que la chaîne s'étend selon la trajectoire de la vitesse du fluide.

En conclusion, nous avons démontré que la machine LBMD est une méthode puissante pour modéliser des processus biologiques hors équilibre où la perturbation provient du flux de fluide. Nos simulations ont montré que lorsque l'on considère les valeurs appropriées du stress de cisaillement ou de la vitesse longitudinale du



**Figure 4.** : Les conformations du complexe ont été échantillonnées le long de la trajectoire de  $0,5 \mu\text{s}$  avec la protéine soumise au flux de Poiseuille,  $= 10^5 \text{ s}^{-1}$ . Le codage des couleurs pour les domaines protéiques est comme dans la Figure 1, l'ancrage du domaine adhésif FimHL à la paroi de la membrane est représenté par un cercle noir.

flux, l'assemblage des trois sous-unités Fim s'allonge sous l'effet de la force de traction. Selon les hypothèses présentées dans la littérature, sous l'effet de cet étirement, le site de fixation avec le récepteur cible est renforcé à cause des changements conformationnels apparaissant au niveau des domaines FimH. Pour explorer en détail cet effet, nous allons poursuivre notre investigation. À savoir, une représentation à plusieurs niveaux de la protéine devrait être construite, avec les deux domaines de FimH modélisés par le champ de force OPEP flexible et en réservant le réseau élastique uniquement aux unités supplémentaires, FimF et FimG. De cette façon, nous pourrions explorer la réorganisation inter-domaine entre FimHL et FimHP. De plus, en raison de l'organisation complexe des trois sous-unités, les directions de cisaillement et de flux pourront également être générées selon les directions Y, ou X-Y, afin d'explorer l'impact de la direction du flux sur l'étirement inter-domaine. Finalement, afin de rendre plus performant l'étude de

notre système, et autres systèmes similaires, nous avons développé une approche multi-échelle pour la résolution de la partie LB. En effet la taille de la maille de la grille cartésienne sur laquelle la dynamique du fluide est intégrée peut être variée en fonction de la distance du système moléculaire. Autour de la protéine la résolution doit être suffisamment précise afin de garantir le couplage correct avec la dynamique moléculaire. Au contraire à plus grande distance la résolution peut être inférieure, en permettant d'utiliser donc un nombre inférieur de points grille, et réduire le temps de calcul. Le maillage de l'espace avec deux ou plusieurs grilles peut être statique, avec une zone définie à haute résolution et une autre à basse résolution, ou peut être dynamique avec la zone à haute résolution redéfinie au vol en fonction des changements des molécules. Cette dernière approche, plus élaborée, permettra de réduire le volume de la grille à haute résolution et donc réduire de manière plus efficace le temps de calcul. ■

## 6 Bibliographie

1. Thomas, W. E., Vogel, V. & Sokurenko, E. Biophysics of Catch Bonds. *Annu. Rev. Biophys.* 37, 399–416 (2008).
2. Springer, T. A. von Willebrand factor, Jedi knight of the bloodstream. *Blood* 124, 1412 (2014).
3. Yakovenko, O., Tchesnokova, V., Sokurenko, E. V & Thomas, W. E. Inactive conformation enhances binding function in physiological conditions. *Proc. Natl. Acad. Sci. {USA}* 112, 9884–9889 (2015).
4. Sauer, M. M. et al. Catch-bond mechanism of the bacterial adhesin FimH. *Nat. Comm.* 7, 10738 (2015).
5. Kalas, V. et al. Evolutionary fine-tuning of conformational ensembles in FimH during host-pathogen interactions. *Sci. Adv.* 3, e1601944 (2017).
6. Sterpone, F., Derreumaux, P. & Melchionna, S. Protein simulations in fluids: coupling the OPEP coarse-grained force field with hydrodynamics. *J. Chem. Theory Comput.* 11, 1843–1853 (2015).
7. Sterpone, F. et al. The OPEP protein model: from single molecules, amyloid formation, crowding and hydrodynamics to DNA/RNA systems. *Chem. Soc. Rev.* 43, 4871–4893 (2014).
8. Sterpone, F., Nguyen, P., Kalimeri, M. & Derreumaux, P. Importance of the ion-pair interactions in the OPEP coarse-grained force field: parametrization and validation. *J. Chem. Theory Comput.* 9, 4574–4584 (2013).
9. M. Chiricotto, P. Derreumaux, F. Sterpone, S. M. Multiscale simulation of molecular processes in cellular environments. *Philos Trans R Soc L. A* 374, 20160225 (2016).
10. Chiricotto, M., Melchionna, S., Derreumaux, P. & Sterpone, F. Hydrodynamics effects on  $\beta$ -amyloid (16–22) peptide aggregation. *J. Chem. Phys.* 145, 035102 (2016).
11. Sing, C. E. & Alexander-Katz, A. Elongational Flow Induces the Unfolding of von Willebrand Factor at Physiological Flow Rates. *Biophys. J.* 98, L35–L37 (2010).
12. Szymczak, P. & Cieplak, M. Proteins in a shear flow. *J. Chem. Phys.* 127, 155106 (2007).
13. Succi, S. *The Lattice Boltzmann Equation for Fluid Dynamics and Beyond*. (Clarendon Press, 2001).
14. Chebaro, Y., Pasquali, S. & Derreumaux, P. The coarse-grained OPEP force field for non-amyloid and amyloid proteins. *J. Phys. Chem. B* 116, 8741–8752 (2012).
15. Bernaschi, M. et al. MUPHY: A parallel MULTi PHYsics/scale code for high performance bio-fluidic simulations. *Comput. Phys. Comm.* 180, 1495–1502 (2009).

## TERA-MESH

### Comité technique : CT6 : Informatique, algorithmique et mathématiques

Auteur



Hugues DIGONNET  
(École Centrale de Nantes / Institut de Calcul Intensif)

### Collaborateurs

Louis DOUTEAU, Luisa SILVA  
(École Centrale de Nantes / Institut de Calcul Intensif)

Mathieu CLOIREC (CINES)

## SOMMAIRE

<b>1</b>	<b>Contexte</b> .....	<b>66</b>	<b>4</b>	<b>Résultats</b> .....	<b>68</b>
	Calculer plus vite et/ou plus gros .....	66		Comparaison entre les partitions haswell et broadwell .....	68
	Combiner les accélérations plutôt que les opposer .....	67		Scalabilité jusqu'à 700 nœuds 800 cœurs d'Occigen .....	16 69
<b>2</b>	<b>Objectif du grand défi</b> .....	<b>67</b>		Exécution sur la quasi-totalité d'Occigen 2017 .....	70
<b>3</b>	<b>Code de calcul</b> .....	<b>68</b>	<b>5</b>	<b>Perspectives</b> .....	<b>70</b>
	Portage.....	68	<b>6</b>	<b>Remerciements</b> .....	<b>71</b>
	Exécution.....	68	<b>7</b>	<b>Références</b> .....	<b>71</b>

## 1 Contexte

Le numérique est aujourd'hui présent dans pratiquement tous les domaines, et plus particulièrement dans le domaine scientifique. Dans le cadre, la simulation numérique reste actuellement un des plus gros consommateurs en puissance de calcul et est toujours limitée par les capacités en celle-ci pour obtenir des calculs de plus en plus précis. Ceux-ci ont pour objectifs l'enrichissement des modèles physiques ou une meilleure discrétisation spatiale et temporelle pour bien capturer des phénomènes très locaux dans des vastes domaines de calcul.

Les supercalculateurs permettent l'accès à une puissance de calcul considérable : supérieur de plusieurs ordres de grandeur à nos ordinateurs personnels, mille à un

million de fois plus performant. Cette puissance de calcul peut aussi être exploitée comme une « puissance temporelle », puisqu'elle permet de réaliser des calculs sur les supercalculateurs plusieurs années avant qu'ils ne soient possibles sur un ordinateur ou calculateur de laboratoire, ce qui amène un avantage considérable d'un point de vue de la recherche.

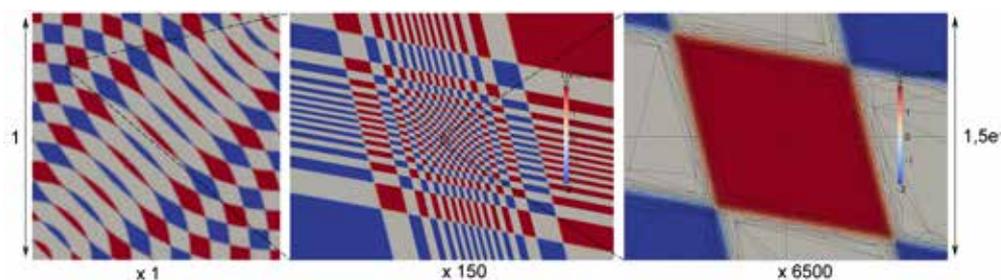
Si l'utilisation du calcul parallèle est aujourd'hui un outil indispensable dans la réduction des temps de calcul et pour la possibilité de réaliser en un temps donné des calculs toujours plus importants et précis, elle n'est qu'une optimisation parmi d'autres pour permettre de calculer plus vite.

### Calculer plus vite et/ou plus gros

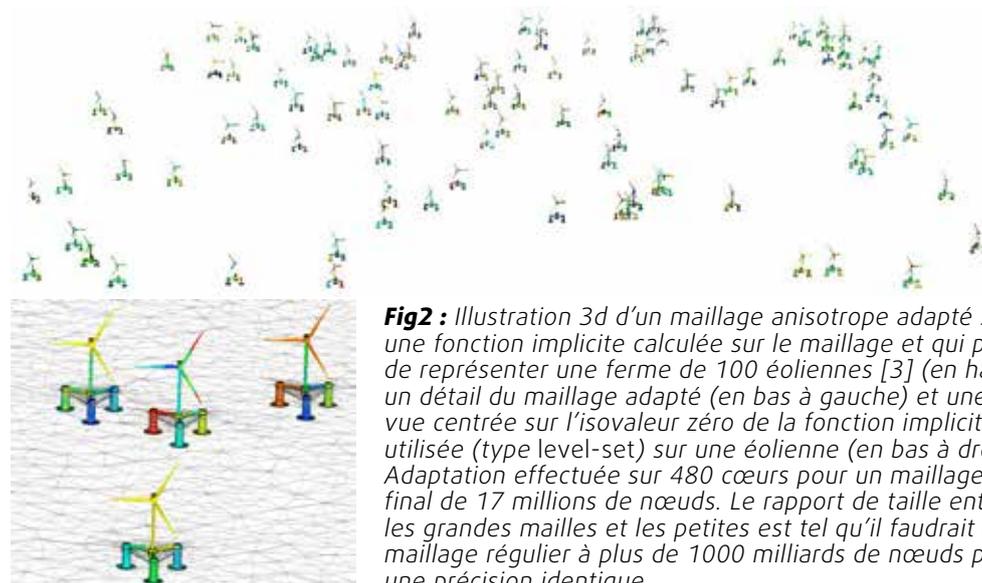
Pour accélérer la résolution d'un problème de taille  $N$ , plusieurs optimisations peuvent être envisagées. Entre autres, nous pouvons citer :

- le calcul parallèle, avec une augmentation du nombre de Flops/s ;
- l'optimisation algorithmique, pour réduire la complexité algorithmique (ex : la méthode multigrille) en diminuant le nombre de Flops nécessaire à la résolution d'un problème de taille donnée  $N$  ;
- l'adaptation de maillage, qui permet de réduire la taille du problème discrétisé en concentrant les mailles là où

elle sont réellement nécessaires. Pour une précision donnée, il y aura alors une diminution de la taille du problème à traiter, avec la nouvelle taille  $M \ll N$ . L'apport de l'adaptation de maillage est très importante, car elle permet de diminuer d'un facteur très important la taille des maillages utilisés, pour une précision identique, comme illustrent les figures suivantes, Fig1 (en 2d) et Fig2 (en 3d). Cette réduction importante de la taille des maillages utilisés pour discrétiser le problème à résoudre entraîne, de facto, une réduction importante des temps de simulation.



**Fig1 :** Illustration 2d d'un maillage anisotrope adapté sur une fonction test analytique [1], avec différentes amplifications. Le maillage adapté contient uniquement 23 millions de nœuds, alors qu'une grille régulière devrait en contenir 1000 milliards pour avoir la même précision.



**Fig2 :** Illustration 3d d'un maillage anisotrope adapté sur une fonction implicite calculée sur le maillage et qui permet de représenter une ferme de 100 éoliennes [3] (en haut), un détail du maillage adapté (en bas à gauche) et une vue centrée sur l'isovaleur zéro de la fonction implicite utilisée (type level-set) sur une éolienne (en bas à droite). Adaptation effectuée sur 480 cœurs pour un maillage final de 17 millions de nœuds. Le rapport de taille entre les grandes mailles et les petites est tel qu'il faudrait un maillage régulier à plus de 1000 milliards de nœuds pour une précision identique.



## Combiner les accélérations plutôt que les opposer

Si l'adaptation de maillage permet de réduire fortement les temps de calcul, il ne faut cependant pas le faire au détriment d'autres optimisations et, a fortiori, de la possibilité d'effectuer les calculs sur des supercalculateurs parallèles. Si l'adaptation de maillage ne constitue la partie la plus consommatrice en temps de calcul (celle-ci reste la résolution des systèmes linéaires) et n'est pas présente dans tous les codes de calcul, elle représente un temps non négligeable lorsqu'elle doit être effectuée (de l'ordre de 25 % du temps de calcul total). Dans ce contexte de calcul massivement parallèle, il est donc indispensable de la paralléliser et de pouvoir l'exécuter sur un grand nombre de cœurs et ainsi de pouvoir cumuler

les optimisations liées à l'utilisation de l'adaptation de maillage et au calcul massivement parallèle.

La parallélisation du maillage se fait par un couplage itératif entre un maillage séquentiel, IciMesh [1-2], maillage topologique non structuré et non hiérarchique qui permet également de remailler sous la contrainte de bloquer des interfaces, avec un repartitionnement dynamique, IciRepart, qui permet d'effectuer l'équilibrage de charges mais aussi, dans le cas de la parallélisation du maillage, le déplacement des interfaces, comme l'illustre dans la Fig 3. Cette stratégie de parallélisation a l'avantage de ne pas être trop intrusive : le maillage IciMesh reste séquentiel, mais est utilisé dans un contexte parallèle. ■

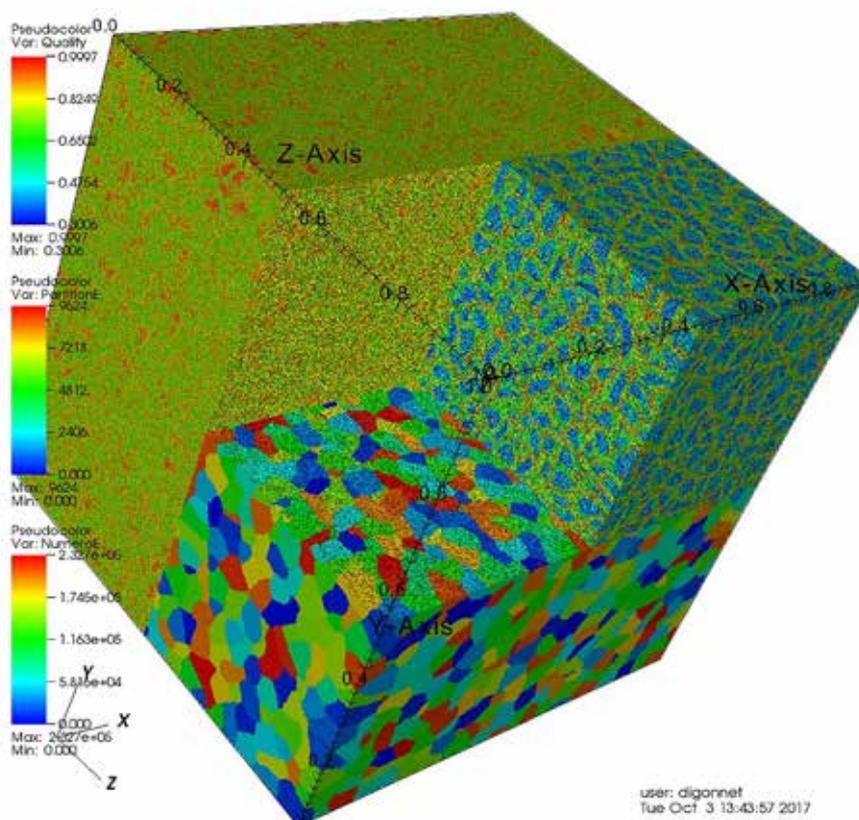
## 2 Objectif du grand défi

L'objectif de ce grand défi TeraMesh est de tester le passage à l'échelle de l'adaptation de maillage sur les supercalculateurs nationaux à plusieurs dizaines de milliers de cœurs. Au vu des capacités de calcul, mais également de la mémoire disponible sur Occigen, ce projet a pour ambition de pouvoir générer des maillages à 1000 milliards d'éléments en 2d et en 3d en utilisant environ 80 000 cœurs. Ceci nécessite l'utilisation de la quasi-totalité du calculateur, que ce soit en termes de cœurs mais également en termes de mémoire vive. À cette échelle, une bonne gestion des entrées/sorties est

également primordiale. En effet, lors de la sauvegarde du maillage, il y a une écriture simultanée par l'ensemble des cœurs d'environ 30 To de données.

Le passage de calcul à cette échelle reste un challenge important d'un point de vue hardware, avec une utilisation très intensive de tout les composants (CPU, RAM ou IO). Également, d'un point de vue de l'application, l'ensemble de la chaîne de calcul doit être parallélisé et l'équilibrage des charges et la minimisation des échanges entre les cœurs prennent une importance considérable lorsque l'on veut utiliser l'ensemble des ressources disponibles. ■

DB: res\_Bin.pvtu



**Fig3 :** Image d'un maillage tridimensionnel, non structuré et non hiérarchique, d'environ 400 millions de nœuds pour 2,25 milliards d'éléments générés sur 9 625 cœurs avec une qualité normalisée minimale de 0,3 pour une moyenne de 0,83 (qualité normalisée entre 0 et 1 : 0 pour un tétraèdre dégénéré et 1 pour un tétraèdre régulier de la taille demandée).

### 3 Code de calcul

Le code de calcul utilisé est IciMesh. C'est un code dédié à l'adaptation de maillage anisotrope pour des maillages simplexiques (triangle en 2d et tétraèdre en 3d), qui

a été couplé à un répartiteur dynamique, IciRepart, pour sa version parallèle.

#### Portage

Le portage du code n'a pas présenté de difficulté particulière, du fait de la faible dépendance du code à des bibliothèques tierces. Ils s'appuie uniquement sur la bibliothèque MPI et, facultativement, sur la bibliothèque Eigen pour optimiser les calculs sur les petites matrices. Cette relative facilité de portage est aussi due au fait

que le code avait déjà été porté sur des architectures variées, telles que : Occigen1 au CINES, Curie au TGCC et Turing (BlueGen/Q) à l'IDRIS ; à plus petite échelle sur XeonPhi KNL (via Frioul au CINES) ou sur ARM (projet européen MontBlanc).

#### Exécution

Lors de la première phase d'installation (extension Occigen2 séparée du calculateur de production Occigen), d'importants problèmes de reproductibilité des temps de calcul lors des exécutions n'ont pas permis d'effectuer correctement la campagne de dimensionnement des cas tests. Ce problème a duré assez longtemps, et à disparu lors de la phase d'intégration d'Occigen2 dans Occigen. Ceci a conduit le démarrage du projet à début 2017.

L'exécution sur la quasi-totalité d'Occigen n'a pu être tentée que deux fois : une fois (en « aveugle ») lors d'une maintenance en mars, où le calcul n'a pas pu être exécuté compte tenu d'un nombre de cœurs disponibles insuffisant (après redémarrage) ; une autre tentative a eu lieu au mois de mai, sur un nombre plus faible de cœurs disponibles, ce qui provoqué un dépassement de la capacité mémoire sur certains nœuds. ■

### 4 Résultats

Les résultats présentés montreront le passage l'échelle de notre code IciMesh sur le supercalculateur Occigen,

version étendue à 85 824 cœurs en 2017.

#### Comparaison entre les partitions haswell et broadwell

Une première analyse du comportement parallèle de notre application a permis de comparer la puissance de calcul délivré par les deux types de processeur présents : Haswell, 24 cœurs, et Broadwell, 28 cœurs (notés, respectivement, HSW et BDW par la suite).

Trois configurations ont été testées :

- a) les 24 cœurs d'un nœud HSW
- b) seulement 24 cœurs d'un nœud BDW
- c) les 28 cœurs d'un nœud BDW

Les configurations a) et b) ont permis de conserver un ratio RAM/coeur constant de 2.66Go/coeur, alors que

la dernière configuration a rendu possible l'exploitation de toute la puissance disponible sur les nœuds BDW. Le cas test consiste à effectuer une adaptation de maillage pour l'obtention d'un maillage final tridimensionnel d'environ 2.7 millions de nœuds et 15.6 millions d'éléments. Les maillages obtenus sur 24 cœurs sont strictement identiques, alors que celui généré sur 28 cœurs diffère légèrement mais reste néanmoins parfaitement correct et très proche des autres.

	Temps Calcul (s)	Temps IO (s)	Total (s)	NbN	NbE
a) 24 HSW	331.6	1.400	333.18	2 671 964	15 565 987
b) 24 BDW	317.41	2.034	319.61	2 671 964	15 565 987
c) 28 BDW	262.06	1.302	264.50	2 659 449	15 502 892

**Tab 1 :** Exécution d'IciMesh sur les deux types de nœuds présents sur Occigen : Haswell 24 cœurs et Broadwell 28 cœurs, pour un total de 3 configurations différentes.

Le tableau précédent, Tab1, montre que, pour notre application, les cœurs HSW ou BDW délivrent une puissance quasi identique, avec BDW 4% plus rapide que le HSW. Par contre, si nous considérons le nœud de calcul complet, avec les 28 cœurs du BDW, le gain devient plus significatif avec une diminution du temps de calcul de 20 %.

Remarque : dans l'avenir, il serait intéressant de pouvoir soumettre des calculs sur un environnement mixte de Haswell 24 cœurs et Broadwell 28 cœurs, pour profiter pleinement de la puissance de calcul d'Occigen.

## Scalabilité jusqu'à 700 nœuds 16 800 cœurs d'Occigen

Nous présentons ici les efficacités parallèles obtenues lors de tests de « weak speed-up » (avec une charge par cœur constante, quel que soit le nombre de cœurs utilisés) réalisés sur Occigen en utilisant de 1 à 700 nœuds de calcul (1 à 16 800 cœurs) sur la partition Haswell. L'utilisation des nœuds Haswell permet d'avoir le rapport RAM/cœur le plus important et de dimensionner la charge maximale atteignable dans le but de générer le plus gros maillage possible.

Plusieurs séries de tests ont été effectuées, avec différentes charges par cœur :

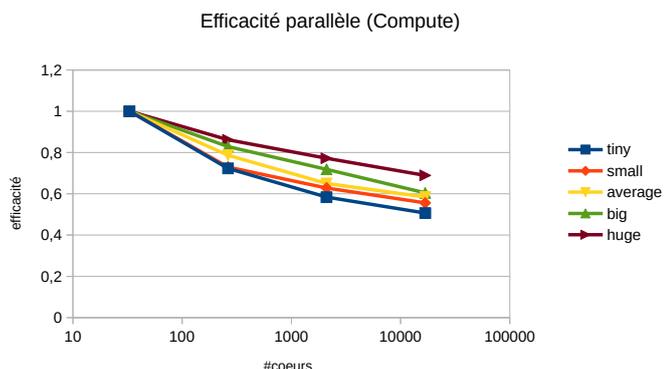
- Tiny: avec une partition de maillage par cœur d'environ 5 000 nœuds et 30 000 éléments
- Small: avec une partition de maillage par cœur d'environ 10 000 nœuds et 60 000 éléments

- Average: avec une partition de maillage par cœur d'environ 80 000 nœuds et 500 000 éléments
  - Big: avec une partition de maillage par cœur d'environ 625 000 nœuds et 3 700 000 éléments
  - Huge: avec une partition de maillage par cœur d'environ 2 230 000 nœuds et 13 000 000 éléments
- Le tableau suivant, Tab2, présente les temps de calcul nécessaires pour générer les maillages sur 33 à 16 800 cœurs. Des exécutions sur 1 et 5 cœurs ont également été faites mais la charge par cœur n'avait pas encore atteint la valeur imposée.

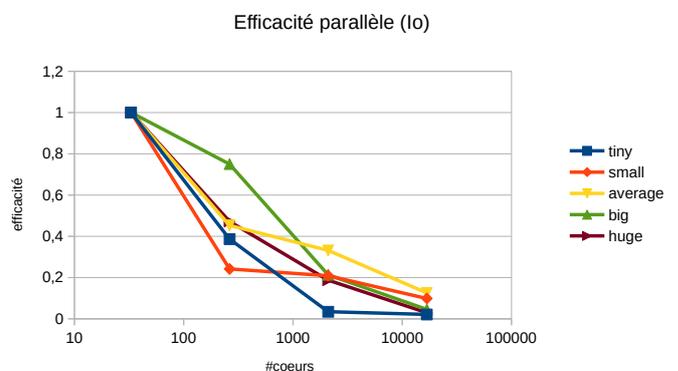
Temps (s)	Calcul / 10			
Nb cœurs	33	263	2100	16 800
tiny	29.58 / 0.017	40.90 / 0.044	50.65 / 0.50	58.38 / 0.81
small	46.15 / 0.14	63.24 / 0.58	73.42 / 0.67	83.02 / 1.42
average	265.5 / 0.53	337.2 / 1.17	408.4 / 1.60	454.2 / 4.22
big	1564 / 1.59	1886 / 2.12	2175 / 7.46	2587 / 34.7
huge	5141 / 8.92	5965 / 18.9	6656 / 47.7	7454 / 308

**Tab2 :** Temps de calcul et des 10 pour des exécutions en utilisant de 33 à 16 800 cœurs pour générer les maillages à charge constante par cœur, et ceci pour 5 tailles de maillage différentes : Tiny, Small, Average, Big et Huge.

Les figures suivantes, Fig4 et Fig5, représentent les efficacités parallèles en « weak speed-up » obtenues sur le cas test précédent à la fois pour le calcul (Compute) et pour les Entrées/Sorties (Io). On observe un assez bon comportement de la partie calcul de l'adaptation de maillage, avec une efficacité parallèle qui varie de 0.5 pour le petit cas à 0.7 pour le cas le plus gros. Pour ce qui est des entrées/sorties parallèles (principalement en écriture), les efficacités parallèles sont beaucoup moins bonnes (inférieures à 0.1), mais ce sorties représentent quand même des débits importants : pour le cas « Huge » l'écriture du maillage représente un volume de 7.3 To (constitué de 1050 fichiers de 7.1Go), écrits en 308 secondes, soit un débit d'écriture de 24Go/s, qui reste très raisonnable.



**Fig4:** Efficacité parallèle de la phase de calcul lors de l'adaptation de maillage pour les 5 catégories.



**Fig5:** Efficacité parallèle de la phase d'entrées/sorties lors de l'écriture des fichiers de maillage pour les 5 catégories.

Dans le tableau suivant, Tab3, nous présentons les caractéristiques des maillages générés sur 16 800 cœurs pour les 5 catégories. Nous remarquons que la qualité des maillages est très correcte et assez indépendante de la taille des maillages générés. Nous avons cependant observé une légère diminution de cette dernière plus le maillage est petit mais, même pour le cas Tiny, la qualité obtenue est plus qu'acceptable. Pour le cas Huge, le maillage généré comporte 37 milliards de nœuds pour 222 milliards d'éléments et a été réalisé sur 16 800 cœurs, ce qui permet d'envisager générer un maillage de 1000 milliards d'éléments en utilisant la quasi-totalité d'Occigen, avec ses environ 77 000 cœurs.

	Nb noeuds	Nb elements	Qualité Min	Qualité Moy
tiny	82 144 000	482 681 073	0.217	0.783
small	172 531 543	1 016 097 339	0.228	0.788
average	1 349 510 245	7 981 339 917	0.232	0.806
big	10 558 271 773	62 651 567 484	0.298	0.822
huge	37 473 972 094	222 714 467 648	0.283	0.830

**Tab3 :** Caractéristiques des maillages générés sur 16 800 cœurs : nombre de noeuds et d'éléments du maillage partitionné en 16 800, ainsi que la qualité normalisé minimale et moyenne

## Exécution sur la quasi-totalité d'Occigen 2017

Une exécution a été tenté sur la totalité d'Occigen en utilisant les deux partitions mais n'a malheureusement pas fonctionné dû à un dépassement de la capacité mémoire sur un nœud. Le calcul initialement prévu pour s'exécuter sur 81 250 cœurs a été lancé sur « uniquement » 77 000 cœurs ce qui a, semble-t-il, créé un léger déséquilibre qui a été fatal pour le code. Ce type de calcul reste donc à faire mais sera retenté lors de prochaine phase de maintenance d'Occigen, de façon à gêner le moins possible les autres utilisateurs.

Dans le contexte de TeraMesh, où l'objectif est de générer un maillage le plus gros possible, l'utilisation uniforme du calculateur avec 24 cœurs par nœud et une limite de 2,6Go de RAM par cœur a été choisie, mais d'autres alternatives sont intéressantes. Par exemple, maximiser l'utilisation du calculateur (28 cœurs sur BDW) et/ou prendre en compte la légère hétérogénéité de la puissance des cœurs de calcul (HSW vs BDW) dans l'application, dans le but de réduire encore le temps de calcul. ■

## 5 Perspectives

Une perspective à court terme est de réussir l'objectif principal de ce grand défi TeraMesh : exécuter des procédures d'adaptation de maillage à l'échelle du supercalculateur Occigen et générer des maillages bi et tri dimensionnels à 1000 milliards d'éléments. Ces exécutions seront faites lors des maintenances du calculateur, pour gêner le moins possible les autres utilisateurs.

À moyens terme, nous souhaitons effectuer des « benchmarks » en prenant en compte l'architecture d'Occigen avec sa double (voir triple) partition des nœuds de calcul : 24HSW/128 Go, 24 HSW/64Go et 28BRW/64Go. Ceci permet d'envisager plusieurs utilisations possibles :

- uniforme, avec 24 cœurs par nœud et une empreinte mémoire inférieure à 2.6 Go par coeur ;
- maximal, avec 24 cœurs sur HSW et 28 cœurs sur BDW mais empreinte mémoire limité à 2.3 Go par coeur ;
- rapide, pour prendre en compte dans l'application de la puissance réel des cœurs de calcul. Dans ce cadre, BDW est légèrement plus rapide que HSW et nous pouvons envisager d'équilibrer la charge de travail en conséquence. Cependant, nous remarquons que ceci va à l'encontre de la mémoire disponible, car les cœurs BDW légèrement plus rapides ont moins de mémoire par cœur que les cœurs HSW.

D'autres exécutions à l'échelle du supercalculateur sont donc envisagées pour évaluer ces différentes configurations.

Ce grand défi avait pour but de tester et valider le noyau d'adaptation de maillage à l'échelle des supercalculateurs de classe Tier1 et de voir vers quelle taille de maillage nous pourrions aller dans le futur. Cette exploration devra bien entendu être suivi d'une phase d'intégration dans une chaîne de calcul plus complexe et plus réaliste, avec la résolution de très grands systèmes linéaires ou non, en utilisant des méthodes multigrille et lors de calculs instationnaires. Il est important de préciser que l'ensemble de cette chaîne de calcul doit être parallélisée et exécutée à distance sur le calculateur : de la génération du maillage de calcul à l'exploitation visuelle des résultats via des logiciels tel que VisIt ou Paraview.

Les cellules de veille technologique de GENCI, avec Frioul et Ouessant permettent d'adapter et d'optimiser nos codes (ou nos minis applications) à ces nouvelles architectures qui seront peut-être au cœur des supercalculateurs de demain ou le nombre d'unité de calcul seront de plusieurs millions. ■

## 6 Remerciements

Nos remerciements vont à Genci et au Cines qui nous ont permis d'effectuer ces grands défis 2017 sur le nouveau calculateur Occigen. Un merci tout particulier aux équipes techniques et au support pour leur réactivité et

leur précieux conseils, ainsi que de m'avoir laissé tenter de lancer des exécutions sur la quasi-totalité d'Occigen (environ 80 000 cœurs).■

## 7 Références

■ [1] «Extreme Scaling of IciPlayer with Components: IciMesh and IciSolve » H. Digonnet - JUQUEEN Extreme Scaling Workshop 2016, 2016

■ [2] « Massively parallel anisotropic mesh adaptation » H Digonnet, T Coupez, P Laure, L Silva - International Journal of High Performance Computing Applications, SAGE Publications, 2017

■ [3] « Simulation numérique d'éolien offshore » L. Douteau, L. Silva, H. Digonnet, T. Coupez, D. Le Touzé, J.C. Gilloteaux - Congrès Français de Mécanique, 2017

# WATT - WEST ASPECT RATIO AND TURBULENT TRANSPORT

Explorer l'impact de ce facteur géométrique du tokamak WEST

Auteur



Philippe GHENDRIH (CEA/Direction de la Recherche Fondamentale/ Institut de Recherche sur la Fusion par confinement Magnétique)

## Collaborateurs

Yuuichi ASAHI (*post-doc*), Nicolas BOUZAT (*thèse*), Elisabetta CASCHERA (*thèse*)  
Guilhem DIFPRADALIER (*thèse*), Peter DONNEL (*thèse*)

Xavier GARBET, Virginie GRANDGIRARD, Guillaume LATU, Chantal PASSERON,  
Yanick SARAZIN (CEA/Direction de la Recherche Fondamentale/Institut de Recherche sur la Fusion par confinement Magnétique)

Jean-Christophe PENALVA (CINES)

## SOMMAIRE

<b>1</b>	<b>Le contexte : ITER, le chemin vers la fusion thermonucléaire par confinement magnétique</b> ...	<b>73</b>
<b>2</b>	<b>WEST et GYSELA en support à ITER</b> .....	<b>74</b>
	Objectif .....	74
<b>3</b>	<b>Code de calcul GYSELA</b> .....	<b>75</b>
<b>4</b>	<b>Préparation de GYSELA pour OCCIGEN2</b> .....	<b>75</b>
<b>5</b>	<b>GYSELA sur OCCIGEN2</b> .....	<b>76</b>
	Les simulations WATT .....	77
<b>6</b>	<b>Observations et résultats scientifiques</b> .....	<b>77</b>
<b>7</b>	<b>Conclusion remerciement</b> ..	<b>79</b>
<b>8</b>	<b>Bibliographie et références</b> .....	<b>79</b>

## 1 Le contexte : ITER, le chemin vers la fusion thermonucléaire par confinement magnétique

L'effort pour la maîtrise de l'énergie de fusion nucléaire à des fins pacifiques est devenu un sujet de recherche ouvert au lendemain du lancement de Spoutnik. Cette source d'énergie a beaucoup d'attrait mais se révèle difficile à mettre en œuvre. L'obtention d'un niveau important de réactions auto entretenues est le verrou que la communauté internationale a décidé de surmonter avec la construction d'ITER à Cadarache. ITER est un dispositif expérimental avec une configuration magnétique torique appelée tokamak. En support à ce défi scientifique et technologique, le CEA, la Fédération de Recherche pour la Fusion par Confinement Magnétique, et les partenaires européens, chinois et japonais, ont modifié le tokamak TORE SUPRA (aussi localisé à Cadarache) pour réaliser un dispositif permettant de mieux soutenir la conception et la construction d'ITER. Ce nouveau dispositif expérimental s'appelle WEST, un clin d'œil amical à nos collègues chinois qui opèrent le tokamak EAST. L'enjeu des expériences à venir sur ITER est de réaliser les conditions requises pour que la matière, dite à l'état plasma, atteigne des températures de l'ordre de 100 000 000 K. Maintenir de telles températures sur terre, sur des échelles de l'ordre du mètre, suppose de mettre en œuvre une capacité d'isolation thermique exceptionnelle. Un champ magnétique intense joue ce rôle et assure le confinement du plasma. L'organisation géométrique du champ magnétique considérée comme la plus favorable est la structure torique appelée tokamak, elle est utilisée pour ITER comme pour WEST, EAST, etc.

Dans le régime de combustion thermonucléaire visé dans ITER, les 2/3 de la puissance nécessaire au chauffage du

plasma seront issus des réactions de fusion nucléaire. La difficulté majeure est que de petites perturbations peuvent limiter la qualité du confinement et la performance lors des expériences. Le contrôle de ces effets est de mieux en mieux maîtrisé dans les dispositifs expérimentaux actuels. L'ambition d'ITER, grâce à un plus grand volume de plasma, est de mettre en œuvre ce savoir pour atteindre le régime de combustion thermonucléaire et le maintenir. Le chemin qui mène à cette réalisation est tracé par un ensemble de données expérimentales, réunies par la communauté internationale, et qui relie la performance d'un plasma aux paramètres de fonctionnement, comme la taille du tokamak. La loi d'échelle empirique qui a ainsi été établie n'a cependant pas de base théorique systématique. Pour augmenter la fiabilité de la loi d'échelle, les expériences avec une géométrie similaire à celle d'ITER ont été privilégiées. En conséquence, l'impact de la géométrie sur la performance n'est pas établi avec la même précision que pour d'autres paramètres. En particulier, le rapport entre le grand rayon et le petit rayon du tore, le paramètre appelé rapport d'aspect, est fixé aux environs de 3. En contraste, en raison de contraintes technologiques, aussi bien EAST que WEST ont un rapport d'aspect plus grand, pouvant atteindre 6. Dans ce contexte, les simulations numériques doivent permettre d'interpréter les propriétés des lois d'échelle et explorer le potentiel d'optimisation par la géométrie. Mettre en perspective les performances à venir dans WEST par rapport aux autres expériences, et donc vis-à-vis de la loi d'échelle empirique, est l'enjeu du projet de simulation WATT. ■

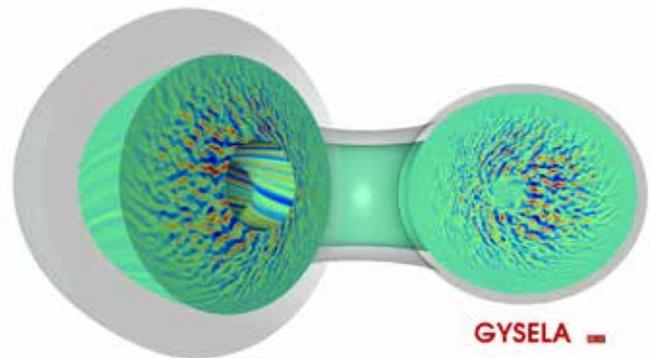


**Figure 1 :** Vue intérieure du tokamak WEST avant les premiers plasmas. En fonctionnement, ce volume torique sera occupé par un plasma confiné par un champ magnétique pouvant être porté à 4 Teslas. La température de ce milieu devrait atteindre 20 000 000 Kelvin (peu différente de la température en °C), mais une pression de l'ordre de 0,15 atmosphère. Dans ce milieu très dilué, les collisions ont un effet faible, et la description fondamentale est nécessairement cinétique.

## 2 WEST et GYSELA en support à ITER

Le paramètre décrit par la loi d'échelle est le temps de confinement de l'énergie. Il mesure le temps caractéristique d'évolution de l'énergie interne du plasma, typiquement la température, lors d'un changement de puissance de chauffage. Il s'agit donc d'un indicateur de la qualité du confinement du plasma. Dans les expériences, une turbulence se développe affectant l'isolation du système. Le niveau de performance dont l'indicateur est le temps de confinement de l'énergie est donc lié à cette turbulence et en particulier au transport de chaleur des ions du plasma. En complément des expériences, la communauté scientifique a initié un effort de simulation numérique pour maîtriser la turbulence plasma (qui n'a pas que des effets négatifs de réduction de performance). Cette activité théorique, riche de nombreux résultats qualitatifs, a essentiellement fourni à ce jour un éclairage parcellaire sur la physique du confinement. L'équipe autour du code GYSELA participe à cet effort et a choisi d'aborder ce problème en réalisant des simulations multi-échelles : depuis l'échelle des structures turbulentes jusqu'à celle de la machine. Cela impose de considérer la géométrie complète du tore, soit des simulations globales du transport de la chaleur par la turbulence, avec un mécanisme de forçage hors de l'équilibre thermodynamique proche des réalisations expérimentales, c'est à dire une source de chaleur déterminant un flux d'énergie orienté vers le puits situé à la périphérie du plasma. En contrepartie, nous avons accepté de simplifier pour le moment la forme des surfaces du tore et de limiter notre étude au

transport de chaleur des ions. Même dans le cadre de ces simplifications, les simulations avec GYSELA nous conduisent à une stratégie de simulations aux limites des calculateurs actuels en taille mémoire et en temps de calcul. La maturité du code, la possibilité d'étudier une propriété globale comme le temps de confinement de l'énergie et la nécessité de situer les performances confinement de WEST par rapport aux autres expériences sous-tendent le projet WATT.

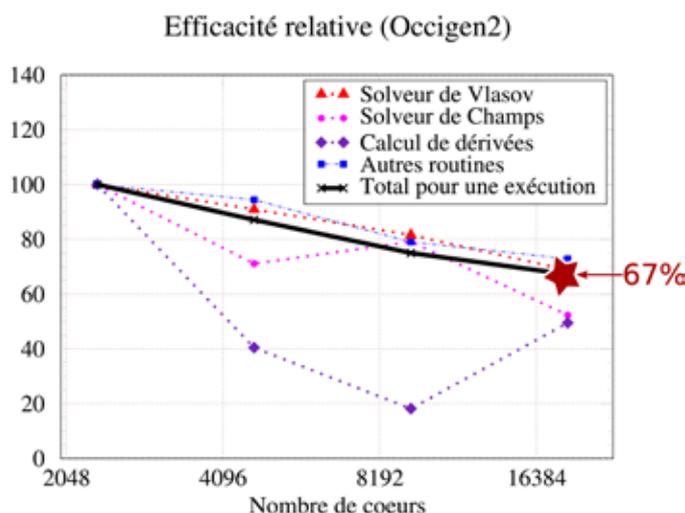


**Figure 2 :** Image reconstruite à partir des simulations GYSELA donnant une carte du champ de potentiel électrique à un instant donné. Deux coupes sont effectuées : (i) le tore est ouvert permettant de visualiser la structure des structures de potentiel, ces dernières sont analogues aux vortex pour la turbulence des fluides neutres, (ii) en évitant la partie centrale, on peut constater l'enroulement de ces structures sur les surfaces toriques.

### Objectif

À notre connaissance, il n'existe pas de tentative pour décrire le temps de confinement de la chaleur dans les plasmas par la simulation même si certaines des propriétés sous-jacentes ont été traitées par ailleurs. Les derniers développements du code GYSELA au niveau du traitement des conditions limites permettent d'aborder

cette question. De plus il est apparu que la variation du rapport d'aspect ne conduisait pas à une modification très importante des conditions de simulation, tout en jouant de manière directe sur un temps caractéristique contrôlant la turbulence plasma. Le projet WATT, pour WEST Aspect ratio and Turbulent Transport, avait ainsi comme objectif de déterminer le temps de confinement de l'énergie et sa dépendance en fonction du rapport d'aspect, couvrant une plage de rapports d'aspect allant de 1.5, expérimenté sur le tokamak anglais MAST, 3 pour TORE SUPRA et ITER et atteignant jusqu'à 6 pour WEST. ■

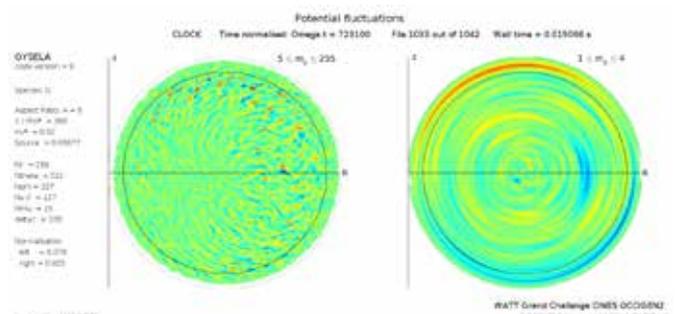


**Figure 3 :** Etude de performance dite de scalabilité (strong scaling) réalisée avec GYSELA sur OCCIGEN2 dans le cadre du grand challenge WATT. Le résultat est correct mais inférieur au niveau atteint avec 33000 cœurs sur BlueGene/Q (87 %).

## 3 Code de calcul GYSELA

Le code de calcul GYSELA est un code de simulation de la turbulence plasma, cette dernière déterminant le transport de chaleur des ions dans le cas de WATT. Le code est adapté aux architectures massivement parallèles et fait l'objet d'une recherche d'amélioration continue dans le cadre de collaborations nationales et internationales. Le modèle pour le plasma est cinétique, c'est-à-dire qu'il traite l'espace des phases, a priori à 6 dimensions sous forme de convection en coordonnées de position et en vitesse sous l'influence de collisions et en présence d'un champ de vitesse turbulent associé au champ électrique. Ce dernier est déterminé en utilisant les équations de Maxwell. Une séparation d'échelle, cohérente avec l'intensité du champ magnétique présent dans les tokamaks, permet de réduire la dimension de l'espace à 5 dimensions. Cette réduction, cruciale pour la réalisation de simulations numériques avec les puissances de calcul actuellement disponibles, se traduit par une plus grande complexité des équations. En effet, dans cette approche, les particules sont transformées en boucles de courant dont la taille est finie et comparable, en général plus grande, à celle du maillage. On introduit ainsi un caractère non-local dans la physique et une nouvelle contrainte sur le numérique. Le code permet de déterminer la fonction de distribution 5D des ions dans le cadre formel appelé gyrocinétique. Un schéma semi-lagrangien réalise l'avancement en temps. Il est original par rapport à la plupart des schémas utilisés dans les codes concurrents, Particle In Cell (PIC) ou Eulérien. En simplifiant, on peut considérer que l'on étudie la déformation du maillage lors des convections dans l'espace des phases et que l'on projette alors la fonction de distribution attachée à ce maillage sur un maillage régulier. On combine ainsi les propriétés de suivi des trajectoires des codes PICs avec l'utilisation d'une grille fixe comme dans le cadre des codes Eulériens. Le code

GYSELA a été testé sur 450 000 cœurs, sur l'un des plus gros calculateurs européens ([http://www.fz-juelich.de/ias/jsc/EN/Expertise/High-Q-Club/\\_node.html](http://www.fz-juelich.de/ias/jsc/EN/Expertise/High-Q-Club/_node.html)), avec une bonne efficacité. La simulation de base utilisée comme point de départ pour le projet est une simulation de 272 109 points de maillage utilisant quelques milliers de cœurs pour un temps de calcul d'un mois environ. Les derniers développements du code, notamment en remplaçant des zones non-physiques pour imposer des conditions limites plus proches de l'expérience, nous incitent maintenant d'aborder les propriétés globales de confinement dans un cadre permettant une confrontation des simulations avec les résultats expérimentaux. ■



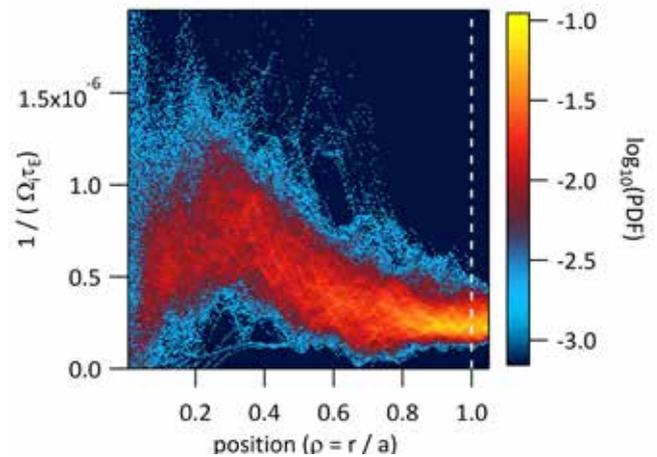
**Figure 4 :** Coupes de la carte du potentiel électrique (voir figure 2) analysée dans le cadre de WATT. Cette figure est utilisée pour construire un film de la simulation. La coupe de gauche représente les petites structures du potentiel électrique caractérisées par le nombre d'onde. Ces fluctuations du potentiel électrique sont à l'origine du transport turbulent de la chaleur dans les tokamaks. La coupe située à droite représente les grandes structures de potentiel, peu étendues dans la direction radiale et qui cisailent les structures turbulentes contribuant ainsi à un contrôle du transport de chaleur. Les paramètres les plus importants de la simulation sont affichés dans la partie gauche.

## 4 Préparation de GYSELA pour OCCIGEN2

Depuis une dizaine d'années, les préoccupations énergétiques ont limité l'augmentation des fréquences d'horloge chez les fondeurs de processeurs et ont conduit à l'ère du *multicore* et du *manycore*. Cela se traduit par une augmentation du parallélisme et du nombre de cœurs au sein de chaque nœud de calcul des supercalculateurs. La technologie de l'hyperthreading permet à deux threads d'être actifs et de réaliser des calculs sur un même cœur. Un certain nombre de ressources se retrouvent ainsi partagées : bus internes, registres, unités de calculs, mémoires cache. En 2009 on employait typiquement 8 threads par nœud : avec OCCIGEN2, pour chacun des nœuds de 24 cœurs, c'est maintenant 48 threads qui sont ainsi actifs. Cette approche permet de maximiser le

nombre d'instructions en calcul flottant qui sont traitées à chaque cycle horloge (et l'approche est beaucoup plus efficace que dans les générations précédentes de processeurs). En effet, il y a deux unités matérielles qui sont à même de réaliser des calculs flottants dans les processeurs Broadwell. En affectant deux threads sur un même cœur, cela permet de proposer deux flux de travaux indépendants qui pourront plus facilement occuper ces unités de calcul. De plus, les unités de calcul sont vectorielles, c'est à dire qu'elles sont capables de réaliser des additions, multiplications sur des petits vecteurs (4 flottants double précision). Dans GYSELA, comme d'autres codes, cette capacité de traitement sur des petits vecteurs augmente les capacités de calcul

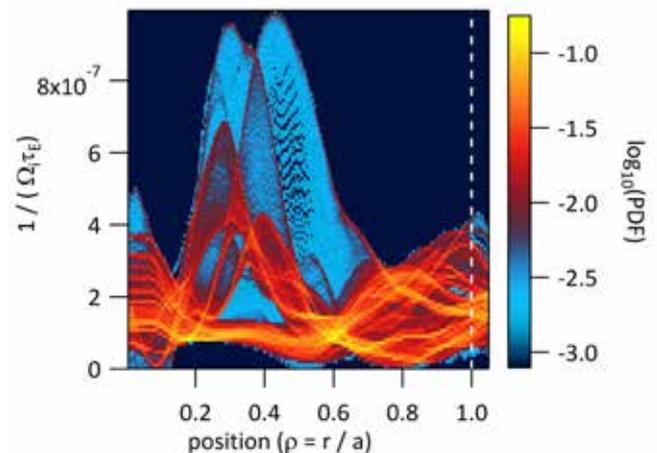
de manière significative. Mais elle n'est effective que si le compilateur possède les qualités nécessaires pour transformer le code source en une série d'opérations vectorielles. Les versions 2017 et 2018 des compilateurs d'INTEL sont caractérisées par un net progrès dans le traitement de cette tâche difficile. Cependant, le programmeur doit mettre en forme le code source pour que la transformation par le compilateur soit la plus efficace possible. C'est ainsi que de nombreux codes, dont GYSELA, sont actuellement dans une phase de réécriture afin d'assister le compilateur dans la détection des opérations vectorielles les plus performantes. Cette démarche augmente significativement le débit des opérations de calcul flottant. Par ailleurs, le nombre élevé de cœurs au sein de chaque nœud nous a conduits à rechercher puis à trouver plus de parallélisme à grain fin en utilisant une grande variété de fonctionnalités du paradigme de parallélisation d'OpenMP. Cette révision de la parallélisation des « nids de boucles » a permis d'occuper les cœurs calculs avec un meilleur équilibrage de charge. Certaines adaptations de l'utilisation combinées



**Figure 5 :** Etude statistique de la performance de confinement en fonction du petit rayon du tore de l'intérieur vers l'extérieur. Dans la partie centrale les fluctuations statistiques sont encore considérables. Au niveau de la ligne verticale, pointillée blanche, un effet de convergence par accumulation permet de tracer la distribution de probabilité de l'indice de performance, le temps de confinement de la chaleur, figure (7)

## 5 GYSELA sur OCCIGEN2

L'utilisation des moyens de calcul rendus disponibles dans le cadre du grand challenge a été organisée de manière graduée avec un portage qui se passe bien suivi des premiers petits cas de production scientifique. C'est dans cette phase que des problèmes de performance sont détectés avec une perte d'un facteur 7 qui affecte de manière aléatoire les simulations. Une interaction efficace se met en place avec le support du CINES. Une version simplifiée de GYSELA est fournie pour faire des tests. L'analyse, assez complexe, révèle finalement un problème informatique ; certaines zones de mémoire allouées lors de l'exécution des calculs sont conservées et ne sont pas toujours libérées en fin d'exécution. Des procédures sont mises en place par le CINES pour circonvenir ce problème. À l'issue de cette phase, avec un cas de 160 109 points de maillage, nous avons réalisé un test de performance entre 3000 et 19000 cœurs avec une efficacité de 67 % (strong scaling) comme indiqué sur l'illustration. Ce résultat est bon mais indique aussi des marges de progression. L'analyse des différentes séquences du code montre en effet une assez forte hétérogénéité. Par ailleurs, de meilleures efficacités ont été réalisées sur d'autres machines, par exemple 87 % sur 33000 cœurs sur BlueGene/Q. Dans la dernière phase du grand challenge, alors que la situation informatique était améliorée, le déroulé des calculs a été ralenti par l'augmentation du temps d'attente pour l'exécution en raison du grand nombre de projets impliqués dans cette phase « Grands challenges ». Dans ce contexte, les plus gros cas, qui devaient être emblématiques de WATT, n'ont pas été réalisés.

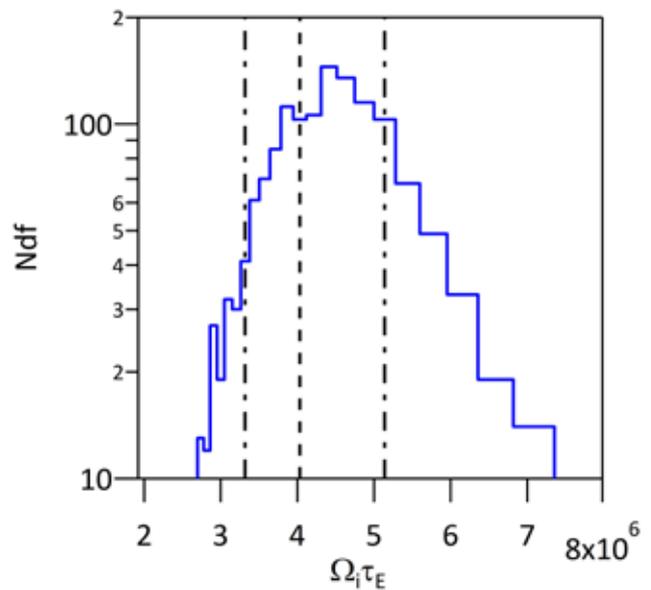


**Figure 6 :** Figure analogue à celle de la figure 5, étude statistique de la performance de confinement en fonction du petit rayon du tore de l'intérieur vers l'extérieur. Ce régime est loin de l'équilibre et correspond à un transitoire. Pendant cette évolution, des lignes d'accumulation de la statistique indiquent l'existence d'états quasi-stationnaires associés une forte hétérogénéité radiale.

## Les simulations WATT

Le code GYSELA a été utilisé dans une configuration bien maîtrisée par l'équipe où les principales inconnues étaient les propriétés d'auto-organisation de la turbulence dans les simulations. Le principe retenu a été d'étudier des plasmas dans un régime où l'énergie interne restait invariante lors des changements de valeurs des paramètres de contrôle. Cela supposait en particulier de varier la source de chaleur pour compenser les variations d'efficacité du confinement. Le point de départ a été donné par une simulation effectuée sur le calculateur COBALT du CCRT où un régime de stationnarité a été obtenu avec une source constante dans le temps et une énergie interne convergeant vers une valeur quasi-stationnaire. Pour tracer la voie dans le programme de simulation nous avons supposé des dépendances lors de la variation des paramètres caractéristiques. Certains ajustements de l'amplitude de la source ont été réalisés notamment pour trouver un encadrement de la solution stationnaire : avec une source d'amplitude trop forte on observe une croissance de l'énergie interne et inversement lorsque la source est trop faible, l'énergie interne décroît. Ce type d'encadrement permet un calcul plus fiable du temps de confinement. Dans la plupart des cas cependant, une dérive qui nous semblait faible a été jugée acceptable. Par ailleurs, nous avons changé le pas de temps numérique de manière analogue à celle projetée pour le temps de confinement global du plasma. Dans certains régimes, cette stratégie est apparue trop optimiste compte tenu de l'existence d'échelles de temps différentes dans le système simulé. Deux aspects ont marqué la réalisation des simulations. L'existence systématique d'un transitoire, caractérisé par une dynamique variant peu d'une simulation à l'autre et d'intérêt réduit pour l'analyse physique. Par ailleurs, une évolution lente des profils est toujours observée sans que l'on puisse dire si elle correspond à l'exploration

de la solution quasi-stationnaire, où un transitoire vers une autre forme d'équilibre. Un questionnement sur les indicateurs nous a permis de quantifier la fiabilité des simulations. Sur un total de 41 simulations, nous avons réalisés trois cas tests et trois simulations ont été arrêtées en cours d'exécution car elles ne semblaient pas fiables. Nous disposons ainsi de 35 simulations permettant de dégager les dépendances paramétriques du temps de confinement telles que déterminées par les simulations. ■



**Figure 7 :** Distribution de probabilité du critère de performance, échelle verticale, le temps de confinement de la chaleur, axe horizontal. On observe un écart à la distribution gaussienne (parabolique dans cette échelle logarithmique). Les traits verticaux indiquent la valeur la plus probable et l'écart type. Ces grandeurs sont calculées à partir de la statistique l'inverse du temps de confinement.

## 6 Observations et résultats scientifiques

La stratégie de réalisation de simulations en régime stationnaire a été un succès. Nous avons pu obtenir ainsi un ensemble de simulations décrivant un domaine de paramètres assez grand où un régime turbulent est atteint. La dynamique présente de fortes fluctuations, y compris pour une quantité intégrale, comme l'énergie interne. En revanche, l'évolution de l'énergie interne sur des temps assez longs ne présente pas une dérive significative. Dans de nombreux cas le choix initial de l'amplitude de la source de chaleur a semblé suffisant pour maintenir un degré de stationnarité souhaité. À contrario nous avons identifié des régimes où l'établissement de conditions stationnaires aurait nécessité un changement de la répartition spatiale de la source. Nous avons adapté notre programme de simulation à la charge de la machine et de ce fait privilégié l'exploration des dépendances paramétriques du temps de confinement de l'énergie sur de « petits » cas. L'étude de l'impact

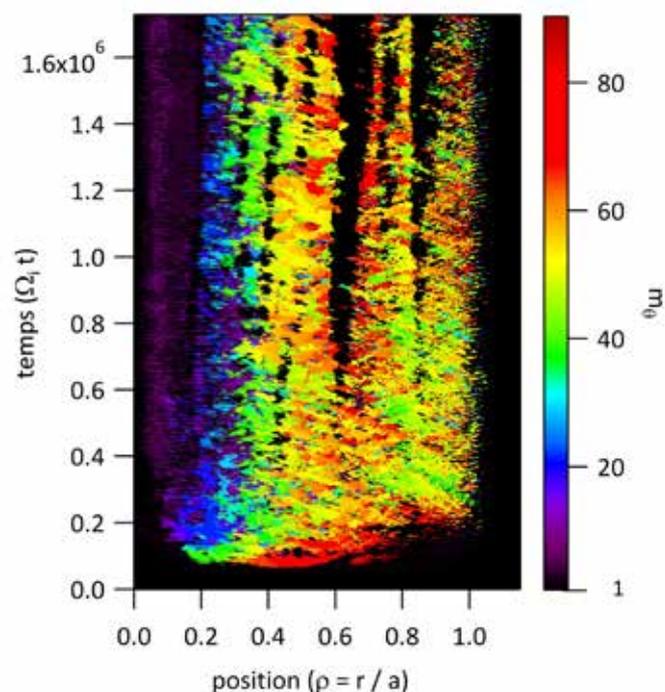
de 3 paramètres clefs sur le temps de confinement a été réalisée.

L'étude de l'effet du rapport d'aspect, compris entre celui d'ITER et celui de WEST, qui était l'objectif premier de WATT, a indiqué une dépendance inverse à celle obtenue de la loi d'échelle empirique. La variation ainsi observée était cependant qualitativement en accord avec celle que nous avons supposée pour conduire les simulations. Celles-ci ont ainsi confirmé l'analyse que nous avons faite quant au rôle de ce paramètre. Aux plus petits rapports d'aspect, la situation n'a pas pu être entièrement clarifiée. En premier lieu, ces simulations sont apparues plus difficiles à conduire en état stationnaire car une variation spécifique du confinement dans la région la plus centrale a été observée. Il est possible que pour ce domaine de rapport d'aspect la dépendance soit plus proche de celle résultant de la loi d'échelle empirique.

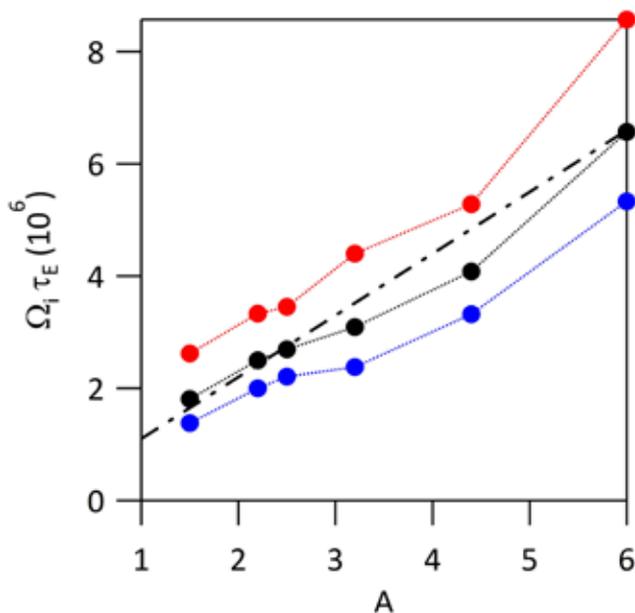
Cela semble assez satisfaisant dans la mesure où les configurations magnétiques de ce type ont été incluses dans la base des données expérimentales. Si ce résultat était confirmé, la dépendance du temps de confinement en fonction du rapport d'aspect présenterait un minimum autour de 3, suggérant une amélioration du confinement sur WEST d'un facteur 2, un effet qui serait facilement identifiable lors des futures expériences qui vont être menées.

En complément à cette étude nous avons réalisé une étude de la dépendance en fonction du paramètre sans dimension évaluant l'efficacité des collisions. Nous avons obtenu une loi de variation compatible avec une loi empirique, mais différant de la loi d'échelle pour ITER. La tendance est une dégradation du confinement avec l'efficacité des collisions, alors que la loi empirique de référence pour ITER ne présente qu'une dépendance négligeable en fonction de ce paramètre.

L'étude la plus exigeante a été celle portant sur un paramètre comparable au nombre de Reynolds et mesurant le nombre de degrés de liberté dans le système physique.



**Figure 9 :** Analyse plus détaillée de la turbulence en fonction du temps, axe vertical, et de la position radiale, axe horizontal. La carte de couleur indique le nombre d'onde du mode de plus forte intensité dans le spectre turbulent. On distingue deux régions, en noir la région où le nombre d'onde vaut 1, ce qui correspond à une très grande structure, et les régions colorées, en vert et jaune où les nombres d'onde sont ceux des modes turbulents qui dominent le transport de chaleur. Le critère utilisé est très sensible et indique avec un fort contraste les régions de turbulence atténuée (en noir) des régions de turbulence forte (colorées). Cette analyse permet de constater que la turbulence continue de s'auto-organiser sur des temps longs. La campagne de simulations WATT est une étape importante, pas un point final.



**Figure 8 :** Le résultat représenté ici est le résultat emblématique de WATT, il montre une augmentation linéaire de la performance de confinement, axe vertical, avec le rapport d'aspect  $A$ , axe horizontal, alors que les lois empiriques indiquent une décroissance. Sur cette figure les points compris entre  $A = 1.5$  et  $A = 3$  doivent être repris et pourraient conduire à une inversion et un minimum vers  $A=3$ ... Affaire à suivre !

Nous avons pu réaliser des simulations intéressantes, dégagant de premières tendances. Pour ce paramètre, les premiers résultats indiquent une dépendance moins forte que celle obtenue empiriquement. Elle serait cependant cohérente avec un transport balistique de la chaleur, en accord avec l'analyse directe du transport que l'on peut faire à partir des simulations effectuées avec GYSELA.

La campagne de simulations WATT est décisive car les lois d'échelle obtenues, qui doivent être confirmées et complétées, permettent déjà de déterminer, à partir des paramètres de contrôle, l'amplitude de la source pour réaliser des simulations en régime proche du régime stationnaire. Cette capacité prédictive est importante pour réduire une partie significative des transitoires vers un régime en équilibre statistique. La deuxième contribution est l'accord ainsi que le désaccord avec les lois empiriques. Cela nous amène à questionner les principes fondamentaux, souvent implicites, qui guident la présentation des lois d'échelle empiriques. La possibilité d'une description sur la base d'une dépendance en loi de puissance semble un point critique. Nos résultats sont également analysés avec un regard critique et doivent être complétés par l'étude des propriétés du transport turbulent de chaleur au niveau microscopique. Ces points structurent notre effort actuel de recherche ainsi que notre programme de simulations. ■

## 7 Conclusion remerciement

Le grand challenge WATT définit de nouvelles perspectives dans le programme de simulations avec le code GYSELA. Les résultats, toujours en cours d'analyse, et complétés par d'autres simulations sur d'autres calculateurs ont permis d'établir une loi d'échelle que l'on peut confronter aux lois d'échelle empiriques. Les prédictions pour WEST sont nettement plus optimistes que celles

construites avec les bases de données expérimentales. La confrontation aux expériences à venir sur WEST s'annonce passionnante.

Ce travail a bénéficié d'une aide du gouvernement français au titre du Programme Investissements d'Avenir, Initiative d'Excellence d'Aix-Marseille Université - A\*MIDEX. ■

## 8 Bibliographie et références

- G. Laval et al., 2007, *La Fusion Nucléaire : de la recherche fondamentale à la production d'énergie ?*, Rapport scientifique et technique n°26 de l'Académie des sciences, EDP Sciences, <https://www.edp-open.org/images/stories/books/fulldl/rapport6.pdf>
- Jean Jacquinet, 2013, *ITER et les recherches sur la fusion par confinement magnétique*, Vol. 32, page 19, Reflets de la physique, DOI : <https://doi.org/10.1051/refdp/201332019>
- E.J. Doyle, W.A. Houlberg, Y.Kamada, V. Mukhovatov, T.H. Osborne, A. Polevoi, G. Bateman, J.W.Connor, J.G. Cordey, T. Fujita, X. Garbet, T.S. Hahm, L.D. Horton, A.E. Hubbard, F. Imbeaux, F.Jenko, J.E. Kinsey, Y. Kishimoto, J. Li, T.C. Luce, Y. Martin, M. Ossipenko, V. Parail, A. Peeters, T.L.Rhodes, J.E. Rice, C.M. Roach, V. Rozhansky, F. Ryter, G. Saibene, R. Sartori, A.C.C. Sips, J.A. Snipes, M. Sugihara, E.J. Synakowski, H. Takenaga, T. Takizuka, K. Thomsen, M.R. Wade, H.R. Wilson, ITPA Transport Physics Topical Group and ITPA Confinement Database and Modelling Topical Group and ITPA Pedestal and EdgeTopical Group, 2007, *Chapter 2 : Plasma confinement and transport*, Nuclear Fusion, Vol. 47, pages S18, <http://stacks.iop.org/0029-5515/47/i=6/a=S02>
- V. Grandgirard, J. Abiteboul, J. Bigo, T. Cartier-Michaud, N. Crouseilles, G. Dif-Pradalier, Ch. Ehrlacher, D. Esteve, X. Garbet, Ph. Ghendrih, G. Latu, M. Mehrenberger, C. Norscini, Ch. Passeron, F. Rozar, Y. Sarazin, E. Sonnendrücker, A. Strugarek, D. Zarzoso, 2016, *A 5D gyrokinetic full-f global semi-Lagrangian code for flux-driven ion turbulence simulations*, Computer Physics Communications, Vol. 207, pages 35-68, DOI : <https://doi.org/10.1016/j.cpc.2016.05.007>
- F. Rozar, G. Latu, J. Roman, V. Grandgirard, 2015, *Achieving Memory Scalability in the GYSELA Code to Fit Exascale*, Concurrency Computat. : Pract. Exper., Vo. 27, pages 994-1009. doi : 10.1002/cpe.3429
- G. Latu, N. Crouseilles, V. Grandgirard, E. Sonnendrücker, 2007, *Gyrokinetic Semi-Lagrangian Parallel Simulation using a Hybrid OpenMP/MPI programming Recent Advances in Parallel Virtual Machine and Message Passing Interface*, Springer, Vol. 4757, pages 356-364 <https://hal.archives-ouvertes.fr/hal-00605748>
- X. Garbet, Y. Idomura, L. Villard, T.H. Watanabe, 2010, *Gyrokinetic simulations of turbulent transport*, Vol. 50, page 043002, DOI <https://doi.org/10.1088/0029-5515/50/4/043002>



Figure 10 : L'équipe GYSELA au CEA Cadarache en octobre 2017.

## Centre Informatique National de l'Enseignement Supérieur

950, rue de Saint-Priest - 34097 Montpellier Cedex 5 - France

Tél. : 04 67 14 14 14 - Fax : 04 67 52 37 63

[www.cines.fr](http://www.cines.fr) - services utilisateurs : [svp@cines.fr](mailto:svp@cines.fr)

Directeur de la publication : Boris DINTRANS

Co-rédacteur en chef : Olivier ROUCHON

