

Un solveur Navier-Stokes massivement parallèle pour l'étude du mascaret et le déferlement des vagues

Stéphane Glockner, Pierre Lubin
I2M – UMR 52 95

*Journée Scientifique du MClA, Université Bordeaux 1
1er février 2013*

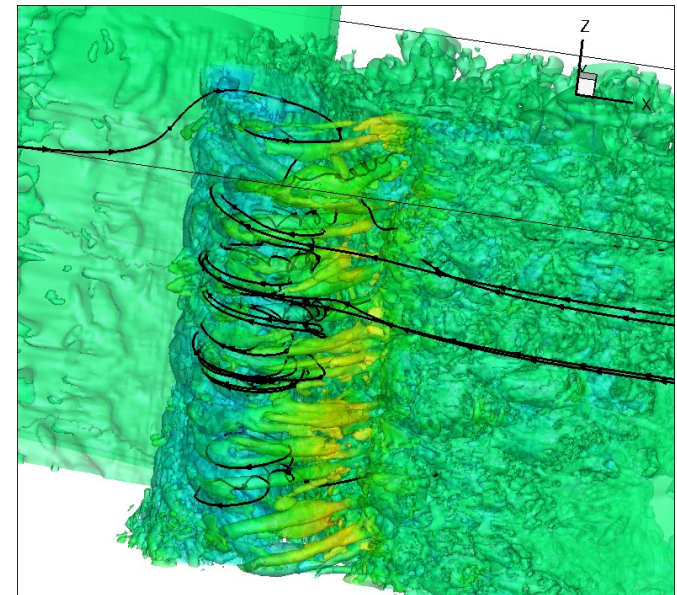
Introduction

Objectif : présenter l'évolution d'un code de calcul pour des simulations 3D à caractère environnemental, dans un contexte de calcul intensif, des calculateurs régionaux aux supercalculateurs européens.

Pourquoi : écoulements turbulents, au large spectre d'échelles spatio-temporelles
→ recours nécessaire au calcul intensif

Plan :

- Les phénomènes physiques
- Thétis
- Calcul parallèle
- I/O
- Visualisation
- Application au déferlement et au mascaret



Pourquoi étudier numériquement le déferlement ?

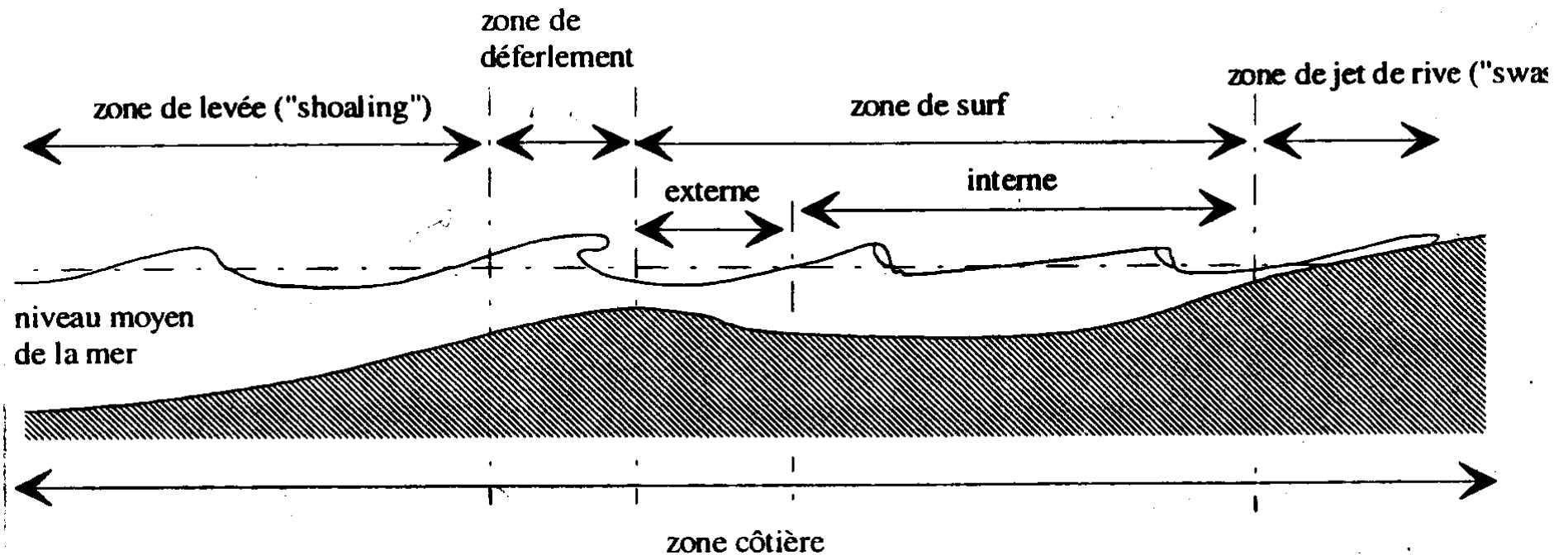
- Principale force dans les processus côtiers (courants, transport sédimentaire, etc.)
- Mesures expérimentales et *in situ* difficiles à réaliser



On étudie l'environnement côtier afin de le comprendre et de le conserver

Étude de la structuration de l'écoulement / Entraînement de l'air

La houle se propage et se transforme à l'approche du rivage



Abadie et al. (1998)

Le déferlement plongeant se produit en plusieurs étapes.

Retournement de l'onde sur elle-même

Éjection d'un jet

Impact du jet sur la face avant de la vague



Le déferlement plongeant se produit en plusieurs étapes.

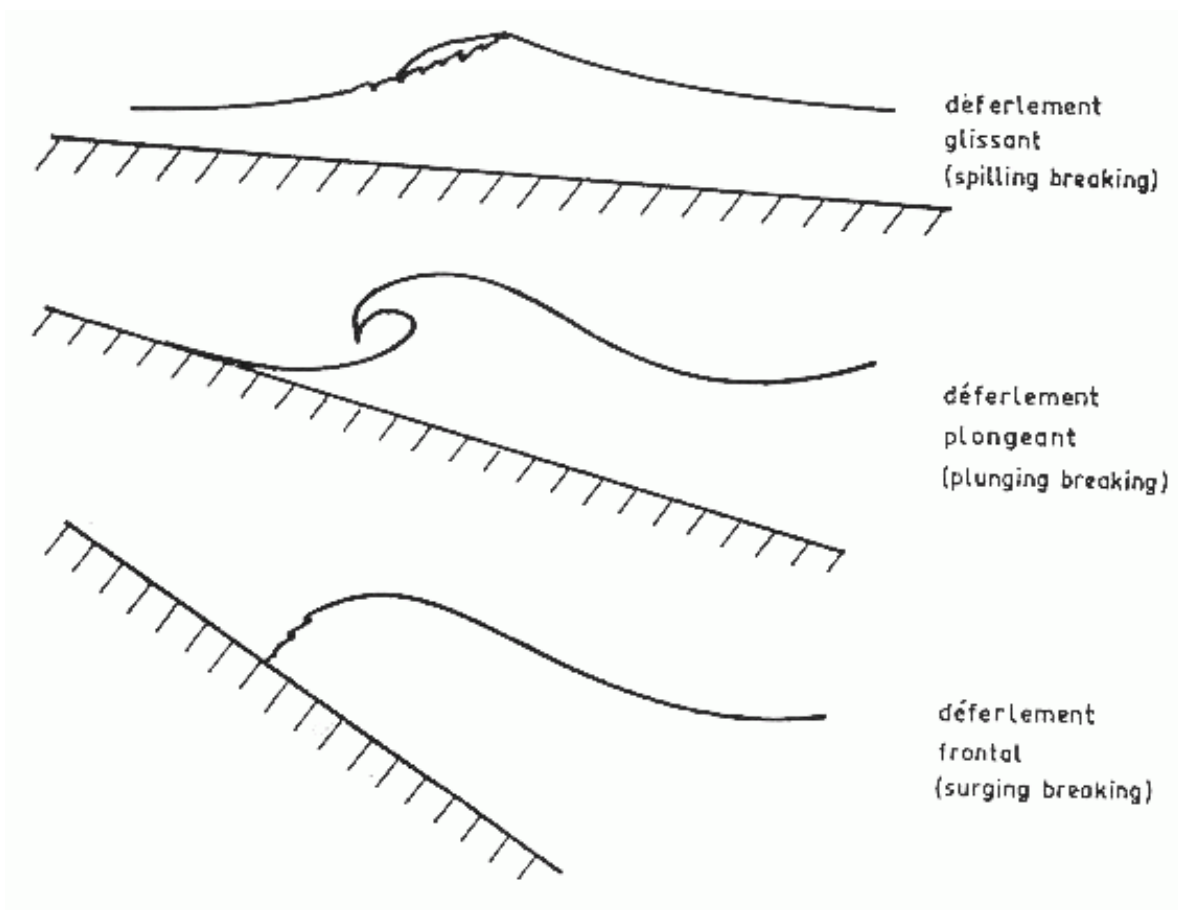
Génération d'un jet secondaire (splash-up)

Entraînement de l'air dans l'eau

Génération de turbulence

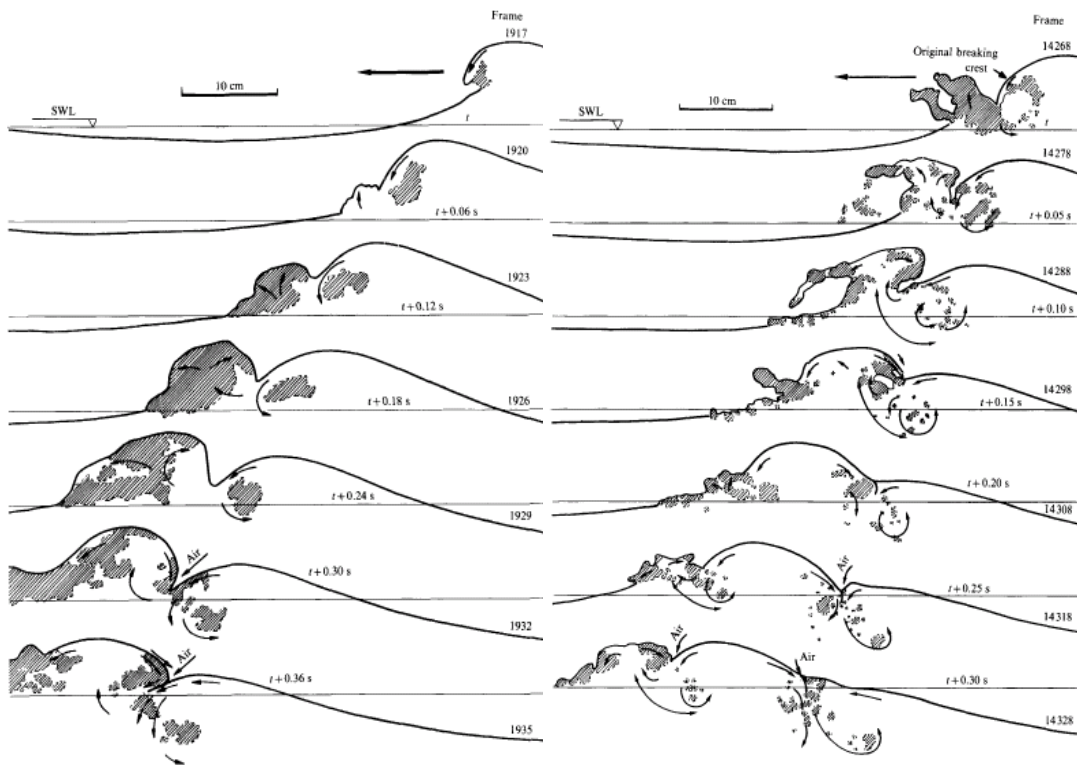


Le déferlement des vagues est classé en plusieurs types

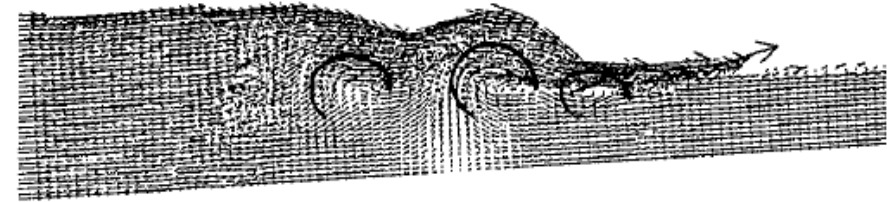


Galvin (1968 & 1972)

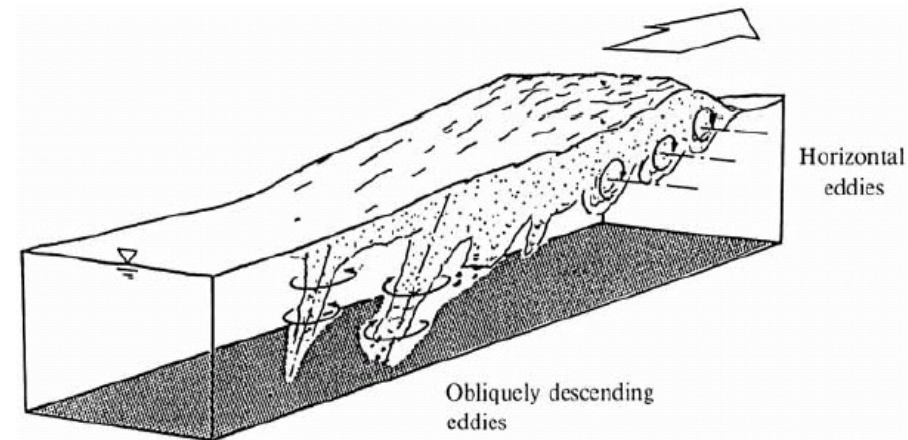
Plusieurs types de structures cohérentes sont observées dans les vagues déferlantes



Bonmarin (1989)

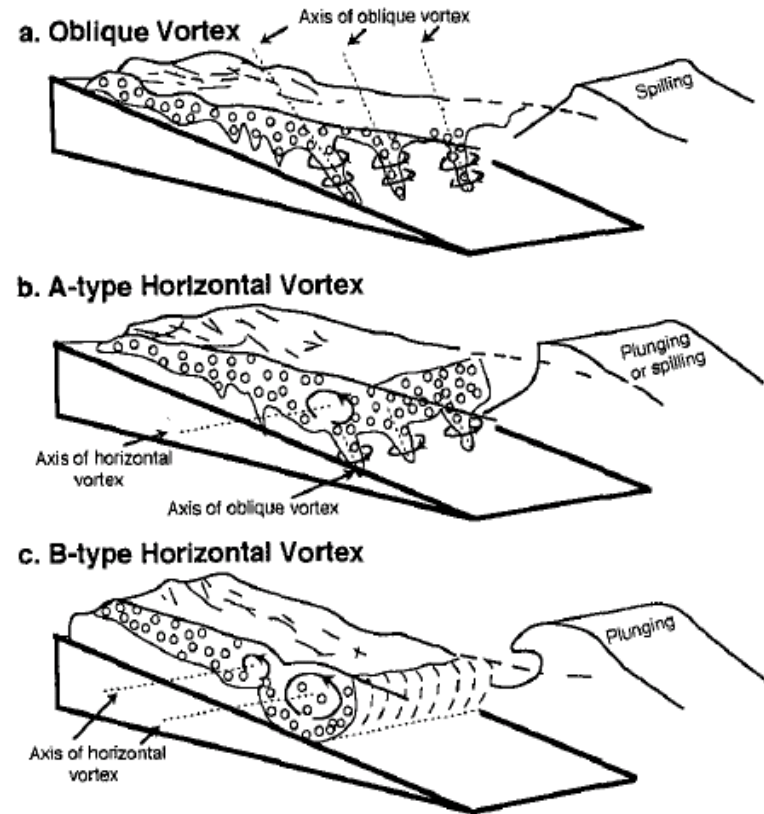


Sakai et al. (1986)



Nadaoka et al. (1989)

Plusieurs types de structures cohérentes sont observées dans les vagues déferlantes



Zhang and Sunamura (1994)

Le 13 septembre 2009...

Un mail et...

Subject: Splotch
From: "Sarhou Arthur" <sarhou@enscbp.fr>
Date: 13/09/2009 11:52
To: "Lubin Pierre" <lubin@enscbp.fr>

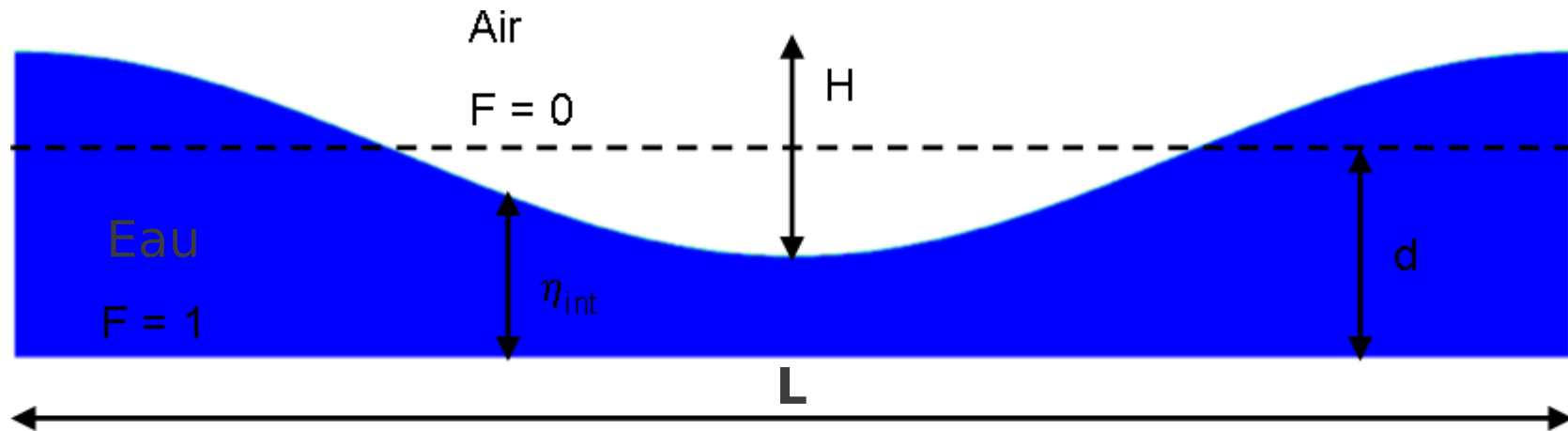
Vortices, maaann

<http://www.youtube.com/watch?v=7B0hDajH0m4>



The screenshot shows a web browser window with the YouTube website. The video player is the central focus, displaying a blue, high-speed shot of a wave. Below the player, the video title is "HD: Super Slo-mo Surfer! - South Pacific - BBC Two" with a view count of 4,651,900. The channel name "BBC" is visible with 12,768 videos and 716,495 subscribers. The video was uploaded on May 6, 2009. To the right of the video player, there is a sidebar with several recommended videos, including "Surf the Big Flowrider", "HD: Tuna Fishing - South Pacific - BBC Two", "HD: Underwater Armageddon - Nature's", "BBC Worldwide", "Hunting a Shark From the Deep - Human Planet:", "HD: Lioness Hunts Zebra - Nature's Great Events:", "Pirate Tuna - Japan", and "Fishing for tuna in the Philippines".

Le déferlement d'une onde sinusoïdale instable est contrôlée par deux paramètres



Domaine périodique

Cambrure initiale $\frac{H}{L}$ Paramètre de dispersion $\frac{d}{L}$

➔ Description du spectre continu des types de déferlement

Quelle stratégie de modélisation pour quels objectifs ?

Écoulement :

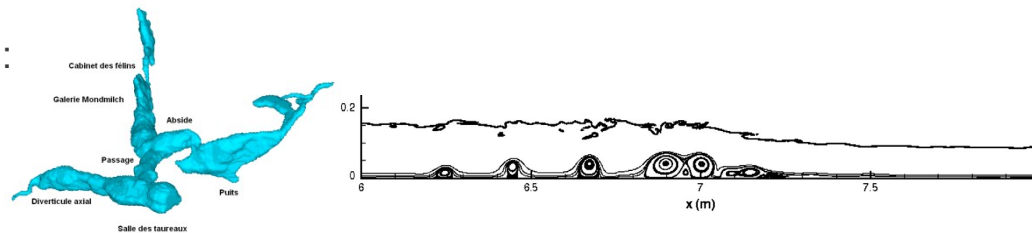
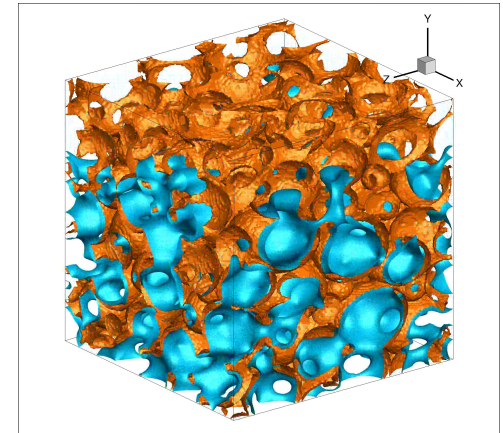
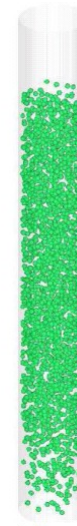
- diphasique
- chaotique
- tridimensionnel



- Description d'un écoulement avec interface
- Déconnexion / reconnexion
- Grande variabilité d'échelles à décrire
- Instationnaire
- Turbulent

Thétis - contexte

- Développé à l'I2M depuis 20 ans (<http://thetis.enscbp.fr>).
- Code à triple usage
 - Recherche
 - Industrie
 - Enseignement
- Environnement multiple
 - Du PC au supercalculateur
 - S'appuie sur la structure pyramidale du calcul intensif
 - Linux et Windows
 - Plusieurs compilateurs (Intel, GNU, PGI)
 - Plusieurs bibliothèques MPI (OpenMPI, MPT, intel, Bull)
 - Documenté, interfacé
 - Contexte de vérification / validation
- Outil d'un équipe
 - +100 ACL depuis 5 ans s'appuyant sur Thétis
 - Projets région Aquitaine (2007, 2012), ANR (2+2), PE (2), coll. institutionnelles et industrielles (ARCELOR, SNECMA, CEA, etc.), projet Lascaux
- Et au-delà, exemples d'utilisation passés ou présents :
 - SIAME, LEMTA, IMFT, ICMCB, LOMC, P', LTN, GEPEA, etc.



Thétis - équations / méthodes

- Navier-Stokes incompressible, modèle 1-fluide, LES
- Reconstruction de l'interface air/eau par VOF-PLIC
- Couplage vitesse/pression
 - Étape de prédiction : éq. vectorielle
 - Étape de correction : éq. d'Helmholtz à coeff. variable
- Maillages structurés par blocs, volumes finis, ordre 2
- Solveurs
 - Étape de prédiction : Bicg, préconditionnement Jacobi diagonal (Thétis)
 - Étape de correction : CG, préconditionnement Multigrille (Hypre)

Calcul Parallèle – historique / objectifs / moyens

- À partir de 2005
 - démocratisation des clusters et explosion des moyens de calcul (en 2007 CINES-jade)
 - → développement de la version parallèle
- Objectifs
 - gagner du temps de calcul → vrai jusqu'à un certain point
 - augmenter la taille des problèmes
- Moyens : s'appuyer sur la structure pyramidale des ressources
 - Cluster de laboratoire → projets région Aquitaine
 - Mesocentre (Avakas) → MCIA
 - GENCI (Jade, Titane, Curie) → quelques Mh CPU / an
 - PRACE (Curie)
 - Preparatory Access PA2012, collaboration avec CINES (N. Audiffren, H. Ouvrard)
 - Regular call 2013 (13 Mh CPU)



Calcul Parallèle – approche choisie, évolution

- Plusieurs possibilités : OpenMP, MPI, GPU
- Objectif principal est d'augmenter la qualité des simulations de problèmes multi-échelles par l'augmentation du maillage
 - → augmentation de la mémoire nécessaire
 - → nécessité de répartir la mémoire sur plusieurs nœuds
 - → EDP, dépendance spatiale de la solution
 - → **MPI** (échanges explicites de messages entre nœuds)
- Stratégie payante, code de production sur qq dizaines de millions de points
- Extension récente à 10^9 points (PA PRACE), mais :
 - Problèmes connexes de transferts de données / stockage / visualisation
 - Les algorithmes/développements/codes prennent du retard sur la puissance de calcul disponible
 - Augmentation des cœurs dans le nœud → MPI / OpenMP (//^{isme} de boucle)
 - MPI / GPU ?

Calcul Parallèle – stratégie de développement

- **Thétis = Résolution d'EDP**

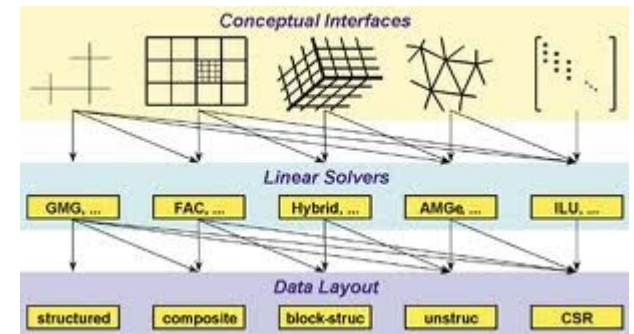
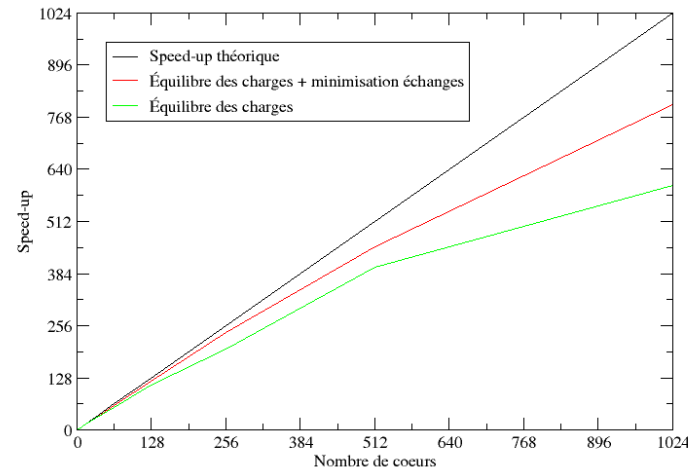
- **Structure du code, modèles, schémas** → I2M

- Partitionnement du domaine
 - Répartir la charge
 - Minimiser les communications
- Discrétisation
- Échanges MPI
- I/O



- **Solveur** → I2M + Hypre

- BiCGStab + Jacobi → I2M (produits mat-vec, produits scalaires)
- CG + préconditionnement Multigrille → Hypre (Livermore)



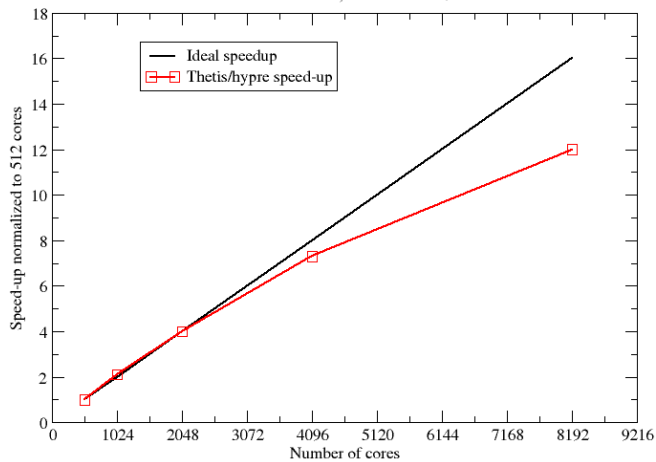
- **Scalabilité forte** : nombre de points du maillage total constant, on augmente le nombre de cœurs, → *attente d'une réduction linéaire du temps de calcul avec le nombre de cœurs*
- **Scalabilité faible** : nombre de point constant par cœur, on augmente le nombre de cœurs et donc la taille globale du problème → *attente d'un temps de calcul constant ou linéaire*

Calcul Parallèle - performances

- Avakas : nœuds Intel Westmere à 12 cœurs, 3 168 cœurs
- Curie : nœud Intel Sandy Bridge à 16 cœurs, 80 640 cœurs

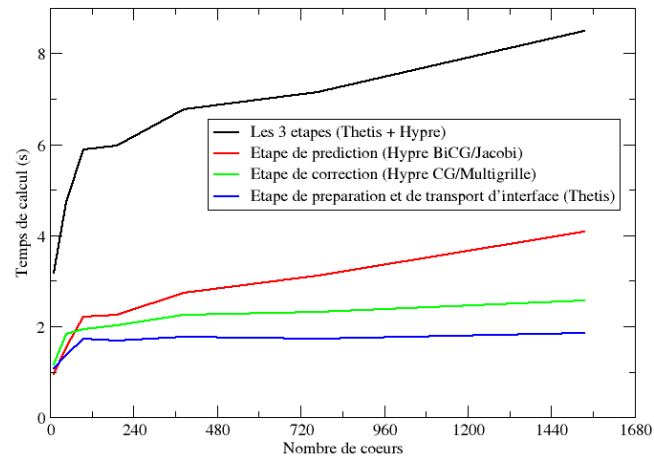
Thétis, Strong Scalability on Curie

540 million mesh cells, from 512 to 8192 cores



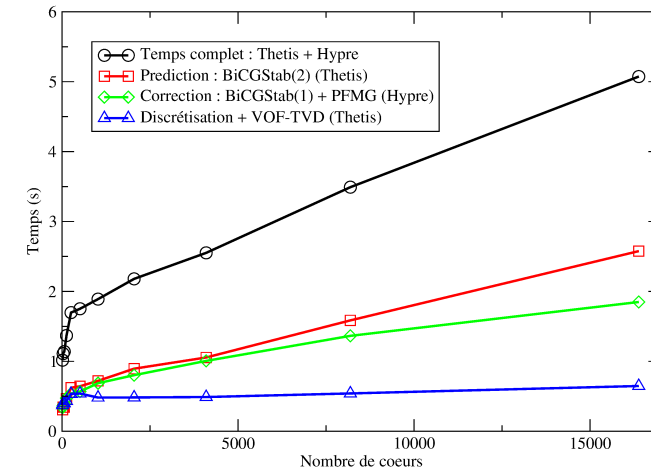
Scalabilité faible de Thétis/Hypre sur Avakas

50^3 cellules par cœur, jusqu'à 1536 cœurs, 200 millions de cellules



Scalabilité Thétis + Hypre (40^3 cellules par cœur)

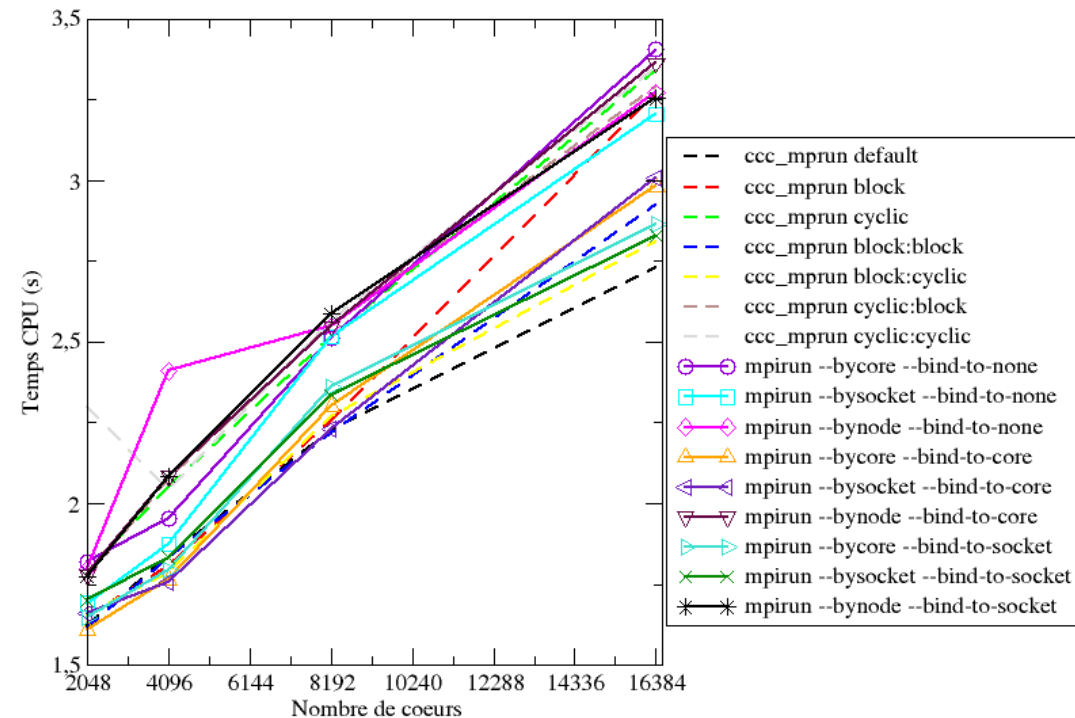
Jusqu'à 1,130 milliard de points et 16384 cœurs.



Calcul Parallèle - performances

- Répartition des process MPI sur Curie (Preparatory access PRACE, avec N. Audiffren CINES)
 - Mode de remplissage des nœuds :
 - par cœur, par socket, par nœud
 - Attachement de la mémoire
 - par cœur, par socket, par nœud
 - Tests effectués sur les mêmes nœuds
 - Effets non négligeables
 - Choix par défaut par toujours le meilleur
 - Dépendant de l'application
 - Pour Thétis/Curie :
 - `ccc_mprun_default`
 - `mpirun -bynode -bind-to-core`

Placement sur Curie (de 2048 à 16 384 cœurs)



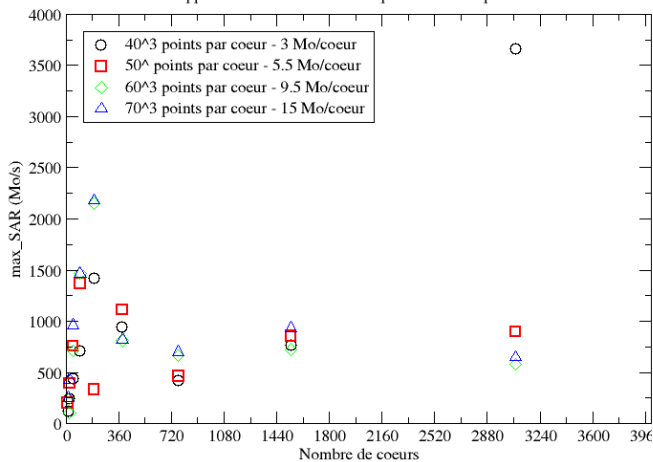
I/O - limites du traitement séquentiel

- 10 millions de points, 400Mo de données
OK par 1 processeur mais inélégant
- Au delà, besoin d'écriture //
 - POSIX : 1 fichier / processus MPI
 - des milliers de fichiers à visualiser
→ logiciel de visualisation adéquat
 - MPI IO : 1 seul fichier pour tous les processus
 - Bibliothèques spécialisées au dessus de MPI IO
→ HDF5 (PA PRACE), netCDF, etc.

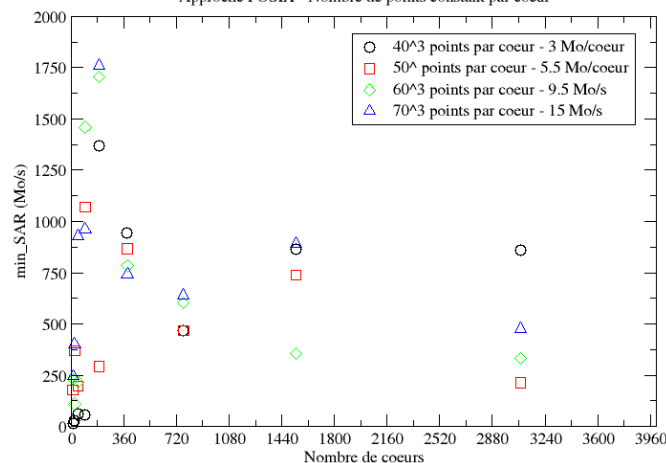
I/O - performances POSIX

- 1 fichier / processus .vts (VTK)
- p cœurs ; chaque cœur écrit N Mo données dans un fichier séparé en T_i s
- Débit soutenu agrégé
 - $\text{max_SAR} = \text{somme}(N/T_i)$
 - $\text{min_SAR} = N.p / \text{max}(T_i)$

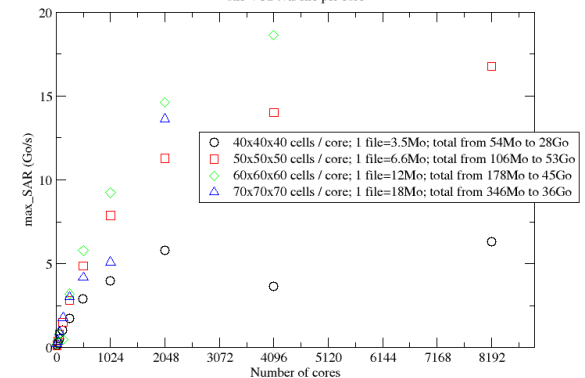
Débit soutenu agrégé maximum (max_SAR) sur Avakas
Approche POSIX - Nombre de points constant par cœur



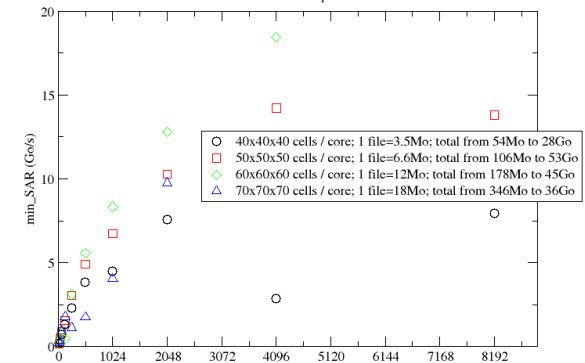
Débit soutenu agrégé minimum (min_SAR) sur Avakas
Approche POSIX - Nombre de points constant par cœur



max_SAR rates on Curie
one VTS .vts file per core



min_SAR rates on Curie
on VTS .vts file per core



Transferts de données

- Simulation de 86 millions de points sur 576 cœurs (Avakas)
 - 400 Go → 60-80Mo/s → 1h30 de transfert
- Simulation de 600 millions de points sur 4096 cœurs (Curie)
 - 6 To → 8Mo/s → 8 jours de transfert > Temps CPU !
 - scp en // (manuel) : 1-2 jours
 - ? bbcp, bbftp, gripftp
- Stockage
- Post-traitement local
- Post-traitement à distance
 - Visualisation à distance. Pas encore dispo au MCIA, CINES, TGCC → 2013
 - Travailler en batch
- Tout visualiser ?
 - Visualisation exploratoire
 - Visualisation sélective

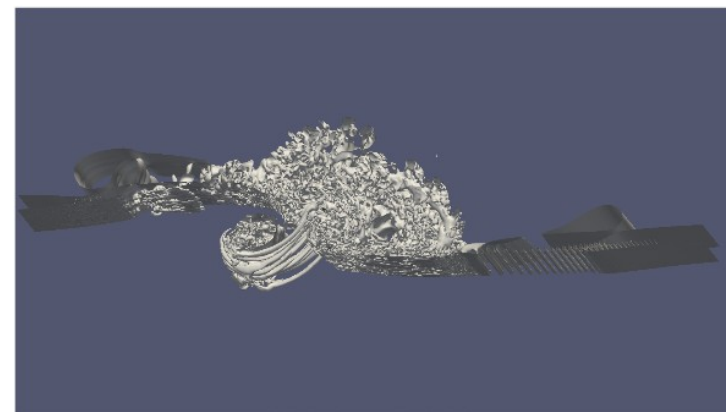
Visualisation

- Logiciels libres Paraview, Visit basés sur VTK
 - Calcul // CPU
 - Version de base pré-compilée // (nb de cœurs max de la machine)
 - Compilation, installation sur cluster
 - Rendu // GPU
- Rendu réaliste : animation avec Blender
 - Extraction d'isosurface avec VTK
 - Préparation du rendu sur une image (parallélisme dans le nœud)
 - Animation de N images
 - Installation de Blender sur cluster
 - Travail en batch, chaque nœud traite N/p images



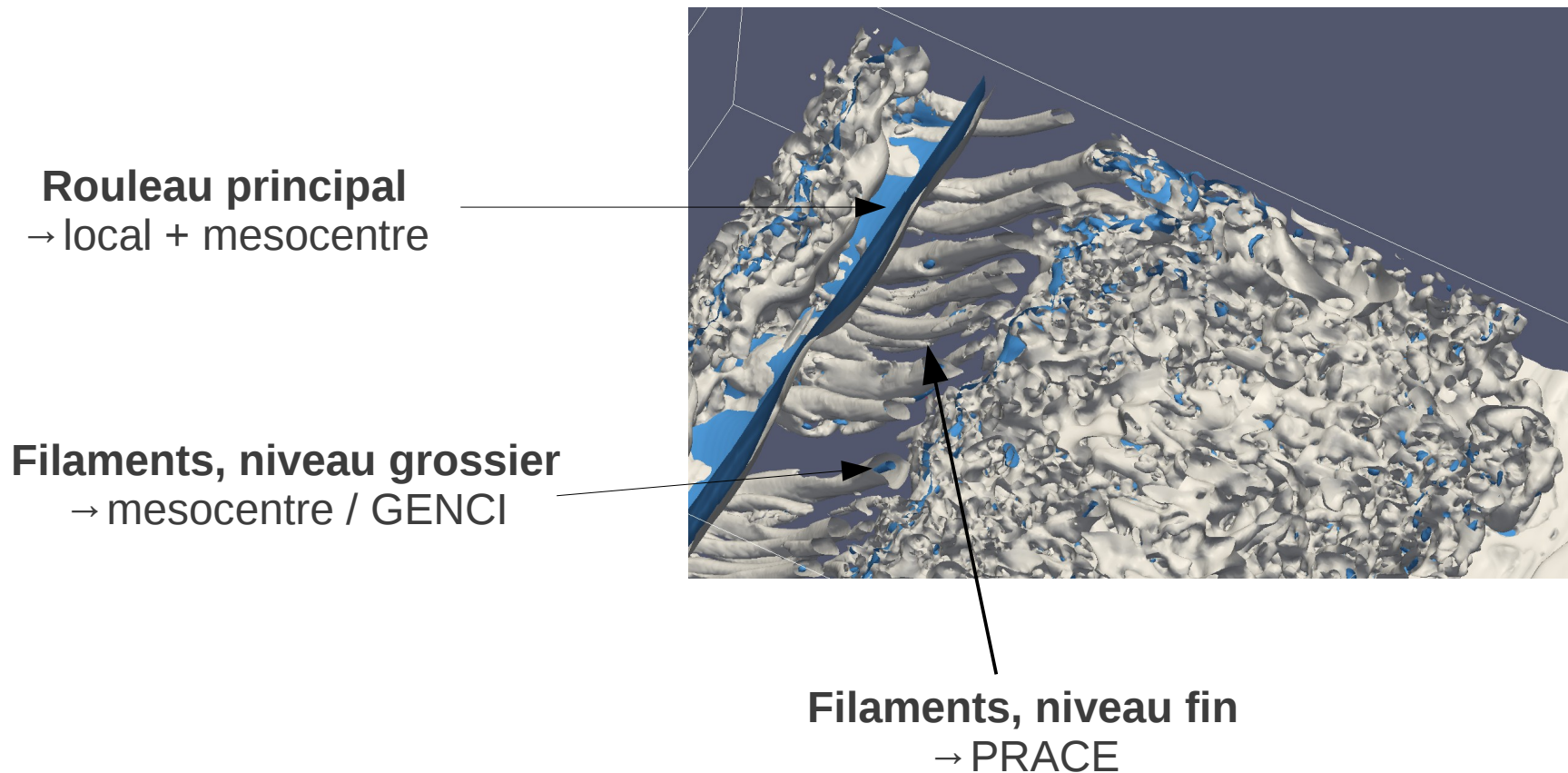
Visualisation – exemple 1

- Simulation de 86 millions de points sur Avakas
- 576 processeurs
- + 100 000 fichiers .vts
- 16 cœurs
- convert, montage, mencoder



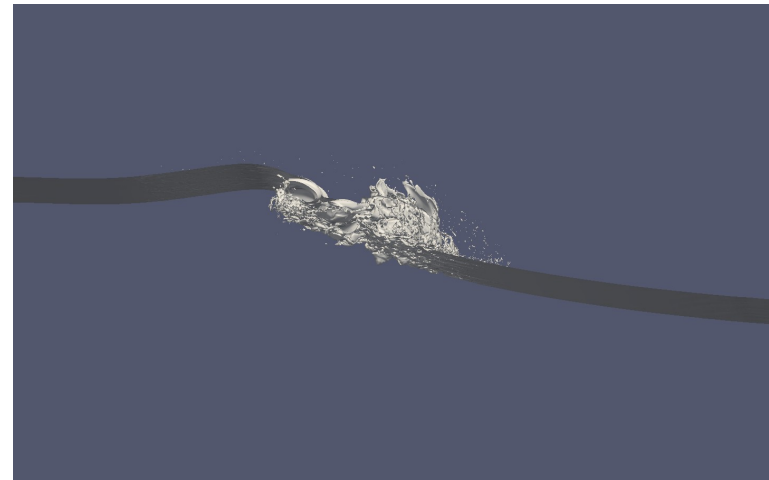
Visualisation – exemple 1

- Mise en évidence de l'intérêt du calcul intensif et de l'architecture pyramidale des moyens



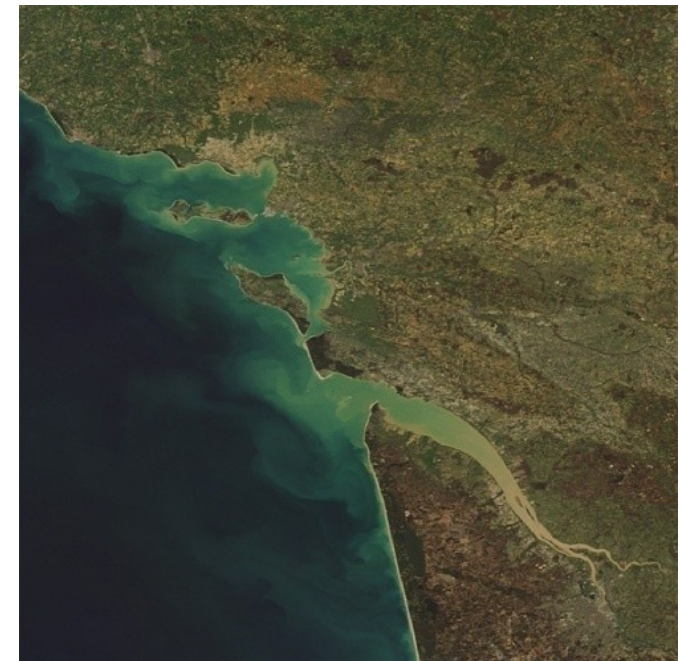
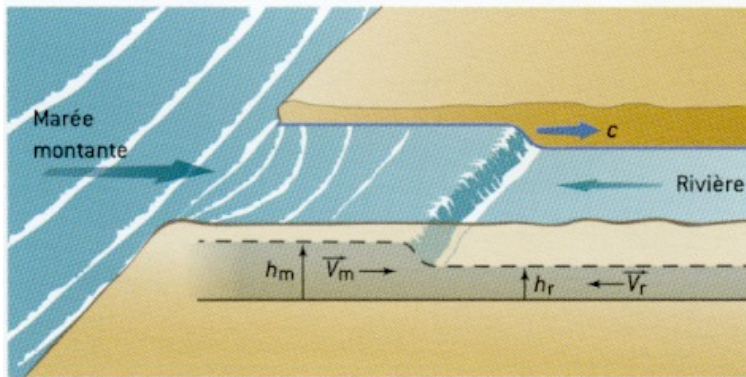
Visualisation – exemple 2

- Simulation de 655 millions de points sur Curie
- 4 096 processeurs
- + 1 000 000 fichiers .vts
- 16 cœurs, lent



Visualisation – exemple 3

- Le mascaret
 - Embouchure d'une rivière (fond plat)
 - Forme convergente
 - Marée montante de grande amplitude



Gironde

Conclusion - Perspectives

- Évolution d'un outil de calcul, du PC aux calculateurs européens
 - Tout doit être parallèle : calcul, IO, transferts, stockage, visualisation
 - Constante adaptation aux évolutions matérielles (many cores, ...)
- Apport du calcul intensif pour une meilleure description et compréhension du déferlement et du mascaret :
 - Étude paramétrique des conditions de formation des filaments
 - Étude de la turbulence dans les mascarets