



HAL
open science

Étude comparative de l'impact d'un codage à précision variable sur des données de simulation en géosciences

Lauriane Bouard, Laurent Duval, Frédéric Payan, Christophe Preux, Marc Antonini

► **To cite this version:**

Lauriane Bouard, Laurent Duval, Frédéric Payan, Christophe Preux, Marc Antonini. Étude comparative de l'impact d'un codage à précision variable sur des données de simulation en géosciences. Colloque CORESA (COmpression et REprésentation des Signaux Audiovisuels), Nov 2021, Sophia Antipolis, France. hal-03414943

HAL Id: hal-03414943

<https://hal.science/hal-03414943>

Submitted on 4 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Étude comparative de l’impact d’un codage à précision variable sur des données de simulation en géosciences

Lauriane Bouard^{1,3}, Laurent Duval², Frédéric Payan³, Christophe Preux², and Marc Antonini³

¹IFP Energies nouvelles, Rond-point de l’échangeur de Solaize, 69360 Solaize, France

²IFP Energies nouvelles, 1-4 Avenue du Bois Préau, 92852 Rueil-Malmaison, France

³Université Côte d’Azur (UCA), CNRS, Laboratoire I3S, 06900 Sophia Antipolis, France

Résumé : *Le volume des données scientifiques produites par simulation numérique est en augmentation croissante. Cette volumétrie incite depuis peu la communauté à employer des méthodes de compression à précision variable. Si l’usage de données scientifiques comprimées peut atténuer les problèmes croissants de stockage primaire, il peut également jouer d’autres rôles à différentes étapes du workflow pour améliorer les performances des outils de calculs, limités notamment en capacité mémoire et en bande passante. Néanmoins ce type d’approche ne doit pas se faire au détriment de la qualité de la simulation. L’étude présentée ici analyse l’impact d’un codage à précision variable (obtenu par combinaison d’un compandeur, d’une transformée en ondelette et d’un codage par arbre-zéro (zerotree)) sur des données de simulation issues des géosciences. Pour valider notre approche nous le comparons à deux codeurs de référence utilisés dans différents domaines scientifiques (SZ [2], ZFP [4]) selon différentes métriques objectives corrélées à la validation subjective de la simulation.*

Mots-clés : Compression, géosciences, maillages, métriques de qualité objectives/subjectives, simulation

1 Introduction

Les supercalculateurs actuels (high performance computers ou HPC en anglais) ont de très forte capacités de calcul, dépassant les centaines de PétaFLOPS (10^{15} opérations en virgule flottante par seconde). Cela permet de simuler des phénomènes complexes avec un très grand réalisme dans de nombreux domaines (cosmologie, climatologie [1], mécanique des fluides, etc.). La contrepartie est l’explosion démesurée de la quantité de données générées, qu’il s’agisse de résultats stockés, de données transférées, ou encore pour sauvegarder et restaurer un état intermédiaire par des checkpoints. En conséquence, l’usage de codeurs à précision variable pour des données numériques simulées se généralise. Il facilite le stockage, l’empreinte mémoire et l’accélération des transferts, et par conséquent le temps d’exécution des simulations. SZ [2] et ZFP [4] (développés notamment au sein des laboratoires d’Argonne et de Lawrence Livermore) sont des codeurs à précision variable ayant fourni des résultats satisfaisants sur différents types de données, notamment en cosmologie (NYX) ou en climatologie (CESM). Par réduction de précision des données codées en nombres flottants, ces codeurs peuvent limiter la taille des données, sans pour autant impacter la qua-

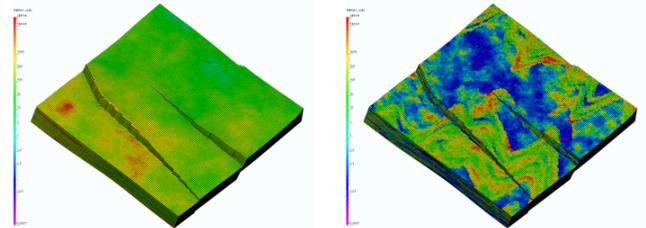


FIGURE 1 – Exemple de maillages hexahédriques en géosciences.

lité visuelle. Des résultats analogues ont été obtenus dans différents contextes de simulation. Ces travaux posent de manière plus générale la question suivante : quelle est la fidélité numérique nécessaire pour l’obtention de résultats précis et réalistes pour la simulation de phénomènes physiques ?

2 Travail proposé

En géosciences par exemple, une simulation d’écoulement s’effectue au travers de maillages hexaédriques (cf. illustrations 1) qui représentent des formations géologiques profondes. Ces simulations sont souvent lancées en grand nombre, pour des études de sensibilité. Elles requièrent une exécution en un temps raisonnable, ce qui a conduit au développement de méthodes d’upscaling, réduisant la résolution des maillages d’origine. Ce type de méthodologie est intégré de manière implicite à différentes résolutions dans le système de compression multi-échelle HexaShrink proposé en [5]. Nous complétons ici ce travail par un codage progressif des propriétés continues associées à ces maillages, notamment la perméabilité. Ce nouveau schéma de codage permet de jouer sur la précision binaire, par l’incorporation d’un codage de type zerotree (ZT) combiné à l’usage d’un compandeur (compresseur-expandeur, notés Λ et V) et d’une transformée en ondelette (9/7 DWT). Notre schéma, qui est donc à précision et à résolution variable, est comparé aux codeurs standards SZ [2] et ZFP [4]. Pour cela, nous utilisons des métriques objectives classiques, mais proposons aussi une nouvelle métrique (Λ -SNR) plus conforme à l’évaluation subjective des résultats de simulation ainsi qu’aux exigences du mé-

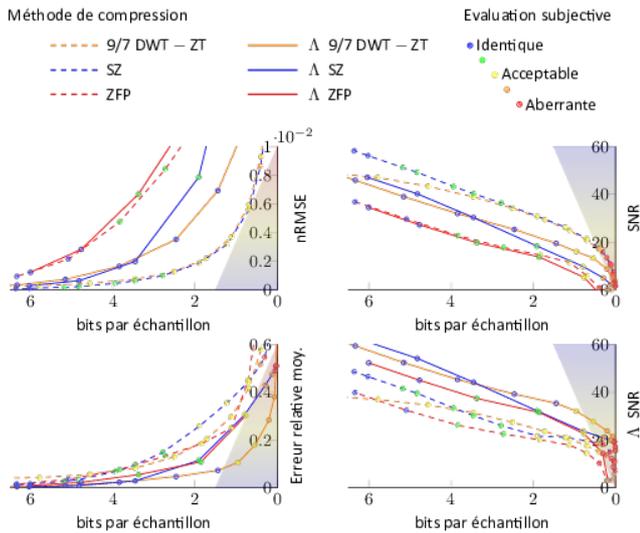


FIGURE 2 – Courbes débit/distorsion obtenues avec SZ, ZFP, et notre codeur progressif (9/7 DWT - ZT) pour des données de perméabilité décompressées à différentes précisions, avec ou sans compandeur (Λ). En ordonnée : différentes métriques objectives (le triangle dégradé symbolise l'échelle de qualité (la base bleue pour les hautes qualités, la pointe rouge pour les plus faibles). La validation subjective des simulations est indiquée par les marqueurs de couleur.

3 Résultats

Nous présentons ici les résultats de compression obtenus avec le codeur progressif proposé, ainsi que SZ et ZFP à différentes précisions. Ces trois codeurs sont utilisés avec ou sans compandeur. Les résultats sont résumés par des courbes débit/distorsion selon 4 mesures (figure 2). Indépendamment, la justesse des simulations obtenues à partir de ces données décompressées a été validée par l'attribution d'un score subjectif (échelonné en cinq valeurs, d'« identique » à « aberrant », et symbolisé par des marqueurs colorés sur les courbes). Nous constatons que :

- les métriques objectives standards ou classiques (nRMSE, SNR) reflètent mal l'évaluation subjective ;
- l'usage d'un compandeur améliore les performances objectives (erreur relative moyenne, Λ -SNR) et subjectives des trois codeurs (en traits pleins) ;
- Sur toute la gamme des résultats de simulation jugés « identiques » aux résultats de référence (marqueurs de couleur bleu), notre méthode atteint des débits inférieurs aux algorithmes SZ [2] et ZFP [4] (voir Λ -SNR). En d'autres termes, de meilleurs taux de compression peuvent être atteints pour une fidélité optimale des résultats de simulation.

4 Conclusion et perspectives

Nous montrons dans cette étude qu'il est possible de maintenir des résultats de simulation quasi-identiques à ceux obtenus à pleine précision tout en réduisant de manière notable la taille des données. Cela s'obtient en jouant sur la résolution et la précision, en exploitant conjointe-

ment la multi-résolution des ondelettes et la profondeur binaire des données numériques par l'utilisation du codage zerotree et d'un compandeur logarithmique. Il est donc possible de réduire drastiquement la quantité de données binaires à traiter, tout en conservant la pleine qualité de la simulation. De plus nous confirmons que l'utilisation d'un compandeur permet d'améliorer la compression de données scientifiques ([3]), dont la grande dynamique rend la compression difficile.

Références

- [1] Allison H. Baker, Dorit M. Hammerling, Sheri A. Mickelson, Haiying Xu, Martin B. Stolpe, Phillippe Naveau, Ben Sanderson, Imme Ebert-Uphoff, Savini Samarasinghe, Francesco De Simone, Francesco Carbone, Christian N. Gencarelli, John M. Dennis, Jennifer E. Kay, and Peter Lindstrom. Evaluating lossy data compression on climate simulation data within a large ensemble. *Geosci. Model Dev.*, 9(12) :4381–4403, dec 2016.
- [2] Franck Cappello, Sheng Di, Sihuan Li, Xin Liang, Ali Murat Gok, Dingwen Tao, Chun Hong Yoon, Xin-Chuan Wu, Yuri Alexeev, and Frederic T. Chong. Use cases of lossy compression for floating-point data in scientific data sets. *Int. J. High Perform. Comput. Appl.*, jul 2019.
- [3] Xin Liang, Sheng Di, Dingwen Tao, Zizhong Chen, and Franck Cappello. An efficient transformation scheme for lossy data compression with point-wise relative error bound. In *2018 IEEE International Conference on Cluster Computing (CLUSTER)*. IEEE, sep 2018.
- [4] Peter Lindstrom. Fixed-rate compressed floating-point arrays. *IEEE Trans. Visual Comput. Graph.*, 20(12) :2674–2683, 12 2014.
- [5] Jean-Luc Peyrot, Laurent Duval, Frédéric Payan, Lauriane Bouard, Lénaïc Chizat, Sébastien Schneider, and Marc Antonini. HexaShrink, an exact scalable framework for hexahedral meshes with attributes and discontinuities : multiresolution rendering and storage of geoscience models. *Computat. Geosci.*, 23 :723–743, Aug. 2019.