HPC Implementations on SEM46: A 3D Modeling and Inversion Code for Anisotropic Visco-Elastic Coupled Acoustic Media

Alizia Tarayoun, Romain Brossier, Jian Cao, Stéphan Jauré, Soline Laforêt, Ludovic Métivier

HPC implementations on SEM46: a 3D modeling and inversion code for anisotropic visco-elastic coupled acoustic media

Alizia Tarayoun[1*], Romain Brossier[1], Jian Cao[1], Stéphan Jauré[2], Soline Laforêt[2], Ludovic Métivier[1,3]

[1] Univ. Grenoble Alpes, ISTerre, Grenoble, France
[2] ATOS, Grenoble, France
[3] CNRS, Univ. Grenoble Alpes, LJK, Grenoble, France

Monday 30th May, 2022

Main objectives

A 3D modeling and inversion code named SEM46 has been developed since few years in the SEISCOPE project. Currently this code implements a coupled visco-elastic/acoustic wave propagation, with generic types of anisotropy in the solid part. For practical usage, such 3D modeling and inversion code needs to be able to exploit efficiently the performance of modern large scale computing facilities. This study aims at presenting the recent HPC implementations and performance optimizations done on this code to achieve this goal.

New aspects covered

(1) Presentation of several optimizations and quantification of the performance gains.
(2) New porting on the TGCC center (AMD Rome partition) and new parallel efficiency assessed to compare acoustic, elastic and acoustic-elastic coupling modeling.

Summary (200 words)

Full Waveform Inversion (FWI) is a high-resolution seismic imaging technique dedicated to the reconstruction of the mechanical properties of the Earth's subsurface. With the developments of hybrid HPC platforms and upcoming Exascale architectures, FWI implementations needs to be adapted to exploit such improved hardware. In this study we present the main HPC related implementations of a code named SEM46, developed in the SEISCOPE project and designed for crustal-scale exploration. It allows to move towards an anisotropic viscoelastic and acoustic engine with an affordable computational cost thanks to several key elements: flexible Cartesian-based deformed mesh, two Message Passing Interface (MPI)-based parallelism levels - one over seismic sources - one over domain decomposition, several choices to store and/or re-compute the incident fields needed for the gradient building allowing an adaptation to the architecture characteristics and finally an optimized modeling kernel for the product of the displacement vector by the stiffness matrix. Different optimizations are investigated, making it possible to significantly improve the computational performances. They combine vectorization and re-arrangement on specific loops to decrease the computational time, and optimizations such as new computations on the fly and use of a fixed point coding approach to decrease the memory bandwidth pressure.

Introduction

Full Waveform Inversion (FWI) is a high-resolution seismic imaging technique dedicated to the reconstruction of the mechanical properties of the Earth's subsurface (see Virieux et al., 2017, for a review). It is based on the iterative minimization of a misfit function related to the distance between observed and calculated seismic data, over a space of model parameters describing the medium (tens of millions of degree of freedom can be explored). FWI offers the possibility to extract high resolution quantitative images of the subsurface, down to half the local wavelength. It is challenging from the high performance computing side, making the total size of a target, the number of sources and the highest frequencies that we can explore, a limitation. This is especially true when moving towards 3D viscoelastic approximation of the wave equation including complex topographies or bathymetries and seismic anisotropy.

In this work, we present latest functionalities and developements of an efficient and accurate 3D modeling and inversion tool that we have developed in the frame of the SEISCOPE project, named *SEM46* (Spectral Element Method for Seismic Imaging at eXploration scale), based on a spectral-element time-domain formulation (Komatitsch and Tromp, 1999). The computation of the gradient of the FWI misfit function is classically performed through the adjoint-state approach (Plessix, 2006). The code has been designed to tackle elastic media reconstruction considering anisotropy (isotropic, VTI, TTI, orthorombic and triclinic) and attenuation. It allows to tackle realistic set-up such as fluid-solid coupling (Cao et al., 2022) or complex topography. The code has been successfully applied to FWI at various scales on field datasets (Trinh et al., 2019a; Lu et al., 2020; Irnaka et al., 2022).

In the following, after a short presentation of the FWI theory, we present the key functionalities making SEM46 an efficient code. Then, we discuss its parallel implementation and illustrate its efficiency on different HPC platforms. Finally we explore implementation optimizations to improve code performances.

FWI formulation

The set of elastic wave equations associated with the modeling part is :

$$\begin{cases} \rho(\mathbf{x})\partial_{tt}u_i(\mathbf{x},t) &= \partial_j\sigma_{ij}(\mathbf{x},t) + f_i(\mathbf{x},t), \\ \sigma_{ij}(\mathbf{x},t) &= C_{ijkl}(\mathbf{x})\varepsilon_{kl}(\mathbf{x},t) + \mathscr{T}_{ij}(\mathbf{x},t), \end{cases} \tag{1}$$

where the density is denoted by ρ, the displacement by u, σ and ε are the stress and strain tensors, respectively, f is the external force and $\mathscr{T}$ denotes the possible stress failure. C_{ijkl} is the elastic stiffness tensor. Following the Voigt indexing and matrix notations, the second-order elastic-wave equation of the displacement field can be written as:

$$\rho\partial_{tt}\mathbf{u} = DCD^T\mathbf{u} + \mathbf{S}, \tag{2}$$

where the spatial derivative operator in the Cartesian space is denoted by D (D^T is for the transpose operator), the elastic stiffness tensor by C and $\mathbf{S}$ the source term. For more details especially on anisotropic visco-elastic or fluid-solid coupling notation, the reader is referred to Trinh et al. (2019b) and Cao et al. (2022), respectively.

The inverse problem consists of minimizing over a set of model ($\mathbf{m}$), the misfit between calculated ($\mathbf{d}_{cal}(\mathbf{m})$) and observed ($\mathbf{d}_{obs}$) seismic data at receiver locations, by following its gradient direction to update appropriately the model parameters. The standard least-squares misfit function is formulated as follows:

$$\chi(\mathbf{m}) = \frac{1}{2}\|\mathbf{d}_{cal}(\mathbf{m}) - \mathbf{d}_{obs}\|^2. \tag{3}$$

A summation over sources and receivers and an integral over time is implicitly considered in this formulation. The gradient of the misfit function with respect to the independent components of the unrelaxed stiffness components (C_{IJ}) is expressed as the zero-lag cross-correlation of the adjoint strain field ($\bar{\boldsymbol{\varepsilon}}$) and the incident strain field ($\boldsymbol{\varepsilon}$):

$$\frac{\partial\chi(\mathbf{m})}{\partial C_{IJ}} = \left(\bar{\boldsymbol{\varepsilon}}, \frac{\partial C}{\partial C_{IJ}}\boldsymbol{\varepsilon}\right)_t. \tag{4}$$

where $(.,.)_t$ stands for integration over time.

HPC oriented wave propagation implementations

The set of Partial Differential Equations describing the physics of wave propagation (2) is solved through a spectral element method. The physical domain is discretized into a set of nonoverlapping hexahedral

Algorithm 1: Forward problem scheme

for $it = 1, \cdots, nt$ **do**
  Newmark Prediction phase
  $\mathbf{u}^{it} = \mathbf{u}^{it-1} + \Delta t \mathbf{v}^{it-1} + \frac{\Delta t^2}{2}\mathbf{a}^{it-1}$;
  $\mathbf{v}^{it-1/2} = \mathbf{v}^{it-1} + \frac{\Delta t^2}{2}\mathbf{a}^{it-1}$;
  Stiffness matrix-vector product
  $\boldsymbol{\varepsilon}^{it} = \mathbf{D}\mathbf{u}^{it}$;
  $\boldsymbol{\sigma}^{it} = C\boldsymbol{\varepsilon}^{it}$;
  $\mathbf{b}^{it} = \mathbf{D}^T\boldsymbol{\sigma}^{it}$;
  Product with inverse mass matrix + source term
  $\mathbf{a}^{it} = -\mathbf{M}^{-1}\mathbf{b}^{it} + \mathbf{S}^{it}$;
  Newmark Correction phase
  $\mathbf{v}^{it} = \mathbf{v}^{it-1/2} + \frac{\Delta t}{2}\mathbf{a}^{it}$;
end

elements. A set of Gauss-Lobatto-Legendre (GLL) integration points is attached to each hexahedron where interpolant basis functions (Lagrange polynomials) are defined. Fourth or fifth order interpolation can be used in SEM46, as it has been shown to provide a good compromise between numerical accuracy and CFL limitation (Komatitsch and Tromp, 1999).

In order to combine an accurate representation of complex topographic and bathymetric interfaces, allowed by finite element mesh, with the easiness of implementation of Finite Difference methods, SEM46 considers a regular Cartesian-based deformed mesh. Thus, the spatial position of each element can be obtained directly from its x, y, z indices, without any additional cost that is otherwise mandatory for unstructured mesh (i.e. neighborhood look-up tables or global mesh searching steps). The complex topography representation is done through a vertically deformed mesh using possibly high-order shape function polynomials.

In addition, variable element-size mesh can be used to reduce the computational cost in case of spatially varying velocity as it is the case in most geophysical targets with velocity increasing with depth. The reduction of the number of elements systematically decreases the computational cost by the same factor, or even more if it induces the relaxation of the CFL stability condition of the explicit time-marching scheme.

Algorithm 1 presents the main steps of the forward modeling part, relying on a second-order explicit Newmark sheme. The most computationally intensive part of this algorithm is related to the stiffness matrix-vector product. This heavy task is factorized into three matrix-free sub-steps. The product with the stiffness tensor is quite challenging for complex anisotropy as 21 stiffness coefficients have to be loaded for each GLL point (see later the optimization of this step). Our implementation relies on Deville et al. (2002) approach, which takes benefit from the tensorial properties of hexaedral elements, the optimization of cache usage, and the combination of efficient loop vectorization and manual loop unrolling. Compared to an explicit sparse matrix-vector product, this matrix free implementation decreases the computational cost by about one order of magnitude (Trinh et al., 2017).

As shown in eq. (4), the gradient computation requires to access simultaneously two wavefields: the incident wavefield backward in time and the adjoint wavefield computed from final-conditions in backward-time. The strategy used to store/access the incident field involves the memory space, IO traffic and computational cost problematics. In the code, two implementations are available. The first one is to store a time-decimated version of the whole incident field in core-memory or disk during the forward simulation, and read/use it during the adjoint simulation at the cost of enough memory and/or high IOs requirement. The second one is to recompute the incident field from the stored wavefield at the boundaries using the reverse propagation. For attenuative media, a Checkpointing-Assisted Reverse-Forward Simulation method (CARFS) has been developed to achieve a better compromise between the memory requirement and the recomputation ratio by doing a smart decision between a reverse propagation and

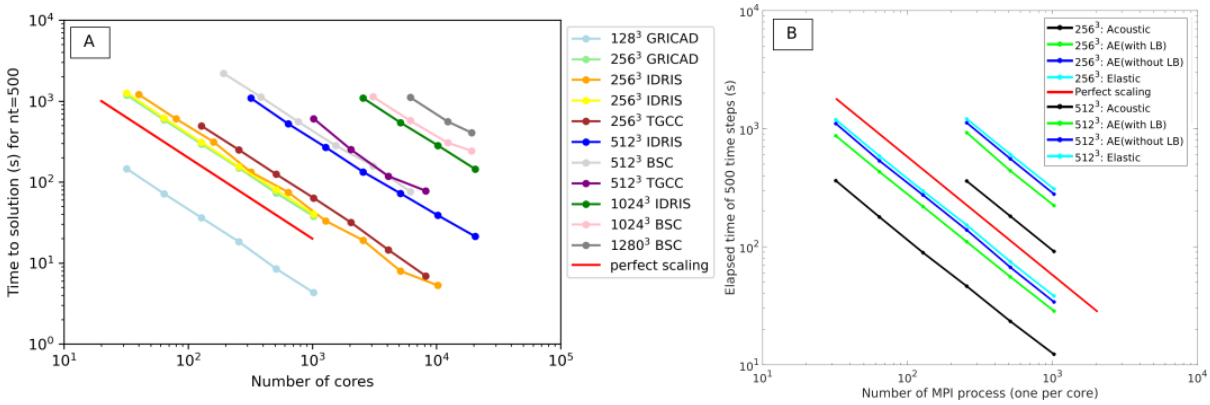

Figure 1: A. Strong scaling on elastic forward modeling on different CPUs-based clusters: a local one in GRICAD center (Dahu), on 2 national centers IDRIS and TGCC (Jean-Zay and Irene Rome machines, respectively) and on 1 international center BSC (MareNostrum4 machine) B. Parallel efficiency for fully acoustic modeling, fully elastic modeling, fluid-solid (AE) coupled modeling (without and with a cost-dependent domain decomposition, LB is for load balancing) on Dahu machine. For both A. and B. panels, the size X^3 of the legend represents the number X of P4 elements in each space direction.

a use of checkpoints based on an energy measure (Yang et al., 2016). In practice, for most of our last applications at various scales, thanks to the HPC platforms improvements with larger node memory and faster access with local SSD disks, the first implementation appears to be the most efficient.

Parallelisation and efficiency

Our implementation relies on a two-level MPI-based parallelization. The inner level is designed on a Cartesian-based domain decomposition, which is performed naturally thanks to the Cartesian-based mesh and provides an efficient load-balancing. The MPI communications are only performed at each domain interface. The outer MPI-level is performed over seismic shots to gather the misfit function and gradient of the different shots, ending up with almost perfect performance on this level.

The parallel efficiency of SEM46 is thus evaluated on the inner level for the domain decomposition. The efficiency has been assessed on different HPC platforms including different CPUs generations and network bands (Fig. 1A). In addition, the fluid-solid coupled case which involves different wave-equations in each target part has been optimized with a wave-equation-cost-dependent domain decomposition, in which the decomposition is performed according to the computational cost rather than the number of elements. The efficiency of this fluid-solid coupled case is displayed in Figure 1B. Up to now, the largest scalability test has been perfomed with a mesh of 1280^3 elements with interpolation order of 4 on the MareNostrum4 machine in Barcelona using 19200 cores (400 nodes).

Performance analyses of SEM46 has been also performed during the 2019 Performance Optimisation and Productivity audit on MareNostrum4 machine on a purely elastic case. The result of this HPC test has shown a parallel code efficiency greater than 94% in modeling and inversion cases. As the computational load per core is well designed, the room for improvement is mostly related to the memory bandwidth side. Indeed, a current limitation for higher performance is linked to the stiffness matrix product where the C tensor (eq. 2), which is composed of 21 coefficients in an anisotropic case, needs to be loaded on each control point and used only once. This leaves improvement possibilities that we address in the following.

Performance improvements: Stiffness matrix-vector product and Newmark scheme

In the frame of the ENERXICO project, several improvements have been performed for the anisotropic orthorombic elastic engine regarding the stiffness matrix product and the Newmark scheme. In order to have better code performance, the idea is to improve arithmetic density by increasing the computational load on CPUs and by decreasing the memory bandwidth needs.
The stiffness matrix product optimization consists in several improvements:

- Vectorization: The matrix-vector product, although already vectorized with Deville et al. (2002)'s approach, can be pushed further by vectorizing computations over multiple elements rather than over a single element integration points. The idea is to optimize cache usages; instead of looping over GLL points one element at a time, to loop over the GLL points for a batch of elements.

Execution types	Forward kernel	Inversion kernel	Newmark scheme update	Total gain
Optimized with vectorization and memory bandwidth (i)	27 %	32 %	39 %	28 %
Optimized with vectorization and memory bandwidth (i) & (ii)	40 %	40 %	41 %	35 %

Table 1: Performance improvement of execution times for three different execution types

- Memory bandwidth (i): depending on the anisotropy type some components of the C tensor can be computed from others, it is not necessary to load every component. The idea is to load only the necessary floats to be able to compute on the fly, within the kernel, the others components of interest. Doing so, only 2 and 5 floats have to be loaded for the isotropic and the VTI anisotropic case, respectively.

- Memory bandwidth (ii) : the C tensor components values do not have a high dynamic range and are not known with a high level of precision in the frame of FWI. Instead of representing them with single precision float, a fixed point coding strategy can be used. The idea is to decrease the data size that need to be loaded. The first step is to encode the initial dataset of float values by decomposing them into N possible integers values over a chosen number of allocated bits. In this work, the mapping is done over 16 bits meaning 65536 possible integer values. A first bijection is made to convert the float values into a [0,1] dataset. A second one is performed to map them into integers coded values. Only the later are used in the compute kernel through a vectorized decoder. The decoding consists in three steps. First, to convert the 16 bits to 32 bits integers. Second, to convert the 32 bits integers into 32 bits floats. Third, thanks to a Fused Multiply Add operation, to perform a bijection between the [0,1] floats dataset into their original space.

The second optimization is related to the time-marching Newmark scheme. In the original version of the algorithm 1, multiple loops over elements occurs for most of the steps (prediction, stiffness matrix-vector product, product with inverse mass, correction), while relatively simple mathematic operations are applied. The idea of improvement is to gather the prediction, product with inverse mass matrix and the corrections steps in the same loop in order to stream to arrays only once. Therefore, the kernel is expressed as the resolution of the acceleration first then the velocity and finally the displacement with a prediction phase used alone only once at the first iteration.

The quantification of performance improvement of those steps is shown in Table 1.

Conclusions

SEM46 is a 3D full waveform modeling and inversion code for visco-elastic and coupled visco-elastic/acoustic media, handling generic anisotropy in the elastic domain. We have presented the key points related to its implementation to make it possible to exploit modern HPC capabilities. The more recent optimizations reduce the memory bandwidth pressure thanks to a fixed-point precision handling of the stiffness tensor coefficients and a reorganization of the main time-stepping loop.

Acknowledgements

The research leading to these results have been partially carried in the frame of the ENERXICO project (https://enerxico-project.eu/) and the Performance Optimisation and Productivity centre (https://pop-coe.eu/) that received funding from the European Union's Horizon 2020 research and innovation programme. This study has also been partially funded by the SEISCOPE consortium (http://seiscope2.osug.fr), sponsored by AKERBP, CGG, CHEVRON, EXXON-MOBIL, GEOLINKS, JGI, PETROBRAS, SHELL, SINOPEC, SISPROBE and TOTALENERGIES. This study was granted access to the HPC resources provided by the GRICAD infrastructure (https://gricad.univ-grenoble-alpes.fr), Cray Marketing Partner Network (https://partners.cray.com) and IDRIS/TGCC under the allocation 046091 made by GENCI.

References

Cao, J., Brossier, R., Górszczyk, A., Métivier, L. and Virieux, J. [2022] 3D multi-parameter full-waveform inversion for ocean-bottom seismic data using an efficient fluid-solid coupled spectral-element solver. *Geophysical Journal International*, **229**(1), 671–703.

Deville, M., Fischer, P. and Mund, E. [2002] *High Order Methods for Incompressible Fluid Flow*. Cambridge University Press.

Irnaka, T.M., Brossier, R., Metivier, L., Bohlen, T. and Pan, Y. [2022] 3D Multi-component Full Waveform Inversion for Shallow-Seismic Target: Ettlingen Line Case Study. *Geophysical Journal International*, **229**(2), 1017–1040.

Komatitsch, D. and Tromp, J. [1999] Introduction to the spectral element method for three-dimensional seismic wave propagation. *Geophysical Journal International*, **139**(3), 806–822.

Lu, Y., Stehly, L., Brossier, R., Paul, A. and Group, A.W. [2020] Imaging Alpine crust using ambient noise wave-equation tomography. *Geophysical Journal International*, **222**(1), 69–85.

Plessix, R.E. [2006] A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, **167**(2), 495–503.

Trinh, P., Brossier, R., Lemaistre, L., Métivier, L. and Virieux, J. [2019a] 3D Elastic FWI with a Non-Linear Model Constraint: Application to a Real Complex Onshore Dataset. In: *Expanded Abstracts, 81th Annual EAGE Conference & Exhibition, London*. EAGE, Th P01 07.

Trinh, P.T., Brossier, R., Métivier, L., Tavard, L. and Virieux, J. [2019b] Efficient 3D time-domain elastic and viscoelastic Full Waveform Inversion using a spectral-element method on flexible Cartesian-based mesh. *Geophysics*, **84**(1), R75–R97.

Trinh, P.T., Brossier, R., Métivier, L., Virieux, J. and Wellington, P. [2017] Bessel smoothing filter for spectral element mesh. *Geophysical Journal International*, **209**(3), 1489–1512.

Virieux, J., Asnaashari, A., Brossier, R., Métivier, L., Ribodetti, A. and Zhou, W. [2017] An introduction to Full Waveform Inversion. In: Grechka, V. and Wapenaar, K. (Eds.) *Encyclopedia of Exploration Geophysics*, Society of Exploration Geophysics, R1–1–R1–40.

Yang, P., Brossier, R., Métivier, L. and Virieux, J. [2016] Wavefield reconstruction in attenuating media: A Checkpointing-assisted reverse-forward simulation method. *Geophysics*, **81**(6), R349–R362.