

A massively parallel frequency-domain full-waveform inversion algorithm for imaging acoustic media: application to a dense OBS data set.

Florent Sourbier*, Stéphane Operto and Jean Virieux, Géosciences Azur, Sophia-Antipolis, France, P. Amestoy, ENSEEIHT-IRIT, France, J. -Y L'Excellent, ENS-Lyon - INRIA, France

SUMMARY

We present a massively parallel algorithm for distributed-memory platform to perform 2D acoustic frequency-domain Full-Waveform Inversion (FWI) of global offset seismic data. Our code is written in Fortran 90 and uses Message Passing Interface (MPI) for parallelism. The linearized inverse problem is solved by a classical gradient method which consists in finding a model perturbation which minimizes the least-squares cost function. A key feature of our code is the direct solver used for frequency-domain full-waveform modeling. Frequency-domain full-waveform modeling requires to solve a huge sparse system of linear equations with complex coefficients whose multiple Right-Hand Side (RHS) terms are the seismic sources and solutions of the system are the pressure wavefield. We use the MUlti frontal Massively Parallel direct Solver (MUMPS) for distributed-memory computer to solve this system (LU factorization). MUMPS includes a functionality to perform parallel multiple-shot resolutions once the LU factors have been distributed over the processors. Once the resolution phase is completed, the multiple-shot solutions are left in core and distributed over the processors allowing straight forward parallelization of subsequent tasks in the FWI program such as gradient computation. We first validated our algorithm with a realistic synthetic test consisting of a dip section of the Overthrust model. We first inverted 7 frequencies successively following a multiresolution strategy. Second, two groups of 4 and 3 frequencies respectively were inverted successively. A fastest and more robust convergence was observed using the first multiresolution strategy. This multiresolution strategy was applied to a real wide-angle 2D dense Ocean Bottom Seismometers (OBS) data set (100 OBS gathers) recorded with a 1-km spacing in the eastern-Nankai trough (Japan) to image the deep crustal structure of the subduction system (105 km X 25 km). A speedup of 14 was obtained using 40 processors of a PC cluster and an Infiniband interconnect. The same parallel strategies was implemented in a 3D full-waveform inversion algorithm.

INTRODUCTION

Full-waveform inversion (FWI) has prompted renewed interest these last years. Especially, the frequency domain formulation of FWI has been shown to be effective to build accurate velocity models for complex structures from global offset acquisition geometries (Pratt, 2004). The potential interest of such approach is to exploit the full aperture range spanned by global offset geometries to image a broad and continuous range of wavelengths in the medium including large to middle wavelengths. For the global offset acquisition geometry, the frequency-domain approach of FWI has been shown to be efficient for several reasons: first, only few discrete frequencies are necessary to develop a reliable image of the medium thanks to the wave number redundancy provided by multi-aperture geometries (Sirgue and Pratt, 2004; Pratt and Worthington, 1990; Pratt, 1999). Second, proceeding sequentially from the low to the high frequencies defines a multi resolution imaging strategy which helps to mitigate the non linearity of the inverse problem. The inverse problem is solved by an iterative local linearized approach using a classic gradient method. At each iteration, the residual wavefield is minimized in the least-squares sense. This process is iterated non linearly, which means that the final model of the current iteration is used as a starting model for the subsequent iteration. Full-waveform modeling of the inverted frequencies is performed with a finite-difference frequency-domain method (Jo et al., 1996; Hustedt

et al., 2004). In the frequency-domain, solving the wave equation requires the resolution of a large sparse system of linear equations. If this system can be solved with a direct solver, the solution for multiple sources (i.e. multiple RHS vectors) can be obtained efficiently by substitutions once the matrix was LU factorized (Marfurt, 1984).

Although the frequency-domain formulation of FWI is attractive for three abovementioned reasons, the computational cost of FWI remains high especially for 2D multiparameter elastic or 3D acoustic inversions. Therefore, it is essential that FWI algorithms take advantage of parallel computing facilities. We present in this paper how parallelism for distributed-memory platform can be implemented in frequency-domain FWI and quantify the speed-up that we obtained with a 2D real data case study. The structure of the parallel FWI algorithm is strongly driven by the parallel functionalities implemented in the MUMPS direct solver (Amestoy et al., 2006, 2007) that we used to solve the forward problem.

FREQUENCY-DOMAIN FULL-WAVEFORM INVERSION

Theoretical aspects

Let's consider the 2D frequency-domain acoustic wave equation

$$\frac{\omega^2}{\kappa(x,z)} p(x,z,\omega) + \frac{1}{\xi_x(x)} \frac{\partial}{\partial x} \left[\frac{b(x,z)}{\xi_x(x)} \frac{\partial p(x,z,\omega)}{\partial x} \right] + \frac{1}{\xi_z(z)} \frac{\partial}{\partial z} \left[\frac{b(x,z)}{\xi_z(z)} \frac{\partial p(x,z,\omega)}{\partial z} \right] = -s(x,z,\omega), \quad (1)$$

where p is pressure, b is buoyancy, ρ is density, ω is angular frequency and s is the source. ξ are 1D damping functions for absorbing PML conditions (Berenger, 1994).

Since the relationship between the pressure wavefield and the source is linear, equation 1 can be recast in matrix form

$$\mathbf{A}(\mathbf{m})\mathbf{p} = \mathbf{s} \quad (2)$$

where A is the so-called impedance matrix whose coefficients depend non linearly on the model parameters \mathbf{m} .

The inverse problem related to the forward problem 2 is solved with a weighted least-squares gradient method. The weighted least-squares cost function is given by

$$\mathcal{C}(\mathbf{m}) = \Delta \mathbf{d}^\dagger \mathbf{W}_d \Delta \mathbf{d}, \quad (3)$$

where $\Delta \mathbf{d}$ is the misfit vector (the difference between the observed data and the data computed in model \mathbf{m}). The superscript \dagger indicates the adjoint (transpose conjugate). \mathbf{W}_d is a weighting operator applied to the data which scales the relative contribution of each component of the vector $\Delta \mathbf{d}$ in the inversion. Minimization of the cost function leads to the following solution for the i^{th} component of the model perturbation $\Delta \mathbf{m}$ after scaling and smoothing of the gradient (see Pratt et al. (1998); Ravaut et al. (2004); Operto et al. (2006) for more details).

$$\Delta \mathbf{m}_i = -\alpha (\text{diag} \mathbf{H}_a + \epsilon I)^{-1} \mathcal{G}_m \text{Re} \left\{ \mathbf{p}^t \left[\frac{\partial \mathbf{A}^t}{\partial m_i} \right] \mathbf{A}^{-1} \mathbf{W}_d \Delta \mathbf{d}^* \right\} \quad (4)$$

A massively parallel frequency-domain full-waveform inversion algorithm

where $\text{diag } \mathbf{H}_a = \text{diagRe}\{ \mathbf{J}^T \mathbf{W}_d \mathbf{J}^* \}$ denotes the diagonal elements of the weighted approximate Hessian \mathbf{H}_a and \mathbf{J} denotes the sensitivity matrix. According to equation 4, one element of the sensitivity matrix is given by

$$\mathbf{J}_{k(m,n),i} = \mathbf{p}_m^t \left[\frac{\partial \mathbf{A}^t}{\partial m_i} \right] \mathbf{A}^{-1} \delta_n. \quad (5)$$

where $k(m,n)$ denotes a source-receiver pair of the acquisition system, m and n denote a shot and a receiver position respectively. δ_n is an impulsive source located at the receiver position n .

The diagonal of the approximate Hessian provides a preconditioner of the gradient which properly scales the perturbation model (Shin et al., 2001). The damping parameter ϵ is used to avoid numerical instability (i.e. division by zero). The matrix \mathcal{G}_m is a smoothing regularization operator. It was implemented in the form of a 2-D Gaussian spatial filter whose correlation lengths are adapted to the inverted frequency component Ravaut et al. (2004). An amplitude gain with offset applied to each seismic trace was used for the operator \mathbf{W}_d (Operto et al., 2006). The inversion code can be applied to vertical geophone data or to hydrophone data generated by explosive sources. Indeed, vertical geophone data can be processed as pressure data thanks to the reciprocity principle (Operto et al., 2006). The inversion is applied in cascade to several groups of discrete frequencies. All the frequencies of one group are inverted simultaneously. The final model obtained close to inversion of one group of frequencies is used as a starting model for the next group of frequencies. For each frequency group, several iterations can be computed.

Equation 4 describes the imaging principle of FWI based on generalized diffraction tomography (Lailly, 1984; ?): the model perturbation is computed by zero-lag correlation between the forward wavefields p with the residuals backpropagated in time, $\mathbf{A}^{-1} \delta \mathbf{d}^*$. Contribution of several shots is taken into account by simple summation. Gradient computation requires two forward problems per shot, one to compute the incident wavefield \mathbf{p} and one to compute backpropagated residuals. The scaling of the gradient in equation 4 requires the explicit estimation of \mathbf{J} and therefore one resolution per non redundant shot and receiver position (see equation 5). In our algorithm, the scaling of the gradient could be estimated once per frequency before the first iteration and kept constant over iterations or re-computed at each iteration. The term $\frac{\partial \mathbf{A}}{\partial m_i}$ is the radiation pattern of the diffraction by the model parameter m_i . In the case of the P-wave velocity, this radiation pattern is that of an explosion. In other words, this matrix reduces to one scalar. In the case of density, the radiation pattern is that of a vertical force for a shot located at the vertices of the model parameter. The source term can be estimated in the FWI algorithm by solving a linear inverse problem (Pratt, 1999).

Parallel implementation of FWI

To solve the system equation 2 for multiple RHSs, we used the massively parallel direct solver MUMPS (Amestoy et al., 2006, 2007) for distributed-memory computer. The direct solver is based on a multifrontal method which is suited for parallelization (Duff and Reid, 1983). The MUMPS algorithm is subdivided in three main steps. First a sequential symbolic analysis step performs re-ordering of the matrix coefficients to minimize the fill-in of the matrix during the subsequent factorization and an estimation of the graph of the matrix. Second, the numerical factorization provides LU factors distributed over all the processors. Third, the resolution is performed in parallel for multiple sources (i.e. multiple RHS vectors) taking advantage the BLAS3 (Basic Linear Algebra subprograms) library. The source vectors for the resolution phase are provided in sparse format on the host processor. After resolution, the multiple solutions are left distributed over processors and each processor stores a spatial sub-domain of all the solutions. We take advantage of the distributed in-core storage of the MUMPS solutions to compute in parallel the gradient and the diagonal Hessian. The gradient of the cost function and the diagonal Hessian

are basically computed by a weighted summation of the forward problem solutions, equation 4, where the weights contain the data residuals and the coefficients of $\frac{\partial \mathbf{A}^t}{\partial m_i}$. Each processor computes a subdomain of the gradient and of the diagonal Hessian according to the domain decomposition of the forward problem solutions performed by MUMPS. At the end of the summations, the gradient and the diagonal Hessian are centralized on the host processor with a collective communication and the gradient is scaled by the diagonal Hessian. The algorithm proceeds with the search of the optimal step length by parabolic fitting. The search of the step length requires to perform two times a factorization and a multi-RHS resolution in order to compute the cost function for two guesses of the step length. These tasks are computed with the parallel functionalities of MUMPS described above.

The parallel FWI algorithm is summarized in Figure 1.

Loop over group of frequencies [lfreqgroup/Nfreqgroup]

Loop over iterations [it/nitermax]

Initialization of gradient, diagonal Hessian and cost function

Loop over frequencies in one group [lfreq/nfreq]

Read starting model and build impedance matrix

Parallel factorization with MUMPS

Hessian computation (if it=1)

Allocate RHS arrays on P0

Allocate solution arrays on P[i],i=1,Nproc

Build Nshot-receiver RHSs on P0

Parallel multi-RHS resolution with MUMPS

Compute sub-domains of diagonal Hessian on P[i],i=1,Nproc

Gradient computation

Allocate RHS arrays on P0

Allocate solution arrays on P[i],i=1,Nproc

Build Nshot RHSs for shot positions on P0

Parallel multi-RHS resolution with MUMPS

Extract wavefields at receiver positions on P[i],i=1,Nproc

Compute data residuals and partial RMS on P[i],i=1,Nproc

Estimate source

Build Nshot RHS for residual positions on P0

Parallel multi-RHS resolution with MUMPS

Compute sub-domains of gradient on P[i],i=1,Nproc

End do over frequencies

Reduce RMS on P0

Centralize the gradient and the diagonal Hessian on P0

Scale the gradient by the diagonal Hessian on P0

Compute step length α

Updated the velocity model on P0

End do over iterations

End do of over group of frequencies

Figure 1: Parallel algorithm for FWI. The tasks in red and black are computed on multi and single processors respectively. $P[i]$ means processor number i . RMS is the root-mean squares.

APPLICATION TO THE SEG/EAGE OVERTHRUST MODEL

We first verified the algorithm using a dip section of the SEG/EAGE Overthrust model (Figure 3a). The test was run on 12 processors of a HP DL 145G2 cluster made of Opteron dual-core bi-processor 2.4 GHz 64 bits nodes with a 921.6 Gflops total peak power. The interconnect is Infiniband. The model grid is 801 x 187 with a 25-m grid interval. Seven frequencies ranging between 3.5 Hz and 20 Hz were inverted. The starting model was obtained by smoothing the true model with a 2D Gaussian filter of vertical and horizontal correlation lengths of 500 m (Figure 3b). During the first test (Test 1), the seven frequencies were inverted successively by proceeding from the lower to the higher frequency. The source wavelet was a Ricker with a 10-Hz dominant frequency. Fifteen iterations were computed

A massively parallel frequency-domain full-waveform inversion algorithm

per frequency leading to 105 velocity models of increasing resolution. The real time required for 15 iterations of 1 frequency inversion was around 10 mn. The final model is shown in Figure 3c. During the last two tests, two group of frequencies were successively inverted (frequencies of one group are inverted simultaneously). The 4 frequencies of the first group range between 3.5 Hz and 10 Hz and the 3 frequencies of the second group range between 13 Hz and 20 Hz. The source wavelet was respectively a Ricker wavelet (same that of Test 1) and a Dirac wavelet for the Tests 2 and 3 respectively. Two different wavelets were used to assess the impact of the relative spectral amplitudes in the multi-frequency FWI behavior. Fifty iterations were necessary for each group of frequencies to obtain a final velocity model with the same resolution that of Test 1 suggesting that the multi-resolution strategy used for Test 1 provides the best preconditioning of the FWI (Figures 3d and 3e). The final model obtained with a Dirac wavelet exhibits a slightly better resolution than the one obtained with a Ricker wavelet. This illustrates the higher weight of the low and high frequencies in FWI when a Dirac wavelet is used.

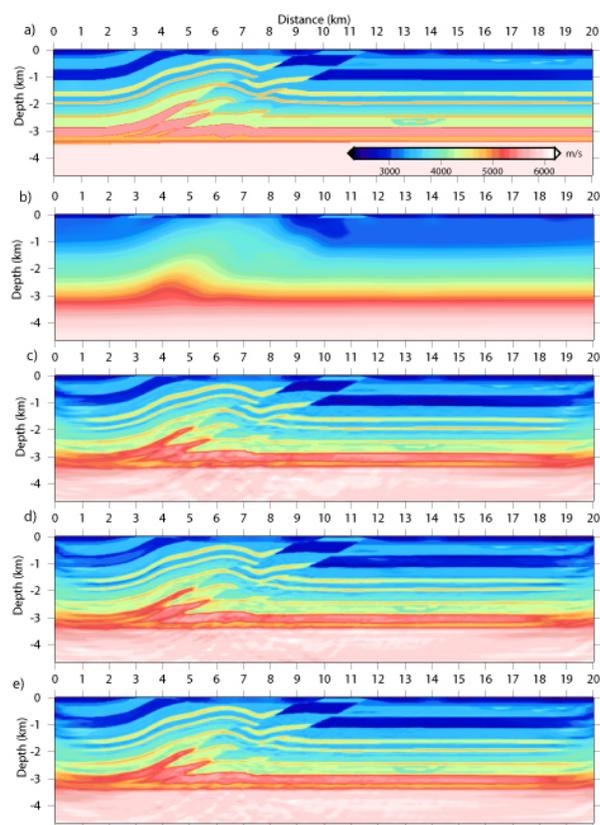


Figure 2: a) Dip section of the Overthrust model. b) Smooth starting model. Note that the true model is set up in the first 100 meters. c) Final model obtained by 7 successive mono-frequency inversions. d) Final model obtained by 2 successive multi-frequency inversions using a Ricker source wavelet. e) Final model obtained by 2 successive multi-frequency inversions using a Dirac wavelet with a 5-Hz dominant frequency.

APPLICATION TO A DENSE OBS DATA SET

We computed the speed-up reached by our parallel algorithm for an application to a real 2D Ocean Bottom Seismometer (OBS) data set. The reader is referred to Operto et al. (2006) for an extensive description of the velocity models obtained by FWI and their interpretation.

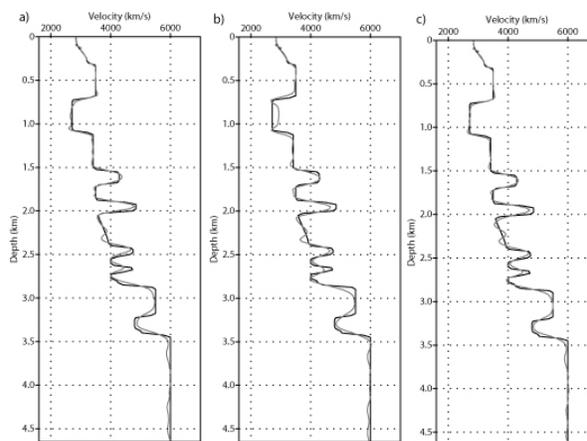


Figure 3: a) Vertical graphs extracted from the true velocity model (black) and the final FWI velocity models (gray) of Test 1 (a), 2 (b) and 3 (c).

In Operto et al. (2006), the application was performed on 12 Xeon processors interconnected with a GBytes network. Only the factorization was computed in parallel. The computing time was more than 20 days. The OBS data presented were collected during the KY0106 cruise, performed by the Institute for Frontier Research on Earth Evolution (IFREE, Japan) between July 29th and September 6th 2001. A dense array with a 1-km spacing of 100 OBS was deployed. Thirteen frequencies uniformly sampled between 3 Hz and 15 Hz (frequency interval: 1 Hz) were inverted successively. Twenty iterations were computed per frequency leading to 260 velocity models of increasing resolution. The velocity model dimensions are 105 km x 25 km corresponding to a finite-difference grid of 4201 x 1001 with a 25-m grid interval. The number of shot per OBS gather was 1005 and the number of inverted OBS gathers was 93. The order of the matrix is about $4.4 \cdot 10^6$ with $40 \cdot 10^6$ non zero coefficients. The diagonal Hessian was computed for a decimated acquisition consisting of 1 shot out of 4. This lead to 354 RHSs in the resolution phase for the Hessian estimation.

The parallel FWI algorithm have been run on several clusters to assess its numerical behavior. Due to the increase of MPI communication in this algorithm, a fast interconnect is required to improve the CPU time. We present here results obtained on a HP DL 145G2 cluster composed of forty eight Opteron dual-core bi-processor 2.4 GHz 64 bits nodes with a 921.6 Gflops total peak power. Each nodes shared 8 Gbytes of RAM memory. Each node is interconnected with an Infiniband network. Four MPI processes were allocated per node. We have also used a HP DL 585 Opteron dual core quadri-processors with 32 Go of RAM to performed our simulation on a single processor.

In parallel computing, speedup refers to how much a parallel algorithm is faster than a corresponding sequential algorithm. It is defined by the ratio between the effective time used by a sequential and parallel execution. In our application and for 40 cores, parallel factorization was 14 times faster than the sequential one. Another important performance metric is the efficiency, defined by the ratio between the speed up and the number of core. The efficiency estimates how well-used the processors in solving the problem, compare to how much effort is wasted in communication and synchronization. Our scalability analysis consists of performing a series of simulations by increasing the number of cores to evaluate the behavior of CPU time and memory. The results are summarized in Figures 4 and 6. The red curve in Figure 4 gives the CPU-time scalability of the overall algorithm for one frequency inversion. The cyan and yellow curves shows the CPU-time scalability of the MUMPS solve phase for respectively 354 and

A massively parallel frequency-domain full-waveform inversion algorithm

93 RHS vectors. Table 1 summarizes the computational complexity of this application. The memory allocated during the factorization is a linear function of the number of core. The amount of memory increases because of increasing overheads during factorization (Figure 6). The optimal number of cores for this application in term of CPU time is 40 because past it, significant MPI communication will be required.

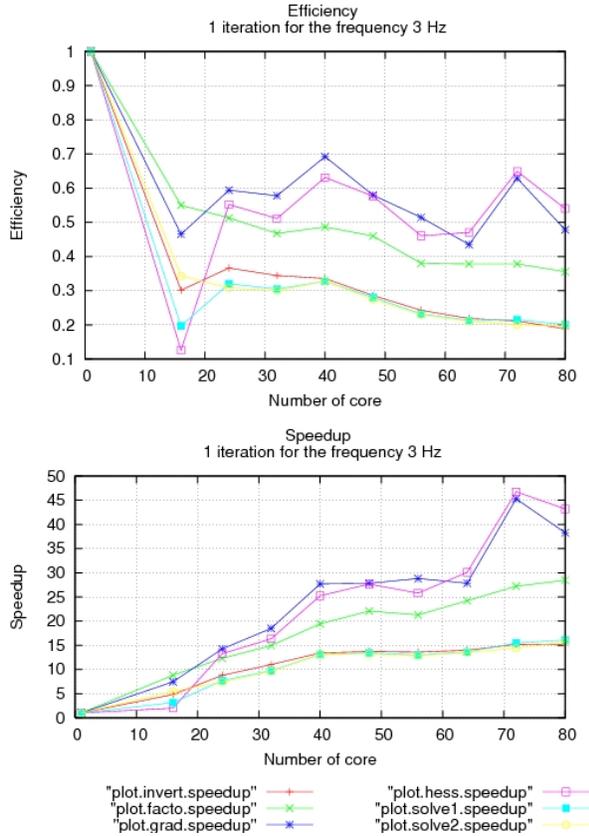


Figure 4: Speed up and efficiency of the main step of the FWI algorithm as a function of the number of core.

Optimal number of cores	40
Number of LU factors (complex)	$645 \cdot 10^6$
Total memory to store LU factors	5 Gb
Total memory to store one solution	32 Mb
Total memory for factorization	19 Gb
Real time for factorization (1 frequency)	19 sec
Real time for 354 resolutions (diagonal Hessian)	106 sec
Real time for 93x2 resolutions (gradient estimation)	2x25 sec
Real time for 1 iteration of 1 frequency inversion	760 sec
Total time for 20 iterations of 13 frequencies inversion	2 days

Table 1: Computational complexity of the parallel FWI algorithm for the seismic imaging of the eastern Nankai trough.

PERSPECTIVE AND CONCLUSIONS

We present a massively parallel algorithm for 2D acoustic full-waveform inversion based on the MUMPS direct solver for distributed-memory

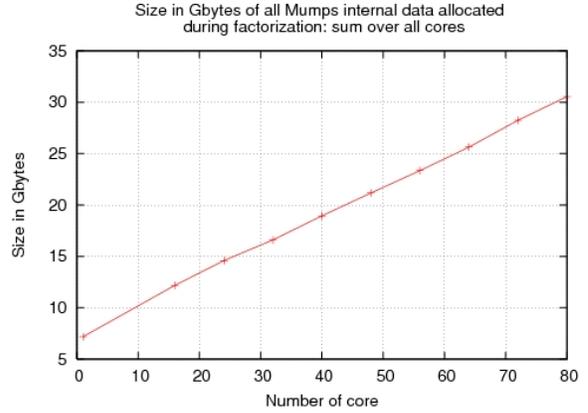


Figure 5: Size in Gbytes of all Mumps internal data allocated during factorization (sum over all processors) as a function of the number of core.

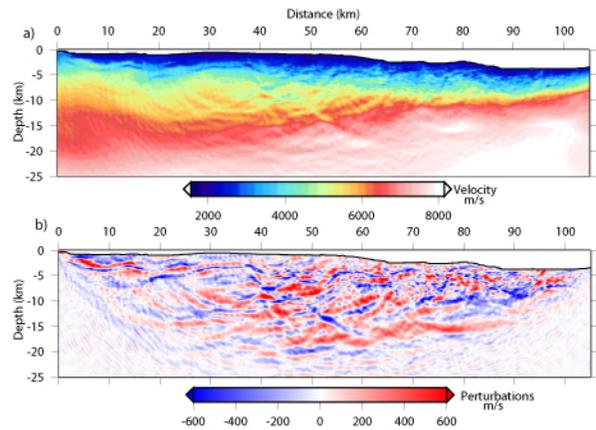


Figure 6: a) Velocity model obtained close of the 5-Hz frequency inversion. b) Corresponding velocity perturbation model (difference between the FWI model and the starting model developed by first-arrival travelt ime inversion).

platform such as High Performance Cluster. Future works will concern the extension to 2D visco-elastic media, 2D TTI anisotropic acoustic media and 3D acoustic media.

ACKNOWLEDGMENTS

Access to the high performance computing facilities of MESOCENTRE SIGAMM computer center provided the required computer resources and we gratefully acknowledge both this facility and the assistance of the support staff. We are particularly grateful to A. Miniussi for his help during the installation of the software on the cluster. This work was carried out within the frame of the SEISCOPE consortium (<http://geoazur.unice.fr/SEISCOPE>) sponsored by BP, CCG, EXXON, TOTAL and SHELL.

A massively parallel frequency-domain full-waveform inversion algorithm

REFERENCES

- Amestoy, P. R., I. S. Duff, J. Y. L'Excellent, and J. Koster, 2007, Multifrontal massively parallel solver (MUMPS version 4.6): Users' guide. <http://enseeiht.fr/apo/MUMPS/>.
- Amestoy, P. R., A. Guermouche, J. Y. L'Excellent, and S. Pralet, 2006, Hybrid scheduling for the parallel solution of linear systems: *Parallel computing*, **32**, 136–156.
- Berenger, J.-P., 1994, A perfectly matched layer for absorption of electromagnetic waves: *Journal of Computational Physics*, **114**, 185–200.
- Duff, I. S. and J. K. Reid, 1983, The multifrontal solution of indefinite sparse symmetric linear systems: *ACM Transactions on Mathematical Software*, **9**, 302–325.
- Hustedt, B., S. Operto, and J. Virieux, 2004, Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling: *Geophysical Journal International*, **157**, 1269–1296.
- Jo, C. H., C. Shin, and J. H. Suh, 1996, An optimal 9-point, finite-difference, frequency-space 2D scalar extrapolator: *Geophysics*, **61**, 529–537.
- Lailly, P., 1984, The seismic inverse problem as a sequence of before stack migrations: inverse scattering. Theory and application, Expanded Abstracts, 206–220, Soc. Ind. appl. Math.
- Marfurt, K., 1984, Accuracy of finite-difference and finite-elements modeling of the scalar and elastic wave equation: *Geophysics*, **49**, 533–549.
- Operto, S., J. Virieux, J. X. Dessa, and G. Pascal, 2006, Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: application to the eastern nankai trough: *Journal of Geophysical Research*, **111**, doi:10.1029/2005JB003835.
- Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part I : theory and verification in a physic scale model: *Geophysics*, **64**, 888–901.
- , 2004, Velocity models from frequency-domain waveform tomography: past, present and future: Presented at the 66th Annual EAGE Meeting, Paris, Eur. Ass. Expl. Geophys.
- Pratt, R. G., C. Shin, and G. J. Hicks, 1998, Gauss-newton and full newton methods in frequency-space seismic waveform inversion: *Geophysical Journal International*, **133**, 341–362.
- Pratt, R. G. and M. H. Worthington, 1990, Inverse theory applied to multi-source cross-hole tomography. Part I: acoustic wave-equation method: *Geophysical Prospecting*, **38**, 287–310.
- Ravaut, C., S. Operto, L. Improtà, J. Virieux, A. Herrero, and P. dell'Aversana, 2004, Multi-scale imaging of complex structures from multi-fold wide-aperture seismic data by frequency-domain full-wavefield inversions: application to a thrust belt: *Geophysical Journal International*, **159**, 1032–1056.
- Shin, C., K. Yoon, K. J. Marfurt, K. Park, D. Yang, H. Y. Lim, S. Chung, and S. Shin, 2001, Efficient calculation of a partial derivative wavefield using reciprocity for seismic imaging and inversion: *Geophysics*, **66**, 1856–1863.
- Sirgue, L. and R. G. Pratt, 2004, Efficient waveform inversion and imaging : a strategy for selecting temporal frequencies: *Geophysics*, **69**, 231–248.