

# Wavefield reconstruction in attenuating media: A checkpointing-assisted reverse-forward simulation method

Pengliang Yang<sup>1</sup>, Romain Brossier<sup>1</sup>, Ludovic Métivier<sup>2</sup>, and Jean Virieux<sup>1</sup>

## ABSTRACT

Three-dimensional implementations of reverse time migration (RTM) and full-waveform inversion (FWI) require efficient schemes to access the incident field to apply the imaging condition of RTM or build the gradient of FWI. Wavefield reconstruction by reverse propagation using final snapshot and saved boundaries appears quite efficient but unstable in attenuating media, whereas the checkpointing strategy is a stable alternative at the expense of increased computational cost through repeated forward modeling. We have developed a checkpointing-assisted reverse-forward simulation (CARFS) method in the context of viscoacoustic wave propagation with a generalized Maxwell body. At each backward reconstruction step, the CARFS algorithm makes a smart decision between forward modeling using checkpoints and reverse propagation based on the minimum time-stepping cost and an energy measure. Numerical experiments demonstrated that the CARFS method allows accurate wavefield reconstruction using less timesteppings than optimal checkpointing, even if seismic attenuation is very strong. For RTM and FWI applications involving a huge number of independent sources and/or applications on architectures with limited memory, CARFS will provide an efficient tool with adequate accuracy in practical implementation.

## INTRODUCTION

Seismic imaging techniques such as reverse time migration (RTM) and time-domain full-waveform inversion (FWI) are computation and memory intensive for 3D implementation. A challenging step of such implementations is the zero-lag crosscorrelation of the incident/

source wavefield and the adjoint/receiver wavefield when applying the imaging condition in RTM, or building the gradient of the misfit function in FWI. In both cases, one needs to access the incident wavefield backward in time while computing the adjoint/receiver wavefield.

Three main classes of method may be identified to do the key step with distinct algorithmic features. (1) The most intuitive strategy is to store the incident wavefield when performing the incident modeling, and accessing it at each time step when computing the adjoint wavefield for imaging condition or building the gradient. This approach could be time consuming due to highly demanding traffic for large data volume unless efficient algorithmic and/or hardware-oriented storage strategies are considered as data compression (Sun and Fu, 2013), local nonvolatile memory storage, or efficient parallel input/output (I/O) procedure (Prabhat and Koziol, 2014). (2) Another option is the checkpointing strategy (Griewank, 1992; Griewank and Walther, 2000; Symes, 2007; Anderson et al., 2012; Komatitsch et al., 2016), which recomputes the incident field forward in time inside a time window started with saved snapshots (checkpoints) at specific times. Each time step has to be recomputed several times from checkpoints, leading to a recomputation ratio (the total number of computed time steps of the incident field to build the RTM image or FWI's gradient, divided by the number of time steps for single incident simulation) much higher than two for moderate memory resources available. (3) A more efficient scheme is the wavefield reconstruction by reverse propagation based on the boundary values and the final-wavefield snapshot. The incident wavefield can be perfectly recomputed, whereas the adjoint wavefield is built on the fly (Clapp, 2008; Dussaud et al., 2008; Brossier et al., 2014). This implementation relies on the reversibility of the wave equation as the incident field is recomputed backward in time. Compared with checkpointing, the reverse-propagation method seems more attractive because of its efficiency, as memory requirements and disk I/O are mitigated. Unfortunately, reverse propagation in attenuative medium is difficult in general, even though it is possible to consider single

Manuscript received by the Editor 10 February 2016; revised manuscript received 22 May 2016; published online 07 September 2016.

<sup>1</sup>Université Grenoble Alpes, ISTerre, France. E-mail: pengliang.yang@univ-grenoble-alpes.fr; romain.brossier@univ-grenoble-alpes.fr; jean.virieux@univ-grenoble-alpes.fr.

<sup>2</sup>Université Grenoble Alpes, LJK, CNRS, France and ISTerre, France. E-mail: ludovic.metivier@univ-grenoble-alpes.fr.

© 2016 Society of Exploration Geophysicists. All rights reserved.

relaxation mechanism (Bai et al., 2014) or filtering with explicit separation of dispersion and dissipation (Zhu, 2015).

Seismic attenuation has been widely observed and investigated by researchers in the global seismology and exploration geophysics community (Futterman, 1962; Caputo, 1967; Liu et al., 1976; Emmerich and Korn, 1987; Carcione et al., 1988b). The significant development of RTM and FWI in complex media, in particular involving attenuation, requires to develop efficient computing methods for reconstructing the incident wavefield in time-reversal manner. In this work, we shall propose a checkpointing-assisted reverse-forward simulation (CARFS) method to solve the time-reversal wavefield reconstruction issue in the context of viscoacoustic medium constructed by the generalized Maxwell body (GMB) (Emmerich and Korn, 1987; Moczo and Kristek, 2005). The proposed CARFS method is a hybrid method that combines the efficiency of reverse propagation and the stability of checkpointing in attenuating medium, using an energy measure to achieve accurate wavefield reconstruction with less time-stepping cost.

The structure of the paper is organized as follows: First, we briefly review the techniques of accessing the incident wavefield for RTM and FWI, i.e., storing the wavefield, the optimal checkpointing strategy, and the reverse-propagation method. Then, we consider the issue of time-reversal wavefield reconstruction in attenuating media within the framework of GMB-based viscoacoustic wave equation, which can be discretized to do forward wavefield extrapolation. The inverse of the forward propagation operator, namely the reverse propagation operator, is unstable due to the amplified numerical errors induced by an exponential factor. To benefit from the efficiency of reverse propagation and the stability of checkpointing, the CARFS scheme is then designed to obtain accurate wavefield reconstruction using much less timesteppings than optimal checkpointing. The 2D Valhall study demonstrates the potential improvement of the efficiency of the CARFS method for RTM and FWI applications in attenuating media. The concluding remarks and perspectives are drawn at the end. Please note that for simplicity, we remain in this study under the isotropic acoustic approximation of the wave equation, but all the developments are also valid for anisotropic and viscoelastic media.

## ACCESSING INCIDENT WAVEFIELD BACKWARD IN NONATTENUATING MEDIA

According to the Newton's law and the constitutive relation, the acoustic approximation of the wave motion can be formulated as a set of a first-order hyperbolic system

$$\begin{cases} \rho \partial_t \mathbf{v} = \nabla p, \\ \partial_t p = \kappa \nabla \cdot \mathbf{v}, \end{cases} \quad (1)$$

or in matrix form

$$\partial_t \underbrace{\begin{bmatrix} \mathbf{v} \\ p \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} \frac{1}{\rho} & 0 \\ 0 & \kappa \end{bmatrix}}_{N(\mathbf{m})} \underbrace{\begin{bmatrix} 0 & \nabla \\ \nabla \cdot & 0 \end{bmatrix}}_{H(\nabla)} \underbrace{\begin{bmatrix} \mathbf{v} \\ p \end{bmatrix}}_{\mathbf{w}}, \quad (2)$$

where the density  $\rho$  and the bulk modulus  $\kappa$  are related to the wave speed  $v$  via  $\kappa = \rho v^2$ , whereas the pressure and the particle velocities are denoted by  $p$  and  $\mathbf{v}$ , respectively. The source/incident wavefield  $\mathbf{w}$  can be simulated by the forward modeling based on zero-valued initial condition, whereas the adjoint wavefield  $\bar{\mathbf{w}}$ , satisfying an

equation close to equations 1 and 2, has to be simulated backward in time based on zero-valued final condition.

In RTM, the most widely used crosscorrelation imaging condition, and the one we consider in this study, requires the multiplication of the source wavefield  $p_s(\mathbf{x}, t)$  and the receiver wavefield  $p_r(\mathbf{x}, t)$  (Symes, 2007) at all the same temporal instance

$$I(\mathbf{x}) = \int_0^T p_s(\mathbf{x}, t) p_r(\mathbf{x}, t) dt, \quad (3)$$

where  $I(\mathbf{x})$  is the migrated image at the point  $\mathbf{x}$ . To retrieve high-resolution model parameters of the subsurface quantitatively (e.g.,  $\mathbf{m} = (\rho, \kappa)$ ), the gradient of the misfit function  $C(\mathbf{m})$  in FWI has to be estimated in each iteration (Virieux and Operto, 2009; Brossier et al., 2014)

$$\nabla C(\mathbf{m}) = \int_0^T \bar{\mathbf{w}}^\dagger(\mathbf{x}, t) \frac{\partial N^{-1}(\mathbf{m})}{\partial \mathbf{m}} \partial_t \mathbf{w}(\mathbf{x}, t) dt, \quad (4)$$

such that the synthetic seismograms fit the field observations. As shown in equations 3 and 4, it is mandatory to access the incident and the receiver wavefields or adjoint wavefield simultaneously to perform crosscorrelation in RTM or build the gradient for FWI. Because the adjoint wavefield is simulated in the opposite time than the incident field, the latter thus has to be accessed backward to correlate with the adjoint wavefield for imaging build-up.

Different strategies exist for mitigating memory and/or computation requirements to achieve this. Between full storage, which could be effective if large memory is available and full-computation in case of limited memory, there is a wide range of possible strategies from full storage with data compression to optimal checkpointing strategies as mentioned previously in the introduction. We shall provide here a quick overview of the approaches for a nonattenuation medium, in which they have been usually considered before considering possible new extension to an attenuating medium.

## Storing the wavefield

To perform the correlation between the incident and the adjoint wavefields at all the same temporal levels in RTM and FWI, one may intuitively store all the snapshots during the evolution of the incident wavefield in memory or on disk. Although the adjoint wavefield is simulated backward in time, the incident wavefield can be loaded synchronously from the memory, leading to frequent I/O traffic and heavy data exchange. The required storage for 3D imaging applications is very large to fit in core memory due to the 4D data volume including three spatial coordinates and one time coordinate. Once this large data volume is stored on disk, one has to access the 3D snapshot at every temporal instant to compute the crosscorrelation between the wavefields, leading to significant I/Os requesting huge network bandwidth requests. To mitigate these requests, some data compression techniques have been recently developed (Sun and Fu, 2013; Boehm et al., 2015).

## Binomial checkpointing strategy

The checkpointing strategy uses a small number of memory units (checkpoints) to store the system state at distinct times. The recomputation of required states that have not been stored is obtained by restarting the forward modeling from the checkpoints. Griewank

(1992) and Griewank and Walther (2000) mathematically prove that for  $N$  steps of forward evolution and  $c$  number of checkpoints, the minimum number of forward timesteppings for reverse mode is given by

$$N' = rN - \beta(c + 1, r - 1), \quad (5)$$

where  $\beta(s, t)$  is a combinatorial number

$$\beta(s, t) := \binom{s+t}{t}, \quad (6)$$

while the number  $r$  is the maximum repetition number for recovering any particular state uniquely determined by  $\beta(c, r - 1) < N \leq \beta(c, r)$ . The recomputation ratio of the optimal checkpointing is equal to

$$R = N'/N, \quad (7)$$

provided the  $c$  checkpoints during  $N$  temporal steps of forward simulations are stored in the following positions:

$$s_{i+1} = s_i + d_{i+1}, i = 0, \dots, c - 1, \quad (8)$$

where the first checkpoint is assumed to start at the initial state  $s_0 = 0$ , and the distance between two neighboring checkpoints  $s_k$  and  $s_{k+1}$  is given by (Stumm and Walther, 2009)

$$d_{i+1} = \begin{cases} \beta(c-i, r-2) & \text{if } N \leq \beta(c, r-1) + \beta(c-i-2, r-1) + s_i, \\ \beta(c-i, r-1) & \text{else if } N \geq \beta(c-i, r) - \beta(c-i-3, r) + s_i, \\ N - s_i - \beta(c-i-1, r-1) - \beta(c-i-2, r-2) & \text{otherwise.} \end{cases} \quad (9)$$

Let us remember that the value  $N'$  for computing the recomputation ratio  $R$  counts the number of time steps in forward and reverse/

adjoint simulation. Then, the multistage ( $r$  level) binomial checkpointing strategy, as shown in Algorithm 1, can be carried out by recursively invoking these forward steps while storing the snapshots in redistributed idle checkpoints.

Several checkpointing techniques have been developed to achieve an acceptable compromise between memory requirement and recomputation increase. The optimal value for the number of snapshots/checkpoints is related to the total number of time steps  $N$  by  $c = \log_2(N)$  according to Griewank and Walther (2000), Symes (2007), and Nguyen and McMechan (2015). The application of checkpointing has been explored in seismic inversion (Blanch et al., 1998). An interesting observation is that for realistic imaging applications, the total number of time steps  $N$  generally ranges within [1000~10,000], giving several stored snapshots  $c$  lower than 14. With such a range of time-steps number, we are able to bound the recomputation ratio  $R \in (3.64, 4.84)$ , which implies generally an affordable but significant computational cost less than five times for the incident field.

### Reverse propagation using final snapshot and stored boundaries

The main idea of the reverse propagation (RP) method is to compute the incident wavefield twice: once in forward time during which the boundary values are stored in core memory (and disk if required) until the final time step. Then, during the computation of the adjoint/receiver field, the incident field is synchronously recomputed backward in time from the final snapshots, as initial condition (Gauthier et al., 1986; Tromp et al., 2005), and the saved boundary condition, acting as Dirichlet boundary condition at each time step (Clapp, 2008; Dussaud et al., 2008; Brossier et al., 2014). The minimum memory requirements for the boundary storage have been investigated in the literature, in particular when involving staggered-grid finite difference (Yang et al., 2014a, 2014b; Nguyen and McMechan, 2015) and different absorbing boundary conditions

#### Algorithm 1. Binomial checkpointing.

- 
1. Distribute checkpoints  $s_0, \dots, s_{c-1}$  in time following the binomial law of optimal checkpointing given in equations 8 and 9;
  2. **for**  $n = 0, \dots, N - 1$  **do**
  3.   compute the wavefield state forward in time:  $F^n : \mathbf{w}^n \rightarrow \mathbf{w}^{n+1}$ ;
  4.   **if**  $n \in s_0, \dots, s_{c-1}$  **then** store the snapshot  $\mathbf{w}^{n+1}$
  5. **end**
  6. **for**  $n = N - 1, \dots, 0$  **do**
  7.   find the closest checkpoint  $s_i$  prior to current time level  $n$  ( $s_i \leq n < s_{i+1}$ ) (the checkpoints  $s_j > n, j = i + 1, \dots, c - 1$  posterior to current time level  $n$  are idle now);
  8.   read the snapshot  $\mathbf{w}^{s_i+1}$  in the checkpoint  $s_i$ ;
  9.   **if**  $n \neq s_i$  **then**
  10.     redistribute the idle checkpoints  $s_{i+1}, \dots, s_{c-1}$  between  $s_i$  and  $n$  according to the binomial law in equations 8 and 9;
  11.     **for**  $k = s_i + 1, n$  **do**
  12.       do forward modeling  $F^k : \mathbf{w}^k \rightarrow \mathbf{w}^{k+1}$ ;
  13.       **if**  $k \in s_{i+1}, \dots, s_{c-1}$  **then** store the snapshot  $k$  at these relocated checkpoint positions
  14.     **end**
  15.   **end**
  16. **end**
-

(Yang et al., 2015). For 3D applications, the memory complexity is associated with one 3D data volume, involving two spatial dimensions and one temporal dimension. This memory request can be significantly decreased up to one or two orders of magnitude based on the temporal sampling determined by Nyquist principle, rather than the more restrictive relation from the Courant-Friedrichs-Lewy (CFL) condition (Sun and Fu, 2013; Yang et al., 2016a, 2016b). Using different interpolation techniques, the missing boundary elements can be accurately recovered, making the in-core boundary storage practically feasible for 3D large-scale applications. Therefore, the RP method appears quite appealing thanks to its efficiency and limited memory requirements.

## VISCOACOUSTIC MODELING AND ITS REVERSE PROPAGATION

It is acknowledged that seismic attenuation is of major importance in many geologic framework, and has to be taken into account to approximate the realistic wave propagation in the earth. The imaging and inversion problems with seismic attenuation have received much attention in recent research literature (Malinowski and Operto, 2008; Prieux et al., 2013; Bai et al., 2014; Dutta and Schuster, 2014; Operto et al., 2015). How to efficiently access the incident wavefield backward in attenuating media is therefore of special importance for time-domain RTM and FWI applications to better delineate the true earth structure (Zhu et al., 2014).

### Viscoacoustic wave propagation

To simulate attenuation for time-domain modeling, many physical models have been proposed to mimic its effects: Maxwell body, Kelvin-Voigt model, standard linear solid model, and their generalizations (Casula and Carcione, 1992), such as GMB (Emmerich and Korn, 1987; Moczo and Kristek, 2005) or its equivalence generalized Zener body (GZB) (Carcione et al., 1988a, 1988b). The GMB/GZB provides a practical characterization of the nearly constant  $Q$  nature of the earth in seismic frequency band (Futterman, 1962; Caputo, 1967; Liu et al., 1976; Kjartansson, 1979) by the superposition of many different mechanisms (Moczo and Kristek, 2005). With a slight abuse of language, in this study, viscous media and viscoacoustic media will refer to this GMB model with superposition of several relaxation mechanisms. Within the framework of GMB/GZB (Moczo et al., 2007b), the viscoacoustic wave equation reads

$$\rho \partial_t \mathbf{v} = \nabla p (\rho \partial_t v_i = \partial_i p), \quad (10a)$$

$$\partial_t p = M_u \left( \nabla \cdot \mathbf{v} - \sum_{l=1}^L Y_l \xi_l \right), \quad (10b)$$

$$\partial_t \xi_l + \omega_l \xi_l = \omega_l \nabla \cdot \mathbf{v}, \quad l = 1, 2, \dots, L, \quad (10c)$$

where the unrelaxed modulus is  $M_u = \kappa = \rho v^2$ ;  $L$  is the memory variables  $\xi_l$ , each one associated with one reference frequency  $\omega_l$ , are included in equation 10 to attenuate the propagating waves. The attenuation is described by the inverse of the quality factor (Moczo et al., 2007b, equation 117)

$$Q^{-1}(\omega) = \sum_{l=1}^L Y_l \frac{\omega_l \omega}{\omega_l^2 + \omega^2} / \left( 1 - \sum_{l=1}^L Y_l \frac{\omega_l^2}{\omega_l^2 + \omega^2} \right). \quad (11)$$

The wave propagation without attenuation can also be obtained by setting all anelastic coefficients to be zeros, say  $Y_l = 0, l = 1, \dots, L$ .

The system (equation 10) is a hybrid of partial differential equation (equations 10a and 10b) and ordinary differential equations (equation 10c). Consider the half-integer time step for particle velocity  $\mathbf{v} = (v_x, v_y, v_z)$  and the integer time step of pressure  $p$  as well as the memory variables  $\xi_l$ . The forward leap-frog timestepping  $F^n$  from level  $n$  to  $n+1$  corresponds to the iteration

$$\begin{cases} v_x^{n+\frac{1}{2}} = v_x^{n-\frac{1}{2}} + \frac{\Delta t}{\rho} \partial_x p^n, \\ v_y^{n+\frac{1}{2}} = v_y^{n-\frac{1}{2}} + \frac{\Delta t}{\rho} \partial_y p^n, \\ v_z^{n+\frac{1}{2}} = v_z^{n-\frac{1}{2}} + \frac{\Delta t}{\rho} \partial_z p^n, \\ \xi_l^{n+1} = e^{-\omega_l \Delta t} \xi_l^n + (1 - e^{-\omega_l \Delta t}) \nabla \cdot \mathbf{v}|_{n+\frac{1}{2}}, l = 1, \dots, L, \\ p^{n+1} = p^n + \Delta t \kappa (\nabla \cdot \mathbf{v}|_{n+\frac{1}{2}} - \sum_{l=1}^L Y_l \xi_l^{n+\frac{1}{2}}), \xi_l^{n+\frac{1}{2}} = \frac{1}{2} (\xi_l^n + \xi_l^{n+1}), \end{cases} \quad (12)$$

which can be abbreviated as

$$\mathbf{w}^{n+1} = F^n \mathbf{w}^n, \quad (13)$$

using the wavefield state vector as  $\mathbf{w}^n = (\mathbf{v}^{n-\frac{1}{2}}, p^n, \xi_1^n, \dots, \xi_L^n)$ . A detailed development of the discretization is provided in Appendix A. Note the coefficient  $e^{-\omega_l \Delta t}$  in front of the field  $\xi_l^n$  in equation 12 is always smaller than one. A complete  $N$ -step forward simulation is then briefly expressed as the composition of the sequence  $F^n, F^{n-1}, F^{n-2}, \dots, F^0$ .

### Unstable viscoacoustic reverse propagation

The reverse propagation of the forward wavefield, say, from time level  $n+1$  to time  $n$ , can be mathematically expressed as the inverse of equation 13

$$\mathbf{w}^n = (F^n)^{-1} \mathbf{w}^{n+1}, \quad (14)$$

where the reverse-propagation operator  $(F^n)^{-1}$  is specified by

$$\begin{cases} \xi_l^n = e^{\omega_l \Delta t} \xi_l^{n+1} + (1 - e^{\omega_l \Delta t}) \nabla \cdot \mathbf{v}|_{n+\frac{1}{2}}, l = 1, \dots, L, \\ p^n = p^{n+1} - \Delta t \kappa (\nabla \cdot \mathbf{v}|_{n+\frac{1}{2}} - \sum_{l=1}^L Y_l \xi_l^{n+\frac{1}{2}}), \xi_l^{n+\frac{1}{2}} = \frac{1}{2} (\xi_l^n + \xi_l^{n+1}), \\ v_x^{n-\frac{1}{2}} = v_x^{n+\frac{1}{2}} - \frac{\Delta t}{\rho} \partial_x p^n, \\ v_y^{n-\frac{1}{2}} = v_y^{n+\frac{1}{2}} - \frac{\Delta t}{\rho} \partial_y p^n, \\ v_z^{n-\frac{1}{2}} = v_z^{n+\frac{1}{2}} - \frac{\Delta t}{\rho} \partial_z p^n. \end{cases} \quad (15)$$

With the final snapshots and saved values at boundaries at hand, a  $N$ -step of reverse propagation is then  $(F^0)^{-1}, (F^1)^{-1}, \dots, (F^{N-1})^{-1}$ . As we see, the reverse propagation (equation 15) is exactly the same as the forward simulation (equation 12) except the sign of time-stepping. In other words, the reverse propagation can be done with ease by only flipping the sign of  $\Delta t$  when invoking the same sub-routines.

However, owing to the factor  $e^{\omega_i \Delta t}$  in front of the memory variables  $\xi_i$ , reverse propagation in attenuating media usually suffers from instability due to the accumulation of exponentially growing errors rather than the discretization of the wave equation based on a rigorous mathematical formulation. The numerical errors are accumulated in the memory variable and then transferred to the pressure  $p$  and particle velocities  $\mathbf{v}$ . This kind of instability significantly depends on whether the digital errors induced by the truncation of the floating point representation in computer are amplified to an intolerable magnitude. Reverse propagation with regularly stored snapshots (RPSS) may find its value for efficient wavefield reconstruction backward in time. However, the required number of snapshots is generally very large and difficult to predict in practice for long duration simulation in attenuating medium.

As a numerical demonstration, we shall perform the viscoacoustic reverse propagation with snapshot strategy in a constant velocity model ( $v = 2000$  m/s) of size  $nx = nz = 301$  for  $N = 1000$  time steps with temporal sampling  $\Delta t = 0.001$  s, using three mechanisms with the specified frequencies  $f_1 = 2$  Hz,  $f_3 = 35$  Hz, and  $f_2 = \sqrt{f_1 f_3}$ . The associated anelastic coefficients are estimated to approximate a constant  $Q$  value ( $Q = 50$ ). We record the total energy (computed according to equation 16) during forward modeling and reverse propagation to check whether they match each other. Figure 1 shows that the energy during reverse propagation is exponentially amplified, whereas the amplitude of a randomly chosen trace in the domain indicates the same problem. One may consider RPSS to make reverse propagation work. As shown in Figure 2, using nine snapshots, the problems have been mitigated but are still

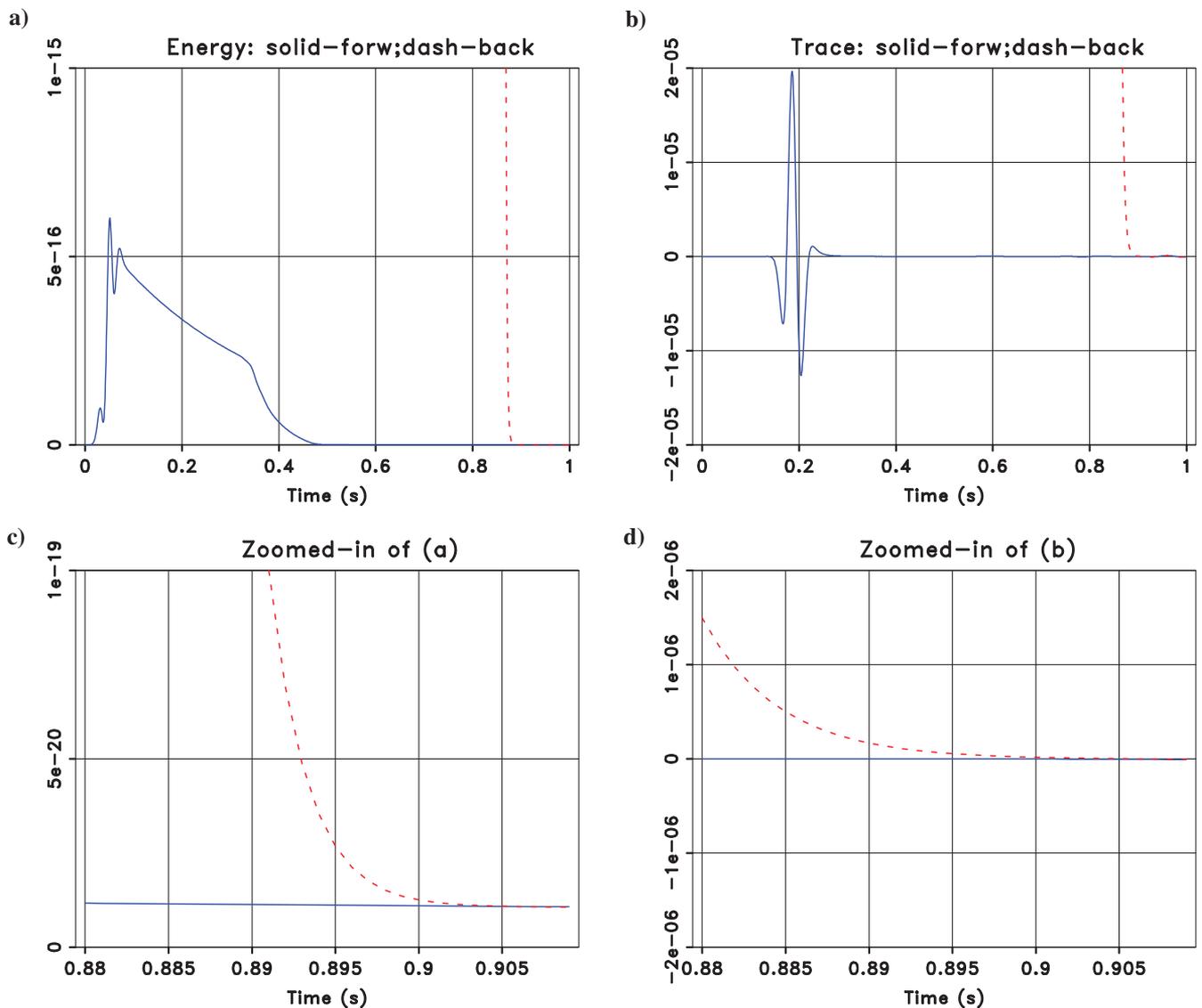


Figure 1. Reverse propagation in a constant viscoacoustic media is amplifying the numerical errors: (a) the energy recorded during reverse propagation cannot match the energy recorded in forward modeling, (b) a randomly chosen trace also shows this mismatch, (c) a magnification of panel (a) shows that the energy is exponentially exaggerated, and (d) after propagation for a while, the error propagated to the trace location and soon the amplitude of the trace is distorted erroneously.

unacceptable. Introducing 11 snapshots is able to obtain a good reconstruction of the incident wavefield because the energy and the trace recorded during backward reconstruction can match quite well their counterparts during forward simulation. However, these 11 snapshots cannot be known prior to the experiment.

## CARFS

### CARFS algorithm

If no snapshots are stored, the reverse-propagation approach enjoying a recomputation ratio of two, although very efficient in nonattenuating media, is generally unstable in the presence of seismic attenuation. Although recomputation ratio might be higher than

two for moderate available memory, checkpointing is stable for attenuating media as only forward in time modeling is performed. To benefit from the efficiency of reverse propagation and the stability of checkpointing in the presence of attenuation, we design a CARFS method. This approach relies on an indicator of the stability of reverse propagation. We choose the total energy in the computing volume  $\Omega$ , giving for the acoustic wave

$$E = \frac{1}{2} \langle \rho \mathbf{v}, \mathbf{v} \rangle_{\Omega} + \frac{1}{2} \left\langle \frac{1}{\kappa} p, p \right\rangle_{\Omega} = \frac{1}{2} \int_{\Omega} \left( \rho \mathbf{v}^2 + \frac{1}{\kappa} p^2 \right) dx. \quad (16)$$

This measure is a good candidate to monitor the stability of the simulation, thanks to its sensitivity to variations in the pressure and

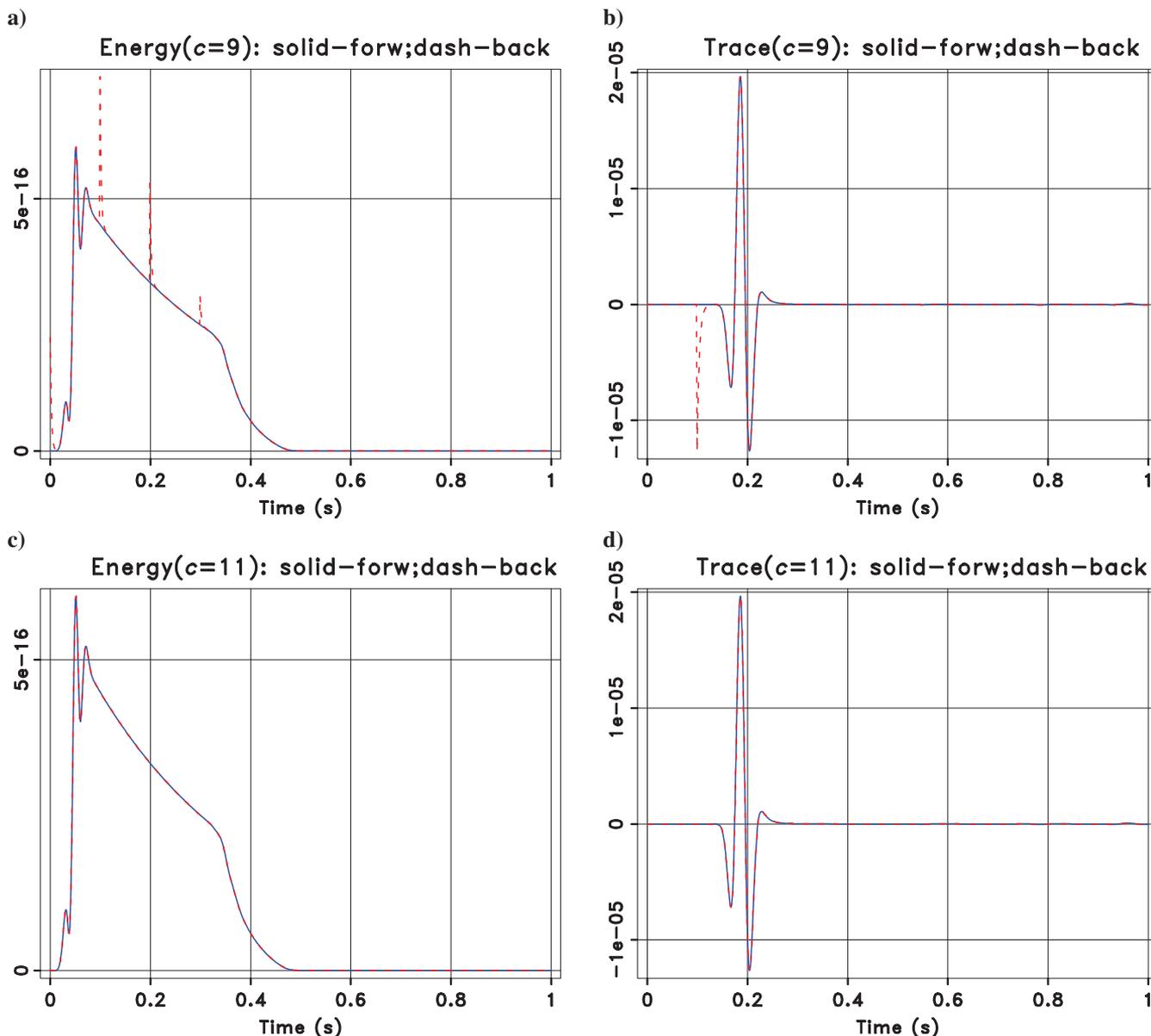


Figure 2. In the constant viscous media, using nine snapshots, (a) the energy and (b) the trace recorded during forward and reverse propagation are still problematic. Using 11 snapshots, (c) the energy and (d) the trace recorded during forward and reverse propagation matches quite well. Note that the discontinuities in panels (a and b) have disappeared in panels (c and d).

the particle velocities (any other measure efficient to compute could be considered as well). The main idea of CARFS is to use, as a default setting, the reverse propagation even in attenuating media, while monitoring the energy of the computation. At any given time step, if the energy of the backward simulation deviates (by a user-defined threshold) from the one of the forward simulation, which has been monitored and kept, the reverse propagation is temporarily stopped, and checkpointing is used starting from the closest snapshot prior to current state to redo some forward timestepping to reinitialize the reverse propagation. The numerical workflow is described in Algorithm 2.

### Relation to RPSS and checkpointing

As shown in Figure 3, at every backward reconstruction step, the proposed CARFS approach tries to make a smart choice between reverse and forward modeling. By preparing the wavefield in advance, reverse propagation has higher priority over forward modeling in checkpointing, but needs to satisfy the tolerance of the energy measure, such that the numerical errors are controlled within a negligible magnitude. The strategy can be seen as an improved reverse propagation method, assisted by checkpointing when reverse propagation becomes unstable. The worst case may be that, when the attenuation is extremely strong, the CARFS approach converges to standard optimal checkpointing with one additional forward simu-

lation to find that the accuracy tolerance is not satisfied at every time step, so that the wavefield has to be computed through checkpointing by repeated forward modeling. In particular, when there is no attenuation, i.e.,  $Y_l = 0, l = 1, \dots, L$ , the CARFS method converges to RPSS, which shares the same number of snapshots specified at the locations by binomial law of checkpointing.

To make the above relation clearer, let us consider some specific configurations to evaluate the computational efficiency of these methods. Let us assume a wavefield state at a noncheckpoint time level  $n$  to recover. As shown in Figure 4, for every noncheckpoint position, RPSS performs backward wavefield extrapolation to compute all the states that have not been stored (Figure 4a), whereas checkpointing performs forward modeling to compute the same states (Figure 4b). It is easy to imagine that the two methods are exactly the same, when the states at all time levels have been stored if large enough memory resource is available on the machine. Let us assume the number of checkpoints are equally spaced in the time coordinate. Only if the closest checkpoint prior to  $n$  is  $n - 1$ , checkpoint attains the same efficiency as RPSS using only one step of forward simulation cost, as shown in Figure 4c. If all checkpoints are equally spaced and the number of checkpoints  $c \geq N/2$ , particularly  $c = N/2$ , the distance to reach step  $n$  is one for checkpointing and RPSS, either from  $n - 1$  to  $n$  or from  $n + 1$  to  $n$ . As shown in Figure 4d, when  $c < N/2$ , the timesteppings required by checkpointing to reach step  $n$  would be usually larger than reverse

### Algorithm 2. CARFS algorithm.

- 
1. Distribute checkpoints  $s_0, \dots, s_{c-1}$  in time following the binomial law of optimal checkpointing given in equations 8 and 9;
  2. **for**  $n = 0, \dots, N - 1$  **do**
  3.   compute the wavefield state forward in time:  $F^n: \mathbf{w}^n \rightarrow \mathbf{w}^{n+1}$  and record the total energy  $E_f^{n+1}$  for  $\mathbf{w}^{n+1}$ ;
  4.   store the boundary of  $\mathbf{w}^{n+1}$  at decimated temporal locations according to Nyquist sampling (Yang et al., 2016b);
  5.   **if**  $n \in s_0, \dots, s_{c-1}$  **then** store the snapshot  $\mathbf{w}^{n+1}$
  6. **end**
  7. **for**  $n = N - 1, \dots, 0$  **do**
  8.   find the closest checkpoint  $s_i$  prior to current time level  $n$  ( $s_i \leq n < s_{i+1}$ ) (the checkpoints  $s_j > n, j = i + 1, \dots, c - 1$  posterior to current time level  $n$  are idle now);
  9.   **if**  $n = s_i$  **then**
  10.     read the snapshot  $\mathbf{w}^{n+1}$  and set the backward energy  $E_b^{n+1} = E_f^{n+1}$ ;
  11.   **else**
  12.     interpolate the boundary of  $\mathbf{w}^{n+1}$  and compute the energy measure  $E_b^{n+1}$ ;
  13.   **end**
  14.   **if**  $|E_f^{n+1} - E_b^{n+1}| > \text{tolerance} \cdot E_f^{n+1}$  **then**
  15.     read the snapshot  $\mathbf{w}^{s_i+1}$  in the checkpoint  $s_i$ ;
  16.     redistribute the idle checkpoints  $s_{i+1}, \dots, s_{c-1}$  between  $s_i$  and  $n$  according to the binomial law in equations 8 and 9;
  17.     **for**  $k = s_i + 1, n$  **do**
  18.       do forward modeling  $F^k: \mathbf{w}^k \rightarrow \mathbf{w}^{k+1}$ ;
  19.       **if**  $k \in s_{i+1}, \dots, s_{c-1}$  **then** store the snapshot  $k$  at these relocated checkpoint positions
  20.     **end**
  21.     override the backward energy  $E_b^{n+1}$  with  $E_f^{n+1}$ :  $E_b^{n+1} = E_f^{n+1}$ ;
  22.   **end**
  23.   perform reverse propagation  $(F^n)^{-1}: \mathbf{w}^{n+1} \rightarrow \mathbf{w}^n$
  24. **end**
-

propagation, which only needs one step backward propagation from step  $n + 1$  to step  $n$ . Compared with RPSS, checkpointing may require more timesteppings in any segment that the distance between two successive checkpoints are larger than two. In other words, the RPSS method can always outperforms checkpointing in terms of efficiency as long as the backward wave propagation is stable. In attenuating medium, it is generally unpredictable whether RPSS can work properly or not, using a predetermined number of checkpoints and the given velocity and attenuation model, whereas checkpointing can always provide a steady reconstruction for the wavefield by repeating forward simulation from previous stored state. The CARFS algorithm provides a viable way to combine the merits from both approaches, using the energy measure to estimate whether the deviation of wavefield is controlled within the tolerance. We highlight that CARFS circumvents the instability issue of RPSS assisted by checkpointing without significant sacrifice of the efficiency. The recomputation ratio curves in Figure 5 provide us a quantitatively evaluation on the efficiency of these methods.

### Comparison with the existing methods

Let us compare all the methods in terms of computation complexity and memory request. Consider  $N$ -steps of wave simulation for a 3D cubic model of size  $n_x^3$ . Storing all the wavefield snapshots has the highest memory consumption without any need for recomputing the incident wavefield. Even if the data compression technique can

be used during the storing procedure (Sun and Fu, 2013), the actual amount of storage, involving some computation for compression and decompression, is still of order  $O(1/a \times N \times n_x^3)$  with only a constant factor  $(1/a)$  depending on the compression rate  $a$  of the algorithm used. The checkpointing strategy needs to store  $c$  checkpoints in which the factor  $c$  grows logarithmically according to the simulation steps  $N$ ,  $c = \log_2(N)$ . The recomputation ratio of checkpointing is generally much higher than two because of the limited memory resource available. Particularly, if all the snapshots are allowed to be stored as checkpoints, the checkpointing strategy converges to the scheme of storing all wavefields, involving only one time of forward modeling ( $R = 1$ ). The reverse propagation method involves six faces of the boundary (front-rear, left-right, and top-bottom) at every time step, with the storage of order  $O(1/b \times N \times n_x^2)$ , in which  $b$  is the decimation rate if the decimation and interpolation techniques are used for storing and restoring the wavefield at boundaries. The recomputation ratio of the reverse propagation is  $R = 2$  because the incident wavefield has to be reverted backward when backpropagating the adjoint wavefield (Yang et al., 2016b). Introducing  $c$  snapshots in  $N$ -steps of reverse propagation leads to RPSS method with a recomputation ratio  $R_{\text{RPSS}} = 1 + (N - c)/N$ . As already shown, RP and RPSS methods using insufficient number of snapshots cannot work in viscous medium. The CARFS algorithm supplies a good tradeoff between memory overhead and the computation complexity. Compared with checkpointing, CARFS needs the extra memory for storing boundary

to perform reverse propagation, resulting in a lower recomputation ratio. The CARFS will be equivalent to RPSS, when every reverse propagation step succeeds. Compared with RPSS, CARFS can work quite well in strong attenuating media using less snapshots because the erroneous wavefield can be overridden by checkpointing strategy from the closest checkpoint prior to current time level, if the energy tolerance is violated. Table 1 and Figure 5 summarize the memory complexity and the recomputation ratios of all the existing methods. It can be seen that the CARFS method using forward and reverse propagation is a practical solution for large-scale imaging applications thanks to the good tradeoff between the computation complexity and the memory requirement.

### NUMERICAL EXAMPLES

In the following, we disclose the potential of the CARFS method in practice with a variety of application settings, based on the 2D Valhall synthetic model with the realistic velocity, density, and quality factor shown in Figure 6. The model is of size  $1671 \times 2813$  with grid spacing  $\Delta x = \Delta z = 3.125$  m. The wavefield reconstruction runs after  $N = 2500$  steps of forward simulation using a seismic source in the middle of the model with a Ricker wavelet. The minimum  $Q$  is 60, in the gas layers, to enforce strong attenuation for the propagation of seismic waves. Three GMB mechanisms are used with the anelastic coefficients  $Y_l$  from least-squares optimized  $Q$  (Blanch et al., 1995) to mimic constant  $Q$  prop-

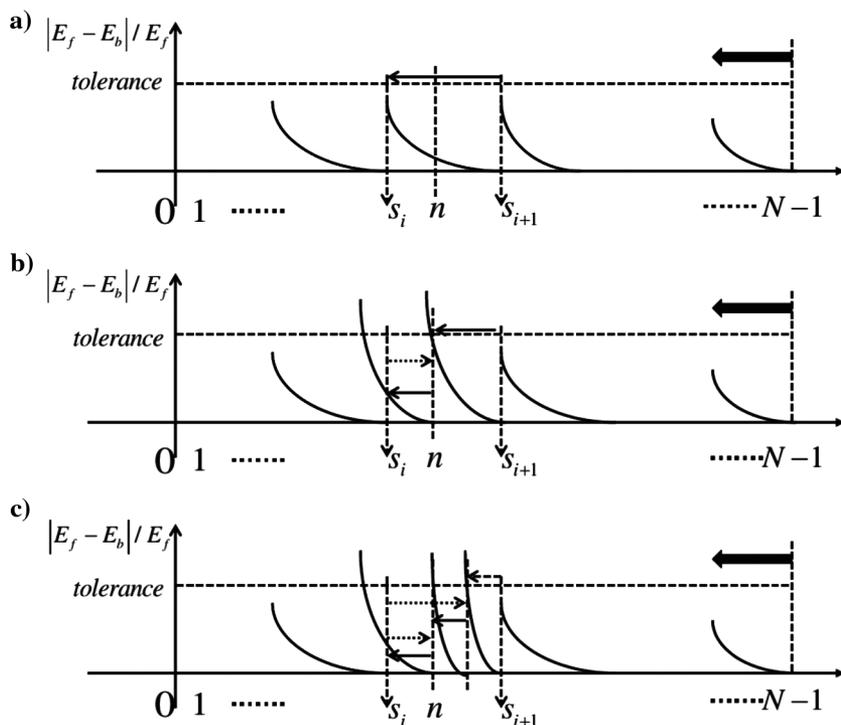


Figure 3. Proposed CARFS method: (a) the best case is that the wavefield at every time step is recovered by reverse propagation without violation of the energy tolerance. The numerical errors are controlled within a negligible magnitude; therefore, there is no need for repeated forward modeling from checkpointing strategy. (b) The CARFS method switches to binomial checkpointing when reverse propagation explodes; i.e., the tolerance of the energy measure is broken. (c) One may need several levels of checkpointing to obtain a backup of the wavefield at time levels between two checkpoints through repeated forward modeling, if the attenuation is strong enough.

erty of the earth. The reference frequencies for GMB mechanisms are discretized logarithmically equidistant (Liu et al., 1976; Casula and Carcione, 1992; Moczo et al., 2007a) over the frequency band of interest. According to Griewank and Walther (2000) and Symes (2007),  $c \approx \log_2(N)$  snapshots are deployed at checkpoint locations following a binomial law. Allowing for accurate wavefield reconstruction, the DFT interpolation technique is used to interpolate the missing boundary values in between the recorded boundary samples (Yang et al., 2016b). The tolerance for CARFS method is fixed to be 1%. As an illustration, we recorded the energy measure and extracted a trace randomly chosen in the computing domain during forward and backward process of CARFS method.

To demonstrate the efficiency of CARFS method without loss of accuracy, the RP method, the optimal checkpointing method, and the proposed CARFS method are tested in acoustic and viscoacoustic cases. Eleven checkpoints are used for checkpointing strategy and CARFS method to achieve a fair comparison. The first test is dedicated to FWI applications, which usually concentrate on relatively low-frequency band because the problem is highly nonlinear with many local minimum, so that the low-frequency background has to be preferred to ensure that the inversion can converge to a correct solution without cycle skipping. Therefore, the reference

frequencies of the three GMB mechanisms are over the range of 2–20 Hz. Allowing for this frequency range, the original 2D Valhall model is allowed to be resampled by a factor of four in lateral and vertical directions without dispersion. The dominant frequency of the Ricker wavelet is  $f_m = 5$  Hz, whereas the temporal sampling is  $\Delta t = 1$  ms. The recomputation ratio shown in Table 2 provides us a quantitative estimation on the efficiency of these methods. The recorded energy measure and extracted trace are shown in Figure 7.

Besides the frequency range of FWI, it is also interesting in a second test to see the behavior of CARFS method over much wider frequency range allowing for the RTM imaging, where high-frequency content plays an important role in delineating the earth structure with high-resolution image. The frequency range within 80 Hz is considered with several representative frequencies. The temporal and spatial samplings are chosen to be  $\Delta t = 0.5$  ms,  $\Delta x = \Delta z = 3.125$  m, such that the numerical simulation using fourth-order finite-difference stencil satisfies the CFL condition without numerical dispersion (approximately five grid points per wavelength) up to 80 Hz. We fix the minimum frequency at 2 Hz, and change the maximum frequency  $f_{\max}$  to see how the efficiency of CARFS method behaves with varying frequency content. The Ricker wavelet is used with the peak frequency  $f_m = f_{\max}/2.5$

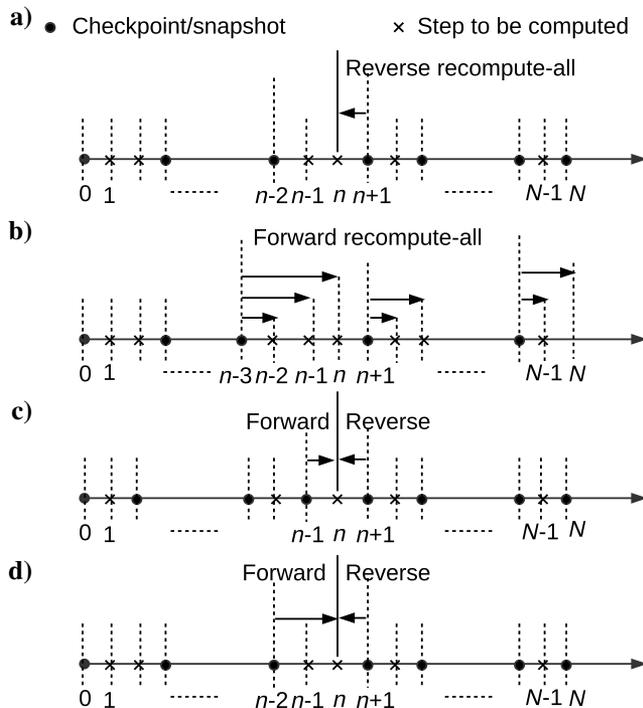


Figure 4. For every noncheckpoint position  $n$ , (a) RPSS performs backward wavefield extrapolation from previous state  $n + 1$  to compute current state, with only one step of reverse simulation cost. (b) Checkpointing strategy tries to recompute the current state  $n$  from the closest checkpoint prior to  $n$ , with generally more than one step of forward simulation cost. (c) Only if the closest checkpoint prior to  $n$  is  $n - 1$ , checkpoint attains the same efficiency as RPSS using only one step of forward simulation cost. If all checkpoints are equally spaced and the number of checkpoints  $c \geq N/2$ , particularly  $c = N/2$ , the distance to reach step  $n$  is 1 for checkpointing and RPSS, either from  $n - 1$  to  $n$  or from  $n + 1$  to  $n$ . (d) Assume the number of checkpoints are equally spaced. When  $c < N/2$ , the timestepping steps required by checkpointing to reach step  $n$  would be larger than RPSS.

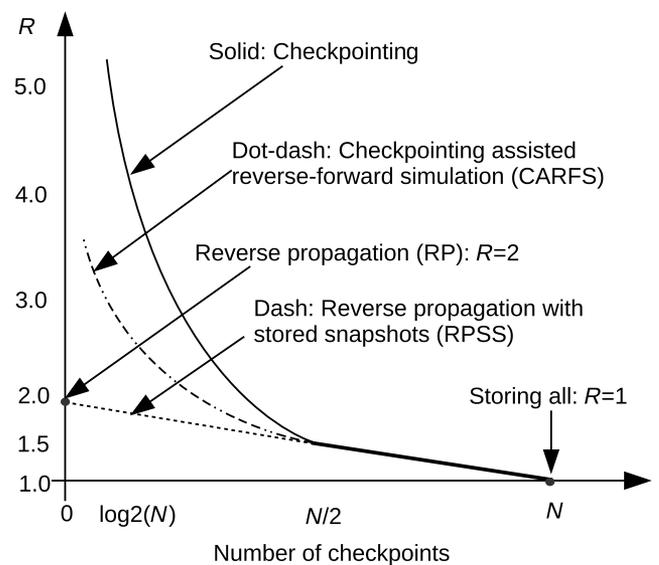


Figure 5. A schematic plot of the recomputation ratios related to the number of checkpoints  $c$  used for reverse propagation, checkpointing, and CARFS methods. The horizontal axis is the number of snapshots stored. Storing all wavefield snapshots leads to a recomputation ratio of  $R = 1$  because only one times of forward simulation was done. The recomputation ratio of reverse propagation approach is  $R = 2$  due to backpropagation of the incident wavefield at each correlation step. In checkpointing strategy, the best tradeoff for the number of checkpoints and repeated forward modeling steps can be achieved at  $c = \log_2(N)$  according to Griewank and Walther (2000) and Symes (2007). For most of application,  $1000 \leq N \leq 10,000$  leads to  $c \leq 14$  and a recomputation ratio bounded by  $R \in (3.6, 4.9)$ . Using  $c < N/2$  equal-spaced checkpoints, RPSS always outperforms checkpointing strategy in terms of recomputation cost. The RPSS and checkpointing can achieve the same recomputation ratio, if  $c \geq N/2$ . Due to smart decision between forward timestepping and reverse propagation, CARFS uses less timesteppings compared with checkpointing (smaller  $N'$ ) to reconstruct the wavefield, implying a smaller recomputation ratio as well as the stability in attenuating medium.

for  $N = 2500$  steps of forward simulation. The resulting number of timesteppings  $N'$  and the recomputation ratio for wavefield reconstruction are shown in Table 3. The energy curves and trace comparisons for two extreme cases with  $f_{\max} = 10$  and 80 Hz are displayed in Figures 8 and 9. As expected, it can be seen that the viscous effects affect much strongly the high-frequency simulations than the lower ones, as shown in Figures 8a and 9a.

## DISCUSSION

As illustrated in the previous example (Table 2), the RP method is efficient to obtain a recomputation ratio of two, but is only applicable in nonattenuating media. The checkpointing strategy treats acoustic and viscoacoustic wave propagation equally with a recomputation much higher than two for classical number of stored checkpoints ( $R = 4.272$  in the example). The CARFS method converges toward the RP method with several snapshots stored in a nonviscous acoustic case, and it is able to benefit from reverse propagation to reduce the number of timestepping of backward reconstruction in the presence of attenuation, resulting in a much smaller recomputation ratio than checkpointing ( $R = 2.664$  in the example). Figure 7 also shows that the energy measure recorded during backward computation matches quite well the energy measure monitored in forward simulation, whereas the extracted trace during backward reconstruction is the same as the trace recorded in forward simulation (the amplitude of the error trace is two orders of magnitude smaller than the original trace). It indicates that CARFS method is efficient and accurate enough for imaging purposes.

According to Table 3, when the maximum reference frequency of GMB is low, CARFS almost performs reverse propagation at every

noncheckpoint step. For example, only 5462 steps are required to do the forward simulation (2500 steps) and backward reconstruction (2942 steps) for  $f_{\max} = 10$  Hz (a complete reverse propagation requires 2500 steps, if there is no stability issue). When the maximum reference frequency becomes higher, more backward timesteppings are required, leading to a larger recomputation ratio. It is very interesting to note that the increment of the recomputation ratio of CARFS becomes smaller as the frequency goes higher; however, it is always much less than the recomputation ratio of optimal checkpointing. No matter what is your maximum reference frequency, the required number of timesteppings of optimal checkpointing is the same in this attenuating medium ( $N' = 10,680$ ), leading to the same recomputation ratio ( $R = 4.272$ ). Therefore, CARFS method achieves the same goal more efficiently than optimal checkpointing, adaptive to the frequency content considered in GMB mechanisms.

It is important to point out that the energy measure we use in CARFS is the total energy for acoustics instead of the one for viscoacoustics (see Appendix A). To detect the instability of reverse propagation in CARFS, one may use some other quantities rather than the energy measure. A good alternative to the energy measure used in this paper should be efficient to compute and sensitive to the variations of pressure and particle velocities in the wavefield. A key ingredient of the CARFS algorithm for successful 3D large-scale application is that the heavy-storage requirement has to be decreased dramatically up to one to two orders of magnitude in terms of Nyquist sampling theorem, whereas the missing values between sampled boundary elements need to be interpolated accurately (Yang et al., 2016a, 2016b). The decimation ratio depends on the velocity contrast. For example, in the previous numerical results of 2D Valhall model, we tested a ratio of four for 2–20 Hz FWI oriented reconstruction, and even much larger ratio of 10 for RTM oriented reconstruction with varying frequency within 80 Hz. The CARFS algorithm works equally well even when the decimation ratio for the boundary values is very large. Besides, the tolerance of the energy mismatch may be chosen differently in CARFS. However, the empirical experiments show that it has little influence on the efficiency of CARFS algorithm.

**Table 1. Comparison of memory and computation complexity and applicability for the methods in different medium.**

Methods	Memory overhead	Recomputation ratio	Nonviscous	Viscous (GMB model)
Storing-all	$\frac{1}{a} n_x^3 N$	$R_S = 1$	Yes	Yes
Checkpointing	$c n_x^3$	$R_C > 1$	Yes	Yes
RP	$\frac{1}{b} n_x^2 N$	$R_{RP} = 2$	Yes	No
RPSS	$c n_x^3 + \frac{1}{b} n_x^2 N$	$R_{RPSS} = 1 + \frac{N-c}{N}$	Yes	Yes/no
CARFS	$c n_x^3 + \frac{1}{b} n_x^2 N$	$R_{RPSS} < R_{CARFS} < R_C$	Yes	Yes

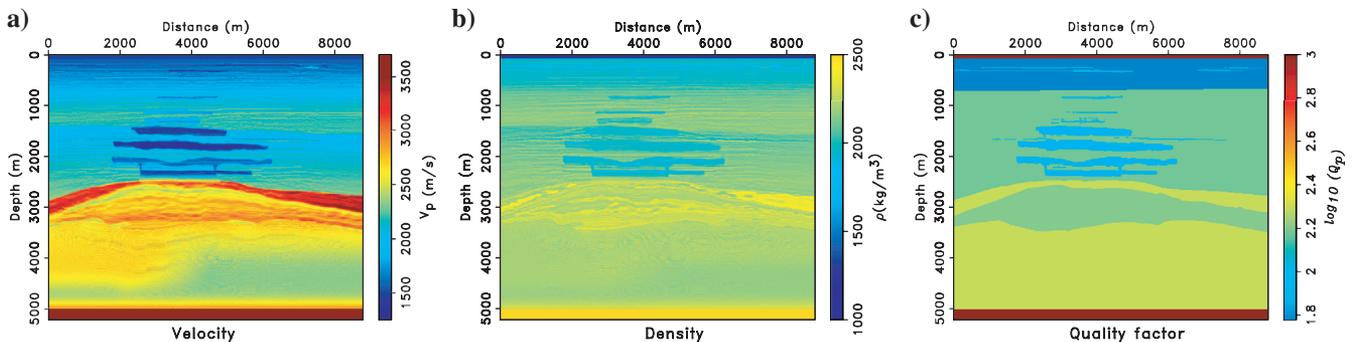


Figure 6. The 2D Valhall synthetic model: (a) the velocity, (b) the density, and (c) the quality factor  $Q$ .

**Table 2. Recomputation ratio for  $N = 2500$  steps simulation in 2D Valhall model.**

Methods	RP	Checkpointing	CARFS
Acoustic	2.0 ( $N' = 5000$ )	4.272 ( $N' = 10,680$ )	1.9956 ( $N' = 4989$ )
Viscoacoustic	–	4.272 ( $N' = 10,680$ )	2.668 ( $N' = 6670$ )

**CONCLUSION**

In this study, we have proposed to combine checkpointing and reserve propagation in an integrated so-called CARFS method to build the incident wavefield in reverse time order in the context of GMB-based viscoacoustic media. We benefit from the efficient reverse propagation as long as the exponentially growing numerical

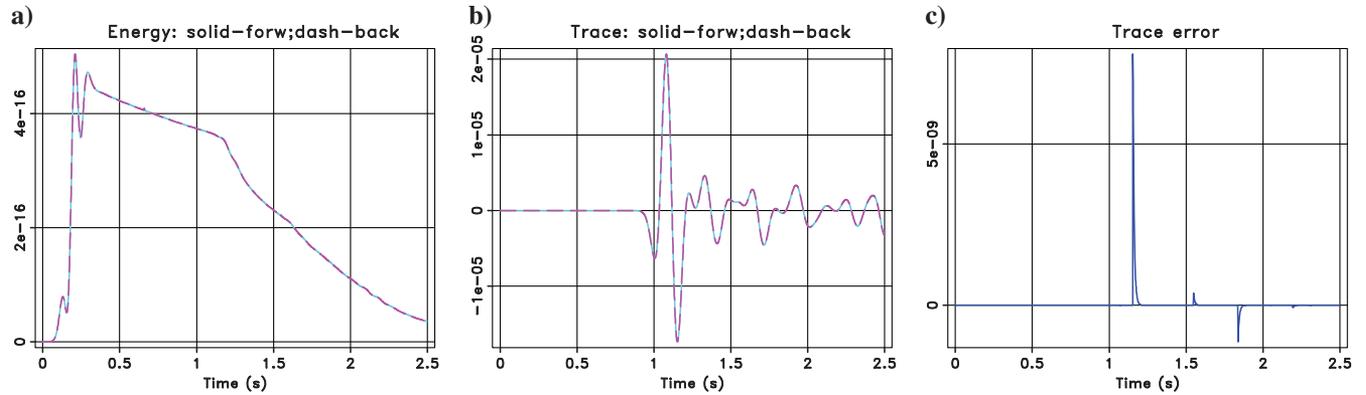


Figure 7. Comparison of the energy measure and randomly extracted seismic trace during the forward simulation and backward computation in resampled 2D Valhall model (2–20 Hz). (a) The energy measure recorded during backward computation matches quite well the energy measure monitored in forward simulation. (b) The CARFS reconstruction works so well that the original traces are essentially indistinguishable. (c) By subtracting the trace recorded in reconstruction with the trace recorded in forward simulation at the same location, the trace error is obtained with four orders of magnitude smaller than the magnitude of the original trace in panel (b).

**Table 3. Number of timesteps and recomputation ratio in CARFS for varying highest reference frequency.**

$f_{\max}$ (Hz)	10	15	20	25	30	40	50	60	70	80
$N'$	5462	5620	5919	6142	6302	6569	6781	6965	7091	7227
$R$	2.1848	2.248	2.3676	2.4568	2.5208	2.6276	2.7124	2.786	2.8364	2.8908

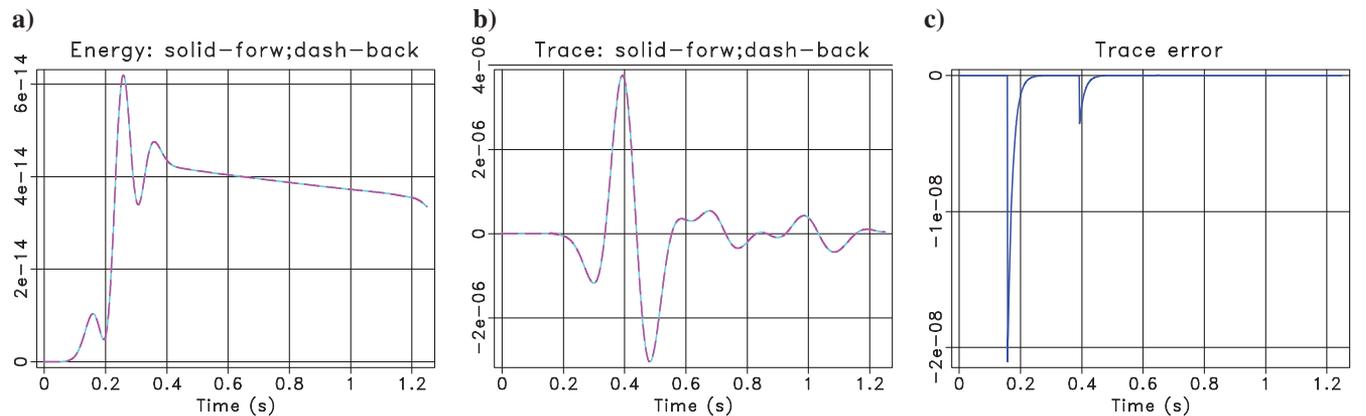


Figure 8. (a) The energy measure, (b) randomly extracted seismic traces, and (c) trace error during the forward simulation and backward computation in the 2D Valhall model ( $f_{\max} = 10$  Hz).

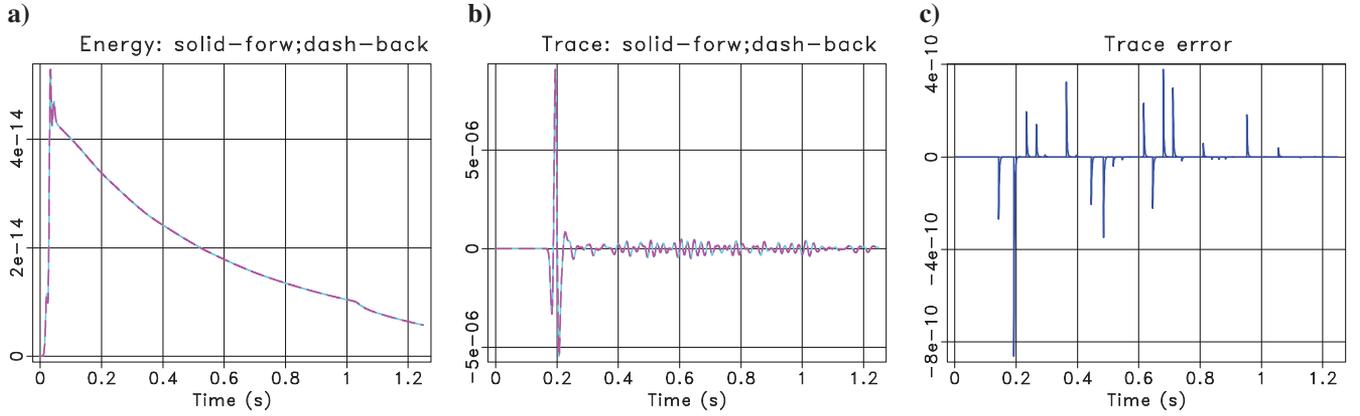


Figure 9. (a) The energy measure, (b) randomly extracted seismic traces, and (c) trace error during the forward simulation and backward computation in the 2D Valhall model ( $f_{\max} = 80$  Hz).

errors within a user-specified tolerance are verified. We automatically switch to checkpointing, once the amplified errors go beyond this threshold. The proposed CARFS approach is a hybrid relative of reverse propagation and optimal checkpointing strategy in the sense that it is equivalent to reverse propagation with stored snapshots for nonattenuating media, and converges to optimal checkpointing if every step is done by forward modeling. The numerical experiment demonstrates that the CARFS method can obtain accurate wavefield reconstruction using fewer time steps than optimal checkpointing, even if seismic attenuation is quite strong. The efficiency of the CARFS can be further improved when explicitly considering the crosscorrelation components in geophysical applications. The proposed CARFS method can be also applied to general anisotropic viscous media, with promising efficiency in practical RTM and FWI implementations.

This integrated algorithmic workflow could nicely complement any storage strategy linked with computer hardware at our disposal for performing the task we wish to achieve. Data compression, efficient I/O procedure, and fast network links could balance the cursor between storage and computational tasks while considering the CARFS method.

## ACKNOWLEDGMENTS

The authors thank W. Symes and J. Anderson for the communication on Griewank's binomial checkpointing strategy. We really appreciate L. Combe for the help to access checkpointing library Revolve with Fortran interface. We thank D. Komatitsch, B. Nguyen, and two anonymous reviewers for valuable comments to improve the paper. This study was partially funded by the SEISCOPE consortium (<http://seiscope2.osug.fr>), sponsored by BP, CGG, CHEVRON, EXXON-MOBIL, JGI, PETROBRAS, SAUDI ARAMCO, SCHLUMBERGER, SHELL, SINOPEC, STATOIL, TOTAL, and WOODSIDE. This study was granted access to the HPC resources of the Froggy platform of the CIMENT infrastructure (<https://ciment.ujf-grenoble.fr>), which is supported by the Rhône-Alpes region (GRANT CPER07\_13 CIRA), the OSUG@2020 labex (reference ANR10 LABX56), and the Equip@Meso project (reference ANR-10-EQPX-29-01) of the programme Investissements d'Avenir supervised by the Agence Nationale pour la Recherche, and the HPC resources of CINES/IDRIS under the allocation 046091 made by GENCI.

## APPENDIX A

### DISCRETIZATION OF THE VISCOACOUSTIC WAVE EQUATION

Consider the half-integer time step for particle velocity  $\mathbf{v} = (v_x, v_y, v_z)$  and the integer time step of pressure  $p$  as well as the memory variables  $\xi_l$ . The timestepping from level  $n$  to  $n+1$  can be done in the following:

$$\begin{cases} v_x^{n+\frac{1}{2}} = v_x^{n-\frac{1}{2}} + \frac{\Delta t}{\rho} \partial_x p^n, \\ v_y^{n+\frac{1}{2}} = v_y^{n-\frac{1}{2}} + \frac{\Delta t}{\rho} \partial_y p^n, \\ v_z^{n+\frac{1}{2}} = v_z^{n-\frac{1}{2}} + \frac{\Delta t}{\rho} \partial_z p^n, \\ p^{n+1} = p^n + \Delta t \kappa \left( \nabla \cdot \mathbf{v}|_{n+\frac{1}{2}} - \sum_{l=1}^L Y_l \xi_l^{n+\frac{1}{2}} \right), \end{cases} \quad (\text{A-1})$$

where  $n$  is the index of time steps,  $\nabla \cdot \mathbf{v}|_{n+(1/2)} = \partial_x v_x^{n+(1/2)} + \partial_y v_y^{n+(1/2)} + \partial_z v_z^{n+(1/2)}$ . Denote  $d = \omega_l \nabla \cdot \mathbf{v}$ . The analytical solutions of the ODEs are given by

$$\xi_l(t) = C(\omega_l) e^{-\omega_l t} + \frac{1}{\omega_l} d, \quad l = 1, \dots, L, \quad (\text{A-2})$$

therefore, the updating of the memory variables are

$$\begin{aligned} \xi_l((n+1)\Delta t) &= e^{-\omega_l \Delta t} \xi_l(n\Delta t) + \frac{1}{\omega_l} (1 - e^{-\omega_l \Delta t}) d|_{n+\frac{1}{2}}, \\ &= e^{-\omega_l \Delta t} \xi_l(n\Delta t) + (1 - e^{-\omega_l \Delta t}) \nabla \cdot \mathbf{v}|_{n+\frac{1}{2}}, l = 1, \dots, L. \end{aligned} \quad (\text{A-3})$$

Of course, the ODEs can also be discretized by

$$\frac{\xi_l^{n+1} - \xi_l^n}{\Delta t} + 0.5\omega_l(\xi_l^{n+1} + \xi_l^n) = d|_{n+\frac{1}{2}}, \quad (\text{A-4})$$

yielding

$$\begin{aligned} \xi_l((n+1)\Delta t) &= \frac{1-0.5\omega_l\Delta t}{1+0.5\omega_l\Delta t} \xi_l(n\Delta t) \\ &+ \frac{\omega_l\Delta t}{1+0.5\omega_l\Delta t} \nabla \cdot \mathbf{v}|_{n+\frac{1}{2}}, \quad l = 1, \dots, L, \end{aligned} \quad (\text{A-5})$$

which is equivalent to the semianalytical discretization in equation A-3, allowing for the following Páde table (Liu and Archuleta, 2006):

$$\exp(z) \approx \frac{1+0.5z}{1-0.5z}, \quad z = -\omega_l\Delta t. \quad (\text{A-6})$$

Note that in equation A-1,  $\xi_l^{n+(1/2)}$  have to be computed to extrapolate the wave propagation. To preserve the second-order accuracy for the time discretization, the following approximation is useful (Moczo et al., 2007a, 2007b):

$$\xi_l^{n+\frac{1}{2}} \approx \frac{1}{2}(\xi_l^n + \xi_l^{n+1}), \quad l = 1, \dots, L. \quad (\text{A-7})$$

As a consequence, we obtain the discretized timestepping shown in equation 12.

### The total energy in viscoacoustic media

The energy of the GMB-based viscoacoustic system in equation 10 is defined by the following nonnegative measure:

$$\begin{aligned} E &= \frac{1}{2} \langle \rho \mathbf{v}, \mathbf{v} \rangle_\Omega + \frac{1}{2} \left\langle \frac{1}{M_r} \left( p - \sum_{l=1}^L \frac{M_u Y_l}{\omega_l} \xi_l \right), p - \sum_{l=1}^L \frac{M_u Y_l}{\omega_l} \xi_l \right\rangle_\Omega \\ &+ \frac{1}{2} \sum_{l=1}^L \left\langle \frac{M_u Y_l}{\omega_l^2} \xi_l, \xi_l \right\rangle_\Omega, \end{aligned} \quad (\text{A-8})$$

where the relaxed modulus  $M_r = M_u(1 - \sum_{l=1}^L Y_l)$  is assumed to be positive. The first derivative of the energy is

$$\begin{aligned} \partial_t E &= \langle \rho \partial_t \mathbf{v}, \mathbf{v} \rangle_\Omega + \left\langle \frac{1}{M_r} \partial_t \left( p - \sum_{l=1}^L \frac{M_u Y_l}{\omega_l} \xi_l \right), \right. \\ &\left. p - \sum_{l=1}^L \frac{M_u Y_l}{\omega_l} \xi_l \right\rangle_\Omega + \sum_{l=1}^L \left\langle \frac{M_u Y_l}{\omega_l^2} \partial_t \xi_l, \xi_l \right\rangle_\Omega. \end{aligned} \quad (\text{A-9})$$

Inserting equation 10c into equation 10b gives

$$\partial_t \left( p - \sum_{l=1}^L \frac{M_u Y_l}{\omega_l} \xi_l \right) = M_u \left( 1 - \sum_{l=1}^L Y_l \right) \nabla \cdot \mathbf{v} = M_r \nabla \cdot \mathbf{v}. \quad (\text{A-10})$$

Taking the inner product over the domain  $\Omega$  for equation 10a with  $\mathbf{v}$ , equation A-10 with  $(1/M_r)(p - \sum_{l=1}^L (M_u Y_l/\omega_l)\xi_l)$ , and equation 10c with  $(M_u Y_l/\omega_l^2)\xi_l$  gives

$$\langle \rho \partial_t \mathbf{v}, \mathbf{v} \rangle_\Omega = \langle \nabla p, \mathbf{v} \rangle_\Omega, \quad (\text{A-11})$$

$$\begin{aligned} &\left\langle \frac{1}{M_r} \partial_t \left( p - \sum_{l=1}^L \frac{M_u Y_l}{\omega_l} \xi_l \right), \left( p - \sum_{l=1}^L \frac{M_u Y_l}{\omega_l} \xi_l \right) \right\rangle_\Omega \\ &= \langle \nabla \cdot \mathbf{v}, p \rangle_\Omega - \left\langle \nabla \cdot \mathbf{v}, \sum_{l=1}^L \frac{M_u Y_l}{\omega_l} \xi_l \right\rangle_\Omega, \end{aligned} \quad (\text{A-12})$$

$$\begin{aligned} &\left\langle \frac{M_u Y_l}{\omega_l^2} \partial_t \xi_l, \xi_l \right\rangle_\Omega = - \left\langle \frac{M_u Y_l}{\omega_l} \xi_l, \xi_l \right\rangle_\Omega \\ &+ \left\langle \nabla \cdot \mathbf{v}, \frac{M_u Y_l}{\omega_l} \xi_l \right\rangle_\Omega, \quad l = 1, 2, \dots, L. \end{aligned} \quad (\text{A-13})$$

According to the zero-valued boundary condition and integration by parts, we have  $\langle \nabla p, \mathbf{v} \rangle_\Omega = -\langle \nabla \cdot \mathbf{v}, p \rangle_\Omega$ . Summing over all the subequations in equation A-11 gives

$$\partial_t E = - \sum_{l=1}^L \left\langle \frac{M_u Y_l}{\omega_l} \xi_l, \xi_l \right\rangle_\Omega \leq 0. \quad (\text{A-14})$$

Therefore, the viscoacoustic system (equation 10) is stable because the energy is a monotonic decreasing function of time for positive  $\omega_l, Y_l$ . Groby and Tsogka (2006) carry out the similar energy analysis based on the variable exchange  $M_u Y_l \xi_l = \omega_l \eta_l = M_r \tilde{Y}_l \xi_l$ , yielding the energy expression and its first derivative

$$\begin{aligned} E &= \frac{1}{2} \langle \rho \mathbf{v}, \mathbf{v} \rangle_\Omega + \frac{1}{2} \left\langle \frac{1}{M_r} \left( p - \sum_{l=1}^L \eta_l \right), p - \sum_{l=1}^L \eta_l \right\rangle_\Omega \\ &+ \frac{1}{2} \sum_{l=1}^L \left\langle \frac{1}{\tilde{Y}_l M_r} \eta_l, \eta_l \right\rangle_\Omega, \end{aligned} \quad (\text{A-15})$$

and

$$\partial_t E = - \sum_{l=1}^L \left\langle \frac{\omega_l}{\tilde{Y}_l M_r} \eta_l, \eta_l \right\rangle_\Omega \leq 0. \quad (\text{A-16})$$

We prefer equation A-9 to equation A-15 because it is much easier to see that the viscoacoustic equation reduces to acoustic case, when  $Y_l = 0$  in terms of equation A-14; there is no attenuation in wave propagation due to  $\partial_t E = 0$ , whereas equation A-16 does not allow  $\tilde{Y}_l = 0$ . With  $Y_l = 0$ , the memory variables  $\xi_l$  for attenuation are then excluded, leading to the acoustic-wave equation 1 and the energy expression 16.

## REFERENCES

- Anderson, J. E., L. Tan, and D. Wang, 2012, Time-reversal checkpointing methods for RTM and FWI: *Geophysics*, **77**, no. 4, S93–S103, doi: [10.1190/geo2011-0114.1](https://doi.org/10.1190/geo2011-0114.1).
- Bai, J., D. Yingst, R. Bloor, and J. Leveille, 2014, Viscoacoustic waveform inversion of velocity structures in the time domain: *Geophysics*, **79**, no. 3, R103–R119, doi: [10.1190/geo2013-0030.1](https://doi.org/10.1190/geo2013-0030.1).
- Blanch, J., J. O. A. Robertson, and W. W. Symes, 1995, Modeling of a constant Q: Methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique: *Geophysics*, **60**, 176–184, doi: [10.1190/1.1443744](https://doi.org/10.1190/1.1443744).

- Blanch, J. O., W. W. Symes, and R. Versteeg, 1998, A numerical study of linear viscoacoustic inversion: Comparison of Seismic Inversion Methods on a Single Real Data Set, *4*, 13–44, doi: [10.1190/1.9781560802082](https://doi.org/10.1190/1.9781560802082).
- Boehm, C., M. Hanzlich, J. de la Puente, and A. Fichtner, 2015, Wavefield compression for adjoint methods in full-waveform inversion: *Geophysics*, **81**, this issue, doi: [10.1190/geo2015-0653.1](https://doi.org/10.1190/geo2015-0653.1).
- Brossier, R., B. Pajot, L. Combe, S. Operto, L. Métivier, and J. Virieux, 2014, Time and frequency-domain FWI implementations based on time solver: Analysis of computational complexities: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, 4200–4204.
- Caputo, M., 1967, Linear models of dissipation whose Q is almost frequency independent-II: *Geophysical Journal International*, **13**, 529–539, doi: [10.1111/j.1365-246X.1967.tb02303.x](https://doi.org/10.1111/j.1365-246X.1967.tb02303.x).
- Carcione, J. M., D. Kosloff, and R. Kosloff, 1988a, Wave propagation simulation in a linear viscoacoustic medium: *Geophysical Journal International*, **93**, 393–401, doi: [10.1111/j.1365-246X.1988.tb02010.x](https://doi.org/10.1111/j.1365-246X.1988.tb02010.x).
- Carcione, J. M., D. Kosloff, and R. Kosloff, 1988b, Wave propagation simulation in a linear viscoelastic medium: *Geophysical Journal International*, **95**, 597–611, doi: [10.1111/j.1365-246X.1988.tb06706.x](https://doi.org/10.1111/j.1365-246X.1988.tb06706.x).
- Casula, G., and J. Carcione, 1992, Generalized mechanical model analogies of linear viscoelastic behaviour: *Bollettino di geofisica teorica ed applicata*, XXXIV, 235–256.
- Clapp, R., 2008, Reverse time migration: Saving the boundaries: Technical Report SEP-136, Stanford Exploration Project.
- Dussaud, E., W. W. Symes, P. Williamson, L. Lemaistre, P. Singer, B. Denel, and A. Cherrett, 2008, Computational strategies for reverse-time migration: 78th Annual International Meeting, SEG, Expanded Abstracts, 2267–2271, doi: [10.1190/1.3059336](https://doi.org/10.1190/1.3059336).
- Dutta, G., and G. T. Schuster, 2014, Attenuation compensation for least-squares reverse time migration using the viscoacoustic-wave equation: *Geophysics*, **79**, no. 6, S251–S262, doi: [10.1190/geo2013-0414.1](https://doi.org/10.1190/geo2013-0414.1).
- Emmerich, H., and M. Korn, 1987, Incorporation of attenuation into time-domain computations of seismic wave fields: *Geophysics*, **52**, 1252–1264, doi: [10.1190/1.1442386](https://doi.org/10.1190/1.1442386).
- Futterman, W., 1962, Dispersive body waves: *Journal Geophysical Research*, **67**, 5279–5291, doi: [10.1029/JZ067i013p05279](https://doi.org/10.1029/JZ067i013p05279).
- Gauthier, O., J. Virieux, and A. Tarantola, 1986, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results: *Geophysics*, **51**, 1387–1403, doi: [10.1190/1.1442188](https://doi.org/10.1190/1.1442188).
- Griewank, A., 1992, Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation: *Optimization Methods and Software*, **1**, 35–54, doi: [10.1080/10556789208805505](https://doi.org/10.1080/10556789208805505).
- Griewank, A., and A. Walther, 2000, Algorithm 799: Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation: *ACM Transactions on Mathematical Software*, **26**, 19–45, doi: [10.1145/347837.347846](https://doi.org/10.1145/347837.347846).
- Grobjy, J.-P., and C. Tsogka, 2006, A time domain method for modeling viscoacoustic wave propagation: *Journal of Computational Acoustics*, **14**, 201–236, doi: [10.1142/S0218396X06003001](https://doi.org/10.1142/S0218396X06003001).
- Kjartansson, E., 1979, Constant Q-wave propagation and attenuation: *Journal of Geophysical Research*, **84**, 4737–4748, doi: [10.1029/JB084iB09p04737](https://doi.org/10.1029/JB084iB09p04737).
- Komatitsch, D., Z. Xie, E. Bozdag, E. S. de Andrade, D. Peter, Q. Liu, and J. Tromp, 2016, Anelastic sensitivity kernels with parsimonious storage for adjoint tomography and full waveform inversion: *Geophysical Journal International*, doi: [10.1093/gji/ggw224](https://doi.org/10.1093/gji/ggw224).
- Liu, H.-P., D. L. Anderson, and H. Kanamori, 1976, Velocity dispersion due to anelasticity: Implications for seismology and mantle composition: *Geophysical Journal International*, **47**, 41–58, doi: [10.1111/j.1365-246X.1976.tb01261.x](https://doi.org/10.1111/j.1365-246X.1976.tb01261.x).
- Liu, P., and R. J. Archuleta, 2006, Efficient modeling of Q for 3D numerical simulation of wave propagation: *Bulletin of the Seismological Society of America*, **96**, 1352–1358, doi: [10.1785/0120050173](https://doi.org/10.1785/0120050173).
- Malinowski, M., and S. Operto, 2008, Quantitative imaging of the Permo-Mesozoic complex and its basement by frequency domain waveform tomography of wide-aperture seismic data from the Polish Basin: *Geophysical Prospecting*, **56**, 805–825, doi: [10.1111/j.1365-2478.2007.00680.x](https://doi.org/10.1111/j.1365-2478.2007.00680.x).
- Moczo, P., and J. Kristek, 2005, On the rheological models used for time-domain methods of seismic wave propagation: *Geophysical Research Letters*, **32**, L01306, doi: [10.1029/2004GL021598](https://doi.org/10.1029/2004GL021598).
- Moczo, P., J. Kristek, M. Galis, P. Pazak, and M. Balazovjeh, 2007a, The finite-difference and finite-element modeling of seismic wave propagation and earthquake motion: *Acta Physica Slovaca*, **52**, 177–406.
- Moczo, P., J. Robertsson, and L. Eisner, 2007b, The finite-difference time-domain method for modeling of seismic wave propagation: Elsevier/Academic Press.
- Nguyen, B. D., and G. A. McMechan, 2015, Five ways to avoid storing source wavefield snapshots in 2D elastic prestack reverse time migration: *Geophysics*, **80**, no. 1, S1–S18, doi: [10.1190/geo2014-0014.1](https://doi.org/10.1190/geo2014-0014.1).
- Operto, S., A. Miniussi, R. Brossier, L. Combe, L. Métivier, V. Monteiller, A. Ribodetti, and J. Virieux, 2015, Efficient 3-D frequency-domain monoparameter full-waveform inversion of ocean-bottom cable data: Application to Valhall in the visco-acoustic vertical transverse isotropic approximation: *Geophysical Journal International*, **202**, 1362–1391, doi: [10.1093/gji/ggv226](https://doi.org/10.1093/gji/ggv226).
- Prabhat and Q. Koziol, 2014, High performance parallel I/O: Chapman and Hall/CRC Inc.
- Prieux, V., R. Brossier, S. Operto, and J. Virieux, 2013, Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from Valhall field. Part 1: Imaging compressional wave speed, density and attenuation: *Geophysical Journal International*, **194**, 1640–1664, doi: [10.1093/gji/ggt177](https://doi.org/10.1093/gji/ggt177).
- Stumm, P., and A. Walther, 2009, Multistage approaches for optimal offline checkpointing: *SIAM Journal on Scientific Computing*, **31**, 1946–1967, doi: [10.1137/080718036](https://doi.org/10.1137/080718036).
- Sun, W., and L.-Y. Fu, 2013, Two effective approaches to reduce data storage in reverse time migration: *Computers & Geosciences*, **56**, 69–75, doi: [10.1016/j.cageo.2013.03.013](https://doi.org/10.1016/j.cageo.2013.03.013).
- Symes, W. W., 2007, Reverse time migration with optimal checkpointing: *Geophysics*, **72**, no. 5, SM213–SM221, doi: [10.1190/1.2742686](https://doi.org/10.1190/1.2742686).
- Tromp, J., C. Tape, and Q. Liu, 2005, Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels: *Geophysical Journal International*, **160**, 195–216, doi: [10.1111/j.1365-246X.2004.02453.x](https://doi.org/10.1111/j.1365-246X.2004.02453.x).
- Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: *Geophysics*, **74**, no. 6, WCC1–WCC26, doi: [10.1190/1.3238367](https://doi.org/10.1190/1.3238367).
- Yang, P., R. Brossier, and J. Virieux, 2016a, Downsampling plus interpolation for wavefield reconstruction by reverse propagation: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, SBT5 08.
- Yang, P., R. Brossier, and J. Virieux, 2016b, Wavefield reconstruction by interpolating significantly decimated boundaries: *Geophysics*, **81**, no. 5, T197–T209, doi: [10.1190/geo2015-0711.1](https://doi.org/10.1190/geo2015-0711.1).
- Yang, P., J. Gao, and B. Wang, 2014a, RTM using effective boundary saving: A staggered grid GPU implementation: *Computers & Geosciences*, **68**, 64–72, doi: [10.1016/j.cageo.2014.04.004](https://doi.org/10.1016/j.cageo.2014.04.004).
- Yang, P., B. Wang, and J. Gao, 2014b, Using the effective boundary saving strategy in GPU-based RTM programming: 84th Annual International Meeting, SEG, Expanded Abstracts, 4014–4021.
- Yang, P., J. Gao, and B. Wang, 2015, A graphics processing unit implementation of time-domain full-waveform inversion: *Geophysics*, **80**, no. 3, F31–F39, doi: [10.1190/geo2014-0283.1](https://doi.org/10.1190/geo2014-0283.1).
- Zhu, T., 2015, Viscoelastic time-reversal imaging: *Geophysics*, **80**, no. 2, A45–A50, doi: [10.1190/geo2014-0327.1](https://doi.org/10.1190/geo2014-0327.1).
- Zhu, T., J. Harris, and B. Biondi, 2014, Q-compensated reverse-time migration: *Geophysics*, **79**, no. 6, S77–S87, doi: [10.1190/geo2013-0344.1](https://doi.org/10.1190/geo2013-0344.1).