Checkpointing-assisted Reverse Forward Simulation: an optimal recomputation method for FWI and RTM

Pengliang Yang, Romain Brossier, Ludovic Metivier, Jean Virieux, Univ. Grenoble Alpes, France.*

SUMMARY

3D implementations of reverse time migration and full waveform inversion requires efficient schemes to access the incident field in order 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, while checkpointing strategy is a stable alternative at the expense of increased computational cost through repeated forward modeling. In this study, we propose a checkpointing-assisted reverse-forward simulation (CARFS) method in the context of viscoacoustic wave propagation with 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 timestepping cost and an energy measure. Numerical experiments demonstrate that the CARFS method allows accurate wavefield reconstruction using less timesteppings than optimal checkpointing, even if seismic attenuation is very strong. The proposed CARFS method can be straightforwardly applied to general anisotropic viscous media, providing an accurate and efficient tool for practical RTM and FWI implementation.

INTRODUCTION

Seismic imaging techniques such as reverse time migration (RTM) and full waveform inversion (FWI) are computationally and memory demanding for 3D applications. One challenging step of such implementations is applying the imaging condition of RTM, equivalent to the gradient building step of FWI: in both cases, one needs to access the incident wavefield backwards in time while computing the adjoint/receiver wavefield. Several strategies have been proposed to conduct this key step: (1) The most naive strategy consists in storing the incident wavefield on disk when performing the forward modeling. Then it is accessed it at each time step when backpropagating the adjoint/receiver wavefield for building the gradient. This approach is time consuming due to highly demanding I/O cost. (2) A more efficient scheme is the wavefield reconstruction by reverse propagation (RP) based on the wavefield values stored at boundaries and on the final wavefield snapshot. The adjoint wavefield is formed 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 backwards in time. (3) Another option is the optimal checkpointing strategy (Griewank and Walther, 2000; Symes, 2007; Anderson et al., 2012), which recomputes the incident field forward in time inside a time window started with saved snapshots (checkpoints) at specific times. The recomputation ratio and the memory request of the third approach is generally higher than the second one.

The development of RTM and FWI in complex media, in particular involving attenuation, requires efficient computing methods for reconstructing the incident wavefield in time-reversal manner. This study proposes a checkpointing assisted reverse-forward simulation (CARFS) method, which combines the efficiency of RP and stability of checkpointing in attenuating medium.

VISCOACOUSTIC MODELING AND ITS REVERSE PROPAGATION

It is acknowledged that seismic attenuation is of major importance in many geological frameworks and has to be taken into account to approximate the realistic wave propagation in the Earth. Many physical models have been proposed to mimic attenuation: Maxwell body, Kelvin-Voigt model, standard linear solid (SLS) model, as well as their generalizations: the generalized Maxwell body (GMB) (Moczo and Kristek, 2005) or its equivalent, the generalized Zener body (GZB) (Carcione et al., 1988). Using GMB or GZB models, the viscoacoustic wave equation reads (Moczo et al., 2007b)

$$\begin{cases} \rho\,\partial_t \mathbf{v} = \nabla p \\ \partial_t p = M_u(\nabla\cdot\mathbf{v} - \sum_{l=1}^{L} Y_l \xi_l) \\ \partial_t \xi_l + \omega_l \xi_l = \omega_l \nabla\cdot\mathbf{v}, \qquad l = 1,2,\ldots,L \end{cases} \tag{1}$$

where the pressure is denoted by p, the particle velocities by $\mathbf{v}$ and the unrelaxed modulus by $M_u = \kappa = \rho v^2$. Equation 1 includes L memory variables ξ_l. Each one is associated to a reference frequency ω_l, allowing for seismic attenuation associated with the quality factor (Moczo et al., 2007a,eq. 117)

$$Q^{-1}(\omega) = \sum_{l=1}^{L} Y_l \frac{\omega_l \omega}{\omega_l^2+\omega^2} / (1 - \sum_{l=1}^{L} Y_l \frac{\omega_l^2}{\omega_l^2+\omega^2}) \tag{2}$$

Considering a leap-frog time integration scheme with half-integer time-step for particle velocity $\mathbf{v} = (v_x, v_y, v_z)$, integer time-step for pressure p and memory variables ξ_l, we end up with the explicit scheme

$$\begin{cases} v_{x/y/z}^{n+\frac{1}{2}} = v_{x/y/z}^{n-\frac{1}{2}} + \frac{\Delta t}{\rho}\partial_{x/y/z} p^n \\ \xi_l^{n+1} = e^{-\omega_l \Delta t}\xi_l^n + (1-e^{-\omega_l\Delta t})\nabla\cdot\mathbf{v}|_{n+\frac{1}{2}}, \quad l = 1,\ldots,L, \\ \xi_l^{n+\frac{1}{2}} = \frac{1}{2}(\xi_l^n + \xi_l^{n+1}), \\ p^{n+1} = p^n + \Delta t\,\kappa\left(\nabla\cdot\mathbf{v}|_{n+\frac{1}{2}} - \sum_{l=1}^{L} Y_l \xi_l^{n+\frac{1}{2}}\right) \end{cases} \tag{3}$$

This scheme can be summarized as $\mathbf{w}^{n+1} = F^n \mathbf{w}^n$ using the wavefield state vector $\mathbf{w}^n = (\mathbf{v}^{n-\frac{1}{2}}, p^n, \xi^n)$. Note that the coefficient $e^{-\omega_l\Delta t}$ in front of the field ξ_l^n in 3 is always smaller than one. With the final snapshots and saved values at boundaries at hand, the reverse propagation of the forward wavefield from

time level $n+1$ to n is given by

$$\begin{cases} \xi_l^n = e^{\omega_l \Delta t}\xi_l^{n+1} + (1-e^{\omega_l \Delta t})\nabla\cdot\mathbf{v}|_{n+\frac{1}{2}}, \quad l=1,\ldots,L, \\ \xi_l^{n+\frac{1}{2}} = \frac{1}{2}(\xi_l^n + \xi_l^{n+1}), \\ p^n = p^{n+1} - \Delta t\kappa\left(\nabla\cdot\mathbf{v}|_{n+\frac{1}{2}} - \sum_{l=1}^{L} Y_l \xi_l^{n+\frac{1}{2}}\right), \\ v_{x/y/z}^{n-\frac{1}{2}} = v_{x/y/z}^{n+\frac{1}{2}} - \frac{\Delta t}{\rho}\partial_{x/y/z} p^n. \end{cases} \tag{4}$$

This is the inverse of 3, abbreviated as $\mathbf{w}^n = (F^n)^{-1}\mathbf{w}^{n+1}$. Owing to the factor $e^{\omega_l \Delta t}$ above, reverse propagation in attenuating media suffers from instability due to the accumulation of exponentially growing errors induced by the truncation of the floating point representation in computer. Note that setting the anelastic coefficients to be zeros ($Y_l = 0, l = 1,\cdots,L$) leads to the acoustic equation which exhibits reversibility.

WAVEFIELD RECONSTRUCTION BY REVERSE PROPAGATION (RP)

The main idea of RP is to compute the forward 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 forward field is synchronously recomputed backwards in time from the final snapshots (as initial condition), and the saved boundary conditions, acting as Dirichlet boundary condition at each time-step (Clapp, 2008; Dussaud et al., 2008; Brossier et al., 2014). The memory request of boundary storage can be dramatically decreased by decimation (Yang et al., 2016a,b). RP is quite appealing thanks to its efficiency and limited memory requirements. Unfortunately, this scheme as described here only works in non-attenuating media.

BINOMIAL CHECKPOINTING STRATEGY

The checkpointing strategy uses a small number of memory units (checkpoints) to store the system state at distinct times. The wavefield is the computed forward in time from each checkpoints. Griewank (1992); Griewank and Walther (2000) proved that for N total steps and c checkpoints, the minimum number of forward timestepping for reverse/adjoint mode is given by $N' = rN - \beta(c+1, r-1)$, where$\beta(s,t) \equiv \binom{s+t}{t}$, with r the maximum repetition number for recovering any particular state, determined by $\beta(c, r-1) < N \leq \beta(c,r)$. The recomputation ratio of the optimal checkpointing strategy is then $R = N'/N$. This is achieved by storing the c checkpoints over N temporal steps following a binomial law (Griewank and Walther, 2000). The optimal value for the number of snapshots/checkpoints is $c = \log(N)$ (Griewank and Walther, 2000; Symes, 2007). For RTM and FWI, Anderson et al. (2012) have shown that an optimal implementation can be achieved when considering explicitly the fields required in the gradient or imaging-condition. While 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.

CHECKPOINTING ASSISTED REVERSE-FORWARD SIMULATION (CARFS)

To benefit from the efficiency of RP and the stability of checkpointing, we design a checkpointing assisted reverse-forward simulation (CARFS) method. This approach relies on an indicator of the stability of RP: the total energy E in the computing volume Ω, which accounts for variations of both pressure and particle velocities

$$E = \frac{1}{2}\int_{\Omega}\left(\rho\mathbf{v}^2 + \frac{1}{\kappa}p^2\right)\mathrm{d}\mathbf{x}. \tag{5}$$

The main idea of CARFS is to use RP, even in attenuating media, while monitoring the energy of the computation. As soon as the energy of the backward simulation deviates from the one of the forward simulation, the RP is stopped, and checkpoints are used starting from the closest snapshot prior to current state to redo some forward timestepping in order to reinitialize the RP. The algorithm is described in algorithm 1 and 1. At every backward reconstruction step, the CARFS approach selects reverse or forward modeling. By preparing the wavefield in advance, reverse propagation has higher priority over forward modeling in checkpointing, but needs to satisfy the energy constraint:

$$|E_f - E_b| < \text{tolerence}\cdot|E_f| \tag{6}$$

When there is no attenuation, i.e., $Y_l = 0, l = 1,\cdots,L$, the CARFS method converges to RP for every non-checkpoint step. The worst case is that when the attenuation is extremely strong, the CARFS approach converges to standard optimal checkpointing with only one additional forward simulation to find that the accuracy tolerance is not satisfied at every time step so that the wavefield has to be computed through repeated forward modeling. In practice, we have experimented that the

Algorithm 1: CARFS algorithm

Distribute checkpoints $s_0,\cdots,s_{c-1}$ in time following the binomial law of optimal checkpointing;
for $n = 0,\cdots,N-1$ **do**
  compute the wavefield state forward in time : $F^n : \mathbf{w}^n \to \mathbf{w}^{n+1}$ and record the total energy E_f^{n+1} for $\mathbf{w}^{n+1}$;
  store the boundary of $\mathbf{w}^{n+1}$ at decimated temporal locations according to Nyquist sampling;
  if $n \in s_0,\cdots,s_{c-1}$ **then** store the snapshot $\mathbf{w}^{n+1}$;
end
for $n = N-1,\cdots,0$ **do**
  find the closest checkpoint s_i prior to current time level n ($s_i \leq n < s_{i+1}$);
  if $n = s_i$ **then**
    read the snapshot $\mathbf{w}^{n+1}$ and set the backward energy $E_b^{n+1} = E_f^{n+1}$;
  else
    interpolate the boundary of $\mathbf{w}^{n+1}$ and compute the energy measure E_b^{n+1};
  end
  if $|E_f^{n+1} - E_b^{n+1}| > tolerance \cdot E_f^{n+1}$ **then**
    read the snapshot $\mathbf{w}^{s_i+1}$ in the checkpoint s_i;
    redistribute the idle checkpoints $s_{i+1},\cdots,s_{c-1}$ between s_i and n according to the binomial law;
    for $k = s_i+1, n$ **do**
      do forward modeling $F^k : \mathbf{w}^k \to \mathbf{w}^{k+1}$;
      if $k \in s_{i+1},\cdots,s_{c-1}$ **then** store the snapshot k at these relocated checkpoint positions
    end
    override the backward energy E_b^{n+1} with E_f^{n+1}: $E_b^{n+1} = E_f^{n+1}$;
  end
  perform RP $(F^n)^{-1} : \mathbf{w}^{n+1} \to \mathbf{w}^n$
end

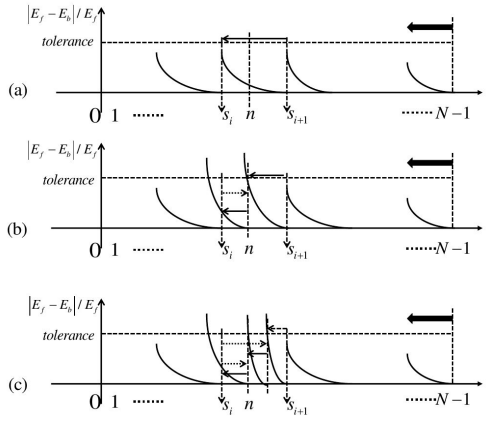

Figure 1: Schematic plot of CARFS method: (a) Best case: the wavefield at every time step is recovered by RP without violation of the energy tolerance, no need for repeated forward modeling from checkpointing. (b) CARFS switches to binomial checkpointing when RP breaks the tolerance of energy measure. (c) Several levels of checkpointing may be required if the attenuation is strong enough.

CARFS approach outperforms checkpointing strategy for most of the realistic models.

NUMERICAL EXAMPLE

We demonstrate the efficiency of CARFS method on the 2D Valhall synthetic model (velocity, density and quality factor are presented in Figure 2). The RP method, the optimal checkpointing method and the proposed CARFS method are tested in acoustic and viscoacoustic cases. The first test is oriented to FWI within the frequency range $2-20Hz$ using resampled Valhall model with a factor 4. We perform $N = 2500$ steps of forward simulation using a Ricker wavelet centered on $f_m = 5Hz$ and record $c = \log_2(N) \approx 11$ snapshots at checkpoint locations only for checkpointing and CARFS methods. Three GMB mechanisms are used to mimic constant Q property over $2-20Hz$ (Blanch et al., 1995). The recomputation ratio shown in Table 1 provides a quantitative estimation on the efficiency of different methods. The RP method is efficient to obtain a recomputation ratio of 2 but only works in the non-attenuating case. The checkpointing strategy treats acoustic and viscoacoustic propagation equally, with a recomputation ratio of 4.272. The CARFS with 1% tolerance is more efficient and works well both in the acoustic and viscoacoustic cases: the CARFS method becomes RP method with several snapshots stored in the acoustic case. 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 ($R = 2.668$).

We recorded the energy measure and extracted a trace randomly chosen in the computing domain during forward and backward process of CARFS method. Figure 3 shows that the energy recorded during backward computation matches the energy computed in forward simulation. The trace extracted during backward reconstruction is in very good agreement with the one recorded in forward simulation: the amplitude of the error trace is 5 orders of magnitude smaller than the original trace. It indicates that CARFS method is both efficient and accurate enough.

It is also interesting test the behavior of CARFS method over much wider frequency range 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 a number of representative frequencies. The temporal and spatial sampling are chosen to be $\Delta t = 0.5ms$, $\Delta x = \Delta z = 3.125m$ such that the numerical simulation using 4-th order finite difference stencil satisfies the CFL condition without numerical dispersion (about 5 grid points per wavelength) up to 80 Hz. We fix the minimum frequency at 2Hz, and change the maximum frequency $f_{\max}$. The Ricker wavelet is employed with the peak frequency $f_m = f_{\max}/2.5$ for $N = 2500$ steps of forward simulation. The resulting number of timesteppings N' and the recomputation ratio for wavefield reconstruction in Table 2 show that the efficiency of CARFS is adaptive to frequency content of interest while always superior to checkpointing strategy: the required number of timestepping is increasing with higher frequency content while still much less than checkpointing ($N' = 10680, R = 4.272$).

CONCLUSION

We have proposed a CARFS method to build the incident wavefield in reverse time order in the context of GMB-based viscoacoustic media. The idea of CARFS approach is to benefit from the efficient reverse propagation as long as the exponentially growing numerical errors within a user-specified tolerance, and automatically switch to checkpointing once the amplified errors explodes. In each backward reconstruction step, the CARFS algorithm makes a smart decision between forward modeling and reverse propagation based on the minimum timestepping cost and an energy measure. The numerical experiment demonstrates that the CARFS method can obtain accurate wavefield reconstruction using less timesteppings than optimal checkpointing, even if seismic attenuation is very strong. The CARFS method can be reduced to standard optimal checkpointing strategy when the tolerance is set to be zero. The proposed CARFS method can be straightforwardly applied to general anisotropic viscous media, with promising efficiency in practical RTM and FWI implementation.

ACKNOWLEDGMENTS

We appreciate W. Symes and J. Anderson for the communication on Griewank's binomial checkpointing strategy. We thank BP Norge AS and their Valhall partner Hess Norge AS, for allowing access to the 2D Valhall synthetic model. 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.

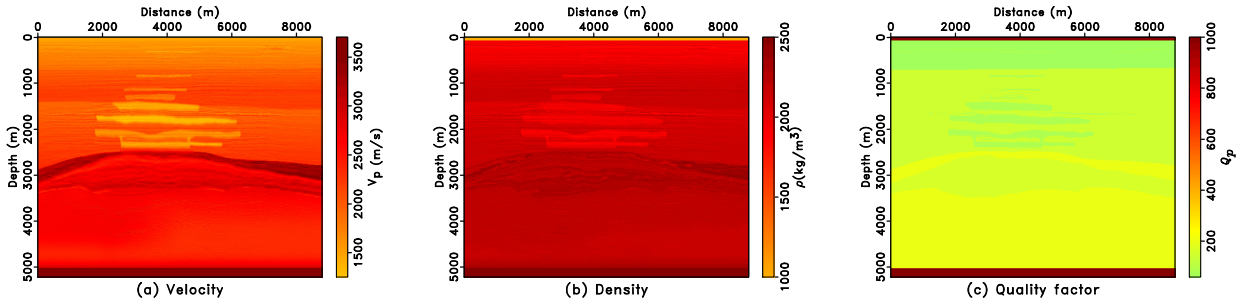

Figure 2: The 2D Valhall synthetic model: (a) the velocity, (b) the density and (c) the quality factor Q.

Table 1: Recomputation ratio R for $N = 2500$ steps simulation in the 2D Valhall model

Methods	RP	Checkpointing	CARFS
Acoustic	2. ($N' = 5000$)	4.272 ($N' = 10680$)	1.9956 ($N' = 4989$)
Viscoacoustic	-	4.272 ($N' = 10680$)	2.668 ($N' = 6670$)

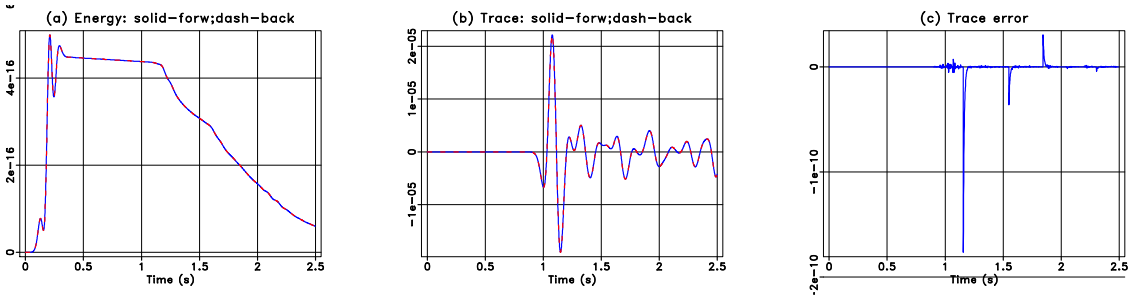

Figure 3: Comparison of the energy measure and randomly extracted seismic trace during the forward simulation and backward computation. (a) The energy measure recorded during backward computation matches quite well the energy measure monitored in forward simulation. (b) The CARFS reconstruction works so well that the original trace 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 5 orders of magnitude smaller than the magnitude of the orignal trace in (b).

Table 2: Number of timesteppings 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

EDITED REFERENCES

Note: This reference list is a copyedited version of the reference list submitted by the author. Reference lists for the 2016 SEG Technical Program Expanded Abstracts have been copyedited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

Anderson, J. E., L. Tan, and D. Wang, 2012, Time-reversal checkpointing methods for RTM and FWI: Geophysics, **77**, no. 4, S93–S103, http://dx.doi.org/10.1190/geo2011-0114.1.

Blanch, J. O., J. O. A. Robertsson, 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, http://dx.doi.org/10.1190/1.1443744.

Brossier, R., B. Pajot, L. Combe, and S. Operto, L. Metivier, 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, http://dx.doi.org/10.3997/2214-4609.20141139.

Carcione, J. M., D. Kosloff, and R. Kosloff, 1988, Wave propagation simulation in a linear viscoacoustic medium: Geophysical Journal International, **93**, 393–401, http://dx.doi.org/10.1111/j.1365-246X.1988.tb02010.x.

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, http://dx.doi.org/10.1190/1.3059336.

Griewank, A., 1992, Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation: Optimization Methods & Software, **1**, 35–54, http://dx.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, http://dx.doi.org/10.1145/347837.347846.

Moczo, P., and J. Kristek, 2005, On the rheological models used for time-domain methods of seismic wave propagation: Geophysical Research Letters, **32**, http://dx.doi.org/10.1029/2004GL021598.

Moczo, P., J. Kristek, M. Galis, P. Pazak, and M. Balazovjech, 2007a, The finite-difference and finite-element modeling of seismic wave propagation and earthquake motion: Acta Physica Slovaca, **52**, 177–406, http://dx.doi.org/10.2478/v10155-010-0084-x.

Moczo, P., J. Robertsson, and L. Eisner, 2007b, The finite-difference time-domain method for modeling of seismic wave propagation, *in* R.-S. Wu and V. Maupin, eds., Advances in wave propagation in heterogeneous earth: Elsevier/Academic Press 48, Advances in Geophysics, 421–516, http://dx.doi.org/10.1016/S0065-2687(06)48008-0.

Symes, W. W., 2007, Reverse time migration with optimal checkpointing: Geophysics, **72**, no. 5, SM213–SM221, http://dx.doi.org/10.1190/1.2742686.

Yang, P., R. Brossier, and J. Virieux, 2016a, Downsampling plus interpolation for wavefield reconstruction by reverse propagation: 78th Annual International Conference and Exhibition, EAGE, Extended Abstracts, http://dx.doi.org/10.3997/2214-4609.201601517.

Yang, P., R. Brossier, and J. Virieux, 2016b, Wavefield reconstruction from significantly decimated boundaries: Geophysics, manuscript submitted for publication.