submitted to *Geophys. J. Int.*

Frozen Gaussian approximation for three-dimensional seismic wave propagation

Lihui Chai

Department of Mathematics, University of California, Santa Barbara.

Ping Tong

Department of Geophysics, Stanford University.

Xu Yang

Department of Mathematics, University of California, Santa Barbara.

SUMMARY

We present a systematic introduction on applying frozen Gaussian approximation (FGA) to compute synthetic seismograms in three-dimensional earth models. In this method, seismic wavefield is decomposed into frozen (fixed-width) Gaussian functions, which propagate along ray paths. Rather than the coherent state solution to the wave equation, this method is rigorously derived by asymptotic expansion on phase plane, with analysis of its accuracy determined by the ratio of short wavelength over large domain size. Similar to other ray-based beam methods (e.g. Gaussian beam methods), one can use relatively small number of Gaussians to get accurate approximations of high-frequency wavefield. The algorithm is embarrassingly parallel, which can drastically speed up the computation with a multicore-processor computer station. We illustrate the accuracy and efficiency of the method by comparing it to the spectral element method for a three-dimensional (3D) seismic wave propagation in homogeneous media, where one has the analytical solution as a benchmark. As another proof of methodology, simulations of high-frequency seis-

mic wave propagation in heterogeneous media are performed for 3D waveguide model and smoothed Marmousi model respectively. The second contribution of this paper is that, we incorporate the Snell's law into the FGA formulation, and asymptotically derive reflection, transmission and free surface conditions for FGA to compute high-frequency seismic wave propagation in high contrast media. We numerically test these conditions by computing traveltime kernels of different phases in the 3D crust-over-mantle model.

Key words: frozen Gaussian approximation, seismic wave propagation, high-frequency wavefield, reflection and transmission, free surface condition.

1 INTRODUCTION

Simulation of seismic wave propagation is a core component of many scientific applications, including investigation of seismic wave propagation in the Earth's interior, exploration of subsurface structures from global to engineering scales, and mitigation of seismic risk with accurate quantitative estimates of seismic hazard (Virieux & Operto 2009). Development of efficient and accurate methods is constantly necessary to improve the modeling of seismic wave propagation in different situations, especially in complex media. So far, two mainstreams of numerical methods have been developed and intensively studied to fulfill this purpose, one of which is direct computational methods including finite difference methods (e.g. Alford et al. 1974; Virieux 1984), pseudospectral methods (e.g. Kolsloff & Baysal 1982; Carcione 1994), boundary integral methods (e.g. Bouchon & Sanchez-Sesma 2007), finite element methods (e.g. Bao et al. 1998) and spectral element methods (e.g. Komatitsch & Tromp 1999; Komatitsch et al. 2005; Tromp et al. 2008). These methods have been widely used with inherent advantages and limitations. For example, finite difference methods are probably the most wildly used in seismic modeling (e.g. Olsen & Archuleta 1996; Graves 1996; Yang et al. 2002), but with limitations from numerical dispersion and finite numerical size of mesh grids as commented in Virieux (1986). Pseudospectral methods approximate the spacial operator up to the Nyquist frequency with high accuracy, but subject to difficulties of implementing free-surface boundary conditions (e.g. Komatitsch et al. 1996). Finite element methods can easily deal with natural boundaries or discontinuous inner interfaces in background media, and thus are suitable for earth models with complex geometric structures. However, they usually have high memory and computational costs, which make them difficult to be used for high dimensional cases or large-scale modeling.

Another category of numerical methods in seismic modeling is based on ray theory (e.g. Cerveny 2001; Popov 2002; Engquist & Runborg 2003). The idea is to decompose the wavefield into elemen-

tary waveforms (e.g. rays or localized Gaussians around rays), propagate the rays and then construct Green's functions or wavefield according to the dynamic information on rays (e.g. path trajectory, amplitude and phase). Compared to direct computational methods as mentioned above, ray-based methods are less restricted by memory load and computational costs for high-frequency (short-wavelength) wave propagation. Famous seismic applications include traveltime seismic tomography (e.g. Aki et al. 1977; Tong et al. 2011), Kirchhoff migration (e.g. Gray 1986; Keho & Beydoun 1988; ten Kroode et al. 1994) and Gaussian beam migration (e.g. Hill 1990, 2001; Nowack et al. 2003; Gray 2005; Gray & Bleistein 2009; Popov et al. 2010). Classic Kirchhoff migration based on dynamic ray tracing is an effective method, however, produces unbounded amplitudes at caustics. Gaussian beam migration (GBM) used localized Gaussian functions around ray paths, and thus can handle multipathing and provide accurate wavefield at caustics. However, when beams spread significantly in the process of wave propagation, one needs to tune the width parameter of Gaussian beams to obtain a good resolution of wavefield (Cerveny et al. 1982; Hill 1990; Fomel & Tanushev 2009). Since the background media is usually heterogeneous in practical applications, the parameter tuning becomes difficult due to the nonlinearity of the Riccati equation involved in the beam construction. The accuracy of GBM relies on a Taylor expansion error determined by the width of beams, and thus the error of the approximation increases when beams spread. This was shown in Qian & Ying (2010) and Lu & Yang (2011) where, even in a simple velocity model, the error of GBM may grow rapidly in time.

Frozen Gaussian approximation (FGA) discussed in this paper has been used in quantum chemistry for propagation of time-dependent Schrödinger equation since the works of Heller (1981); Herman & Kluk (1984), with systematic justifications by Kay (1994, 2006); Swart & Rousse (2009). Later on, the formula was generalized to study general linear hyperbolic systems by Lu & Yang (2011, 2012a,b), with preliminary applications in two-dimensional (2D) seismic wave propagation (Yang et al. 2013) and 2D multi-level particle swarm optimization for seismic inversion (Li et al. 2015). Compared to 2D simulations, 3D seismic high-frequency wave propagation is rather computationally challenging due to the drastic increase in both simulation time and storage memory, which is the main motivation of this paper. The main idea of FGA is to decompose seismic wavefield into frozen (fixed-width) Gaussian functions and propagate them along ray paths. This overcomes the difficulty of tuning the width parameter in GBM while maintains the accuracy at caustics. Despite its superficial similarity to coherent state method (CSM) previously applied in seismic imaging (Foster & Huang 1991; Albertin et al. 2001; Foster et al. 2002), FGA is different at a fundamental level: CSM directly applied coherent state transform to wave operator which yielded complex-valued eikonal equation with complex source and receiver positions; FGA uses fixed-width Gaussian functions (also called coherent states in quantum chemistry) as building blocks and approximate wavefield by asymptotic expansion

on phase plane. This actually does not require to solve the eikonal equation, and produces essentially the same set of ray equations used in GBM except the dynamics of amplitude factor. We will illustrate these differences with more details during the derivation of FGA given in the following sections.

In this paper, we aim to introduce the frozen Gaussian approximation (FGA) systematically for the modeling of 3D seismic wave propagation. As a proof of accuracy and efficiency, we compare FGA to the spectral element method in homogeneous media (Komatitsch & Tromp 1999; Komatitsch et al. 2005; Tromp et al. 2008), where one can have an analytical solutions as a benchmark. The implemented algorithm is embarrassingly parallel, and thus drastically speed up the computation using a multicore-processor computer station. We also demonstrate the performance of the method by simulating high-frequency seismic wave propagation in 3D heterogeneous waveguide model and smoothed Marmousi model respectively. In addition, free surface, reflection and transmission conditions of FGA will be derived asymptotically for simulating high-frequency seismic wave propagation in high contrast media, and numerically tested by computing traveltime kernels of different phases in the 3D crust-over-mantle model.

2 MATHEMATICAL THEORY

We introduce the frozen Gaussian approximation (FGA) formula for the following 3D scalar wave equation,

$$\partial_t^2 u - c^2(\boldsymbol{x})\Delta_{\boldsymbol{x}} u = 0, \quad \boldsymbol{x} = (x, y, z) \in \mathbb{R}^3, \tag{1}$$

with initial wavefield given by

$$\begin{cases} u(0, \boldsymbol{x}) = f_0^k(\boldsymbol{x}), \\ \partial_t u(0, \boldsymbol{x}) = f_1^k(\boldsymbol{x}), \end{cases} \tag{2}$$

where $c(\boldsymbol{x})$ is sound velocity, $\Delta_{\boldsymbol{x}}$ is the Laplace operator in $\boldsymbol{x}$ and the wave number $k \gg 1$ indicates that we are considering seismic wave propagation of high-frequency (corresponding to short wavelengths).

Solution ansatz. FGA approximates the wavefield u in (1) by the following integral,

$$\begin{aligned} u_{\mathrm{F}}^k(t, \boldsymbol{x}) = &\iint \frac{a_+ \psi_+^k}{(2\pi/k)^{9/2}} e^{\mathrm{i}k\boldsymbol{P}_+ \cdot (\boldsymbol{x} - \boldsymbol{Q}_+) - \frac{k}{2}|\boldsymbol{x} - \boldsymbol{Q}_+|^2} \, \mathrm{d}\boldsymbol{q} \, \mathrm{d}\boldsymbol{p} \\ &+ \iint \frac{a_- \psi_-^k}{(2\pi/k)^{9/2}} e^{\mathrm{i}k\boldsymbol{P}_- \cdot (\boldsymbol{x} - \boldsymbol{Q}_-) - \frac{k}{2}|\boldsymbol{x} - \boldsymbol{Q}_-|^2} \, \mathrm{d}\boldsymbol{q} \, \mathrm{d}\boldsymbol{p}, \end{aligned} \tag{3}$$

where

$$\psi_{\pm}^{k}(\boldsymbol{q},\boldsymbol{p}) = \int u_{\pm,0}^{k}(\boldsymbol{y},\boldsymbol{q},\boldsymbol{p}) e^{-\mathrm{i}k\boldsymbol{p}\cdot(\boldsymbol{y}-\boldsymbol{q})-\frac{k}{2}|\boldsymbol{y}-\boldsymbol{q}|^2}\,\mathrm{d}\boldsymbol{y}, \tag{4}$$

$$u_{\pm,0}^{k}(\boldsymbol{y},\boldsymbol{q},\boldsymbol{p}) = \frac{1}{2}\Big(f_0^k(\boldsymbol{y}) \pm \frac{\mathrm{i}}{kc(\boldsymbol{q})|\boldsymbol{p}|} f_1^k(\boldsymbol{y})\Big). \tag{5}$$

In (3), $\mathrm{i} = \sqrt{-1}$ is the imaginary unit, and we use superscripts "$+$" and "$-$" to indicate two wave branches respectively, and superscripts k to indicate quantities depending on wave number k. Remark that $e^{\mathrm{i}k\boldsymbol{p}\cdot(\boldsymbol{y}-\boldsymbol{q})-\frac{k}{2}|\boldsymbol{y}-\boldsymbol{q}|^2}$ can be understood as complex localized wave-packet centered at $\boldsymbol{q}$ with propagation vector $\boldsymbol{p}$, and $\psi_{\pm}^{k}$ is the projection of initial wavefield onto each wave-packet computed by (4) with $\boldsymbol{y}$ serving as the dummy variable in the integration. Then $u_{\mathrm{F}}^{k}(t,\boldsymbol{x})$ in (3) can be treated as an approximation to seismic wavefield by a summation of dynamic localized wave-packets with fixed width. Associated with each frozen Gaussian wave-packet, the *time-dependent* quantities are the position center $\boldsymbol{Q}_{\pm}$, propagation vector $\boldsymbol{P}_{\pm}$, and amplitude $a_{\pm}$.

Formulation. The evolution of $\boldsymbol{Q}_{\pm}(t,\boldsymbol{q},\boldsymbol{p})$ and $\boldsymbol{P}_{\pm}(t,\boldsymbol{q},\boldsymbol{p})$ satisfies the ray tracing equations

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{Q}_{\pm}}{\mathrm{d}t} = \pm c(\boldsymbol{Q}_{\pm})\dfrac{\boldsymbol{P}_{\pm}}{|\boldsymbol{P}_{\pm}|}, \\ \dfrac{\mathrm{d}\boldsymbol{P}_{\pm}}{\mathrm{d}t} = \mp \partial_{\boldsymbol{Q}} c(\boldsymbol{Q}_{\pm})|\boldsymbol{P}_{\pm}|, \end{cases} \tag{6}$$

with initial conditions

$$\boldsymbol{Q}_{\pm}(0,\boldsymbol{q},\boldsymbol{p}) = \boldsymbol{q} \quad \text{and} \quad \boldsymbol{P}_{\pm}(0,\boldsymbol{q},\boldsymbol{p}) = \boldsymbol{p}. \tag{7}$$

Note that (4) is in a form of coherent state transform (also called FBI transform, Martinez 2002), and (6) are the ray equations corresponding to the Hamiltonian of two wave branches,

$$H_{\pm}(\boldsymbol{Q}_{\pm},\boldsymbol{P}_{\pm}) = \pm c(\boldsymbol{Q}_{\pm})|\boldsymbol{P}_{\pm}|\,. \tag{8}$$

Remark that, if one applies the WKB expansion used in geometric optics (e.g. Engquist & Runborg 2003),

$$u(t,\boldsymbol{x}) = A(t,\boldsymbol{x})\exp\big(\mathrm{i}kS(t,\boldsymbol{x})\big),$$

then $S(t,\boldsymbol{x})$ satisfies the real-valued eikonal equation,

$$S_t^2 - c^2|\nabla_{\boldsymbol{x}} S|^2 = 0,$$

which actually has two Hamiltonians given by (8) with corresponding ray equations given by (6) on phase plane.

Evolution of the amplitude factor $a_{\pm}(t,\boldsymbol{q},\boldsymbol{p})$ is given by

$$\frac{\mathrm{d}a_{\pm}}{\mathrm{d}t} = a_{\pm}\frac{\partial_{\boldsymbol{P}_{\pm}} H_{\pm}\cdot\partial_{\boldsymbol{Q}} H_{\pm}}{H_{\pm}} + \frac{a}{2}\mathrm{Tr}\left(Z_{\pm}^{-1}\frac{\mathrm{d}Z_{\pm}}{\mathrm{d}t}\right), \tag{9}$$

with initial condition $a_{\pm}(0,\boldsymbol{q},\boldsymbol{p}) = 2^{d/2}$. Note that the choice of initial conditions for a is to guarantee

(3) is consistent with (2) initially, i.e., $u_{\mathrm{F}}^k = u,\ \partial_t u_{\mathrm{F}}^k = \partial_t u$ at $t = 0$ (e.g. Folland 1989). In (9), we have used the shorthand notations

$$\partial_{\boldsymbol{z}} = \partial_{\boldsymbol{q}} - \mathrm{i}\partial_{\boldsymbol{p}}, \qquad Z_\pm = \partial_{\boldsymbol{z}}(\boldsymbol{Q}_\pm + \mathrm{i}\boldsymbol{P}_\pm). \tag{10}$$

Here $\partial_{\boldsymbol{z}}\boldsymbol{Q}$ and $\partial_{\boldsymbol{z}}\boldsymbol{P}$ are understood as matrices with the (j,k) component of matrix $\partial_{\boldsymbol{z}}\boldsymbol{Q}$ given by $\partial_{z_j}\boldsymbol{Q}_k$. Matrices $\partial_{\boldsymbol{z}}\boldsymbol{Q}$ and $\partial_{\boldsymbol{z}}\boldsymbol{P}$ can be solved by the following dynamic ray-tracing equations:

$$\frac{\mathrm{d}(\partial_{\boldsymbol{z}}\boldsymbol{Q})}{\mathrm{d}t} = \partial_{\boldsymbol{z}}\boldsymbol{Q}\frac{\partial^2 H}{\partial \boldsymbol{Q}\partial \boldsymbol{P}} + \partial_{\boldsymbol{z}}\boldsymbol{P}\frac{\partial^2 H}{\partial \boldsymbol{P}^2}, \tag{11}$$

$$\frac{\mathrm{d}(\partial_{\boldsymbol{z}}\boldsymbol{P})}{\mathrm{d}t} = -\partial_{\boldsymbol{z}}\boldsymbol{Q}\frac{\partial^2 H}{\partial \boldsymbol{Q}^2} - \partial_{\boldsymbol{z}}\boldsymbol{P}\frac{\partial^2 H}{\partial \boldsymbol{P}\partial \boldsymbol{Q}}. \tag{12}$$

with a componentwise form as

$$\frac{\mathrm{d}(\partial_{\boldsymbol{z}}\boldsymbol{Q})_{jk}}{\mathrm{d}t} = \partial_{\boldsymbol{z}_j}\boldsymbol{Q}_l\frac{\partial^2 H}{\partial \boldsymbol{Q}_l\partial \boldsymbol{P}_k} + \partial_{\boldsymbol{z}_j}\boldsymbol{P}_l\frac{\partial^2 H}{\partial \boldsymbol{P}_l\boldsymbol{P}_k},$$
$$\frac{\mathrm{d}(\partial_{\boldsymbol{z}}\boldsymbol{P})_{jk}}{\mathrm{d}t} = -\partial_{\boldsymbol{z}_j}\boldsymbol{Q}_l\frac{\partial^2 H}{\partial \boldsymbol{Q}_l\boldsymbol{Q}_k} - \partial_{\boldsymbol{z}_j}\boldsymbol{P}_l\frac{\partial^2 H}{\partial \boldsymbol{P}_l\partial \boldsymbol{Q}_k},$$

where Einstein's index summation convention has been applied. Note that (11)-(12) are exactly the dynamic ray tracing equations used in Gaussian beam method (Cerveny et al. 1982), but serve as a rather different purpose here for computing $\dfrac{\mathrm{d}Z}{\mathrm{d}t}$ in (9).

The key component of FGA is to derive the dynamics (9) of amplitude factors $a_\pm$, which distinguishes it from coherent state method and Gaussian beam method. We leave the details of derivation to Appendix A, and refer interested readers to Lu & Yang (2012a) for a non-dimensional mathematical version of derivation.

Remark. A dimensional version of (3)-(4) can be given by in the following generalized FGA integral, with width-tuning parameters $\sigma_1 = 1\ \mathrm{km}^{-1}$ and $\sigma_2 = 1\ \mathrm{km}^{-1}$,

$$\begin{aligned} u_{\mathrm{F}}^k(t,\boldsymbol{x}) = &\iint \frac{a_+\psi_+^k}{(2\pi/k)^{9/2}} e^{\mathrm{i}k\boldsymbol{P}_+\cdot(\boldsymbol{x}-\boldsymbol{Q}_+)-\frac{k\sigma_1}{2}|\boldsymbol{x}-\boldsymbol{Q}_+|^2}\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p} \\ &+ \iint \frac{a_-\psi_-^k}{(2\pi/k)^{9/2}} e^{\mathrm{i}k\boldsymbol{P}_-\cdot(\boldsymbol{x}-\boldsymbol{Q}_-)-\frac{k\sigma_1}{2}|\boldsymbol{x}-\boldsymbol{Q}_-|^2}\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p}, \end{aligned} \tag{13}$$

where

$$\psi_\pm^k(\boldsymbol{q},\boldsymbol{p}) = \sigma_2^{3/2}\int u_{\pm,0}^k(\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})e^{-\mathrm{i}k\boldsymbol{p}\cdot(\boldsymbol{y}-\boldsymbol{q})-\frac{k\sigma_2}{2}|\boldsymbol{y}-\boldsymbol{q}|^2}\,\mathrm{d}\boldsymbol{y}, \tag{14}$$

and k is in dimension km^{-1}, $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{q}$ are in km, and $\boldsymbol{p}$ is the dimensionless propagation vector.

The FGA formulation for (13)-(14) can be derived similarly as in Appendix A, which is the same as (6)-(12) but with (10) modified to

$$\partial_{\boldsymbol{z}} = \sigma_2^{-1}\partial_{\boldsymbol{q}} - \mathrm{i}\partial_{\boldsymbol{p}}, \qquad Z_\pm = \partial_{\boldsymbol{z}}(\sigma_1\boldsymbol{Q}_\pm + \mathrm{i}\boldsymbol{P}_\pm). \tag{15}$$

Asymptotic accuracy. With proper smoothness assumptions on sound velocity $c(\boldsymbol{x})$ and high-frequency initial conditions (2), energy estimate of the wave equation (1) can show that FGA has the first order accuracy of k^{-1}, which implies FGA has better accuracy for high-frequency (large k) seismic wave propagation, and mathematically, one has

$$\|u(t,\boldsymbol{x}) - u_F^k(t,\boldsymbol{x})\|_E \leq C(kL)^{-1}, \quad \text{for } t > 0, \tag{16}$$

where $\|\cdot\|_E$ stands for the scaled energy norm, and $C > 0$ is a constant depending on final prorogation time T, gradients of velocity c and scaled initial wave energy. Including the computational domain size L in (16) is to make its right-hand-side dimensionless, and we refer interested readers to (Lu & Yang 2012a, Theorem 4.1) for the details of mathematical proof.

Remark. The choice of k in (13)-(14) does not have to be exactly the wave number, but can be any constant times wave number. Indicated by (16), choosing a larger k provides better accuracy, but increases the number of wave packets to discretize (13) since the Gaussian wave packet is narrower. In general, large k is preferred for simulating wave pulse propagation because the wavefield is rather localized; see Section 5.5 for example.

3 ALGORITHMS

The main FGA algorithm consists of three steps:

Step 1. *Initial decomposition.* Choose the sets $G_\pm$ of $(\boldsymbol{q},\boldsymbol{p})$ pair to discretize (3) and calculate $\psi_\pm^k$ defined in equation (4).

To choose the proper sets $G_\pm$, we use the fast FBI transform algorithm introduced in (Yang et al. 2013, Page 2) to decompose the initial wavefields into Gaussians. For a purpose of convenience, we summarize the algorithm as below in consistent notations with this paper.

Define a Fourier transform of a function $g(\boldsymbol{y})$ as

$$\hat{g}(\boldsymbol{p}) = \int g(\boldsymbol{y}) e^{-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{y}}\,\mathrm{d}\boldsymbol{y},$$

then (4) can be rewritten as

$$\psi_\pm^k(\boldsymbol{q},\boldsymbol{p}) = \widehat{g_{\boldsymbol{q},\pm}^k}(k\boldsymbol{p}), \tag{17}$$

with

$$g_{\boldsymbol{q},\pm}^k(\boldsymbol{y}) = u_{\pm,0}^k(\boldsymbol{y}+\boldsymbol{q},\boldsymbol{q},\boldsymbol{p})\exp(-\tfrac{k}{2}|\boldsymbol{y}|^2). \tag{18}$$

Notice that $g_{\boldsymbol{q},\pm}^k$ decays exponentially in $\boldsymbol{y}$, and thus one can neglect the function value outside a small domain of size $\mathcal{O}(k^{-1/2})$ centered at zero, e.g., a small box

$$B = [-\ell/2,\ell/2]^3 \subset \mathbb{R}^3, \quad \text{with} \quad \ell \sim \mathcal{O}(k^{-1/2}).$$

Then $\psi_{\pm}^k$ can be efficiently computed by applying Fast Fourier Transform of $g_{\boldsymbol{q},\pm}^k$ restricted on the small box B_k, and the sets $G_{\pm}$ can be selected by applying a simple thresholding where $\psi_{\pm}^k$ have relatively large values. In practice, we chose a threshold ratio so that the approximation to the FGA formula (3) at $t = 0$ by the selected Gaussians gives a good approximation to the initial conditions (2).

Step 2. *Time propagation.* Solve (6)-(12) numerically by, e.g., the fourth-order Runger-Kutta method. Note that the 3-by-3 matrices $\partial_{\boldsymbol{z}}\boldsymbol{Q}$ and $\partial_{\boldsymbol{z}}\boldsymbol{P}$ can also be solved at each time by divided difference instead of using (11)-(12), which would save some computational cost for not solving ordinary differential equations (ODEs), but introduce more numerical errors from divided difference.

Step 3. *Reconstruction.* Compute the wave field at time T by the following approximation,

$$u_{\mathrm{F}}^k(t,\boldsymbol{x}) \approx \sum_{(\boldsymbol{q},\boldsymbol{p})\in G_+} \frac{a_+\psi_+^k}{(2\pi/k)^{9/2}} e^{\mathrm{i}k\boldsymbol{P}_+\cdot(\boldsymbol{x}-\boldsymbol{Q}_+)-\frac{k}{2}|\boldsymbol{x}-\boldsymbol{Q}_+|^2}\delta\boldsymbol{q}\delta\boldsymbol{p} + \sum_{(\boldsymbol{q},\boldsymbol{p})\in G_-} \frac{a_-\psi_-^k}{(2\pi/k)^{9/2}} e^{\mathrm{i}k\boldsymbol{P}_-\cdot(\boldsymbol{x}-\boldsymbol{Q}_-)-\frac{k}{2}|\boldsymbol{x}-\boldsymbol{Q}_-|^2}\delta\boldsymbol{q}\delta\boldsymbol{p}.$$

Remark. Notice that (6)-(12) are uncoupled ODEs, and therefore one can solve them by parallel implementation in a multicore-processor computer station, which will drastically speed up the computation. Moreover, solving the ODEs does not impose any stability limitation on the ratio of time step over spacial resolution, and thus FGA can be used for large domain simulation.

4 INTERFACE, FREE SURFACE AND TRANSPARENT BOUNDARY CONDITIONS

To simulate seismic wave propagation on a bounded domain, one needs to impose reflection and transmission conditions at interfaces of background media, free surface condition on top boundary, and transparent boundary condition on side and bottom boundaries.

Reflection and transmission conditions. When waves hit the interface where the wave speed $c(\boldsymbol{x})$ is discontinuous, proper interface conditions should be incorporated to capture reflected and transmitted waves in the FGA formulation. Without loss of generality, we assume the interface is located at $z = z_0$, with

$$c(\boldsymbol{x}) = \begin{cases} c_1(\boldsymbol{x}), & \text{if } z > z_0, \\ c_2(\boldsymbol{x}), & \text{if } z < z_0, \end{cases} \tag{19}$$

where c_1 and c_2 are two smooth functions. Denote the FGA solutions for the incident, reflected, and transmitted waves by u^{in}, u^{re}, and u^{tr} respectively. For different wave indexes $w = \mathrm{in},\ \mathrm{re},\ \mathrm{tr}$, define

$$u^w(t,\boldsymbol{x}) = \iiint \frac{a^w u_0}{(2\pi/k)^{9/2}} e^{\mathrm{i}k\Phi^w(t,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})}\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p}, \tag{20}$$

with $\Phi^w(t,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})$ given by

$$\Phi^w = \boldsymbol{P}^w\cdot(\boldsymbol{x}-\boldsymbol{Q}^w) - \boldsymbol{p}\cdot(\boldsymbol{y}-\boldsymbol{q}) + \frac{\mathrm{i}}{2}|\boldsymbol{x}-\boldsymbol{Q}^w|^2 + \frac{\mathrm{i}}{2}|\boldsymbol{y}-\boldsymbol{q}|^2.$$

Suppose at time t, an incident Gaussian hits the interface at (x,y,z_0) from above, then

$$\begin{aligned}&\boldsymbol{Q}^{\mathrm{re}} = \boldsymbol{Q}^{\mathrm{tr}} = \boldsymbol{Q}^{\mathrm{in}} = (x,y,z_0),\\ &\boldsymbol{P}^{\mathrm{in}} = (p_x,p_y,p_z^{\mathrm{in}}),\quad \boldsymbol{P}^{\mathrm{re}} = (p_x,p_y,p_z^{\mathrm{re}}),\quad \boldsymbol{P}^{\mathrm{tr}} = (p_x,p_y,p_z^{\mathrm{tr}}).\end{aligned} \tag{21}$$

Denote $C_1 = \lim\limits_{z\to z_0^+} c_1(\boldsymbol{x})$ and $C_2 = \lim\limits_{z\to z_0^-} c_2(\boldsymbol{x})$, then Snell's law gives

$$C_1|\boldsymbol{P}^{\mathrm{in}}| = C_1|\boldsymbol{P}^{\mathrm{re}}| = C_2|\boldsymbol{P}^{\mathrm{tr}}|,$$

which implies

$$p_z^{\mathrm{re}} = -p_z^{\mathrm{in}},\ \text{and}\ p_z^{\mathrm{tr}} = p_z^{\mathrm{in}}\sqrt{\frac{C_1^2}{C_2^2} + \left(\frac{C_1^2}{C_2^2}-1\right)\frac{p_x^2+p_y^2}{\left(p_z^{\mathrm{in}}\right)^2}}. \tag{22}$$

The continuity of wavefield and its normal derivative produces

$$\begin{aligned}u^{\mathrm{in}}(t,x,y,z_0) + u^{\mathrm{re}}(t,x,y,z_0) &= u^{\mathrm{tr}}(t,x,y,z_0),\\ \partial_z u^{\mathrm{in}}(t,x,y,z_0) + \partial_z u^{\mathrm{re}}(t,x,y,z_0) &= \partial_z u^{\mathrm{tr}}(t,x,y,z_0),\end{aligned} \tag{23}$$

then (A.10)-(A.11) and (20)-(23) yield, after ignoring high order terms in asymptotic expansion,

$$\begin{aligned}a^{\mathrm{in}} + a^{\mathrm{re}} &= a^{\mathrm{tr}},\\ p_z^{\mathrm{in}}a^{\mathrm{in}} - p_z^{\mathrm{in}}a^{\mathrm{re}} &= p_z^{\mathrm{tr}}a^{\mathrm{tr}}.\end{aligned} \tag{24}$$

Therefore

$$a^{\mathrm{re}} = Ra^{\mathrm{in}},\ \text{and}\ a^{\mathrm{tr}} = Ta^{\mathrm{in}}, \tag{25}$$

with the reflection coefficient R and transmission coefficient T given by

$$R = \frac{p_z^{\mathrm{in}} - p_z^{\mathrm{tr}}}{p_z^{\mathrm{in}} + p_z^{\mathrm{tr}}},\ \text{and}\ T = \frac{2p_z^{\mathrm{in}}}{p_z^{\mathrm{in}} + p_z^{\mathrm{tr}}}. \tag{26}$$

Note that when $\frac{C_1^2}{C_2^2} + \left(\frac{C_1^2}{C_2^2}-1\right)\frac{p_x^2+p_y^2}{\left(p_z^{\mathrm{in}}\right)^2} < 0$, waves hitting the interface $z=z_0$ will be totally reflected, and then one needs to modify (20)-(23) as

$$u^{\mathrm{in}}(t,x,y,z_0) + u^{\mathrm{re}}(t,x,y,z_0) = 0, \tag{27}$$

which implies $R=-1$ and $T=0$ in this case.

For convenience, we define

$$\Theta_r = \left\{\boldsymbol{P}\ \middle|\ \frac{C_1^2}{C_2^2} + \left(\frac{C_1^2}{C_2^2}-1\right)\frac{p_x^2+p_y^2}{\left(p_z^{\mathrm{in}}\right)^2} < 0\right\}, \tag{28}$$

and summarize R and T for reflection and transmission conditions

$$R = \begin{cases} -1, & \text{if } \boldsymbol{P}^{\rm in} \in \Theta_r, \\ \dfrac{p_z^{\rm in} - p_z^{\rm tr}}{p_z^{\rm in} + p_z^{\rm tr}}, & \text{otherwise,} \end{cases} \quad \text{and } T = \begin{cases} 0, & \text{if } \boldsymbol{P}^{\rm in} \in \Theta_r, \\ \dfrac{2p_z^{\rm in}}{p_z^{\rm in} + p_z^{\rm tr}}, & \text{otherwise.} \end{cases} \tag{29}$$

To compute FGA at the interface, one also needs interface conditions for (11) and (12) as below,

$$\begin{aligned} \partial_{\boldsymbol{z}} \boldsymbol{Q}^{\rm re} &= \partial_{\boldsymbol{z}} \boldsymbol{Q}^{\rm in}\, F, \\ \partial_{\boldsymbol{z}} \boldsymbol{P}^{\rm re} &= \partial_{\boldsymbol{z}} \boldsymbol{P}^{\rm in}\, F + \frac{|\boldsymbol{P}^{\rm in}|^2}{C_1 p_z^{\rm in}} \left(\partial_{\boldsymbol{z}} \boldsymbol{Q}^{\rm re} - \partial_{\boldsymbol{z}} \boldsymbol{Q}^{\rm in}\right) \cdot \nabla c_1(\boldsymbol{Q}^{\rm in}) \boldsymbol{e}_3, \\ \partial_{\boldsymbol{z}} \boldsymbol{Q}^{\rm tr} &= \partial_{\boldsymbol{z}} \boldsymbol{Q}^{\rm in}\, G, \\ \partial_{\boldsymbol{z}} \boldsymbol{P}^{\rm tr} &= \partial_{\boldsymbol{z}} \boldsymbol{P}^{\rm in}\, W - \frac{|\boldsymbol{P}^{\rm tr}|}{C_2 p_z^{\rm tr}} \left(\partial_{\boldsymbol{z}} \boldsymbol{Q}^{\rm tr} \cdot \nabla c_2(\boldsymbol{Q}^{\rm tr}) - \partial_{\boldsymbol{z}} \boldsymbol{Q}^{\rm in} \cdot \nabla c_1(\boldsymbol{Q}^{\rm in})\right) \boldsymbol{e}_3, \end{aligned} \tag{30}$$

where $\boldsymbol{e}_3 = (0, 0, 1)$ is a row vector, $F = \text{diag}\{1, 1, -1\}$, G and W are two 3×3 matrices given by

$$G = \begin{bmatrix} 1 & & \\ & 1 & \\ \left(\frac{C_2^2}{C_1^2} - 1\right) \frac{p_x^{\rm in}}{p_z^{\rm in}} & \left(\frac{C_2^2}{C_1^2} - 1\right) \frac{p_y^{\rm in}}{p_z^{\rm in}} & \frac{C_2^2}{C_1^2} \frac{p_z^{\rm tr}}{p_z^{\rm in}} \end{bmatrix}, \qquad W = \begin{bmatrix} 1 & & \left(\frac{C_1^2}{C_2^2} - 1\right) \frac{p_x^{\rm in}}{p_z^{\rm tr}} \\ & 1 & \left(\frac{C_1^2}{C_2^2} - 1\right) \frac{p_y^{\rm in}}{p_z^{\rm tr}} \\ & & \frac{C_1^2}{C_2^2} \frac{p_z^{\rm in}}{p_z^{\rm tr}} \end{bmatrix}.$$

Remark. 1. One needs to use the Eulerian formulation of frozen Gaussian approximation (Lu & Yang 2012a) to derive the interface conditions (30), which is quite technically involved, and thus we leave the details to Appendix B.

2. For a reflecting geometry of general shape, one needs to apply the formulation (29)-(30) in the local tangent-normal coordinates by treating the tangential direction as the local flat horizontal interface.

Free surface conditions. The free surface boundary condition is equivalent to modify (20)-(23) as

$$\partial_z u^{\rm in}(t, x, y, z_0) + \partial_z u^{\rm re}(t, x, y, z_0) = 0, \tag{31}$$

which implies $R = 1$ and $T = 0$ as a special case of reflection and transmission conditions.

The *transparent boundary conditions* can be handled easily as follows: If at computational time step, a Gaussian exits the boundary, then we remove this Gaussian from computation.

5 NUMERICAL EXAMPLES

In this section, five numerical tests are given to confirm the computational performance of frozen Gaussian approximation (FGA) method. First, to show accuracy and efficiency, we compare it to the spectral element method (e.g. Komatitsch & Tromp 1999; Komatitsch et al. 2005; Tromp et al. 2008) in a homogeneous media, where one can have an analytical solution as a benchmark. We also present the performance of FGA in a 3D heterogeneous waveguide model and a smoothed Marmousi model

respectively, and numerically test the free surface, reflection and transmission conditions of FGA by computing traveltime kernels of different phases in the 3D crust-over-mantle model. All numerical examples presented in this section are implemented on a computer equipped by Intel(R) Xeon(R) E5-2670 (2.60GHz) with a 62GB RAM.

5.1 Homogeneous media

We consider the wave equation (1) with $c(\boldsymbol{x}) \equiv c_0$ as a constant, in which analytical solution is available as a benchmark. Denote $\phi(t)$ to be a function that rapidly decays, with ϕ' as its derivative. For $t_0 > 0$ and $\boldsymbol{x}_0 \in \mathbb{R}^3$, if one takes f_0^k and f_1^k in (2) as below, with $\phi(t)$ compactly supported on $[-t_0, t_0]$,

$$\begin{aligned} f_0^k(\boldsymbol{x}) &= \frac{1}{4\pi|\boldsymbol{x}-\boldsymbol{x}_0|}\phi\left(t_0 - \frac{|\boldsymbol{x}-\boldsymbol{x}_0|}{c_0}\right), \\ f_1^k(\boldsymbol{x}) &= \frac{1}{4\pi|\boldsymbol{x}-\boldsymbol{x}_0|}\phi'\left(t_0 - \frac{|\boldsymbol{x}-\boldsymbol{x}_0|}{c_0}\right), \end{aligned} \tag{32}$$

then the solution to (1)-(2) can be given by

$$u^k(t,\boldsymbol{x}) = \frac{1}{4\pi|\boldsymbol{x}-\boldsymbol{x}_0|}\phi\left(t_0 + t - \frac{|\boldsymbol{x}-\boldsymbol{x}_0|}{c_0}\right). \tag{33}$$

Also if ϕ decays fast enough (nearly compactly supported), one can still use (33) as a reference solution to (1)-(2).

Accuracy. We shall take ϕ as an exponential function as below and $c_0 = 3.2$ km/s in (32), and thus can use (33) as a benchmark solution,

$$\phi(t) = \exp\left(-\frac{t^2}{2\sigma^2}\right)\cos\big(2\pi f t\big), \tag{34}$$

with frequency $f = 2.0372$ Hz, pulse width $\sigma = 0.625$ s, and pulse center $\boldsymbol{x}_0 = (64, 64, 64)$ km.

We solve (1)-(2) using FGA from time $t = 0$ s to $t = 10$ s to get the approximate solution $u_{\mathrm{F}}^k(t,\boldsymbol{x})$, and compute the numerical error between the FGA solution and the reference solution (33) by

$$\mathrm{Error}(t) = \frac{1}{D(t)}\left(\iiint |u^k(t,\boldsymbol{x}) - u_{\mathrm{F}}^k(t,\boldsymbol{x})|^2 \,\mathrm{d}\boldsymbol{x}\right)^{\frac{1}{2}},$$

where $D(t)$ is a normalized factor as

$$D(t) = \left(\iiint |u^k(t,\boldsymbol{x})|^2 \,\mathrm{d}\boldsymbol{x}\right)^{\frac{1}{2}}.$$

The total number of Gausians used in the FGA simulation is 1161188, the time step used in solving the ODEs is chosen as $\Delta t = 0.01$ s, and we take $\sigma_1 = \sigma_2 = 1$ km^{-1} and $k = 2^8$ km^{-1} in (13)-(14).

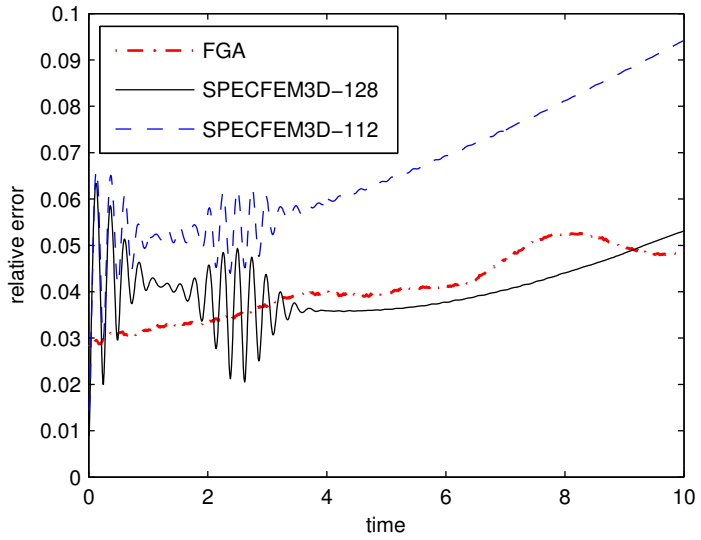

Figure 1. Accuracy of FGA and SPECFEM3D. SPECFEM3D-n refers to the simulation by SPECFEM3D with n elements in each spatial direction.

An average computational time on 8 processors is 172.03 s, and the error is shown as a function of time t in Figure 1.

In comparison, we also simulate the same problem by using the spectral element software package SPECFEM3D (https://geodynamics.org/cig/software/specfem3d/). We use 5^3 nodes in each element for the SPECFEM3D and solve (1)-(2) in a domain centered at $\boldsymbol{x}_0$ of size 128 km $\times$ 128 km $\times$ 128 km. The time step is chosen as the same as that of FGA, i.e. $\Delta t = 0.01$ s. As shown in Figure 1, to achieve a comparable accuracy as FGA, SPECFEM3D needs 128 elements in each spatial direction, which requires an average computational time on 8 processors of 698.38 s – around three times longer than that of FGA. Decreasing the number of elements in each direction from 128 to 112 reduces computational time to 489.90 s (still longer than FGA), but increases the error larger than FGA. Further coarsening the mesh may reduce computational time but deteriorate accuracy because of lack of resolution of the initial conditions.

Parallelizability. FGA approximates the seismic wavefield by a set of Gaussians, which propagate according to the ODEs (6)-(12). Notice that there is no coupling between the ODEs for any two different Gaussians, and thus FGA is embarrassingly parallel, i.e., except the initial decomposition and final summation of Gaussians, no communication is needed between the parallel ODE solvers.

In this subsection, we investigate the same problem as in homogeneous media with different number of processors and the computational time is shown in Table 1. One can see that, as the number of processors increases, there is a significant reduction in computational time for FGA. Denoting T_N as

the computational time that a method runs with N processors, we use the following to compute the speed-up ratio,

$$S_N = \frac{T_N}{T_{N/2}}.$$

The ideal speed-up ratio is 2, and in practice S_N decays as N increases due to the time spent on communications between different cores. One can see that, in Table 1 $S_8 \approx 1.9387$ and $S_{16} \approx 1.8983$ for FGA, which are slightly smaller than 2 indicating an almost perfectly parallel efficiency. On the other hand, as a comparison, SPECFEM3D is used with 128 elements in each spatial direction to achieve a comparable accuracy to FGA. The speed-up ratio for the SPECFEM3D solver is around 1.5668 and 1.4224 for $N = 8$ and $N = 16$ respectively, which are smaller than those of FGA. This is because SPECFEM3D solves (1) on a parallel computer with N processors by partitioning the whole domain into N slabs with each processor solving the equation in each slab. Therefore, for each time step, each processor needs to communicate with its neighbors to get necessary boundary information, which decreases the speed-up ratio. Thus FGA has the advantage of parallel computing when the problem is of large domain size and for high-frequency wave propagation.

5.2 High-frequency wave propagation

For high-frequency seismic waves, i.e. when $kL \gg 1$ with L as the computational domain size, direct numerical methods for (1) require the mesh size resolve the short wavelength of initial wavefield (2). This leads to unaffordable computational cost in 3D simulations, i.e. the total number of grid points is of the order $(kL)^3$. Moreover, the storage of data is another big computational load for high-frequency seismic wave propagation. While for FGA, since it asymptotically approximates the solution to (1)-(2), mesh is in general not required for resolution of wavelength. The computational cost depends on the number of Gaussians in the initial wavefield decomposition. In general, a good accuracy can be achieved with a much smaller number of Gaussians compared to $(kL)^3$.

In this subsection, we test FGA for initial conditions with different frequencies and present its performance when frequency f becomes large by comparing it to SPECFEM3D. For each f, FGA uses $k = 128\pi f/c_0$ km^{-1} for an initial decomposition with a relative error less than 4%, and SPECFEM3D uses $\pi L f/(2c_0)$ elements in each spatial direction in order to resolve the short wavelength of initial wavefield for a comparable accuracy to FGA. In Figure 2, we plot the one-step computational time with respect to different number f, from which one can observe that the computational time spent by FGA increases more or less linearly in f, while that of SPECFEM3D increases like f^3. Due to the restriction of mesh size required by SPECFEM3D, the storage of wavefield data exceeds the size of RAM of our current computer station when $f \geq 4.0744$ Hz, and thus we only estimate the computational time

Table 1. Runing time of FGA and SPECFEM3D with respect to number of processors.

FGA simulation

Number of processors	4	8	16
cpu time	337.55	168.30	90.338
cpu time	334.26	173.45	96.030
cpu time	328.36	172.70	88.958
cpu time	335.00	175.58	90.430
cpu time	328.37	167.75	90.362
cpu time	335.47	167.57	89.382
cpu time	335.65	178.86	88.862
average time	333.52	172.03	90.623
Speed up	–	1.9387	1.8983

SPECFEM3D simulation

Number of processors	4	8	16
cpu time	1344.3	707.75	639.60
cpu time	1020.0	877.94	422.78
cpu time	1011.5	645.68	420.35
cpu time	1318.0	703.97	537.70
cpu time	1044.3	557.44	412.60
cpu time	1017.7	695.77	437.44
cpu time	903.72	700.14	566.50
average time	1094.2	698.38	491.00
Speed up	–	1.5668	1.4224

of SPECFEM3D for $f \geq 4.0744$ Hz by extrapolation. By this test, one may conclude that FGA has the advantage of saving both computational time and memory for simulating high-frequency seismic wave propagation compared to SPECFEM3D.

5.3 A waveguide model

To study the performance of FGA for seismic wave propagation in 3D heterogeneous media, we first test it in a 3D waveguide model, where one can use a solution to 1D wave equation as an approximation.

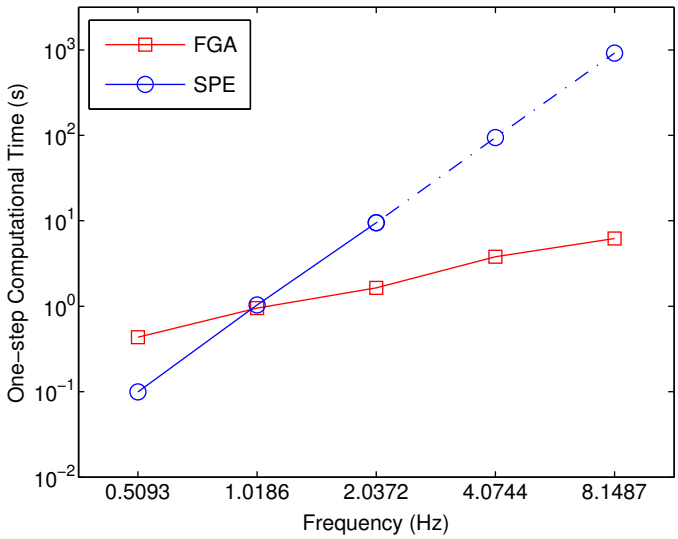

Figure 2. Dependence of one-step computational time on frequency. The horizontal axis is the frequency f (Hz) and the vertical axis is the one-step computational time of the solvers. The square line stands for the FGA simulations and the circle line stands for the SPECFEM3D simulations. Due to the limitation of memory, SPEFEM3D can not run for $f \geq 4.0744$ Hz, and the dashed circle line is obtained by extrapolation.

We consider (1) and choose the velocity c isotropic in y and z directions, i.e. ,

$$c(\boldsymbol{x}) = c(x, y, z) = c(x) = c_0 \left(\frac{1}{2} \sin(\pi x / L) + 1 \right), \text{ where } c_0 = 3.2 \text{ km/s and } L = 64 \text{ km}, \quad (35)$$

with initial conditions

$$\begin{cases} f_0^k(\boldsymbol{x}) = f^k(x, y, z) = u_0^k(x)\varphi(y/L)\varphi(z/L), \\ f_1^k(\boldsymbol{x}) = f_1^k(x, y, z) = u_1^k(x)\varphi(y/L)\varphi(z/L), \end{cases} \quad (36)$$

and

$$\begin{aligned} u_0^k(x) &= \exp\Big(-25(x-L)^2/L^2 + \mathrm{i}k(x-L)/64\Big), \\ u_1^k(x) &= \mathrm{i}k \exp\Big(-25(x-L)^2/L^2 + \mathrm{i}k(x-L)/64\Big). \end{aligned} \quad (37)$$

In (36), φ is taken as

$$\varphi(s) = \frac{(\tanh(\gamma(s-a)) + 1)(\tanh(\gamma(b-s)) + 1)}{4},$$

with $a = 0.5$, $b = 1.5$ and a relatively large $\gamma = 50$. Therefore, φ is a smoothed indicator function which is approximately 1 on $(0.5,\ 1.5)$ and 0 elsewhere.

With this choice of initial data (see Figure 3) and velocity field, the wavefield of (1) at $(y, z) =$

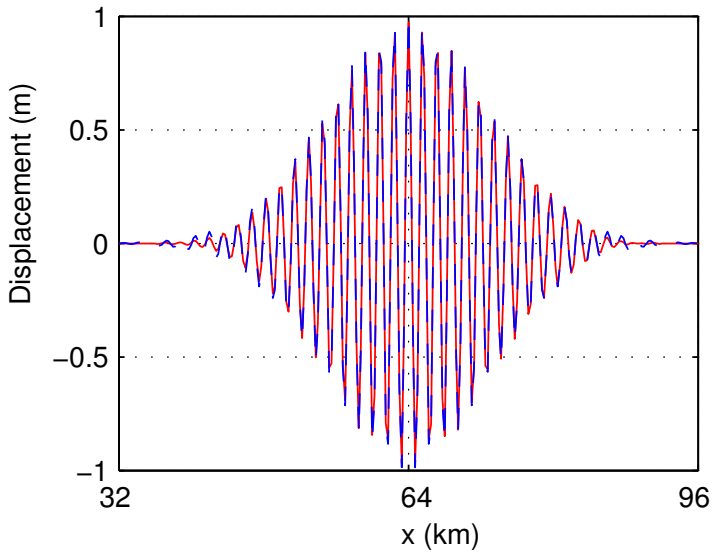

Figure 3. Waveguide initial wavefiled: a 1D cross-section at $y = 64$ km, $z = 64$ km. The blue dashed line is the exact initial wavefield and the red solid line is given by FGA.

$(64, 64)$ km can be treated as a solution to the following 1D wave equation for short-time propagation,

$$\begin{aligned} \partial_t^2 u - c^2(x)\partial_{xx} u =& 0, \\ u(0,x) =& u_0^k(x), \\ \partial_t u(0,x) =& u_1^k(x), \end{aligned} \tag{38}$$

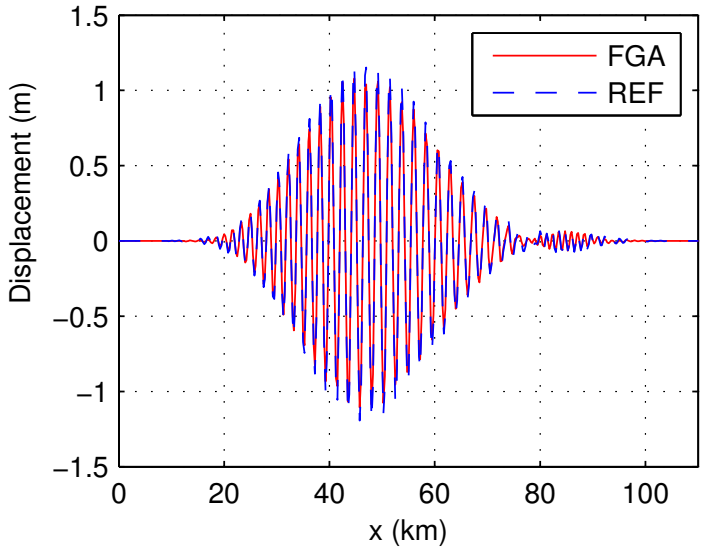

Figure 4. Waveguide wavefiled at $t = 5$ s: a 1D cross-section at $y = 64$ km, $z = 64$ km. The blue dashed line comes from a numerical solution of (38) by a finite difference method with $\Delta x = 2^{-6}$ km and $\Delta t = 2^{-11}$ s, and the red solid line is given by FGA.

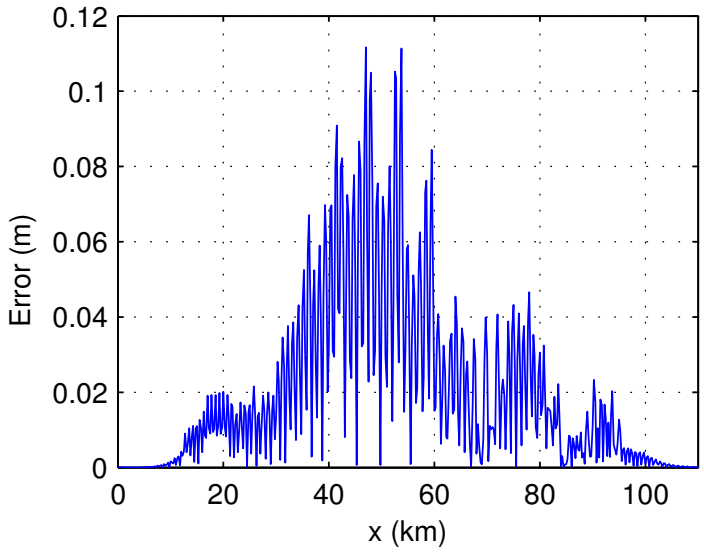

Figure 5. Accuracy of FGA for waveguide model: a 1D cross-section at $y = 64$ km, $z = 64$ km

which can be easily solved by a finite difference method with fine mesh. We denote the solution to (38) as $u^k_{\rm 1d}$.

We solve (1) with the choice of velocity field and initial conditions in (35) - (37) using FGA for $k = 2^8$ km^{-1}. The total number of Gaussians used in the simulation is 310464, which gives an relative L^2 error at 0.0426.

We denote $\tilde{u}^k_{\rm F}(x) = u^k_{\rm F}(x, 64, 64)$ as the FGA solution restricted at $(y, z) = (64, 64)$ km and compare it to $u^k_{\rm 1d}$ at time $t = 5$ s. In Figure 4, we plot the FGA solution $\tilde{u}^k_{\rm F}$ and $u^k_{\rm 1d}$, which shows a good agreement in these two solutions. In Figure 5, we plot the difference of the two solutions $|\tilde{u}^k_{\rm F} - u^k_{\rm 1d}|$, and one can see that has a good accuracy, with the relative L^2 error

$$\left(\int |u^k_{\rm 1d}|^2 {\rm d}x\right)^{-1/2} \left(\int |\tilde{u}^k_{\rm F} - u^k_{\rm 1d}|^2 {\rm d}x\right)^{1/2}$$

at time $t = 5$ s equal to 8% approximately. We show several 2D cross-section snapshots of the FGA solutions at $z = 64$ km in Figure 6.

5.4 Smoothed Marmousi model

We then consider a smoothed Marmousi model[⋆] to demonstrate the performance of FGA for computing seismic wave propagation in 3D heterogeneous media. Here we choose the velocity field $c(x, y, z)$

[⋆] The smooth model is obtained by convolving a 'hard' model a spatial Hanning ($\cos^2$) filter of radius 150 m. See http://www.caam.rice.edu/∼benamou/testproblem.html for details.

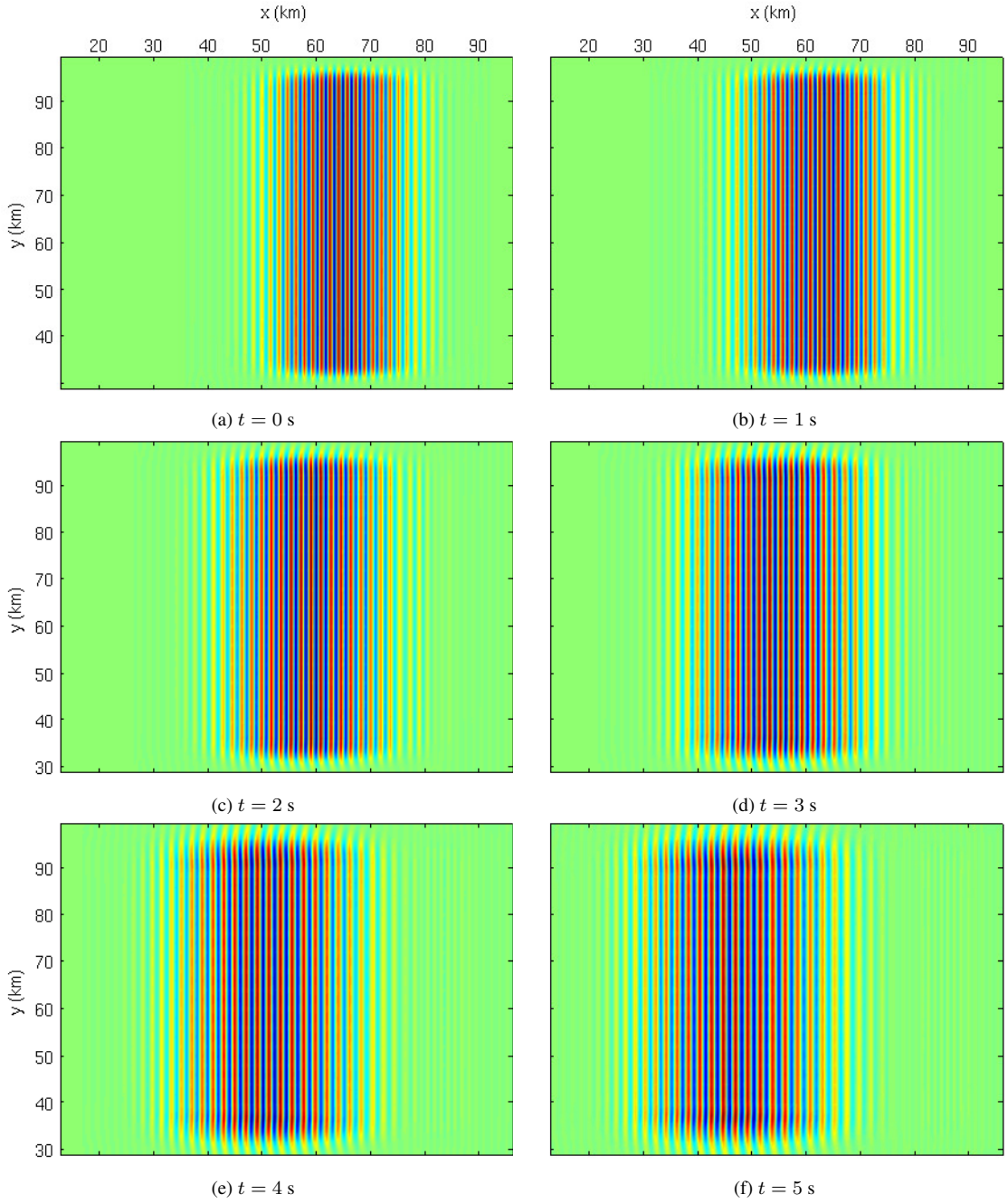

Figure 6. Wavefield of waveguide model: 2D cross-section snapshots at $z = 64$ km.

isotropic in the y-direction, with $c(x, y, z) = c(x, z)$ given by the 2D Marmousi velocity model at position (x, z). Figure 7 shows the structure of $c(x, z)$ with the velocity varying form $1.5 \sim 4.5$ km/s in a domain given by 3.984 km $\leq x \leq 8.016$ km, 0 km $\leq y \leq 4.032$ km, and $0 \leq z \leq 2.904$ km, where x and y are the two horizontal directions and z is the depth.

Unfortunately no analytical solution is available for 3D smoothed Marmousi model, and as studied

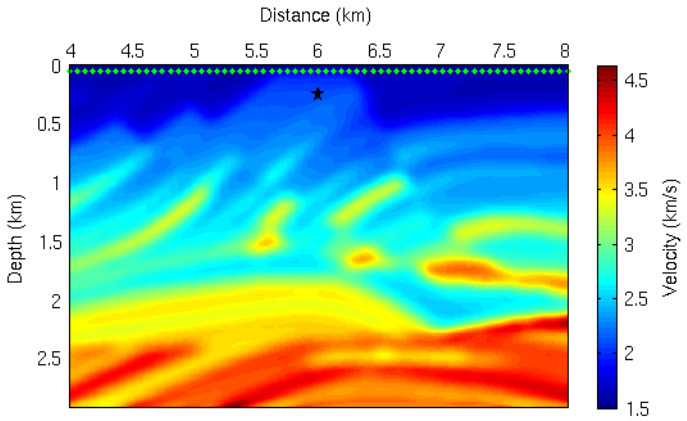

Figure 7. Marmousi Model: Smoothed velocity. The solid pentagram indicates the location of $\boldsymbol{x}_0$, and the green dots indicate the station locations.

in Sections 5.1-5.2, computation of reference solution using SPECFEM3D is extremely costly and memory-demanding for high-frequency seismic wave propagation in large domain. Therefore, we first choose to compare FGA to SPECFEM3D in a small region for the purpose of verifying the accuracy of FGA by using wave-guide initial conditions of lower frequency as below.

$$\begin{cases} f_0^k(\boldsymbol{x}) = \exp\left(-\dfrac{100|\boldsymbol{x}-\boldsymbol{x}_0|^2}{L^2}\right)\cos\left(\dfrac{256x}{L}\right), \\ f_1^k(\boldsymbol{x}) = 256\exp\left(-\dfrac{100|\boldsymbol{x}-\boldsymbol{x}_0|^2}{L^2}\right)\cos\left(\dfrac{256x}{L}\right), \end{cases} \tag{39}$$

where $\boldsymbol{x}_0 = (4.992, 0.504, 0.5)$ km, and $L = 1.008$ km. SPECFEM3D solves this problem in the domain 3.984 km $\leq x \leq 6.0$ km, 0 km $\leq y \leq 1.008$ km, and 0 km $\leq z \leq 1.008$ km. The element size is taken uniformly as 15.75 m. The FGA solves the problem with 21368 initial Gaussians, and $k = 256$ km^{-1} and $\sigma_1 = \sigma_2 = 1$ km^{-1} in (13)-(14). In Figure 8 we compare the FGA solutions with the SPECFEM3D solutions and display snapshots at different time, which shows that FGA produces reasonable good resolution of seismic wavefield.

Then we study the smooth Marmousi model in a larger domain and seismic wave propagation at a higher frequency using FGA simulation. The initial conditions are now chosen as the form of (32) with $c_0 = c(\boldsymbol{x}_0) = 2.017$ km/s and

$$\phi(t) = \exp\left(-\frac{t^2}{\sigma^2}\right)\cos\left(2\pi f t\right), \tag{40}$$

where $f = 81.4873$ Hz, $\sigma = 2^{-7}$ s, $t_0 = 0.03125$ s, and $\boldsymbol{x}_0 = (6, 6, 0.25)$ km.

In Figure 9 and Figure 10 we plot the snapshots of wavefield generated by FGA at different time.

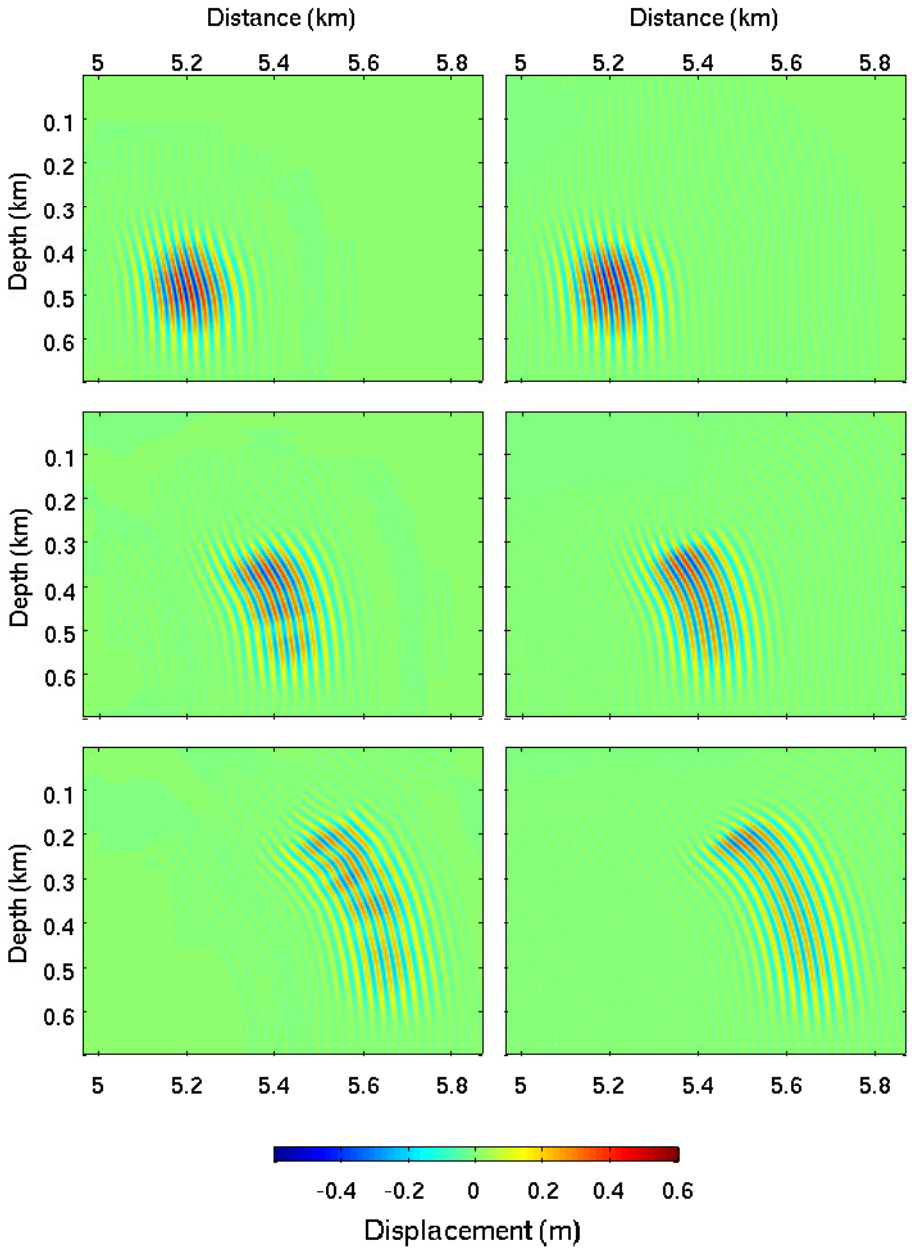

Figure 8. Marmousi Model benchmark test: Wavefield snapshots at $y = 2.016$ km. Left column: FGA solutions; Right column: SPECFEM3D solutions. Top row: time $t = 0.1$ s; Middle row: time $t = 0.2$ s; Bottom row: time $t = 0.3$ s.

We put a series of stations along the line $y = 2.016$ km, $z = 0.05$ km and record the seismogram $u(t)$ at each station. The seismograms recorded by stations are shown in Figure 11. Thanks to the high-frequency seismic modeling, the direct arrivals and free surface reflections are decoupled and

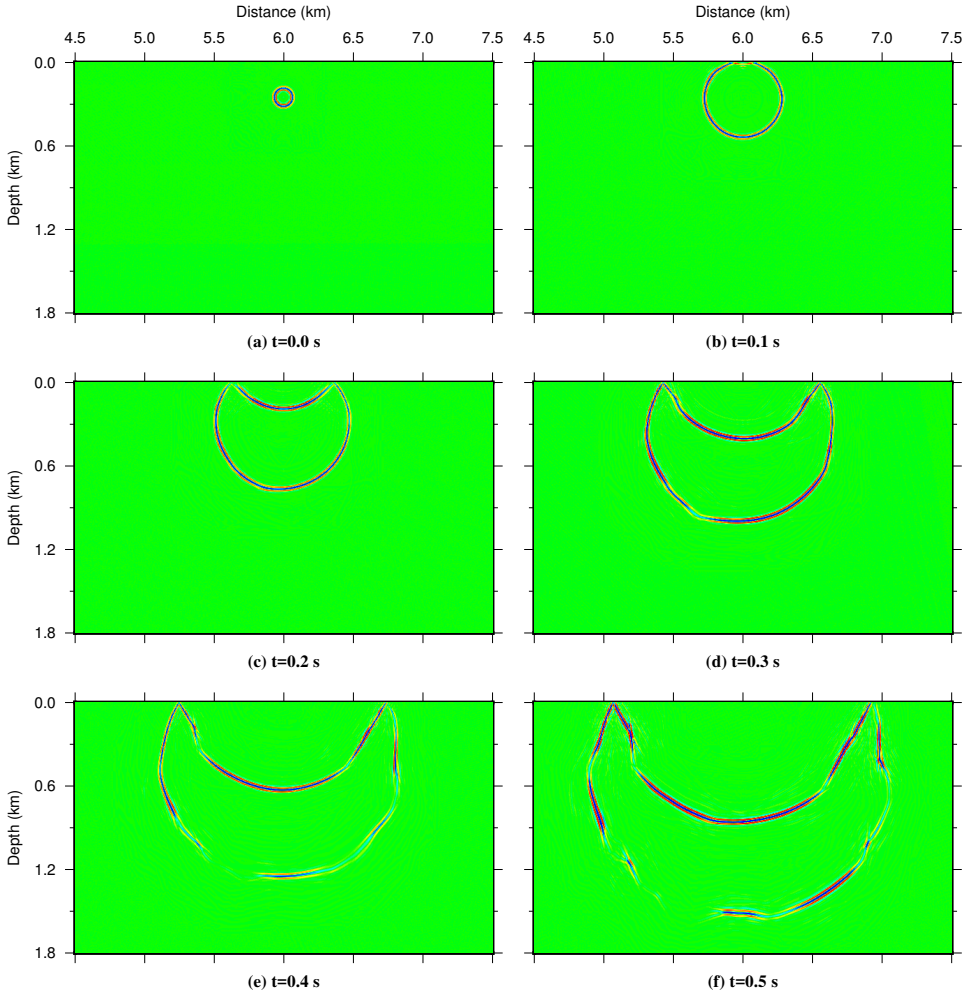

Figure 9. Wavefield cross-section snapshots at $y = 2.016$ km.

can be easily identified on most of the seismograms even though the source is only 250 meters away from the free surface.

5.5 Crust-over-mantle model

We consider a two-layer velocity model with the Moho discontinuity at a depth of 30.0 km and use FGA to compute synthetic seismograms and sensitivity kernels for different phases. The *P*-wave velocities in the crust and the mantle are 5.8 km/s and 7.8 km/s respectively. An earthquake is placed at the horizontal distance $(x, y) = (50.0, 50.0)$ km and the depth $z = 12.0$ km. In order to model the earthquake, we choose the initial conditions as the form of (32) with $c_0 = c(\boldsymbol{x}_0) = 5.8$ km/s and

$$\phi(t) = \exp\left(-\frac{t^2}{\sigma^2}\right)\cos\left(2\pi f t\right), \tag{41}$$

where $f = 9.0541$ Hz, $\sigma = 0.0703$ s, $t_0 = 0.2813$ s, and $\boldsymbol{x}_0 = (50.0, 50.0, 12.0)$ km. Figure

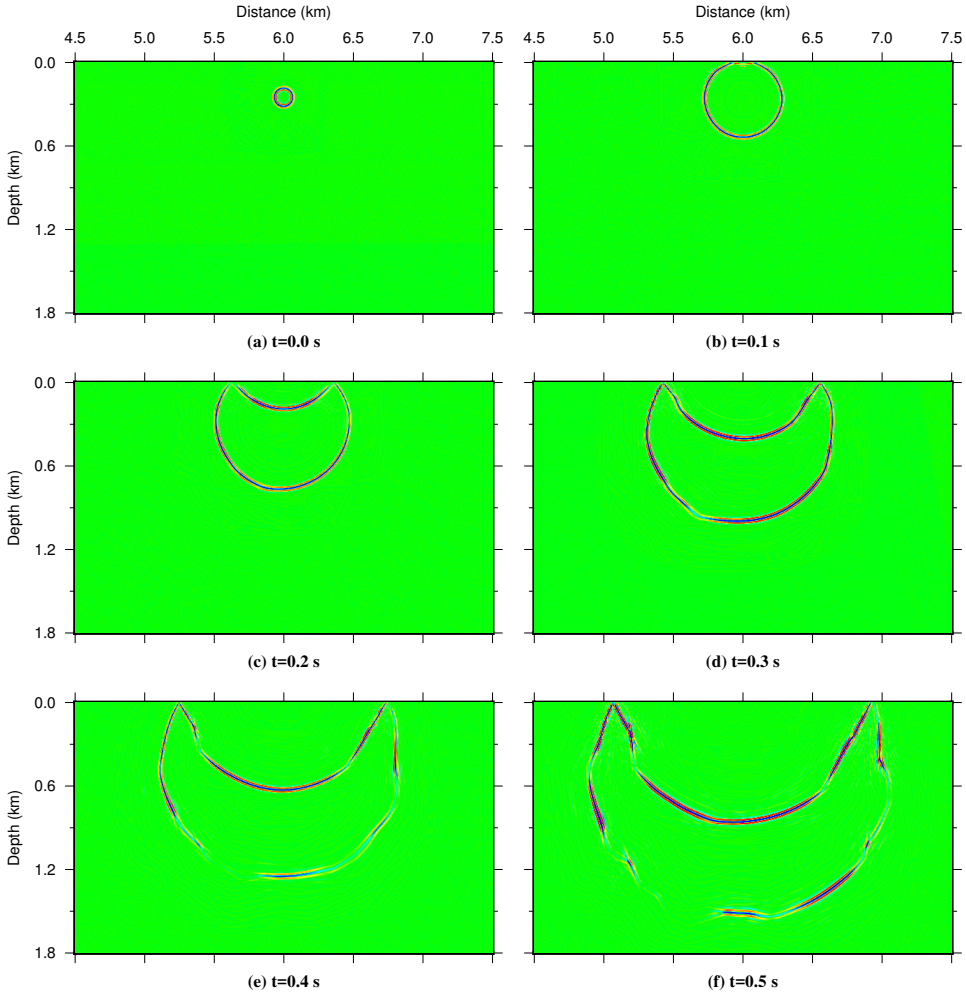

Figure 10. Wavefield cross-section snapshots at $x = 2.016$ km.

13 demonstrates the recorded seismograms at the two receivers at $(x, y) = (14.0, 50.0)$ km and $(x, y) = (86.0, 50.0)$ km on the surface. Three different phases as the direct *P* arrival, Moho reflected phase (PmP), and surface reflection (pPmP) are clearly shown on the two seismograms. We use FGA to calculate the traveltime kernels for the three different phases (e.g. Tromp et al. 2005; Tong et al. 2014a). In Figure 14, we can observe that the kernels for different phases are centered by the geometrical ray paths of the corresponding phase and each kernel is zero along the ray path. This is consistent with the previous observation that 3D traveltime kernels have zero sensitivity along the ray paths (Tromp et al. 2005). We can also see that there are isochrons in the kernels for the reflected phases (PmP, pPmP). It means that any velocity perturbations on the isochrons would delay/advance the related phase at the receiver. The ability of calculating 3D traveltime kernels makes it straightforward to conduct full waveform inversion using FGA.

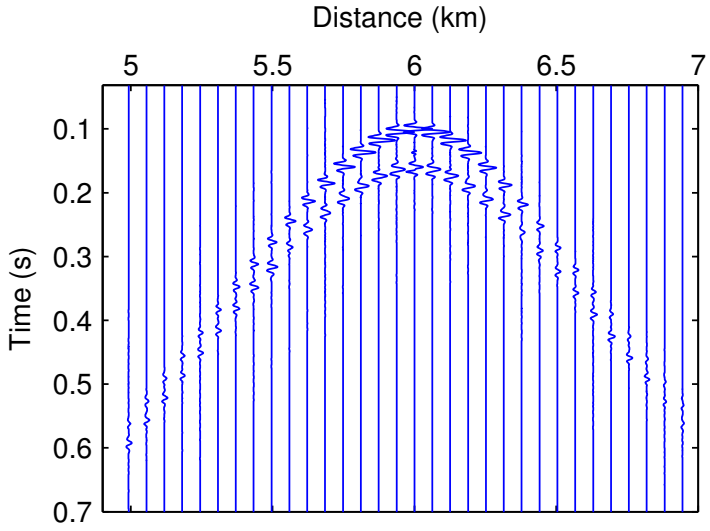

Figure 11. Seismograms recorded by each station.

6 DISCUSION AND CONCLUSIONS

We proposed an efficient and accurate numerical method, frozen Gaussian approximation (FGA), for simulating high-frequency seismic waves in 3D earth models. In this method, one decomposes seismic wavefield into frozen (fixed-width) Gaussian functions, which propagate along ray paths. Inheriting the merit of ray-based beam methods (e.g. Kirchhoff migration and Gaussian beam methods), one can use relatively small number of Gaussians to get accurate solutions of high-frequency wavefield. The algorithm can be drastically speeded up by parallel implementation in a multicore-processor computer station. Moreover, there is no stability limitation on the ratio of time step over spacial resolution, and thus can be used for large domain simulation. This method will greatly help us understand the interactions between high-frequency seismic waves and fine subsurface structures. Reversely, by analyzing high-frequency seismic data and with the aid of the efficient FGA solver, we can reveal detailed subsurface structures and have a deep understanding of the Earth's interior at a variety of scales.

Standard FGA can only be used to simulate seismic wave propagation in smoothly varying velocity models without strong velocity contrasts. In this paper, we incorporated the Snell's law into the FGA formulation, and asymptotically derived free surface, reflection and transmission conditions of FGA for high-frequency seismic wave propagation in high contrast media. These conditions were numerically tested by computing traveltime kernels of different phases in the 3D crust-over-mantle model. We remark that, the application of FGA is limited to the computation of high-frequency/large domain seismic wave propagation only. As suggested by (16), the accuracy of FGA is inversely proportional to wave number. Meanwhile, the proposed method is for solving 3D acoustic wave equations.

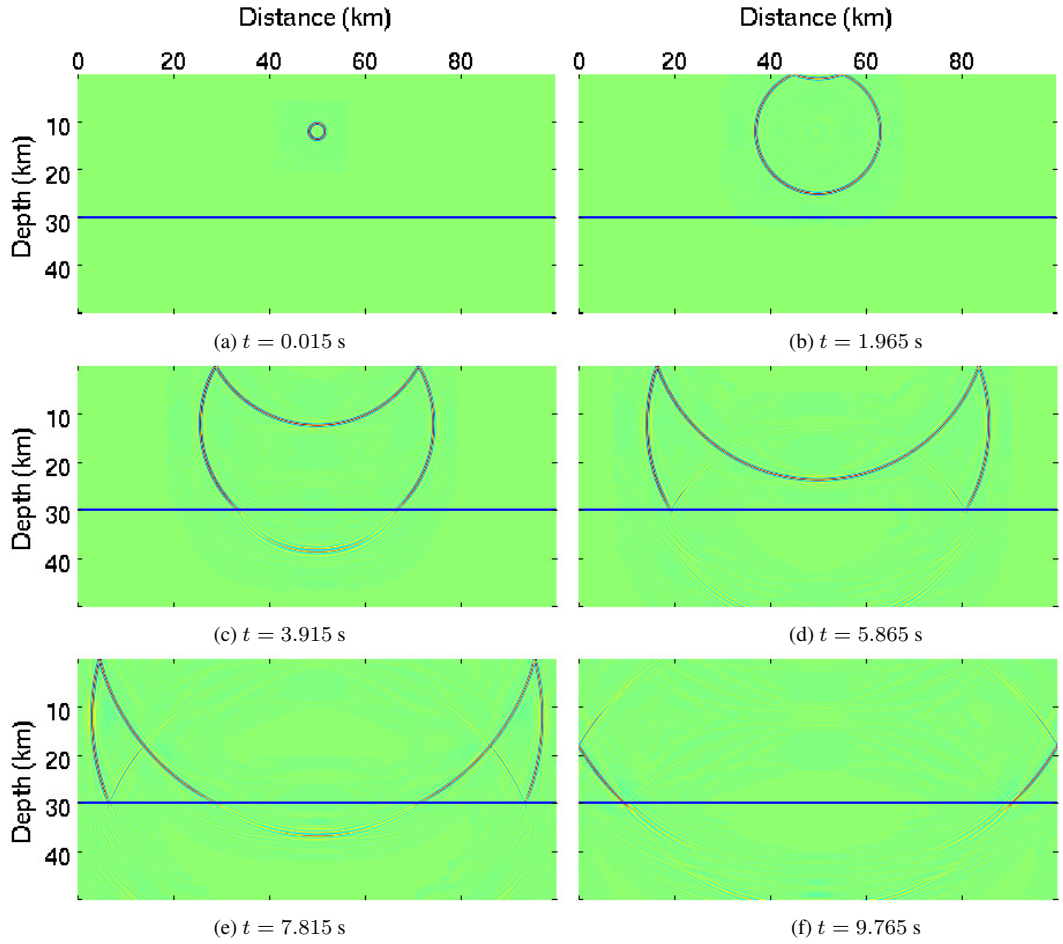

Figure 12. Wavefield 2D cross-section snapshots at $y = 50.0$ km.

Sometimes we need to take into account complex wave propagation phenomena, such as wave conversion, coupling of different phases, anisotropy, and attenuation. The extension to FGA for solving 3D elastic/visoc-elastic wave equations becomes necessary. However, derivation of FGA formulation for elastic/visoc-elastic wave equations are quite technically involved, and we shall leave them as future work.

In seismic imaging, the reverse time migration (RTM) and full waveform inversion (FWI) methods are considered to be the two of the most promising imaging techniques of nowadays (Fei et al. 2015; Virieux & Operto 2009; Tong et al. 2014a). RTM is a high-fidelity algorithm for accurate imaging in and below areas with structural complexities and has become the state-of-the-art technique to image subtle and complex geologic features (Zhang et al. 2014). FWI is able to generate high-resolution subsurface images at half the propagated wavelength. It is potentially the last-course procedure of extracting the information of the Earth's interior from seismograms (Virieux & Operto 2009). The key ingredient of both RTM and FWI is to efficiently and accurately solve seismic wave equations.

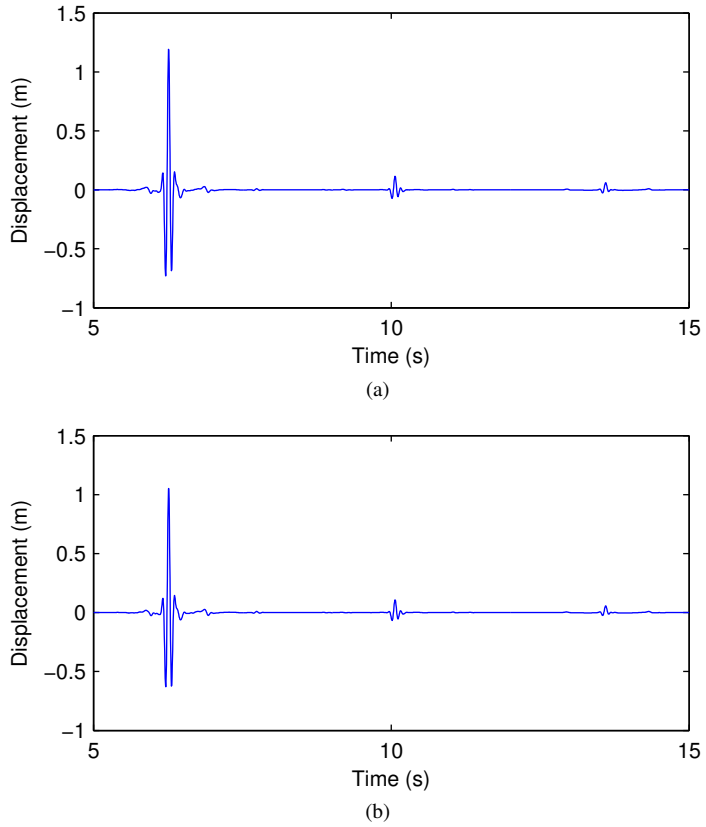

Figure 13. Seismograms recorded by the stations located at (a) $(x, y) = (14, 50)$ km and (b) $(x, y) = (86, 50)$ km on the surface .

Except for RTM and FWI, the implementation of some other imaging techniques such as adjoint tomography using cross-correlation traveltime measurement (Tromp et al. 2005) and wave-equation-based traveltime tomography (Tong et al. 2014b) also heavily relies on fully solving wave equations. In addition, wave-equation-based earthquake location methods require the numerical solutions of wave equations as well (e.g. Liu et al. 2004; Tong et al. 2016). Due to its computational efficiency, FGA can be used as the forward modeling tool of all the mentioned imaging and earthquake location methods. FGA can not only increase the efficiency of the numerical modelings at current scales but also make these wave-equation-based imaging and earthquake location methods applicable to seismic data at high frequency. This is essential for generating higher-resolution subsurface structures. To confirm this, a crosswell full waveform inversion study using FGA as the forward modeling tool is under our investigation.

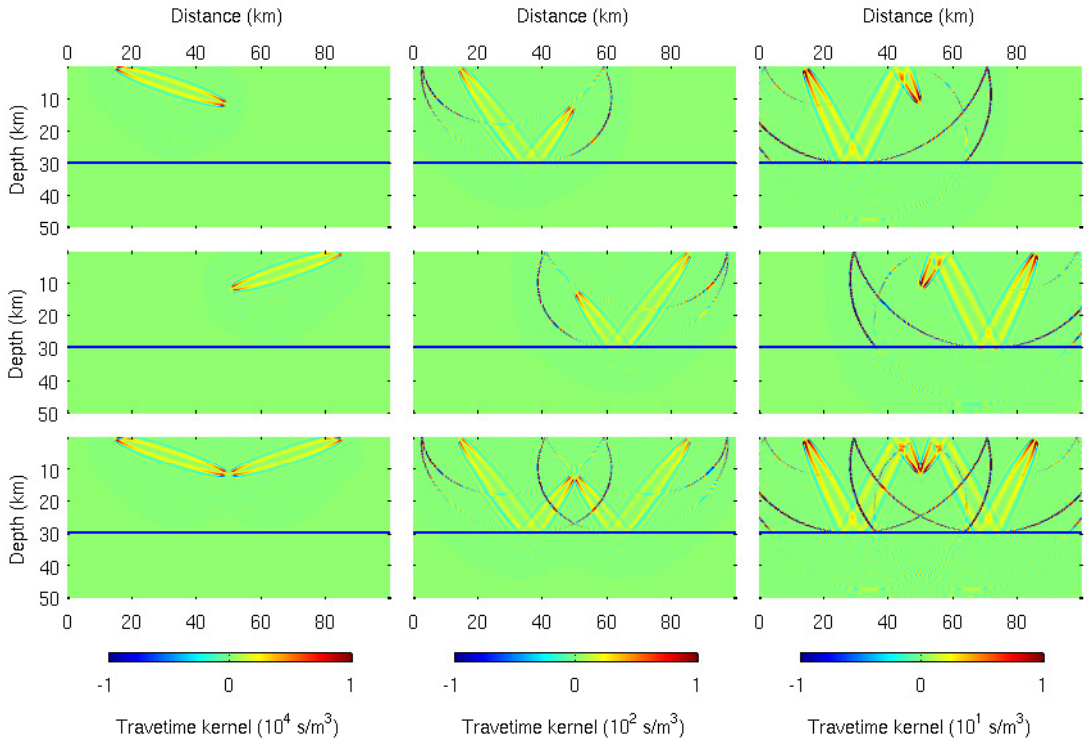

Figure 14. Traveltime kernels of three different phases at $y = 50.0$ km. Left column: direct P arrival; Middle column: Moho reflected phase (PmP); Right column: sureface reflection (pPmP).

APPENDIX A: DERIVATION OF THE FROZEN GAUSSIAN APPROXIMATION

For a sake of simplicity, we will only derive the equation for a_+ by keeping the "+" integral in (3). We shall also omit the subscript "+" in the derivation for convenience. Note that the derivation for a_- will be essentially the same.

By substituting (4) in (3) and combing the phase function, one can rewrite (3) as,

$$u_{\mathrm{F}}^k(t,\boldsymbol{x}) = \iiint \frac{a u_0^k}{(2\pi/k)^{9/2}} e^{\mathrm{i}k\Phi(t,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})}\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{q}\,\mathrm{d}\boldsymbol{p}, \tag{A.1}$$

with $\Phi(t,\boldsymbol{x},\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})$ given by

$$\Phi = \boldsymbol{P}\cdot(\boldsymbol{x}-\boldsymbol{Q}) - \boldsymbol{p}\cdot(\boldsymbol{y}-\boldsymbol{q}) + \frac{\mathrm{i}}{2}|\boldsymbol{x}-\boldsymbol{Q}|^2 + \frac{\mathrm{i}}{2}|\boldsymbol{y}-\boldsymbol{q}|^2, \tag{A.2}$$

and u_0^k given in (5). Taking derivatives of (A.1) produces

$$\partial_t^2 u_{\mathrm{F}}^k = \iiint \frac{u_0^k e^{\mathrm{i}k\Phi}}{(2\pi/k)^{9/2}}\Big(\partial_t^2 a + 2\mathrm{i}k\partial_t a\partial_t\Phi + \mathrm{i}ka\partial_t^2\Phi - k^2 a(\partial_t\Phi)^2\Big)\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}, \tag{A.3}$$

and

$$\Delta_{\boldsymbol{x}} u_{\mathrm{F}}^k = \iiint \frac{a u_0^k e^{\mathrm{i}k\Phi}}{(2\pi/k)^{9/2}}\Big(\mathrm{i}k\Delta_{\boldsymbol{x}}\Phi - k^2|\partial_{\boldsymbol{x}}\Phi|^2\Big)\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}. \tag{A.4}$$

Direct calculations using (A.2) and (6) yield

$$(\partial_t\Phi)^2 = c^2|\boldsymbol{P}|^2 + \Big((\boldsymbol{x}-\boldsymbol{Q})\cdot\Big(|\boldsymbol{P}|\partial_{\boldsymbol{Q}}c + \mathrm{i}c\frac{\boldsymbol{P}}{|\boldsymbol{P}|}\Big)\Big)^2 + 2c|\boldsymbol{P}|(\boldsymbol{x}-\boldsymbol{Q})\cdot\Big(|\boldsymbol{P}|\partial_{\boldsymbol{Q}}c + \mathrm{i}c\frac{\boldsymbol{P}}{|\boldsymbol{P}|}\Big), \tag{A.5}$$

$$\partial_t^2\Phi = c\boldsymbol{P}\cdot\partial_{\boldsymbol{Q}}c + \mathrm{i}c^2 - (\boldsymbol{x}-\boldsymbol{Q})\cdot\Big(-\partial_{\boldsymbol{Q}}c\boldsymbol{P}\cdot\partial_{\boldsymbol{Q}}c + c\partial_{\boldsymbol{Q}}^2c\cdot\boldsymbol{P} - \mathrm{i}c\partial_{\boldsymbol{Q}}c + 2\mathrm{i}c\boldsymbol{P}\frac{\boldsymbol{P}\cdot\partial_{\boldsymbol{Q}}c}{|\boldsymbol{P}|^2}\Big), \tag{A.6}$$

$$|\partial_{\boldsymbol{x}}\Phi|^2 = |\boldsymbol{P}|^2 + 2\mathrm{i}\boldsymbol{P}\cdot(\boldsymbol{x}-\boldsymbol{Q}) - |\boldsymbol{x}-\boldsymbol{Q}|^2, \tag{A.7}$$

$$\Delta\Phi = 3\mathrm{i}. \tag{A.8}$$

To proceed further the asymptotic expansion on phase plane for the equation of a, one needs to relate different powers of $\boldsymbol{x}-\boldsymbol{Q}$ to the order of k, which is given by the following identities (Lu & Yang 2011, Lemma 3.2).

Let us first introduce the notation $f\sim g$ to mean that

$$\iiint f e^{\mathrm{i}k\Phi}\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q} = \iiint g e^{\mathrm{i}k\Phi}\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}, \tag{A.9}$$

then for any vector $\boldsymbol{v}(\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})$ and matrix $M(\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})$, we have

$$\boldsymbol{v}(\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})\cdot(\boldsymbol{x}-\boldsymbol{Q}) \sim -k^{-1}\partial_{z_k}(v_jZ_{jk}^{-1}), \tag{A.10}$$

$$(\boldsymbol{x}-\boldsymbol{Q})\cdot M(\boldsymbol{y},\boldsymbol{q},\boldsymbol{p})(\boldsymbol{x}-\boldsymbol{Q}) \sim k^{-1}\partial_{z_l}Q_jM_{jk}Z_{kl}^{-1} + k^{-2}\partial_{z_m}\big(\partial_{z_l}(M_{jk}Z_{kl}^{-1})Z_{jm}^{-1}\big), \tag{A.11}$$

where Einstein's summation convention and the following short hand notations have been used

$$\partial_{\boldsymbol{z}} = \partial_{\boldsymbol{q}} - \mathrm{i}\partial_{\boldsymbol{p}}, \qquad Z = \partial_{\boldsymbol{z}}(\boldsymbol{Q}+\mathrm{i}\boldsymbol{P}). \tag{A.12}$$

Here $\partial_{\boldsymbol{z}}\boldsymbol{Q}$ and $\partial_{\boldsymbol{z}}\boldsymbol{P}$ are understood as matrices, with the (j,k) component of a matrix $\partial_{\boldsymbol{z}}\boldsymbol{Q}$ given by $\partial_{\boldsymbol{z}_j}\boldsymbol{Q}_k$.

Expanding $c^2(\boldsymbol{x})$ around $\boldsymbol{x}=\boldsymbol{Q}$,

$$c^2(\boldsymbol{x}) = c^2 + 2c\partial_{\boldsymbol{Q}}c\cdot(\boldsymbol{x}-\boldsymbol{Q}) + (\partial_{\boldsymbol{Q}}c\cdot(\boldsymbol{x}-\boldsymbol{Q}))^2 + c(\boldsymbol{x}-\boldsymbol{Q})\cdot\partial_{\boldsymbol{Q}}^2c(\boldsymbol{x}-\boldsymbol{Q}) + \mathcal{O}(|\boldsymbol{x}-\boldsymbol{Q}|)^3, \tag{A.13}$$

and substituting (A.3)-(A.8) into (1) yield

$$2\mathrm{i}kc|\boldsymbol{P}|\partial_t au \sim \mathrm{i}ka(c\boldsymbol{P}\cdot\partial_{\boldsymbol{Q}}c - 2c^2\mathrm{i})u - k^2a(\boldsymbol{x}-\boldsymbol{Q})\cdot M(\boldsymbol{x}-\boldsymbol{Q})u, \tag{A.14}$$

where

$$M = (|\boldsymbol{P}|\partial_{\boldsymbol{Q}}c - \mathrm{i}c\boldsymbol{P}/|\boldsymbol{P}|)\otimes(|\boldsymbol{P}|\partial_{\boldsymbol{Q}}c - \mathrm{i}c\boldsymbol{P}/|\boldsymbol{P}|) + c^2I - |\boldsymbol{P}|^2\partial_{\boldsymbol{Q}}c\otimes\partial_{\boldsymbol{Q}}c - |\boldsymbol{P}|^2c\partial_{\boldsymbol{Q}}^2c. \tag{A.15}$$

Here $\otimes$ means the tensor product of two 3D vectors.

Applying (A.11) to (A.14) brings, after ignoring higher order terms,

$$\partial_t a = \frac{a}{2}\Big(\frac{\boldsymbol{P}}{|\boldsymbol{P}|}\cdot\partial_{\boldsymbol{Q}}c - \frac{2\mathrm{i}}{|\boldsymbol{P}|}c\Big) + \frac{a}{2}\mathrm{Tr}\bigg(Z^{-1}\partial_{\boldsymbol{z}}\boldsymbol{Q}\Big(2\frac{\boldsymbol{P}}{|\boldsymbol{P}|}\otimes\partial_{\boldsymbol{Q}}c - \frac{\mathrm{i}c}{|\boldsymbol{P}|}\Big(\frac{\boldsymbol{P}\otimes\boldsymbol{P}}{|\boldsymbol{P}|^2} - I\Big) - \mathrm{i}|\boldsymbol{P}|\partial_{\boldsymbol{Q}}^2 c\Big)\bigg). \quad \text{(A.16)}$$

Notice that, by (10) and (6),

$$\begin{aligned}\frac{\mathrm{d}Z}{\mathrm{d}t} =&\partial_{\boldsymbol{z}}\left(\frac{\mathrm{d}\boldsymbol{Q}}{\mathrm{d}t} + \mathrm{i}\frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t}\right) = \partial_{\boldsymbol{z}}\left(c\frac{\boldsymbol{P}}{|\boldsymbol{P}|} - \mathrm{i}\partial_{\boldsymbol{Q}}c|\boldsymbol{P}|\right)\\ =&\partial_{\boldsymbol{z}}\boldsymbol{Q}\frac{\partial_{\boldsymbol{Q}}c\otimes\boldsymbol{P}}{|\boldsymbol{P}|} + c\partial_{\boldsymbol{z}}\boldsymbol{P}\Big(\frac{I}{|\boldsymbol{P}|} - \frac{\boldsymbol{P}\otimes\boldsymbol{P}}{|\boldsymbol{P}|^3}\Big) - \mathrm{i}\partial_{\boldsymbol{z}}\boldsymbol{Q}\partial_{\boldsymbol{Q}}^2 c|\boldsymbol{P}| - \mathrm{i}\partial_{\boldsymbol{z}}\boldsymbol{P}\frac{\boldsymbol{P}\otimes\partial_{\boldsymbol{Q}}c}{|\boldsymbol{P}|}, \quad \text{(A.17)}\end{aligned}$$

and

$$-\frac{2\mathrm{i}}{|\boldsymbol{P}|}c = \mathrm{Tr}\bigg(Z^{-1}(\partial_{\boldsymbol{z}}\boldsymbol{Q} + \mathrm{i}\partial_{\boldsymbol{z}}\boldsymbol{P})\frac{\mathrm{i}c}{|\boldsymbol{P}|}\Big(\frac{\boldsymbol{P}\otimes\boldsymbol{P}}{|\boldsymbol{P}|^2} - I\Big)\bigg),$$

then one can reformulate (A.16) as

$$\frac{\mathrm{d}a}{\mathrm{d}t} = a\frac{\boldsymbol{P}}{|\boldsymbol{P}|}\cdot\partial_{\boldsymbol{Q}}c + \frac{a}{2}\mathrm{Tr}\left(Z^{-1}\frac{\mathrm{d}Z}{\mathrm{d}t}\right). \quad \text{(A.18)}$$

Here we have used the fact that (A.11) has a quadratic form, which implies

$$\mathrm{Tr}\bigg(Z^{-1}\partial_{\boldsymbol{z}}\boldsymbol{Q}\frac{\boldsymbol{P}}{|\boldsymbol{P}|}\otimes\partial_{\boldsymbol{Q}}c\bigg) = \mathrm{Tr}\bigg(Z^{-1}\partial_{\boldsymbol{z}}\boldsymbol{Q}\frac{\partial_{\boldsymbol{Q}}c}{|\boldsymbol{P}|}\otimes\boldsymbol{P}\bigg).$$

Moreover, differentiating (6) with respect to $\boldsymbol{z}$ implies a set of equations for $\partial_{\boldsymbol{z}}\boldsymbol{Q}$ and $\partial_{\boldsymbol{z}}\boldsymbol{P}$,

$$\frac{\mathrm{d}(\partial_{\boldsymbol{z}}\boldsymbol{Q})}{\mathrm{d}t} = \pm\partial_{\boldsymbol{z}}\boldsymbol{Q}\frac{\partial_{\boldsymbol{Q}}c\otimes\boldsymbol{P}}{|\boldsymbol{P}|} \pm c\partial_{\boldsymbol{z}}\boldsymbol{P}\left(\frac{I}{|\boldsymbol{P}|} - \frac{\boldsymbol{P}\otimes\boldsymbol{P}}{|\boldsymbol{P}|^3}\right), \quad \text{(A.19)}$$

$$\frac{\mathrm{d}(\partial_{\boldsymbol{z}}\boldsymbol{P})}{\mathrm{d}t} = \mp\partial_{\boldsymbol{z}}\boldsymbol{Q}\partial_{\boldsymbol{Q}}^2 c|\boldsymbol{P}| \mp \partial_{\boldsymbol{z}}\boldsymbol{P}\frac{\boldsymbol{P}\otimes\partial_{\boldsymbol{Q}}c}{|\boldsymbol{P}|}, \quad \text{(A.20)}$$

which can be used for solving $\dfrac{\mathrm{d}Z}{\mathrm{d}t}$ in (A.18).

APPENDIX B: DERIVATION OF THE INTERFACE CONDITIONS FOR $\partial_{\boldsymbol{z}}\boldsymbol{Q}$ AND $\partial_{\boldsymbol{z}}\boldsymbol{P}$

First we define the Liouville operator

$$\mathcal{L}_\pm = \partial_t + \partial_{\boldsymbol{p}}H_\pm\cdot\partial_{\boldsymbol{q}} - \partial_{\boldsymbol{q}}H_\pm\cdot\partial_{\boldsymbol{p}},$$

and recall the 3D Eulerian formulation of the frozen Gaussian approximation in Lu & Yang (2012b),

$$u_{\mathrm{E}}^k(t,\boldsymbol{x}) = \frac{1}{(2\pi/k)^{9/2}}\iint\left(a_+(t,\boldsymbol{q},\boldsymbol{p}) + a_-(t,\boldsymbol{q},\boldsymbol{p})\right)e^{\mathrm{i}k\boldsymbol{p}\cdot(\boldsymbol{x}-\boldsymbol{q})-\frac{k}{2}|\boldsymbol{x}-\boldsymbol{q}|^2}\,\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{q}. \quad \text{(B.1)}$$

Consider the auxiliary functions

$$\boldsymbol{\phi}_\pm(t,\boldsymbol{q},\boldsymbol{p}) = (\phi_{\pm,1},\phi_{\pm,2},\phi_{\pm,3}),$$

which satisfy

$$\mathcal{L}_{\pm}\boldsymbol{\phi}_{\pm} = 0, \quad \text{with} \quad \boldsymbol{\phi}_{\pm}(0,\boldsymbol{q},\boldsymbol{p}) = \boldsymbol{p} + \mathrm{i}\boldsymbol{q}, \tag{B.2}$$

then the evolutionary equation for $a_{\pm}$ is given by

$$\mathcal{L}_{\pm}a_{\pm} = \pm a_{\pm}\frac{\boldsymbol{p}}{|\boldsymbol{p}|}\cdot\partial_{\boldsymbol{q}}c + \frac{a_{\pm}}{2}\mathrm{Tr}\left(Z_{\pm}^{-1}\frac{\mathrm{d}Z_{\pm}}{\mathrm{d}t}\right), \tag{B.3}$$

where $Z_{\pm} = (\partial_{\boldsymbol{p}}\boldsymbol{\phi}_{\pm})^T - \mathrm{i}(\partial_{\boldsymbol{q}}\boldsymbol{\phi}_{\pm})^T$. Theorem 4.1 in Lu & Yang (2012a) provides

$$\partial_{\boldsymbol{z}}\boldsymbol{Q}_{\pm} = (\partial_{\boldsymbol{p}}\boldsymbol{\phi}_{\pm})^T, \text{ and } \partial_{\boldsymbol{z}}\boldsymbol{P}_{\pm} = -(\partial_{\boldsymbol{q}}\boldsymbol{\phi}_{\pm})^T. \tag{B.4}$$

Suppose the interface is located at $z = z_0$, and denote $\boldsymbol{q}^{\mathrm{re}} = \boldsymbol{q}^{\mathrm{tr}} = \boldsymbol{q}^{\mathrm{in}} = (x, y, z_0)$, $\boldsymbol{p}^{\mathrm{in}} = (p_x, p_y, p_z^{\mathrm{in}})$, $\boldsymbol{p}^{\mathrm{re}} = (p_x, p_y, p_z^{\mathrm{re}})$, and $\boldsymbol{p}^{\mathrm{tr}} = (p_x, p_y, p_z^{\mathrm{tr}})$, where

$$p_z^{\mathrm{re}} = -p_z^{\mathrm{in}}, \text{ and } p_z^{\mathrm{tr}} = p_z^{\mathrm{in}}\sqrt{\frac{C_1^2}{C_2^2} + \left(\frac{C_1^2}{C_2^2} - 1\right)\frac{p_x^2 + p_y^2}{\left(p_z^{\mathrm{in}}\right)^2}}.$$

Without loss of generality, we consider the "+" branch and omit the subscript "+" in the derivation for convenience. In order to give a simple and clear presentation of the idea, we assume the initial wave comes form the above of the interface, and hit the interface only once.

The $\boldsymbol{\phi}^{\mathrm{re,tr}}$ and $a^{\mathrm{re,tr}}$ shall satisfy the same evolutionary equations as $\boldsymbol{\phi}$ and a in (B.2) and (B.3) with the corresponding interface conditions at $q_z = z_0$, which are given as follows:

$$\begin{aligned} a^{\mathrm{re}}(t,\boldsymbol{q}^{\mathrm{re}},\boldsymbol{p}^{\mathrm{re}}) &= R(\boldsymbol{p}^{\mathrm{in}})a^{\mathrm{re}}(t,\boldsymbol{q}^{\mathrm{in}},\boldsymbol{p}^{\mathrm{in}}),\\ a^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{tr}},\boldsymbol{p}^{\mathrm{tr}}) &= T(\boldsymbol{p}^{\mathrm{in}})a^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{in}},\boldsymbol{p}^{\mathrm{in}}), \end{aligned} \tag{B.5}$$

$$\begin{aligned} \boldsymbol{\phi}^{\mathrm{re}}(t,\boldsymbol{q}^{\mathrm{re}},\boldsymbol{p}^{\mathrm{re}}) &= \boldsymbol{\phi}^{\mathrm{re}}(t,\boldsymbol{q}^{\mathrm{in}},\boldsymbol{p}^{\mathrm{in}}),\\ \boldsymbol{\phi}^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{re}},\boldsymbol{p}^{\mathrm{re}}) &= \boldsymbol{\phi}^{\mathrm{tr}}(t,\boldsymbol{q}^{\mathrm{in}},\boldsymbol{p}^{\mathrm{in}}), \end{aligned} \tag{B.6}$$

where the reflection coefficient $R(\boldsymbol{p}^{\mathrm{in}})$ and transmission coefficient $T(\boldsymbol{p}^{\mathrm{in}})$ are given by the Snell's Law of Refraction,

$$R(\boldsymbol{p}^{\mathrm{in}}) = \begin{cases} -1, & \text{if } \boldsymbol{P}^{\mathrm{in}} \in \Theta_r,\\ \dfrac{p_z^{\mathrm{in}} - p_z^{\mathrm{tr}}}{p_z^{\mathrm{in}} + p_z^{\mathrm{tr}}}, & \text{otherwise,} \end{cases} \text{ and } T(\boldsymbol{p}^{\mathrm{in}}) = \begin{cases} 0, & \text{if } \boldsymbol{P}^{\mathrm{in}} \in \Theta_r,\\ \dfrac{2p_z^{\mathrm{in}}}{p_z^{\mathrm{in}} + p_z^{\mathrm{tr}}}, & \text{otherwise,} \end{cases}$$

where Θ_r is given in (28).

The interface conditions (30) for $\partial_{\boldsymbol{z}}\boldsymbol{Q}$ and $\partial_{\boldsymbol{z}}\boldsymbol{P}$ are implied by (B.6) and (B.4). In fact, taking the transmission for example, since

$$p_z^{\mathrm{tr}} = p_z^{\mathrm{in}}\sqrt{\frac{C_1^2}{C_2^2} + \left(\frac{C_1^2}{C_2^2} - 1\right)\frac{p_x^2 + p_y^2}{\left(p_z^{\mathrm{in}}\right)^2}},$$

and $\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm tr},\boldsymbol{q}^{\rm tr})=\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm in},\boldsymbol{q}^{\rm in})$, then by the chain rule

$$\begin{aligned}
\partial_{p_x}\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm tr},\boldsymbol{p}^{\rm tr}) &= \left(\frac{C_2^2}{C_1^2}-1\right)\frac{p_x}{p_z^{\rm in}}\partial_{p_z}\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm in},\boldsymbol{p}^{\rm in})+\partial_{p_x}\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm in},\boldsymbol{p}^{\rm in}),\\
\partial_{p_y}\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm tr},\boldsymbol{p}^{\rm tr}) &= \left(\frac{C_2^2}{C_1^2}-1\right)\frac{p_y}{p_z^{\rm in}}\partial_{p_z}\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm in},\boldsymbol{p}^{\rm in})+\partial_{p_y}\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm in},\boldsymbol{p}^{\rm in}),\\
\partial_{p_z}\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm tr},\boldsymbol{p}^{\rm tr}) &= \frac{C_2^2}{C_1^2}\frac{p_z^{\rm tr}}{p_z^{\rm in}}\partial_{p_z}\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm in},\boldsymbol{p}^{\rm in}).
\end{aligned}\tag{B.7}$$

Together with (B.4), (B.7) brings exactly the interface condition (30) for $\partial_{\boldsymbol{z}}Q$.

By $\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm tr},\boldsymbol{q}^{\rm tr})=\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm in},\boldsymbol{q}^{\rm in})$, one has

$$\partial_t\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm in},\boldsymbol{p}^{\rm in})=\partial_t\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm tr},\boldsymbol{p}^{\rm tr}),$$

then the Liouville equation yields, at the interface,

$$\left[\nabla_{\boldsymbol{p}}H\cdot\nabla_{\boldsymbol{q}}\boldsymbol{\phi}^{\rm tr}\right]_{q_z=z_0}=\left[\nabla_{\boldsymbol{q}}H\cdot\nabla_{\boldsymbol{p}}\boldsymbol{\phi}^{\rm tr}\right]_{q_z=z_0},\tag{B.8}$$

where $[\cdot]$ denotes the jump function. Notice that

$$\partial_{q_x}\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm in},\boldsymbol{p}^{\rm in})=\partial_{q_x}\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm tr},\boldsymbol{p}^{\rm tr}),\ \text{and}\ \partial_{q_y}\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm in},\boldsymbol{p}^{\rm in})=\partial_{q_y}\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm tr},\boldsymbol{p}^{\rm tr}),\tag{B.9a}$$

then

$$\begin{aligned}
\partial_{q_z}\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm tr},\boldsymbol{p}^{\rm tr}) &= \frac{C_1^2p_z^{\rm in}}{C_2^2p_z^{\rm tr}}\partial_{q_z}\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm in},\boldsymbol{p}^{\rm in})\\
&\quad+\left(\frac{C_1^2}{C_2^2}-1\right)\left(\frac{p_x}{p_z^{\rm tr}}\partial_{q_x}\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm in},\boldsymbol{p}^{\rm in})+\frac{p_y}{p_z^{\rm tr}}\partial_{q_y}\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q}^{\rm in},\boldsymbol{p}^{\rm in})\right)\\
&\quad+\frac{|\boldsymbol{p}^{\rm tr}|}{C_2\partial_z^{\rm tr}}\left[|\boldsymbol{p}|\nabla_{\boldsymbol{q}}c(\boldsymbol{q})\cdot\nabla_{\boldsymbol{p}}\boldsymbol{\phi}^{\rm tr}(t,\boldsymbol{q},\boldsymbol{p})\right]_{q_z=z_0}.
\end{aligned}\tag{B.9b}$$

Therefore by (B.4), (B.9) gives exactly the interface condition (30) for $\partial_{\boldsymbol{z}}P$.

ACKNOWLEDGMENTS

This work was partially supported by the NSF grants DMS-1418936 and DMS-1107291, and Hellman Family Foundation Faculty Fellowship, UC Santa Barbara. We are grateful to the Editor Rene-Edouard Plessix and anonymous referees for their valuable comments and suggestions.

REFERENCES

Aki, K., Christofferson, A., & Husebye, E., 1977. Determination of the three-dimensional seismic structure of the lithosphere, *J. Geophys. Res.*, **82**, 277–296.

Albertin, U., Yingst, D., & Jaramillo, H., 2001. Comparing common-offset Maslov, Gaussian beam, and coherent state migrations, in *71st Ann. Internat. Mtg*, pp. 913–916, Soc. of Expl. Geophys.

Alford, R. M., Kelly, K. R., & Boore, D. M., 1974. Accuracy of finitedifference modelling of the acoustic wave equation, *Geophysics*, **39**(6), 834–842.

Bao, H., Bielak, J., Ghattas, O., Kallivokas, L. F., O'Hallaron, D. R., Shewchuk, J. R., & Xu, J., 1998. Large-scale simulation of elastic wave propagation in heterogeneous media on parallel computers, *Comp. Meth. appl. Mech. Eng.*, **152**, 85–102.

Bouchon, M. & Sanchez-Sesma, F. J., 2007. Boundary integral equations and boundary elements methods in elastodynamics, *Adv. Geophys.*, **48**(06), 157–189.

Carcione, J., 1994. The wave equation in generalized coordinates, *Geophysics*, **59**(12), 1911–1919.

Cerveny, V., 2001. *Seismic ray theory*, Cambridge University Press, Cambridge.

Cerveny, V., Popov, M. M., & Psencik, I., 1982. Computation of wave fields in inhomogeneous media – Gaussian beam approach, *Geophys. J. Roy. Astr. Soc.*, **70**, 109–128.

Engquist, B. & Runborg, O., 2003. Computational high frequency wave propagation, *Acta Numer.*, **12**, 181–266.

Fei, T., Luo, Y., Yang, J., Liu, H., & Qin, F., 2015. Removing false images in reverse time migration: The concept of de-primary, *Geophysics*, **80**(6), S237–S244.

Folland, G. B., 1989. *Harmonic analysis in phase space*, Annals of Mathematics Studies, no. 122, Princeton University Press, Princeton.

Fomel, S. & Tanushev, N., 2009. Time-domain seismic imaging using beams, in *79th Ann. Internat. Mtg*, pp. 2747–2752, Soc. of Expl. Geophys.

Foster, D., Wu, R., & Mosher, C., 2002. Coherent-state solutions of the wave equation, in *72nd Ann. Internat. Mtg*, pp. 1348–1351, Soc. of Expl. Geophys.

Foster, D. J. & Huang, J.-I., 1991. Global asymptotic solutions of the wave equation, *Geophys. J. Int.*, **105**, 163–171.

Graves, R. W., 1996. Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences, *Bull. Seismol. Soc. Am.*, **86**, 1091–1106.

Gray, S. & Bleistein, N., 2009. True-amplitude gaussian-beam migration, *Geophysics*, **74**(2), S11–S23.

Gray, S. H., 1986. Efficient traveltime calculations for Kirchhoff migration, *Geophysics*, **51**, 1685–1688.

Gray, S. H., 2005. Gaussian beam migration of common-shot records, *Geophysics*, **70**, S71–S77.

Heller, E. J., 1981. Frozen Gaussians: A very simple semiclassical approximation, *J. Chem. Phys.*, **75**, 2923–2931.

Herman, M. F. & Kluk, E., 1984. A semiclassical justification for the use of non-spreading wavepackets in dynamics calculations, *Chem. Phys.*, **91**, 27–34.

Hill, N. R., 1990. Gaussian beam migration, *Geophysics*, **55**, 1416–1428.

Hill, N. R., 2001. Prestack Gaussian-beam depth migration, *Geophysics*, **66**, 1240–1250.

Kay, K., 1994. Integral expressions for the semi-classical time-dependent propagator, *J. Chem. Phys.*, **100**, 4377–4392.

Kay, K., 2006. The Herman-Kluk approximation: Derivation and semiclassical corrections, *Chem. Phys.*, **322**,

3–12.

Keho, T. H. & Beydoun, W. B., 1988. Paraxial ray Kirchhoff migration, *Geophysics*, **53**, 1540–1546.

Kolsloff, D. & Baysal, E., 1982. Forward modelling by a Fourier method, *Geophysics*, **56**, 231–241.

Komatitsch, D. & Tromp, J., 1999. Introduction to the spectral element method for three-dimensional seismic wave propagation, *Geophys. J. Int.*, **139**(3), 806–822.

Komatitsch, D., Coutel, F., & Mora, P., 1996. Tensorial formulation of the wave equation for modelling curved interfaces, *Geophys. J. Int.*, **127**, 156–168.

Komatitsch, D., Tsuboi, S., & Tromp, J., 2005. The spectral-element method in seismology, in *Seismic Earth: Array Analysis of Broadband Seismograms: Geophysical monograph. 157, AGU*, pp. 205–228, eds Levander, A. & Nolet, G., Washington DC, USA.

Li, J., Lin, G., & Yang, X., 2015. A frozen gaussian approximation-based multi-level particle swarm optimization for seismic inversion, *J. Comput. Phys.*, **296**, 58–71.

Liu, Q., Polet, J., Komatitsch, D., & Tromp, J., 2004. Spectral-element moment tensor inversions for earthquakes in southern California, *Bull. Seismol. Soc. Am.*, **94**(5), 1748–1761.

Lu, J. & Yang, X., 2011. Frozen Gaussian approximation for high frequency wave propagation, *Commun. Math. Sci.*, **9**, 663–683.

Lu, J. & Yang, X., 2012a. Convergence of frozen Gaussian approximation for high frequency wave propagation, *Comm. Pure Appl. Math.*, **65**, 759–789.

Lu, J. & Yang, X., 2012b. Frozen Gaussian approximation for general linear strictly hyperbolic systems: Formulation and Eulerian methods, *Multiscale Model. Simul.*, **10**, 451–472.

Martinez, A., 2002. *An introduction to semiclassical and microlocal analysis*, Springer-Verlag, New York.

Nowack, R., Sen, M., & Stoffa, P., 2003. Gaussian beam migration for sparse common-shot and common-receiver data, in *73rd Ann. Internat. Mtg.*, pp. 1114–1117, Soc. of Expl. Geophys.

Olsen, K. B. & Archuleta, R. J., 1996. 3-D simulation of earthquakes on the Los Angeles fault system, *Bull. seism. Soc. Am.*, **86**, 575–596.

Popov, M. M., 2002. *Ray theory and Gaussian beam method for geophysicists*, EDUFBA.

Popov, M. M., Semtchenok, N. M., Popov, P. M., & Verdel, A. R., 2010. Depth migration by the Gaussian beam summation method, *Geophysics*, **75**, S81–S93.

Qian, J. & Ying, L., 2010. Fast multiscale Gaussian wavepacket transforms and multiscale Gaussian beams for the wave equation, *Multiscale Model. Simul.*, **8**, 1803–1837.

Swart, T. & Rousse, V., 2009. A mathematical justification of the Herman-Kluk propagator, *Commun. Math. Phys.*, **286**, 725–750.

ten Kroode, A., Smit, D.-J., & Verdel, A., 1994. Linearized inversed scattering in the presence of caustics, in *Mathematical Methods in Geophysical Imaging II*, vol. 2301, pp. 28–42.

Tong, P., Zhao, D., & Yang, D., 2011. Tomography of the 1995 kobe earthquake area: comparison of finite-frequency and ray approaches, *Geophys. J. Int.*, **187**(1), 278–302.

Tong, P., Chen, C.-W., Komatitsch, D., Basini, P., & Liu, Q., 2014a. High-resolution seismic array imaging

based on an SEM-FK hybrid method, *Geophys. J. Int.*, **197**(1), 369–395.

Tong, P., Zhao, D., Yang, D., Yang, X., Chen, J., & Liu, Q., 2014b. Wave-equation-based travel-time seismic tomography—Part 1: Method, *Solid Earth*, **5**, 1151–1168.

Tong, P., Yang, D., Liu, Q., Yang, X., & Harris, J., 2016. Acoustic wave-equation-based earthquake location, *Geophys. J. Int.*, **205**, 464–478.

Tromp, J., Tape, C., & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, **160**(1), 195–216.

Tromp, J., Komatitsch, J., & Liu, Q., 2008. Spectral-element and adjoint methods in seismology, *Comm. in Comput. Phys*, **3**, 1–32.

Virieux, J., 1984. SH-wave propagation in heterogeneous media: velocity-stress finite difference method, *Geophysics*, **49**, 1933–1942.

Virieux, J., 1986. P-SV wave propagation in heterogeneous media: velocity-stress finite-difference method, *Geophysics*, **51**, 889–901.

Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics*, **74**, WCC1–WCC26.

Yang, D. H., Liu, E., Zhang, Z., & Teng, J., 2002. Finite-difference modelling in two-dimensional anisotropic media using a fluxcorrected transport technique, *Geophys. J. Int.*, **148**, 320–328.

Yang, X., Lu, J., & Fomel, S., 2013. Seismic modeling using the frozen Gaussian approximation, *SEG Technical Program Expanded Abstracts 2013*, pp. 4677–4682.

Zhang, Y., Ratcliffe, A., Roberts, G., & Duan, L., 2014. Amplitude-preserving reverse time migration: From reflectivity to velocity and impedance inversion, *Geophysics*, **79**(6), S271–S283.