Bulletin of the Seismological Society of America, Vol. 89, No. 1, pp. 69–79, February 1999

3D Displacement Finite Differences and a Combined Memory Optimization

by Peter Moczo, Mária Lucká, Jozef Kristek, and Miriam Kristeková

Abstract We present a general optimization technique for the three-dimensional finite-difference (FD) modeling of seismic-wave propagation and earthquake ground motion. Our combined memory optimization (CDMO) naturally comprises core memory optimization and disk memory optimization. While core memory optimization is based on keeping only a limited number of model planes in core memory at a given time, disk memory optimization is based on data compression in the wavelet domain. CDMO enables significant reduction of both computer core and disk memory requirements. CDMO is general: It is applicable to any explicit finite-difference scheme on a conventional or staggered grid.

CDMO is presented on the example of the displacement finite-difference scheme. Accuracy of the scheme was tested through numerical comparisons with the discrete-wavenumber method. The scheme was shown to be capable to account for the position of the material discontinuity more accurately than other recent finite-difference schemes.

Extensive numerical experiments were carried out in order to find proper parameters of the wavelet compression and investigate effects of the compression on synthetics.

Introduction

The finite-difference (FD) method has recently proven to be a powerful tool for numerical three-dimensional (3D) modeling of earthquake ground motion. This is documented by numerous studies presenting simulations of the earthquake ground motion due to real or anticipated earthquakes in major populated areas (see, for example, Olsen and Schuster, 1992; Frankel, 1993; Graves, 1993; Yomogida and Etgen, 1993; Olsen *et al.*, 1995; Pitarka *et al.*, 1998; Wald and Graves, 1998). Most of the recent 3D FD simulations are based on the velocity-stress formulation of the equation of motion.

Consider a Cartesian coordinate system (x_1, x_2, x_3). Denote $\rho(\mathbf{x})$ density; $\lambda(\mathbf{x})$ and $\mu(\mathbf{x})$ Lamé elastic coefficients; $\mathbf{u}(\mathbf{x}, t)$ displacement vector; $\tau_{ij}(\mathbf{x}, t)$, where $i, j \in \{1, 2, 3\}$, stress tensor; and $\mathbf{f}(\mathbf{x}, t)$ body force per unit volume. Wave propagation in a linear isotropic perfectly elastic medium can be described by equations in the displacement-stress formulation, that is,

$$\rho u_{i,tt} = \tau_{ij,j} + f_i,$$

$$\tau_{ij} = \lambda u_{k,k}\delta_{ij} + \mu(u_{i,j} + u_{j,i}),$$

or velocity-stress formulation, that is,

$$\rho \dot{u}_{i,t} = \tau_{ij,j} + f_i,$$

$$\tau_{ij,t} = \lambda \dot{u}_{k,k}\delta_{ij} + \mu(\dot{u}_{i,j} + \dot{u}_{j,i})$$

where $\dot{u}_i$ denotes the ith particle velocity component. In both equations, $i, j, k \in \{1, 2, 3\}$, and the summation convention for repeated subscripts are assumed. $u_{i,tt} = \partial^2 u_i/\partial t^2$, $\tau_{ij,j} = \partial \tau_{ij}/\partial x_j$, and so on.

A parsimonious FD scheme based on the displacement-stress formulation was suggested for the *P–SV* case by Luo and Schuster (1990) and extended to the 3D case by Olsen and Schuster (1992) and Ohminato and Chouet (1997). The displacement-stress FD scheme uses the stress components only as temporary quantities that need not be stored. Therefore, in the 3D case, the displacement-stress scheme requires only 75% of memory required by the velocity-stress scheme. Given this advantage, the displacement-stress scheme should receive more attention in future 3D modeling.

In this article, we present and test the FD scheme that is based on the displacement formulation of the equation of motion; that is,

$$\rho u_{i,tt} = (\lambda u_{k,k})_i + (\mu u_{i,j})_j + (\mu u_{j,i})_j + f_i. \qquad (1)$$

While in the two previous cases, it is enough to apply a standard central-difference formula to approximate the first spatial derivatives, in the displacement formulation, it is necessary to find a proper approximation of terms with the non-mixed and mixed second derivatives, such as $\partial(a\ \partial f/\partial x_1)/\partial x_1$

and $\partial(a\ \partial f/\partial x_1)/\partial x_3$. The approximation of the nonmixed derivative is relatively easy. A proper approximation of the mixed term is not so obvious. Different approximations lead to schemes with different accuracy. We apply the approximation suggested by Zahradník (1995).

It is well known that the displacement schemes (in the *P–SV* and 3D cases) have problems with high Poisson's ratio and very large velocity contrasts. Velocity-stress and displacement-stress schemes do not have these problems (see, e.g., Virieux, 1986). We will show, however, that our displacement scheme is in certain media more accurate than the two other schemes.

Generally, no scheme was shown to be the best in all possible modeling problems. In other words, no FD scheme (including ours) can be recommended for really universal use. We want to stress two things here: (1) The reason why we present and use the displacement FD scheme is a remarkable accuracy of the scheme in a certain class of media, as shown later. (2) The combined memory optimization, which we present in this article, is in no way restricted to the particular FD scheme.

One major drawback of the FD method is considerable computer time and memory requirements as the recent ground-motion simulations clearly indicate. Obviously, an improvement can be achieved by progress in computer technology. It would not be wise, however, to wait for it and rely only on it. The advance in technology may solve today's problems but not necessarily those in future seismological research. Therefore, developments of optimized computational algorithms and sophisticated computer codes have crucial importance for application of the FD method to modeling seismic-wave propagation and earthquake ground motion.

Several approaches have been developed in order to make the FD calculations more efficient. Schemes on spatial grids with varying size of grid spacings (e.g., Moczo, 1989; Pitarka, 1999) and combined grids (e.g., Jastram and Behle, 1992; Moczo *et al.,* 1996, 1997; Aoi and Fujiwara, 1998) reduce total number of grid points and thus memory requirements and number of operations compared to schemes on regular grids. Reduction in computer memory and time requirements is also achieved by fourth-order schemes compared to second-order ones (e.g., Levander, 1988; Graves, 1996). Core memory optimization reduces number of grid points residing at a given time in core memory and makes use of disk memory (e.g., Olsen and Schuster, 1992; Graves, 1996). Parallel programming allows to reduce computational time compared to conventional programming (e.g., Olsen *et al.,* 1995).

Core memory optimization (as described by Graves, 1996, for 3D fourth-order velocity-stress scheme) consists in keeping only a limited number of grid planes in core memory at one time and performing a maximum possible number of time updates for these planes. Core memory requirements are significantly reduced; however, what had to be stored in core memory in computation without core memory optimization has to be stored in disk if core memory optimization is applied. It is easy to estimate that while in the case of a realistic large model, we can keep core memory requirements relatively low (say, several hundreds of megabytes), tens of gigabytes of the hard-disk memory are needed. Disk storage is cheap now but the time necessary for the input/output (I/O) operations in the case of the big disk storage is significant and creates a bottleneck of the computation. Moreover, subroutines performing the I/O operations cannot be well parallelized, though a major part of the code can be. A significant consequence is that hard-disk memory rather than core memory determines a limit for a model size and maximum frequency.

Therefore, we suggest a combined memory optimization (CDMO) that comprises the above-described core memory optimization and disk memory optimization based on data compression in the wavelet domain.

Equation of Motion

The system of the second-order partial differential equations (1) can be rewritten (replacing x_1, x_2, x_3 by x, y, z, and u_1, u_2, u_3 by u, v, w, respectively) in a more readable form

$$\begin{aligned}
\rho u_{tt} &= (\lambda u_x)_x + 2(\mu u_x)_x + (\mu u_y)_y + (\mu u_z)_z + (\lambda v_y)_x \\
&\quad + (\lambda w_z)_x + (\mu v_x)_y + (\mu w_x)_z + f_x, \\
\rho v_{tt} &= (\mu v_x)_x + (\lambda v_y)_y + 2(\mu v_y)_y + (\mu v_z)_z + (\mu u_y)_x \\
&\quad + (\lambda u_x)_y + (\lambda w_z)_y + (\mu w_y)_z + f_y, \qquad (2) \\
\rho w_{tt} &= (\mu w_x)_x + (\mu w_y)_y + (\lambda w_z)_z + 2(\mu w_z)_z + (\mu u_z)_x \\
&\quad + (\mu v_z)_y + (\lambda u_x)_z + (\lambda v_y)_z + f_z,
\end{aligned}$$

where, for example, $(\lambda v_y)_x = \partial(\lambda \partial v/\partial y)/\partial x$.

Finite-Difference Scheme

Let U^m_{ikl} be a discrete approximation to $u^m_{ikl} = u(x_i, y_k, z_l, t_m)$, where i, k, l are spatial indices and m is a time index. Similarly, let V^m_{ikl} and W^m_{ikl} approximate v^m_{ikl} and w^m_{ikl}, respectively. Let h be the spatial grid spacing in all three directions and Δt a time step. The approximation of the nonmixed derivative was suggested by Tikhonov and Samarski (e.g., Mitchell and Griffiths, 1980), and it has been used since Boore (1972) introduced it in seismology. Approximation of the mixed derivative is more problematic. Several ways were suggested (see, e.g., Kummer *et al.,* 1987; Sochacki *et al.,* 1991; Zahradník, 1995) and applied to the *P–SV* modeling. Theoretical analysis of consistency on material discontinuities and numerical tests showed very good accuracy of the *P–SV* scheme of Zahradník (Zahradník, 1995; Zahradník and Priolo, 1995; Moczo *et al.,* 1997). Therefore, we apply Zahradník's approach. For equations (2), we can obtain the

explicit second-order (in space and time) FD scheme. The scheme is given in the Appendix.

Combined Memory Optimization

Core Memory Optimization

Graves (1996) presented a good example of the core memory optimization for a fourth-order velocity-stress scheme on a staggered grid. While only a limited number of planes (e.g., *xz* planes) resides in core memory at one time, a maximum possible number of time updates is performed for these planes. The subset of planes repeatedly moves throughout the entire model space until the desired time window is computed. We apply such a memory optimization to our displacement scheme.

While the optimization significantly reduces problem with core memory, it may impose considerable disk memory requirements. This is because what had to be stored in core memory in the case with no optimization now has to be stored in disk. Disk memory requirements can reach easily tens of gigabytes if realistic large-scale model should be computed. There is, however, a possibility to reduce also disk memory requirements significantly by combining the foregoing core memory optimization with a disk memory optimization.

Disk Memory Optimization

A considerable amount of disk memory, say NBWF, is occupied by two time levels of the displacement values in all grid points: NBWF $= p \cdot 2 \cdot MX \cdot MY \cdot MZ \cdot 3$, where p is the number of bytes for the used real-number precision, 2 stands for two time levels, *MX, MY,* and *MZ* are the total numbers of grid points in the *x,y,* and *z* directions, respectively, and 3 stands for the three displacement components. These values are successively (plane by plane) and periodically overwritten as the subset of *NP xz* planes moves throughout the entire model space. It would be, therefore, reasonable to store and overwrite not directly $2 \cdot MX \cdot MZ \cdot 3$ displacement values for each plane but only $2 \cdot 3$ streams of zeros and ones that we could obtain as a result of data compression and that would occupy considerably less memory.

A two-dimensional array of the displacement-component values in one plane usually has a relatively very large information entropy, that is, no values are significantly more common than others. Usual compression methods, for example, a well-known adaptive Lempel-Ziv coding (UNIX command "compress"), in such a case fail. Therefore, we suggest to use the wavelet transform to decrease an information entropy in data and perform compression in the wavelet domain. We can call such an approach a wavelet compression.

Two-Dimensional Wavelet Transform

A one-dimensional (1D) continuous wavelet transform of a real function $f(t)$ with respect to an analyzing wavelet $\psi(t)$ is defined as

$$W(f, \psi; a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} \psi^* \left(\frac{t - b}{a}\right) f(t) dt. \quad (3)$$

Parameters a and b are called dilation and translation parameters, respectively, and ψ^* is a complex conjugate to ψ. A set of functions

$$\psi^{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi \left(\frac{t - b}{a}\right)$$

differing from one another in a and b values is the wavelet basis. In other words, the basis is obtained by dilation and translation of a single function $\psi(t)$. $W(a,b)$ for given values of a and b is the wavelet coefficient.

Function $f(t)$ can be reconstructed from its wavelet transform using the inverse wavelet transform

$$f(t) = C_\psi^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{da\, db}{a^2} W(a, b) \psi^{a,b}(t)$$

where $C_\psi = 2\pi \int_{-\infty}^{\infty} d\xi |\hat{\psi}(\xi)|^2 |\xi|^{-1} < \infty$ and $\hat{\psi}(\xi)$ denotes the Fourier transform of $\psi(\xi)$.

A discrete wavelet transform (DWT) can be obtained from definition (3) by restricting parameters a and b to discrete values $a = a_0^r$ and $b = sb_0a_0^r$, where r and s are integers and $a_0 > 1$, $b_0 > 0$ are some constants. The wavelet coefficients $W_{r,s}$ are then defined as

$$W_{r,s} = a_0^{-r/2} \int_{-\infty}^{\infty} f(t) \psi(a_0^{-r} t - sb_0) dt$$

and can be computed by hierarchical and fast scheme, a so-called subband filtering scheme (see, e.g., Daubechies, 1992; Press *et al.,* 1992). For most of the applications to the discrete data, the DWT is more convenient than the continuous transform.

A standard way of how to obtain the wavelet transform of a 2D array A_{ij} ($i = 1, \ldots, m$ and $j = 1, \ldots, n$) is to apply the 1D DWT twice. First, we apply the DWT to row vectors $A_{1j}, A_{2j}, \ldots, A_{mj}$ ($j = 1, \ldots, n$). As a result, we obtain a new array $\tilde{A}_{ij}^W$ ($i = 1, \ldots, m$ and $j = 1, \ldots, n$). Then we apply the DWT to column vectors $\tilde{A}_{i1}^W$, $\tilde{A}_{i2}^W, \ldots, \tilde{A}_{in}^W$, ($i = 1, \ldots, m$). The resulting array A_{ij}^W, ($i = 1, \ldots, m$ and $j = 1, \ldots, n$) is the array of the coefficients of the 2D wavelet transform of the original array A_{ij}.

The subband filtering scheme requires that a length of an input vector be 2^K; K being an integer. In practice, a spatial FD grid covering a given model of the medium has dimensions that are not equal to a power of 2. We found that adding zeros leads to poor compression. It is much better to divide array A_{ij} into M subarrays with dimensions $2^{K_\varepsilon} \times$

2^{L_ε} ($\varepsilon = 1, \ldots, M$) and to apply the two-dimensional wavelet transform to each of them separately.

The DWT enables to represent input data by a small number of the large-value wavelet coefficients that contain a substantial part of information, and by a large number of small or zero coefficients.

Thresholding, Quantization, and Encoding

Thresholding. Small coefficients with absolute values below the chosen threshold can be truncated to zero without any substantial distortion of input data. We choose the threshold THR as THR $= |W_{\max}|/PT$, where $W_{\max}$ is the wavelet coefficient with the maximum absolute value and PT is a parameter. The smaller PT, the better compression, but the more probable data distortion.

Quantization. After thresholding, we can map the real wavelet coefficients from the interval $\langle -W_{\max}, W_{\max}\rangle$ into integers from interval $\langle -2^{N-1} + 1, 2^{N-1} - 1\rangle$. Thus, the number of possible values of the coefficients is restricted. The maximum number of bits needed to represent any of the integer coefficients is N.

Encoding. The wavelet transformation, thresholding, and quantization decrease the information entropy of our data. Encoding translates the integer representation into a bit stream consisting of bit patterns. A chain of zeros is represented by bit value 0 followed by a bit pattern meaning the number of zeros in the chain. A bit pattern for any other number starts with the bit value 1, which is followed by a bit indicating a sign of the number and $N - 1$ bits representing an absolute value of the number.

Decoding. A bit stream is decoded into integer wavelet coefficients. If a bit pattern begins with a bit value 0, then the following bit pattern determines number of zeros in the chain. If the first bit of a bit pattern has value 1, the next N bits determine the value of a coefficient.

Dequantization. Integer wavelet coefficients are mapped into real values within interval $\langle -W_{\max}, W_{\max}\rangle$.

Algorithm of the Combined Optimization

The whole computation is performed during a certain number of passages of the subset of NP planes throughout the entire model space. One passage consists of three procedures that may be called (Graves, 1996) the ROLL-IN, CASCADE, and ROLL-OUT procedures.

ROLL-IN: First $NP + 1$ planes are read into core memory. The first time level is updated successively plane by plane starting from plane NP and ending with plane 1. The second time level is then updated successively for planes $NP - 1, \ldots, 1$. Next, time levels are updated until the $NP -$ 2nd time level is updated successively for planes 3, 2, and 1. Plane 1 (only displacement values) is then compressed and written into disk.

CASCADE: Displacements of plane 2 are compressed and written into disk. Displacements of plane $NP + 2$ in the compressed form are read in from disk and decompressed. Original displacement values of plane 2 in core memory are overwritten by decompressed displacements of plane $NP + 2$. Medium parameters of plane 2 residing in core memory are overwritten by parameters of plane $NP + 2$ that are read into core memory. Then, the first time level is updated for plane $NP + 1$, the second time level for plane NP, and so on, until the $NP -$ 2nd time level is updated for plane 4. The subset of NP planes moves throughout the entire model space, and the CASCADE computations are performed until the subset reaches plane $MY - 1$, where MY is the total number of xz planes.

ROLL-OUT: This is similar to the ROLL-IN procedure. At the end of the ROLL-OUT, all xz planes are updated up to $NP -$ 2nd time level. The sequence of the three procedures is then repeated until the desired number of time levels (which has to be a multiple of $NP - 2$) is computed.

Compression and Decompression. Compression is applied to each displacement component separately. While 2D DWT is applied to each of M subarrays, thresholding, quantization, and encoding are applied to the whole xz plane. A resulting bit stream and a wavelet coefficient with the maximum absolute value ($W_{\max}$) are written into disk. After a bit stream and $W_{\max}$ are read in from disk, the bit stream is decoded and dequantizated. An inverse 2D DWT is applied to each of M subarrays of the wavelet coefficients. As a result, displacement values are obtained.

Memory Requirements

Let the model be inhomogeneous between the free surface ($l = 1$) and horizontal grid plane with index $l = LIN$; $1 < LIN < MZ$, and homogeneous between plane LIN and bottom boundary ($l = MZ$), l being the spatial index in the z (downward vertical) direction. Let $COREM$ be the number of bytes of core memory and $DISKM$ the number of bytes of disk memory. Then memory needed to store material parameters plus displacements in two time levels are

No optimization

$$COREM = p\ MX\ MY[7\ (LIN - 1) + 6\ MZ] \quad (4a)$$

$$DISKM = 0 \quad (4b)$$

Core memory optimization

$$COREM = p\ MX\ NP[7\ (LIN - 1) + 6\ MZ] \quad (5a)$$

$$DISKM = p\ MX\ MY[7\ (LIN - 1) + 6\ MZ] \quad (5b)$$

Combined memory optimization

$$COREM = p\ MX\ NP[7\ (LIN - 1) + 6\ MZ] \quad (6a)$$

$$DISKM = p\ MX\ MY[7\ (LIN - 1) + 6\ MZ/CR] \quad (6b)$$

CR is compression ratio defined as NBWF/NBWF^W, where NBWF $= p \cdot 2 \cdot MX \cdot MY \cdot MZ \cdot 3$ is the number of bytes of disk memory occupied by displacement values (without compression) and NBWF^W the number of bytes occupied by streams of zeros and ones representing displacement values after the wavelet compression. Because the wavelet compression is applied separately to each displacement component at one plane at one time level, the compression ratio CR in equations (6) represents, in fact, a minimum of compression ratios for all displacement components, planes, and time levels. Note that formulas (4) to (6) give the upper estimates since we assume that, theoretically, the medium in its inhomogeneous part can change point to point.

Test Computations

Accuracy of the Finite-Difference Scheme

First, we tested accuracy of the FD scheme without applying wavelet compression. We made numerical comparisons of the FD synthetics with those obtained by the discrete-wavenumber (DWN) method (Bouchon, 1981; computer code by O. Coutant). From numerous computations, we present here some examples for models of a homogenous half-space (HH, HP1, HP2, and HP3) and a single layer over half-space (SL1, SL2, SEN1, and SEN2). Their parameters are given in Table 1. Source was simulated using a body-force term and method suggested by Frankel (1993). Gabor signal, $s(t) = \exp\{-[\omega(t - t_s)/\gamma]^2\} \cos[\omega(t - t_s) + \theta]$, was used as a source time function. Here, $\omega = 2\pi f_p$, $t \in \langle 0, 2t_s \rangle$, f_p is predominant frequency, γ controls the width of the signal, θ is a phase shift, and $t_s = 0.45\gamma/f_p$. The source parameters are given in Table 2.

Table 1
Model Parameters

Model	v_P (m/sec)	v_S (m/sec)	ρ (kg/m^3)	H (m)	h_S (m)	h (m)	Δt (sec)	f_{ac} (Hz)
HH	4000	2300	1800	∞	2511	93	0.018	2.06
SL1	2600	1400	1500	1100	2585	55	0.010	2.12
	4000	2300	1800	∞				
SL2	2600	1400	1500	1100	385	55	0.010	2.12
	4000	2300	1800	∞				
HP1	1600	400	1700	∞	128	16	0.008	2.08
HP2	1200	400	1700	∞	128	16	0.010	2.08
HP3	800	400	1700	∞	128	16	0.005	2.08
SEN1	1125	625	1600	200	300	50	0.007	1.04
	5468	3126	1800	∞				
SEN2	1125	625	1600	225	300	50	0.007	1.04
	5468	3126	1800	∞				

v_P, *P*-wave velocity; v_S, *S*-wave velocity; ρ, density; H, thickness of a layer; h_s, source depth; h, grid spacing; Δt, time step; f_{ac} ($= v_{S,\min}/12h$), the frequency up to which a computation should be sufficiently accurate.

Figure 1 compares velocigrams calculated by the FD and DWN methods for model HH. Figures 2 and 3 show velocigrams for models SL1 and SL2, respectively. While the source is located in the half-space in the case of model SL1, it is inside the layer in model SL2. The source inside the

Table 2
Source Parameters

M_0 (kg m^2/sec^2)	Φ_S (deg)	δ (deg)	λ (deg)	γ	f_P	θ	t_s
10^{16}	0	45	90	0.5	0.225	0.0	1.0

M_0, scalar seismic moment; Φ_S, strike; δ, dip; λ, rake; γ, f_P, θ and t_s, parameters of Gabor signal.

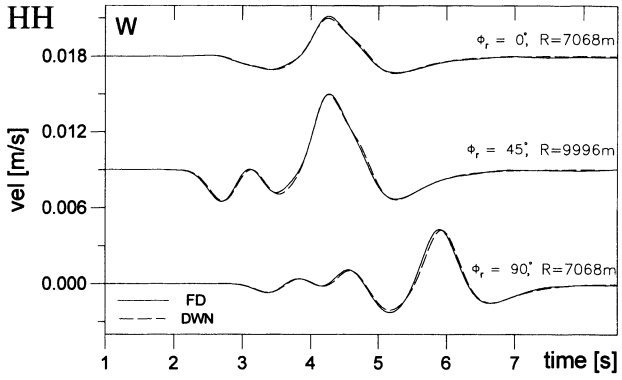

Figure 1. Comparison of the finite-difference (FD) and discrete-wavenumber (DWN) velocigrams (vertical-component *W*) on the free surface of the homogeneous half-space (HH). ϕ_r is the receiver azimuth, R is the epicentral distance.

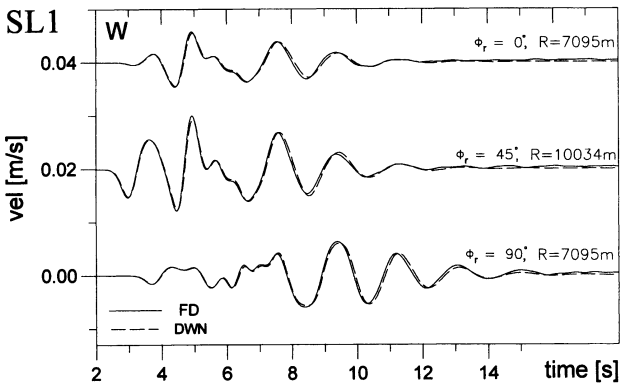

Figure 2. Similar comparison as in Figure 1 but for the single layer over half-space (SL1) with the source located in the half-space.

layer is efficient in exciting surface waves. Propagation of relatively intensive Rayleigh waves in model SL2 serves mainly as a severe test of the free-surface simulation. All three figures show a very good agreement between the FD and DWN solutions.

It is well known that the displacement schemes generally have problems in media with higher Poisson's ratio. This is illustrated in Figure 4, which shows the FD and DWN velocigrams for three models of a homogeneous half-space, HP1, HP2, and HP3. They only differ in the *P*-wave velocity and consequently by Poisson's ratio. The v_P/v_S ratios are 4, 3, and 2 in models HP1, HP2, and HP3, respectively. It is clear from Figure 4 that accuracy of the FD velocigrams decreases as the v_P/v_S ratio increases. These and other calculations show that the FD scheme gives sufficiently accurate results for $v_P/v_S \leq 2$.

The last numerical example demonstrates remarkable property of our displacement FD scheme—sensitivity to the position of a material discontinuity. In the first model of a single layer over half-space, SEN1, the material discontinuity (i.e., interface between the layer and half-space with the velocity contrast 1:5) lies exactly on the fifth horizontal

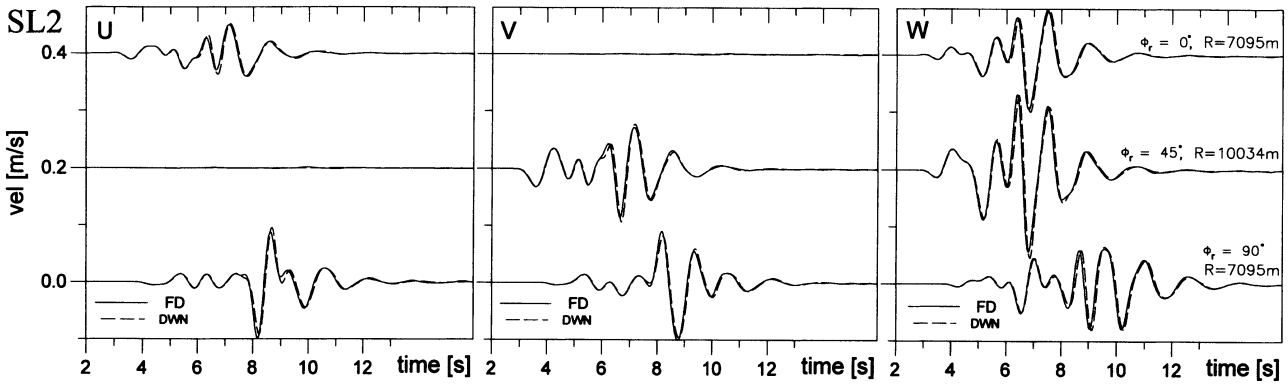

Figure 3. Comparison of the FD and DWN velocigrams on the free surface of the single layer over half-space (SL2) with the source located inside the layer and exciting intensive Rayleigh waves. *U* and *V* are the horizontal components; *W* is the vertical component.

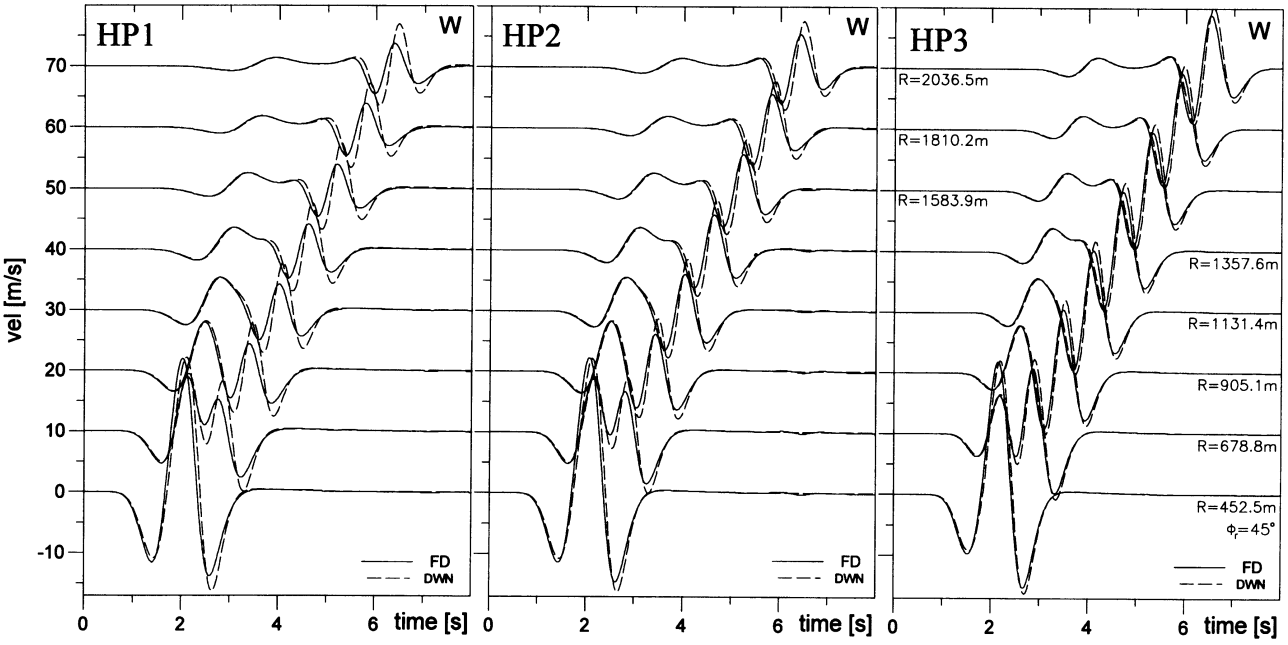

Figure 4. Comparison of the FD and DWN velocigrams (*W* component) in three models of a homogeneous half-space (HP1, HP2, HP3) differing in the v_P/v_S ratio. Accuracy of the FD velocigrams decreases for $v_P/v_S > 2$.

grid plane (see Fig. 5). The point dislocation source is located in the half-space. The FD and DWN synthetics for the model are shown in Figure 5 by the thick solid and dashed lines, respectively. The agreement between the two methods is good. Let us note that the agreement between the two methods would be better for lower layer–half-space velocity contrast and slightly worse for a higher contrast.

The second model, SEN2, slightly differs from SEN1: The thickness of the layer is larger by half grid spacing, which means that the layer–half-space interface (material discontinuity) is now located exactly halfway between the fifth and sixth horizontal grid planes. Without any change of the spatial grid, our scheme is capable to reflect the new position of the material discontinuity. The FD and DWN synthetics for model SEN2 are shown in Figure 5 by the thin solid and dashed lines, respectively. As far as we know, none of the velocity-stress and displacement-stress schemes was shown to be capable to distinguish between the two (SEN1 and SEN2) positions of the material discontinuity without changing the spatial grid. At the same time, it is clear from Figure 5 that the synthetics for the two models differ considerably.

Wavelet Compression

In numerous numerical calculations, we tested performance of the wavelet compression. The effects of the wavelet compression on synthetics are illustrated on the examples for model SL1. Figure 6 shows the effect of the wavelet compression on the synthetics as well as values of compression ratio (successively for all performed compressions) during the computations for two different threshold ratios $1/PT$ and fixed number of NP planes in a moving subset of planes. For example, threshold ratio $1/PT = 1/2000$ means that all wavelet coefficients whose absolute values are smaller than 1/2000 of the maximum absolute value (at a given plane and time level) are truncated to zero. It is obvious that the minimum compression ratio CR for the threshold ratio $1/PT = 1/4000$ is smaller than that for $1/PT = 1/2000$. It is clear from Figure 6 that the distortion of the synthetics is smaller for the smaller threshold ratio.

Figure 7 shows the effect of the wavelet compression on synthetics in the case of the larger (than in Fig. 6) number NP of the planes in the moving subset of planes and for four values of the threshold ratio $1/PT$. The larger NP means smaller number of performed compressions and consequently smaller distortion of the synthetics. Practically no distortion (in other words, the agreement within thickness of the line) is reached in the case of $1/PT = 1/16{,}000$. Let us note, however, that practically the same level of agreement is also reached with $NP = 12$ if $1/PT = 1/16{,}000$ (not shown in Fig. 6).

Reductions of core and disk memory in the foregoing computations for the simple model of the layer over half-space, SL1, are clear from the following comparison of core and disk memory requirements (*COREM* and *DISKM*):

No optimization
COREM = 310.3 MB
DISKM = 0 MB

Core memory optimization
COREM = 11.3 MB ($NP = 12$), 142.5 MB ($NP = 152$)
DISKM = 310.3 MB

Combined memory optimization
COREM = 11.3 MB ($NP = 12$), 142.5 MB ($NP = 152$)
DISKM = 18.3 MB (minimum CR = 17)

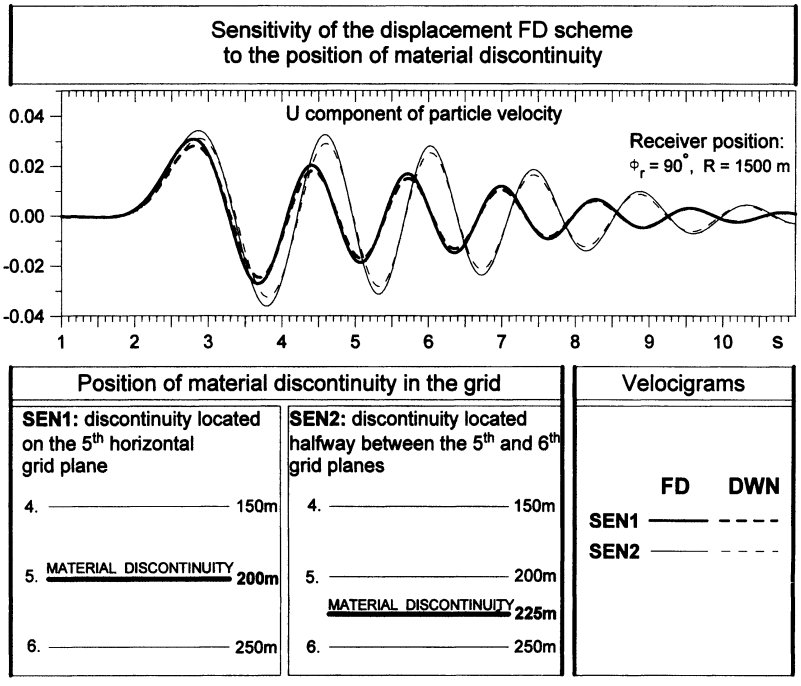

Figure 5. The FD and DWN velocigrams for two models of a single layer over half-space, SEN1 and SEN2. The figure demonstrates the capability of the displacement FD scheme to account for the position of the material discontinuity (layer–half-space interface). In the SEN2 model, the discontinuity is located halfway between the two horizontal grid planes.

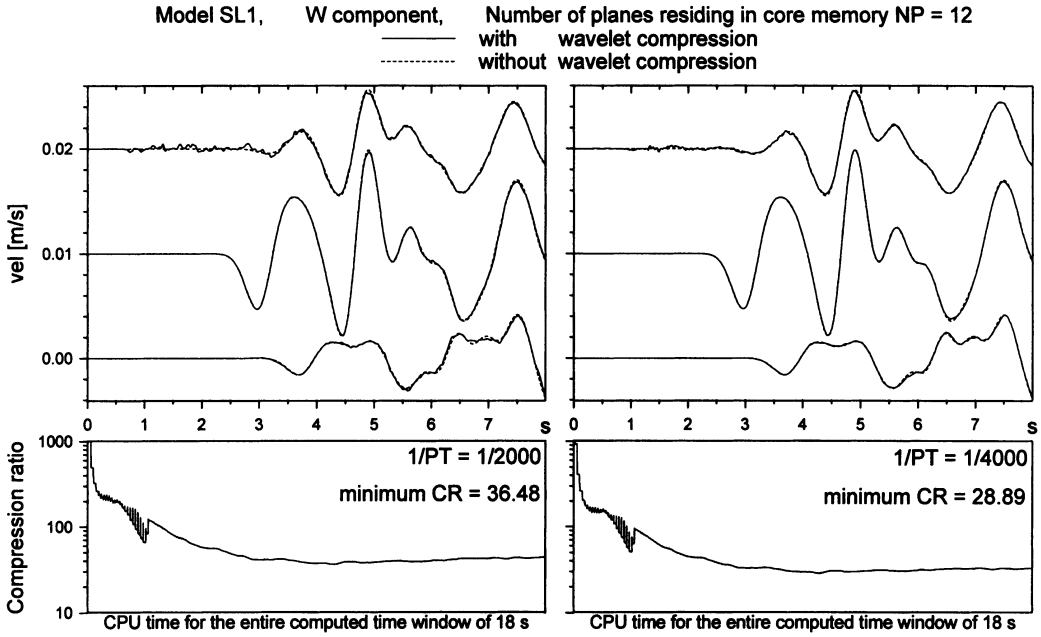

Figure 6. Effects of the wavelet compression on the FD velocigrams. FD velocigrams calculated with the wavelet compression (solid line) are compared with the reference FD velocigrams calculated without the wavelet compression. Variation of the compression ratio is also shown. Two different threshold ratios ($1/PT$ = 1/2000, 1/4000) led to two different minimum compression ratios CR (36.48 and 28.89).

Results of our tests can be summarized as follows:

a. Dimensions of subarrays ($2^{K_\varepsilon}, 2^{L_\varepsilon}; \varepsilon = 1, \ldots, M$) should be larger than 2^5 in order to achieve a reasonable compression.
b. Because compression is performed after every $NP - 2$ time levels for each xz plane, it is advantageous to use as large NP as possible.
c. Compression is less sensitive to the choice of a wavelet basis than to parameters PT (which determines threshold) and N (which determines interval of integers that wavelet coefficients are mapped into). Assuming not-too-small NP, PT should be larger than 4000. N has to be larger than $1 + \log_2 PT$; otherwise, an additional thresholding is introduced.
d. It is important (and useful) that any distortion of the seismograms due to inappropriate choice of the compression parameters was clearly recognizable in all our numerical experiments. Moreover, distortion of the synthetics can be easily removed by filtering. This property is very important because it allows the use of relatively small NP and relatively large $1/PT$, if necessary.
e. The increase of the CPU time due to one passage of the subset of planes with compression was always smaller than 0.75% of the time necessary for one passage without compression. The increase is proportional to the number of passages, that is, inversely proportional to NP.

Conclusions

1. We presented a 3D finite-difference scheme that is based on the displacement formulation of the equation of motion. The scheme is second-order accurate both in time and space. Accuracy of the scheme was tested through numerical comparisons with the discrete-wavenumber method. The scheme gives sufficiently accurate results for $v_P/v_S < 2$. The scheme is capable to account for the position of the internal material discontinuity more accurately than the recent velocity-stress and displacement-stress schemes.
2. We have developed a combined memory optimization (CDMO) for the 3D finite-difference modeling of seismic-wave propagation and earthquake ground motion. CDMO comprises core and disk memory optimizations and significantly reduces both core and disk memory requirements. Core memory optimization is based on keeping only a limited number of model planes in a core memory at one time and performing the maximum possible time updates for these planes. Disk memory optimization is based on data compression in the wavelet domain. CDMO is in no way restricted to the particular displacement finite-difference scheme. CDMO is general and applicable to any second- or fourth-order finite-difference scheme on conventional or staggered grid.

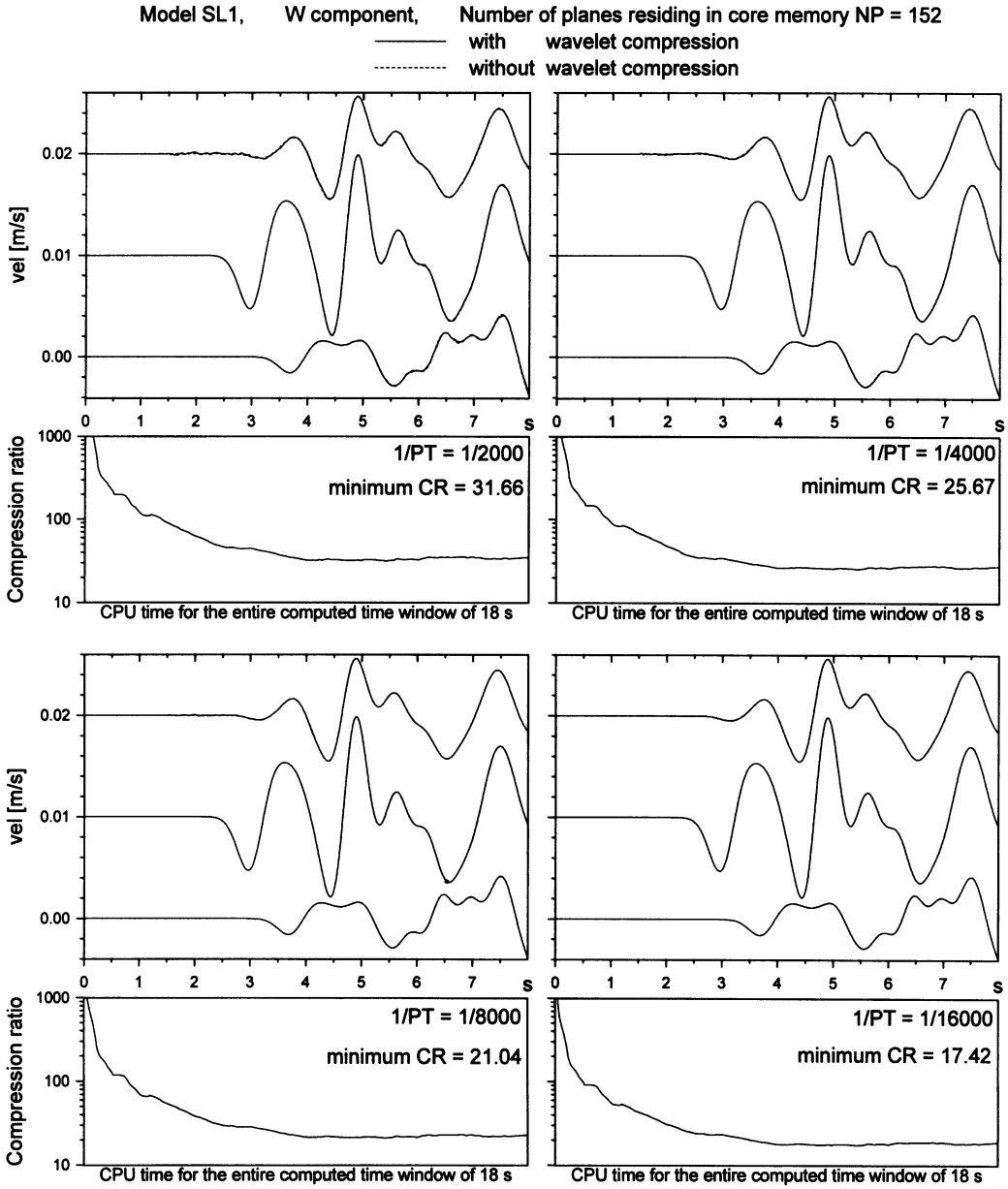

Figure 7. Similar comparison as in Figure 6 but for the larger number of planes residing in core memory ($NP = 152$) and four different threshold ratios $1/PT$.

3. Application of CDMO to the fourth-order schemes on nonuniform grids should be, in our opinion, the next step in order to increase efficiency of the finite-difference modeling.
4. Because inclusion of realistic attenuation in the finite-difference simulations requires large additional memory, CDMO makes such inclusion more affordable.

Acknowledgments

The reviews by K. B. Olsen and an anonymous reviewer helped us to improve the article. This work was supported in part by Grant 2/5131/98, VEGA, Slovak Republic, Grant 205/96/1743, Grant Agency of Czech Republic, and INCO-COPERNICUS Grant PL963311. Part of the numerical computations was performed on the SGI CRAY Origin 2000 parallel computer during a visit of one of us (M. L.) in Parallab, University of Bergen, Norway. The visit was supported by the Research Council of Norway.

References

Aboudi, J. (1971). Numerical simulation of seismic sources, *Geophysics* **36,** 810–821.

Alterman, Z. S. and D. Loewenthal (1970). Seismic waves in a quarter and three-quarter plane, *Geophys. J. R. Astr. Soc.* **20,** 101–126.

Aoi, S. and H. Fujiwara (1998). 3-D finite-difference method using discontinuous grids, *Bull. Seism. Soc. Am.*, submitted.

Boore, D. M. (1972). Finite difference methods for seismic wave propagation in heterogeneous materials, in *Methods in Computational Physics 11,* B. A. Bolt (Editor), Academic, New York, 1–38.

Bouchon, M. (1981). A simple method to calculate Green's functions for elastic layered media, *Bull. Seism. Soc. Am.* **71,** 959–971.

Daubechies, I. (1992). *Ten Lectures on Wavelets,* SIAM, Philadelphia.

Frankel, A. (1993). Three-dimensional simulations of ground motions in the San Bernardino Valley, California, for hypothetical earthquakes on the San Andreas fault, *Bull. Seism. Soc. Am.* **83,** 1020–1041.

Graves, R. W. (1993). Modeling three-dimensional site response effects in the Marina district basin, San Francisco, California, *Bull. Seism. Soc. Am.* **83,** 1042–1063.

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

Higdon, R. L. (1991). Absorbing boundary conditions for elastic waves, *Geophysics* **56,** 231–241.

Jastram, C. and A. Behle (1992). Acoustic modeling on a grid of vertically varying spacing, *Geophys. Prosp.* **40,** 157–170.

Kummer, B., A. Behle, and F. Dorau (1987). Hybrid modeling of elastic-wave propagation in two-dimensional laterally inhomogeneous media, *Geophysics* **52,** 765–771.

Levander, A. (1988). Fourth-order finite-difference P-SV seismograms, *Geophysics* **53,** 1425–1436.

Luo, Y. and G. Schuster (1990). Parsimonious staggered grid finite-differencing of the wave equation, *Geophys. Res. Lett.* **17,** 155–158.

Mikumo, T. and T. Miyatake (1987). Numerical modeling of realistic fault rupture processes, In *Seismic Strong Motion Synthetics,* B. A. Bolt (Editor), Academic, New York, 91–151.

Mitchell, A. R. and D. F. Griffiths (1980). *The Finite Difference Method in Partial Differential Equations,* Wiley, New York.

Moczo, P. (1989). Finite-difference technique for SH-waves in 2-D media using irregular grids—application to the seismic response problem, *Geophys. J. Int.* **99,** 321–329.

Moczo, P., P. Labák, J. Kristek, and F. Hron (1996). Amplification and differential motion due to an antiplane 2D resonance in the sediment valleys embedded in a layer over the half-space, *Bull. Seism. Soc. Am.* **86,** 1434–1446.

Moczo, P., E. Bystrický, J. Kristek, J. M. Carcione, and M. Bouchon (1997). Hybrid modeling of P-SV seismic motion at inhomogeneous viscoelastic topographic structures, *Bull. Seism. Soc. Am.* **87,** 1305–1323.

Ohminato, T. and B. A. Chouet (1997). A free-surface boundary condition for including 3D topography in the finite-difference method, *Bull. Seism. Soc. Am.* **87,** 494–515.

Olsen, K. and G. T. Schuster (1992). Seismic hazard analysis in Salt Lake Valley by finite-difference simulation of three dimensional elastic wave propagation, in *Scientific Excellence in High Performance Computing: The 1990 IBM Price Papers,* vol. I, sec. 6, Baldwin Press, Athens, Georgia, 135–165.

Olsen, K. B., R. J. Archuleta, and J. R. Matarese (1995). Magnitude 7.75 earthquake on the San Andreas fault: three-dimensional ground motion in Los Angeles, *Science* **270,** 1628–1632.

Pitarka, A. (1999). 3D elastic finite-difference modeling of seismic motion using staggered grids with nonuniform spacing, *Bull. Seism. Soc. Am.* **89,** 54–68.

Pitarka, A., K. Irikura, T. Iwata, and H. Sekiguchi (1998). Three-dimensional simulation of the near-fault ground motion for the 1995 Hyogoken Nanbu (Kobe), Japan, earthquake, *Bull. Seism. Soc. Am.* **88,** 428–440.

Press, W. H., S. A. Teukolsky, W. T. Wetterling, and B. P. Flannery (1992). *Numerical Recipies in Fortran: The Art of Scientific Computing,* Cambridge Press, New York.

Sochacki, J. S., J. H. George, R. E. Ewing, *et al.* (1991). Interface conditions for acoustic and elastic wave propagation, *Geophysics* **56,** 168–181.

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

Wald, D. and R. W. Graves (1998). The seismic response of the Los Angeles Basin, California, *Bull. Seism. Soc. Am.* **88,** 337–356.

Yomogida, K. and J. T. Etgen (1993). 3-D wave propagation in the Los Angeles basin for the Whittier-Narrows earthquake, *Bull. Seism. Soc. Am.* **83,** 1325–1344.

Zahradník, J. (1995). Simple elastic finite-difference scheme, *Bull. Seism. Soc. Am.* **85,** 1879–1887.

Zahradník, J. and E. Priolo (1995). Heterogeneous formulations of elastodynamic equations and finite-difference schemes, *Geophys. J. Int.* **120,** 663–676.

Zahradník, J., P. Moczo, and F. Hron (1993). Testing four elastic finite-difference schemes for behaviour at discontinuities, *Bull. Seism. Soc. Am.* **83,** 107–129.

Appendix

Finite-Difference Scheme for Interior Grid Points

$$
\begin{aligned}
U_{ikl}^{m+1} &= 2U_{ikl}^{m} - U_{ikl}^{m-1} \\
&\quad + \frac{\Delta^2 t}{\rho_{ikl}} [L_{xx}(\lambda, U) + 2\, L_{xx}\,(\mu, U) + L_{yy}\,(\mu, U) \\
&\quad + L_{zz}(\mu, U) + L_{yx}(\lambda, V) + L_{zx}(\lambda, W) \\
&\quad + L_{xy}(\mu, V) + L_{xz}(\mu, W) + F_{ikl}^{x,m}], \\
V_{ikl}^{m+1} &= 2V_{ikl}^{m} - V_{ikl}^{m-1} \\
&\quad + \frac{\Delta^2 t}{\rho_{\mathrm{ikl}}} [L_{xx}\,(\mu, V) + L_{yy}\,(\lambda, V) + 2L_{yy}\,(\mu, V) \\
&\quad + L_{zz}\,(\mu, V) + L_{yx}\,(\mu, U) + L_{zy}\,(\lambda, U) \\
&\quad + L_{zy}\,(\lambda, W) + L_{yz}\,(\mu, W) + F_{iki}^{y,m}], \\
W_{ikl}^{m+1} &= 2W_{ikl}^{m} - W_{ikl}^{m-1} \\
&\quad + \frac{\Delta^2 t}{\rho_{ikl}} [L_{xx}\,(\mu, W) + L_{yy}\,(\mu, W) + L_{zz}\,(\lambda, W) \\
&\quad + 2L_{zz}\,(\mu, W) + L_{zx}\,(\mu, U) + L_{zy}\,(\mu, V) \\
&\quad + L_{xz}\,(\lambda, U) + L_{yz}\,(\lambda, V) + F_{ikl}^{z,m}],
\end{aligned}
\tag{A1}
$$

where operator $L_{\gamma\gamma}\,(a, \Phi)$, ($\gamma \in \{x, y, z\}$, $a \in \{\lambda, \mu\}$, $\Phi \in \{U, V, W\}$), has the form

$$
L_{\gamma\gamma}\,(a, \Phi) = \frac{1}{\mathrm{h}^2} [a^{\gamma}\,(\Phi_{+}^{m} - \Phi^{\mathrm{m}}) - a_{-}^{\gamma}\,(\Phi^{m} - \mathrm{gF};_{-}^{m})]
$$

and subscripts $\pm$ stand for $i \pm 1kl$, $ik \pm 1l$ or $ikl \pm 1$ if $\gamma = x$, y, or z, respectively. No subscript means ikl. The effective parameters a^{γ} and a_{-}^{γ} are defined as

$$
a^{\gamma} = h \left[\int_{\gamma_{\mathrm{n}}}^{\gamma_{\mathrm{n}+1}} \frac{d\gamma}{a} \right]^{-1} \quad \text{and} \quad a_{-}^{\gamma} = h \left[\int_{\gamma_{\mathrm{n}-1}}^{\gamma_{\mathrm{n}}} \frac{d\gamma}{a} \right]^{-1}, \tag{A2}
$$

where n stands for i, k, or l if $\gamma = x$, y, or z, respectively. Operator $L_{\gamma\eta}$, $\gamma \neq \eta$ and $\gamma, \eta \in \{x, y, z\}$, has the form

$$L_{\gamma\eta}(a, \Phi) = \frac{1}{4h^2}\Big[a^{\eta}(\Phi^m_{2_+} + \Phi^m_{3_+} - \Phi^m_{2_-} - \Phi^m_{3_-}) - a^{\eta}_{-}(\Phi^m_{1_+} + \Phi^m_{2_+} - \Phi^m_{1_-} - \Phi^m_{2_-})\Big],$$

where subscripts $1_\pm$, $2_\pm$, and $3_\pm$ stand for indices as follows:

	xz	xy	yz	zx	yx	zy
$1_\pm$	$i \pm 1kl-1$	$i \pm 1k-1l$	$ik \pm 1l-1$	$i-1kl \pm 1$	$i-1k \pm 1l$	$ik-1l \pm 1$
$2_\pm$	$i \pm 1kl$	$i \pm 1k\ \ l$	$ik \pm 1l$	$i\ \ kl \pm 1$	$i\ \ k \pm 1l$	$ik\ \ l \pm 1$
$3_\pm$	$i \pm 1kl+1$	$i \pm 1k+1l$	$ik \pm 1l+1$	$i+1kl \pm 1$	$i+1k \pm 1l$	$ik+1l \pm 1$

If the interface between two layers/blocks of media lies exactly on a grid plane, the corresponding effective parameters have to be evaluated according to formulas (A2) for an arithmetic average of the elastic coefficients in the two layers/blocks of media. This follows from both theoretical analysis of consistency of the displacement schemes (Zahradník *et al.,* 1993; Zahradník and Priolo, 1995) and numerical tests (e.g., Zahradník and Priolo, 1995; Moczo *et al.,* 1997).

Aboudi (1971) found sufficient stability condition for the displacement FD scheme in a homogeneous medium:

$$\Delta t \leqq h/\sqrt{2(\alpha^2 + 2\beta^2)}.$$

As follows from numerical computations of other investigators (e.g., Mikumo and Miyatake, 1987) and ours, stability is satisfied under less restrictive condition

$$\Delta t \leqq h/\sqrt{\alpha^2 + \beta^2},$$

that is, stability condition for the *P–SV* displacement scheme (Alterman and Loewenthal, 1970).

Second-order schemes on both conventional and staggered grids require at least 10 grid spacings per minimum wavelength, which is to be propagated without significant grid dispersion. It follows from our numerical experiments that 12 grid spacings per minimum wavelength is a good choice.

Various absorbing boundary conditions can be used with the FD scheme. Numerical experience suggests that there is no best absorbing condition, and the user of the FD method should be ready to use an alternative condition if some condition does not give good results. In our computations, we used Higdon's (1991) condition.

Finite-Difference Scheme for Grid Points on the Flat Free Surface

Scheme (A1) can be used also for the flat free surface. Let point ikl be located on the free surface. Then operators L_{xx}, L_{yy}, L_{xy}, and L_{yx} have the same form as the operators for the interior grid points. The other operators follow from the application of the vacuum formalism ($\lambda = 0$ and $\mu = 0$ above the free surface) to the second type of approximation of the mixed derivative suggested by Zahradník (1995). They are

$$L_{zz}(a, \Phi) = \frac{1}{h^2} a^z_{ikl}(\Phi^m_{ikl+1} - \Phi^m_{ikl}),$$

$$L_{z\eta}(a, \Phi) = \frac{1}{4h^2}[a^{\eta}(\Phi^m_{2_+} + \Phi^m_{3_+} - \Phi^m_{2} - \Phi^m_{3}) - a^{\eta}_{-}(\Phi^m_{1_+} + \Phi^m_{2_+} - \Phi^m_{1} - \Phi^m_{2})],$$

and

$$L_{\eta z}(a, \Phi) = \frac{1}{4h^2}[a^{z}(\Phi^m_{3} + \Phi^m_{3_+} - \Phi^m_{2} - \Phi^m_{2_+}) + a^{z}_{-}(\Phi^m_{2} + \Phi^m_{2_+} - \Phi^m_{1} - \Phi^m_{1_+})],$$

where subscripts 1, 2, 3, 1_+, 2_+, and 3_+ stand for indices as follows:

	$\eta = x$	$\eta = y$		$\eta = x$	$\eta = y$
1	$i - 1kl$	$ik - 1l$	1_+	$i - 1kl + 1$	$ik - 1l + 1$
2	$i\ \ kl$	$ik\ \ l$	2_+	$i\ \ kl + 1$	$ik\ \ l + 1$
3	$i + 1kl$	$ik + 1l$	3_+	$i + 1kl + 1$	$ik + 1l + 1$

Effective parameters are $a^{\eta} = a^x_{ikl+1/2}$, $a^{\eta}_{-} = a^x_{i-1kl+1/2}$, $a^z = a^z_{i+1/2kl}$, and $a^z_{-} = a^z_{i-1/2kl}$, if $\eta = x$, and $a^{\eta} = a^y_{ikl+1/2}$, $a^{\eta}_{-} = a^y_{ik-1l+1/2}$, $a^z = a^z_{ik+1/2l}$, and $a^z_{-} = a^z_{ik-1/2l}$, if $\eta = y$.

As follows from both theoretical analysis of consistency of the displacement schemes (Zahradník *et al.,* 1993; Zahradník and Priolo, 1995) and numerical tests (e.g., Zahradník and Priolo, 1995; Moczo *et al.,* 1997), only half-values of density and Lamé's coefficients along the free surface have to be considered:

1. Only half-value of density ρ has to be considered in schemes (A1) when applied to the grid points on the free surface.
2. Only half-values of parameters a^x and a^y (a being either λ or μ) have to be considered in the operators L_{xx}, L_{yy}, L_{xy}, and L_{yx} when applied to the grid points on the free surface.

Geophysical Institute
Slovak Academy of Sciences
Dúbravská cesta 9
842 28 Bratislava, Slovak Republic

Manuscript received 13 April 1998.