# Acoustic wave-equation traveltime and waveform inversion of crosshole seismic data

Changxi Zhou*, Wenying Cai*, Yi Luo‡, Gerard T. Schuster*, and Sia Hassanzadeh§

## ABSTRACT

A hybrid wave-equation traveltime and waveform inversion method is presented that reconstructs the interwell velocity distribution from crosshole seismic data. This inversion method, designated as WTW, retains the advantages of both full wave inversion and traveltime inversion; i.e., it is characterized by reasonably fast convergence which is somewhat independent of the initial model, and it can resolve detailed features of the velocity model. In principle, no traveltime picking is required and the computational cost of the WTW method is about the same as that for full wave inversion.

We apply the WTW method to synthetic data and field crosshole data collected by Exxon at their Friendswood, Texas, test site. Results show that the WTW tomograms are much richer in structural information relative to the traveltime tomograms. Subtle structural features in the WTW Friendswood tomogram are resolved to a spatial resolution of about 1.5 m, yet are smeared or completely absent in the traveltime tomogram. This suggests that it might be better to obtain high quality (distinct reflections) crosshole data at intermediate frequencies, compared to intermediate quality data (good quality first arrivals, but the reflections are buried in noise) at high frequencies.

Comparison of the reconstructed velocity profile with a log in the source well shows very good agreement within the O-200 m interval. The 200-300 m interval shows acceptable agreement in the velocity fluctuations, but the tomogram's velocity profile differs from the sonic log velocities by a DC shift. This highlights both the promise and the difficulty with the WTW method; it can reconstruct both the intermediate and high wavenumber parts of the model, but it can have difficulty recovering the very low wavenumber parts of the model.

## INTRODUCTION

Among the various seismic inversion methods there are two extremes, traveltime inversion (Dines and Lytle, 1979; Paulsson et al., 1985; Ivansson, 1985; Bishop et al., 1985; Lines, 1988; and many others) and full wave inversion (Tarantola, 1986, 1987; Mora, P. 1987; Crase et al., 1992; and others). In traveltime tomography, the time of flight information is inverted for the smooth features of the velocity model, while waveform inversion inverts the amplitude and phase information for the fine details of the earth model. Both methods have complementary strengths and weaknesses.

A weakness of traveltime inversion is that it employs a high frequency approximation and so it can fail when the earth's velocity variations have nearly the same wavelength as does the source wavelet. In addition, the resolution of the reconstructed model from traveltime inversion is much less than that from full wave inversion. On the other hand, the traveltime misfit function (sum of squared errors between observed and calculated traveltimes) can be quasi-linear (Luo and Schuster, 1991b) with respect to the normed difference between the starting and actual velocity models. This means that successful inversion can be achieved even if the starting model is far from the actual model.

The characteristics of full wave inversion are complementary to those of traveltime inversion. While sensitive to the choice of starting model or noisy amplitudes, full wave inversion can sometimes reconstruct a highly resolved earth model. This is because there are no high-frequency assump-

tions about the data, and almost all seismic events are embedded in the misfit function. The problem with full wave inversion, however, is that its misfit function (normed squared error between the observed and synthetic seismograms) can be highly nonlinear with respect to the velocity model (Gauthier et al., 1986; Luo and Schuster, 1991b). In this case, a gradient method will tend to get stuck in local minima if the starting model is too far from the actual model.

To exploit the strengths and ameliorate the weaknesses of both traveltime and waveform inversion, this paper presents a hybrid inversion method that minimizes a weighted combination of first arrival traveltime (Luo and Schuster, 1991a) and seismogram (Tarantola, 1987) misfit functions. The main benefits are a convergence rate that is somewhat insensitive to the starting model, a high degree of model resolution, no high-frequency approximations to the data, and a robustness in the presence of data noise. Synthetic and real data tests show that this hybrid method, designated wave-equation traveltime and waveform inversion (WTW), provides a tomogram with both the smooth and the detailed parts of the velocity model.

## THEORY

The following analysis assumes that the propagation of seismic waves honors the 2-D acoustic wave equation. Let $p(\mathbf{x}_r, t|\mathbf{x}_s)_{obs}$ be the observed pressure at time t observed at the receiver location $\mathbf{x}_r(r = 1, 2, \ldots, N_r)$ as a result of a source at $\mathbf{x}_s(s = 1, 2, \ldots, N_s)$. The source is always assumed to be initiated at zero time and this time variable will be suppressed in our notation. For a given velocity model, $p(\mathbf{x}_r, t|\mathbf{x}_s)_{cal}$ denotes the calculated seismogram that honors the 2-D wave equation

$$\frac{1}{\kappa(\mathbf{x}_r)} \frac{\partial^2 p(\mathbf{x}_r, t|\mathbf{x}_s)}{\partial t^2} - \nabla \cdot \left[ \frac{1}{\rho(\mathbf{x}_r)} \nabla p(\mathbf{x}_r, t|\mathbf{x}_s) \right]$$
$$= s(\mathbf{x}_r, t|\mathbf{x}_s), \quad (1)$$

where $\rho(\mathbf{x}_r)$ is the density, $\kappa(\mathbf{x}_r)$ is the bulk modulus, and $s(\mathbf{x}_r, t|\mathbf{x}_s)$ is the source function at $x_s$. The forward modeling problem is defined as finding everywhere the pressure field that satisfies equation (1) with the given boundary + initial conditions. In practice, we find this pressure field by a fourth-order finite-difference solution (Levander, 1988) to the first-order equations of motion.

We will define the inverse problem as finding the velocity model $c(\mathbf{x}) = \sqrt{\kappa(\mathbf{x})/\rho(\mathbf{x})}$ that both predicts the observed seismograms $p(\mathbf{x}_r, t|\mathbf{x}_s)_{obs}$ and minimizes the following hybrid misfit function:

$$S = \frac{1}{2} \sum_s \sum_r [\delta\tau_{rs}]^2 + \frac{1}{2} \sum_s \sum_r \int dt \, \delta p_{rs}(t) w \delta p_{rs}(t). \quad (2)$$

Here $\delta p_{rs}(t)$ is the seismogram residual

$$\delta p_{rs}(t) = p(\mathbf{x}_r, t|\mathbf{x}_s)_{obs} - p(\mathbf{x}_r, t|\mathbf{x}_s)_{cal}, \quad (3)$$

and $\delta\tau_{rs} = \tau_{obs}(\mathbf{x}_r, \mathbf{x}_s) - \tau_{cal}(\mathbf{x}_r, \mathbf{x}_s)$ is the traveltime residual, or the difference between the observed and calculated first arrival times for a source at $x_s$ and a receiver at $x_r$. In practice, these traveltimes can be found either by picking the first arrival times or by some automated method such as

cross-correlating the observed seismogram with the calculated seismograms (Luo and Schuster, 1991a). The w (discussed in Luo and Schuster, 1990) is a weighting factor used to balance out the contribution from these two residuals. Implicit in our definition of the inverse problem is that we assume that the density distribution is known. This assumption is appropriate if the densities can be obtained from either well log data or by a simple empirical relation between velocity and density.

For simplicity we choose a steepest descent method to minimize equation (2), with the understanding that a conjugate gradient method is used in practice. To update the velocity model, the steepest descent method gives,

$$c_{k+1}(\mathbf{x}) = c_k(\mathbf{x}) + \alpha_k \cdot \gamma_k(\mathbf{x}), \quad (4)$$

where $\gamma_k(\mathbf{x})$ is the steepest descent direction for the misfit function S, x represents any location between the wells, $\alpha_k$ is the step length, and $k$ denotes the kth iteration.

Representing the first term in equation (2) as $S_1$, the second term as $S_2$, and taking the Frechet derivative of S with respect to velocity, yields

$$\gamma(\mathbf{x}) = -\frac{\partial S}{\partial c(\mathbf{x})} = -\frac{\partial S_1}{\partial c(\mathbf{x})} - \frac{\partial S_2}{\partial c(\mathbf{x})} = \gamma_1(\mathbf{x}) + \gamma_2(\mathbf{x}). \quad (5)$$

From equation (8) in Luo and Schuster (1991a)

$$\gamma_1(\mathbf{x}) = \frac{1}{c^3(\mathbf{x})} \sum_s \int dt \, \dot{p}(\mathbf{x}, t|\mathbf{x}_s)_{cal} \dot{p}'_1(\mathbf{x}, t|\mathbf{x}_s), \quad (6)$$

where

$$p'_1(\mathbf{x}, t|\mathbf{x}_s) = \sum_r g(\mathbf{x}, -t|\mathbf{x}_r, 0) * \delta\tau(\mathbf{x}_r, t|\mathbf{x}_s), \quad (7)$$

and the symbol $*$ represents temporal convolution, $\dot{p}$ represents the time derivative of p, and $g(\mathbf{x}, -t|\mathbf{x}_r, 0)$ is the Green's function associated with equation (1) for the velocity field $c_k(\mathbf{x})$. Here $\delta\tau$ is the pseudo-traveltime residual defined by

$$\delta\tau(\mathbf{x}_r, t|\mathbf{x}_s) = -\frac{2}{E} \dot{p}(\mathbf{x}_r, t + \Delta\tau_{rs}|\mathbf{x}_s)_{obs} \cdot \Delta\tau_{rs}, \quad (8)$$

with the normalization factor $E$ defined as

$$E = \int dt \, \dot{p}(\mathbf{x}_r, t + \Delta\tau_{rs}|\mathbf{x}_s)_{obs} \dot{p}(\mathbf{x}_r, t|\mathbf{x}_s)_{cal}. \quad (9)$$

For a single source, the interpretation of equation (6) is that the forward-modeled field $\dot{p}(\mathbf{x}, t|\mathbf{x}_s)_{cal}$ is cross-correlated with the back-projected field $\dot{p}'_1(x, t|\mathbf{x}_s)$ to yield the gradient value at x. From equation (7), the back-projected field $\dot{p}'_1(\mathbf{x}, t|\mathbf{x}_s)$ is found by back-projecting the pseudo-traveltime residual $\delta\tau(\mathbf{x}_r, t|\mathbf{x}_s)$, where $\delta\tau(\mathbf{x}_r, t|\mathbf{x}_s)$ is formed by weighting the observed seismogram at the $\mathbf{x}_r$ receiver with its associated traveltime residual $\Delta\tau_{rs}$ and normalization value $E$. In other words, the $\gamma_1(\mathbf{x})$ term is calculated by reverse time migration of the weighted observed seismograms, where the traveltime residuals serve as scalar weighting factors. This is the gradient for the wave-equation traveltime-inversion (WT.) method of Luo and Schuster

(1991a). In practice, all waveforms except those of the transmitted arrivals are muted out prior to back-projection.

It is shown in Luo and Schuster (1991a) that the seismogram misfit gradient $\gamma_2(\mathbf{x})$ is exactly the same as the traveltime misfit gradient $\gamma_1(x)$ except the pseudo-traveltime residual $\delta\tau(\mathbf{x}_r, t|\mathbf{x}_s)$ in equation (7) is replaced by the normalized seismogram residual $2 \cdot \delta p_{rs}(t)/E$, where $\delta p_{rs}(t)$ is defined in equation (3). Similar to the interpretation of $\gamma_1(\mathbf{x})$, $\gamma_2(\mathbf{x})$ is calculated by reverse time migration of the normalized seismogram residuals.

A useful strategy for inversion is to first reconstruct the long wavelength parts of the velocity model with the traveltime gradient and then reconstruct the short wavelength features with the waveform gradient (see appendix). This strategy can be implemented by re-weighting the hybrid gradient in equation (5)

$$\gamma(\mathbf{x}) = a \cdot \gamma_1(\mathbf{x}) + (1 - a) \cdot \gamma_2(\mathbf{x}), \qquad (10)$$

where the scalar weight $a$ is used to emphasize either the traveltime misfit gradient or the seismogram misfit gradient. The criteria we use is to set $a = 1$ for $\Delta\tau > T/4$, and $a = 0$ for $\Delta\tau <= T/4$, where $T$ is the period corresponding to the peak frequency of the wavelet. In practice, we find that using the traveltime tomogram as a starting model will allow the waveform inversion to converge to the true model. In the following numerical tests, each of the seismograms is normalized to an amplitude of one.

## NUMERICAL EXAMPLES

The WTW method will be applied to both synthetic and field crosshole seismic data. Figure l(a) depicts the synthetic model. The field crosshole data were collected by Exxon (Chen et al., 1990) near 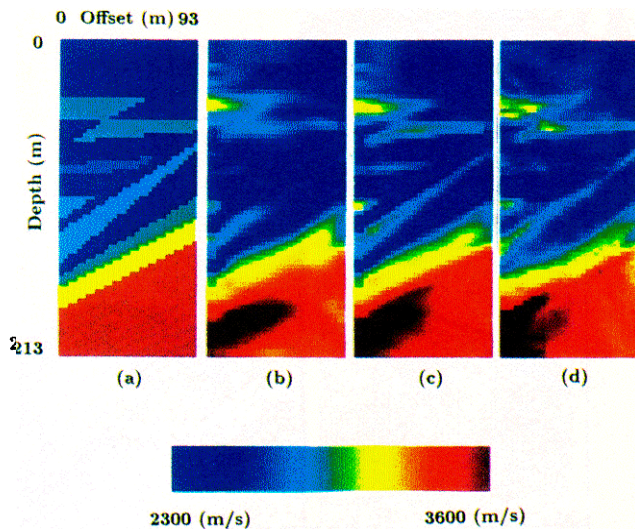Friendswood, Texas. The inversion scheme is a nonlinear conjugate gradient method with preconditioning (Beydoun and Mendes, 1989) and a constant step length.

### Synthetic crosshole data

The WTW method will now be applied to the Figure l(a) fault model used in Luo and Schuster (1991a). The "observed" seismograms in this case were generated by a fourth-order finite difference solution to the 2-D acoustic wave equation (with constant density). The fault model is discretized onto a mesh with 31 x 71 gridpoints, with 18 line sources and 36 receivers along the left side and right side of the model, respectively; a 40-gridpoint wide absorbing sponge zone is added along each boundary. The source wavelet is a Ricker wavelet having a peak frequency of 80 Hz, and the starting velocity model is a uniform medium with a velocity of 3000 m/s.

Figure l(b) shows the wave equation traveltime inversion (WT) tomogram after 10 iterations. After the tenth iteration the traveltime gradient in equation (10) was turned off [by setting $a = 0$ in equation (10)] and the waveform gradient was turned on for another four iterations to give the WTW tomogram in Figure 1(c). For comparison, Figure 1(d) shows the "best" traveltime tomogram after more than 50 iterations. The traveltime tomogram was reconstructed by a first arrival tomography method using ray tracing and a smoothed gradient field. Note that the WTW tomogram in Figure l(c) shows much better interface definition than the WT [Figure l(b)] or ray tracing traveltime [Figure 1(d)] tomograms. This is not surprising because the WT method seeks to fit only the first arrival traveltimes compared to fitting all of the waveforms in the seismograms for the WTW method. This means that the first arrival times can be accounted for by a smooth velocity model, while only a highly detailed fault model will account for the entire seismograms. It is to be noted that Luo and Schuster (1991a) showed that standard full wave inversion failed for this fault model if the starting model was a homogeneous velocity model. This underscores the important feature of the WTW algorithm which first inverts for the long wavelength velocity features by traveltime inversion, and then extracts the model's detailed features by waveform inversion.

Figure 2 depicts shot gathers associated with the actual fault model [Figure l(a)], the WT tomogram [Figure l(b)], and the



FIG. 1. Synthetic data test results. (a) Fault model discretized onto a mesh of 3 1 x 71 gridpoints with a gridpoint spacing of 3 m. (b) WTW tomogram after 10 iterations, except only the traveltime gradient is used. (c) WTW tomogram after 14 iterations, where the waveform gradient was exclusively used after the 10th iteration. (d) Raytracing traveltime tomogram after 50 iterations.
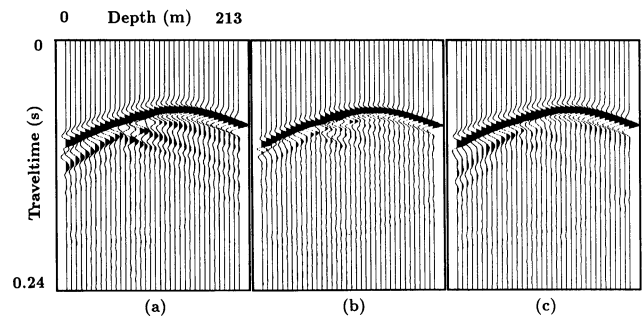


FIG. 2. Common shot gathers (for a source depth at 108 m) computed for (a) the Figure l(a) fault model, (b) the Figure l(b) WT tomogram, and (c) the WTW tomogram in Figure l(c).

WTW tomogram [Figure l(c)]. It is observed that the WTW shot gather contains many of the scattering and reflection events in the fault model shot gather in Figure 2(a). In comparison, the WT shot gather does not correlate well with the actual shot gather, except for the direct arrivals and some reflection events.

The WTW method can be robust in the presence of random noise. This is demonstrated by WTW inversion of the shot gathers from the fault model, except random noise has been added to the data. In this example the signal-to-noise ratio is 4:1, and Figure 3 depicts the shot gather in Figure 2(a) after the addition of random noise. Applying the WTW method to the noisy fault model data yields the WTW tomogram after 16 iterations in Figure 4. The tomogram's accuracy is quite acceptable. Other examples of applying the WTW method to synthetic data are given in Luo and Schuster (1990).

### Friendswood  crosshole  data

The WTW method is applied to a crosshole seismic data set collected by Exxon near their Friendswood, Texas, test site (Chen et al., 1990). The source and receiver wells are 180 m apart, the wells extend to a depth of approximately 300 m, and the source and receiver intervals are 3.05 m so that there are 98 sources and 96 receivers. The source consists of a small amount of dynamite, and the seismic data have a usable bandwidth of 80 to 600 Hz. A typical unprocessed shot gather at intermediate depth is shown in Figure 5(a).
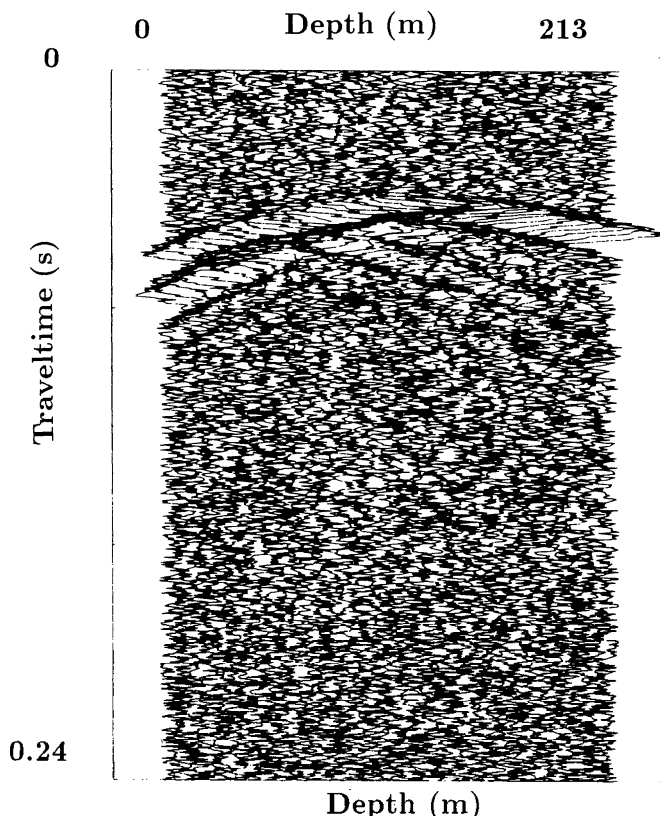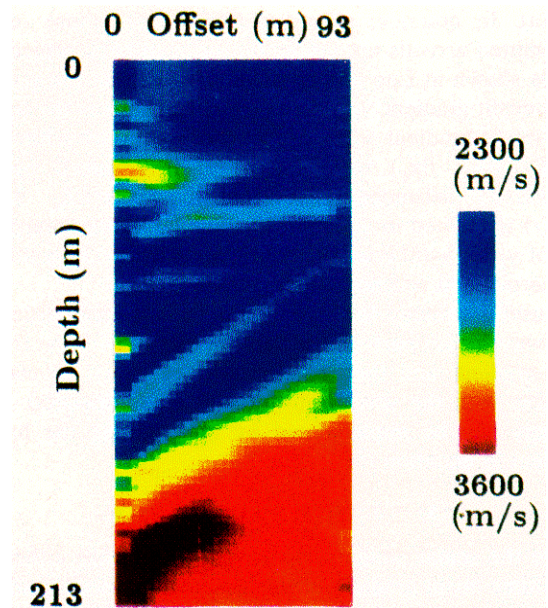


FIG. 4. WTW tomogram (after 16 iterations) inverted from the noisy fault model data (Figure 3).



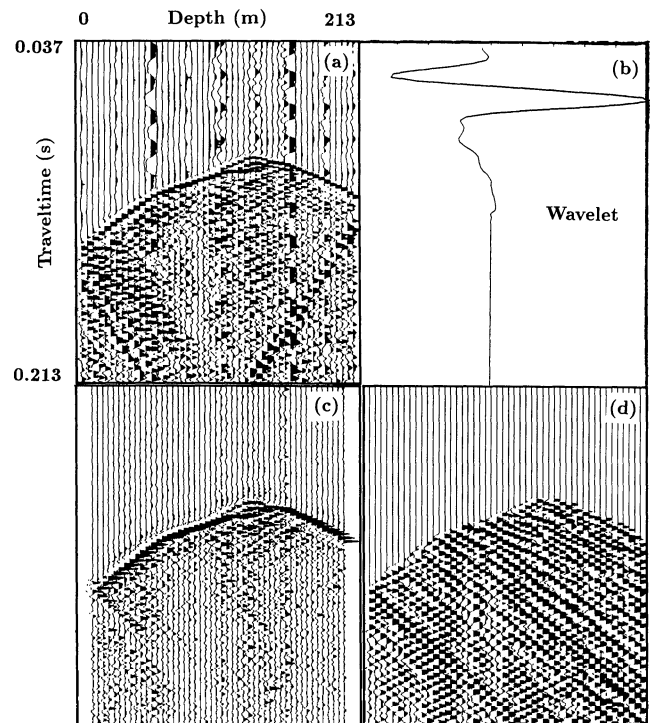FIG. 3. The same shot gather as in Figure 2(a), except the random noise has been added to the data.



FIG. 5. A typical shot gather of Fnendswood crosshole data collected by Exxon near their Friendswood, Texas, test site. The source depth is 156 m. (a) Raw shot gather. (b) First arrival wavelet extracted from the Figure 5(a) shot gather. (c) Figure 5(a) shot gather data after signal processing. (d) Downgoing waves after f-k filtering of the Figure 5(a) shot gather.

Processing steps applied to the shot gathers include (Cai and Schuster, 1993) the following: (1) tube waves are eliminated by median filtering; (2) free-surface reflections are muted out in the time domain; (3) an 80-600 Hz bandpass filter is applied to the data; (4) each trace is normalized to its maximum amplitude value; and (5) direct arrivals are muted out after the waveform inversion is turned on. Each forward modeled shot gather used a source wavelet extracted from the corresponding observed shot gather; e.g., Figure 5(b) shows the first arrival source wavelet associated with a trace at intermediate depth. To accommodate the 80-600 Hz bandwidth of the data, a 2-D finite-difference mesh of 301 x 501 gridpoints is used for the forward modeling and back-projection, with the same well geometry as in the Friendswood experiment. Well deviations in the source and receiver wells were corrected by shifting the coordinates of deviated sources and receivers to the appropriate position in the vertical plane of the source-receiver wells. The data were corrected to a 2-D format by multiplying the filter $\sqrt{i/\omega}$ by the spectrum of the observed seismograms and scaling the data by $\sqrt{t}$ to approximate geometric spreading. The final processed shot gather associated with Figure 5(a) is shown in Figure 5(c). Figure 5(d) shows the same gather except the upgoing and transmitted waves have been eliminated by f-k filtering.

The WTW method is now applied to the 98 shot gathers of the processed Friendswood data, where the starting model is a homogeneous velocity model of 1800 m/s. Figures 6(a), 6(b), and 6(c) show the WTW tomograms after 10, 16, and 46 iterations, respectively. Figure 6(d) shows the Figure 6(c) tomogram after applying an edge detection algorithm (Bilbro et al., 1992). For these WTW tomograms, the waveform gradient is turned on and the traveltime gradient is turned off after the tenth iteration. This means that Figure 6(a) is the WT tomogram after 10 iterations, whereas Figures 6(b)-6(d) can be considered as full waveform inversion tomograms that use the Figure 6(a) WT tomogram as the starting velocity model. We assign the density model by a simple empirical formula (Gardner et al., 1974) that relates density to the P-wave velocity. Note that the final WTW tomogram in Figure 6(c) provides a much finer layer and velocity resolution compared to the Figure 6(a) WT tomogram. This is verified in Figure 7 which compares the smoothed sonic log (solid line) in the source hole to a vertical slice of the final WTW and WT tomograms. These slices were taken from the tomograms along a vertical line 12 m from the sonic log (or source well on the left side) and demonstrate that the vertical resolution of the WTW tomogram is about 1.5-3.0 m compared to the 6.0-12.0 m
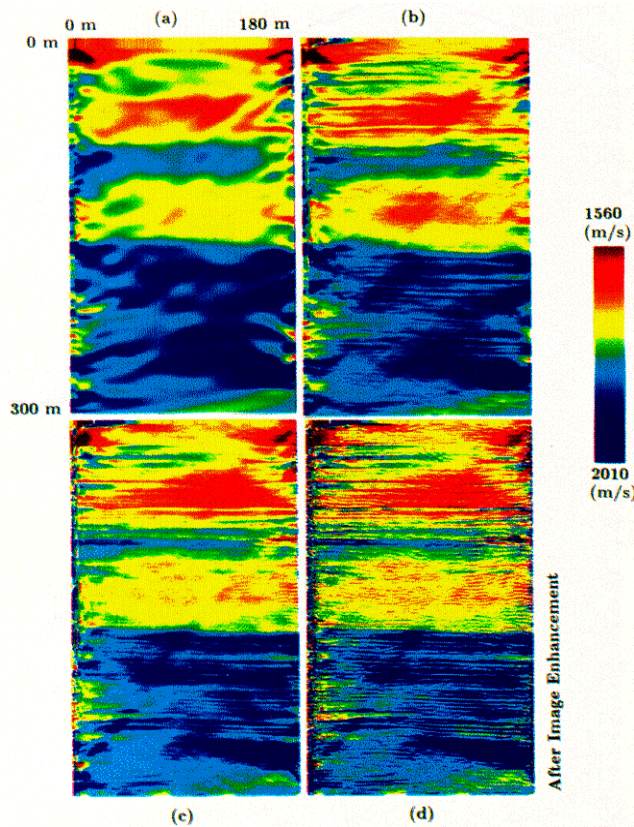


FIG. 6. The WTW tomograms inverted from the Friendswood crosshole data. (a) WT inversion after 10 iterations. Figures 6(b), 6(c), and 6(d) are WTW tomograms after 16 iterations, 46 iterations, and image enhancement of the Figure 6(c) tomogram, respectively.
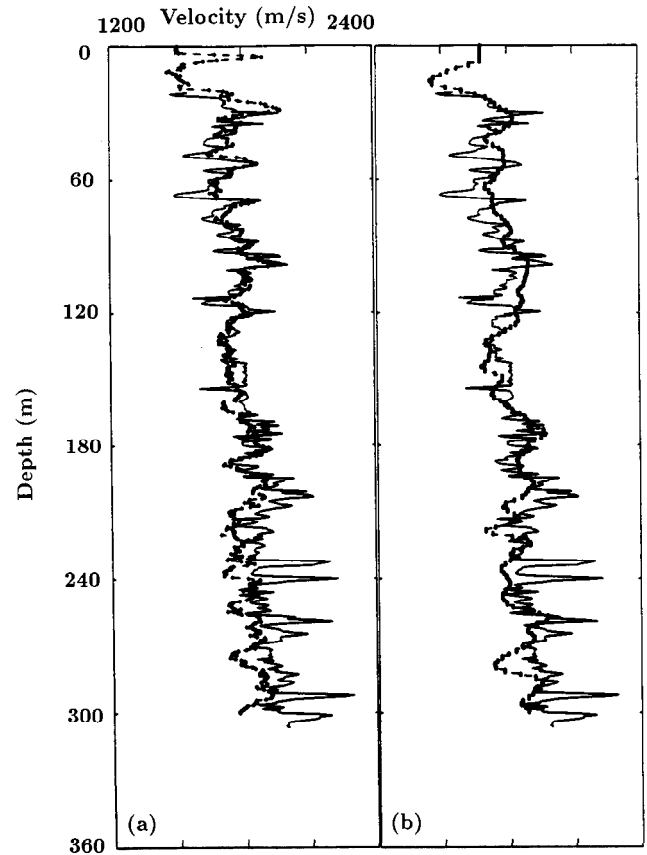


FIG. 7. Tomogram velocity profiles (dots spaced at 0.6 m) compared to the sonic log (solid line) in source well. The profiles are extracted from the tomograms 12 m from the source well. (a) Figure 6(c) WTW tomogram velocity profile and sonic log, (b) Figure 6(a) WT tomogram velocity profile and sonic log.

resolution of the WT tomogram. Note that the WTW velocity profile differs from the sonic log profile by a DC shift in the 200-300 m interval. This highlights a weakness in the WTW method. It is effective at reconstructing both the high and intermediate wavenumber parts of the model but it can have difficulty in recovering the low wavenumber part of the model.

As a final accuracy check, Figures 8(a) and 8(b) show the synthetic shot gathers computed from the velocity field in the 10th iteration WT tomogram [Figure 6(a)] and the final WTW tomogram [Figure 6(c)], respectively. In the WTW synthetic shot gather, both direct waves and reflection arrivals show significant correlations with their counterparts in the observed shot gather shown in Figure 8(c). For example, downgoing reflection events A, B, C, and D are present in both the real and synthetic shot gathers. Also note that the primary and secondary arrivals around zone G in the field data are matched quite well by similar events in the synthetic WTW gather. Reflection events F and E in the field records are not present in the synthetic gather because the free surface and subwell reflectors were deliberately excluded from the synthetic velocity model.

The events not accounted for in the synthetic acoustic gather are depicted in the shot gather residual shown in Figure 8(d). Note the large residual associated with the direct wave; this indicates that the synthetic source wavelet did not exactly match the shape of the actual source wavelet. Part of the mismatch can be attributed to the acoustic finite-difference modeling which does not take into account P-S conversions or viscoelastic losses in the field data.

In comparison with the WT results, the WT synthetics in Figure 8(a) show a good correlation with the actual direct arrivals in Figure 8(c). However, the secondary arrivals are almost completely absent because of the smooth nature of the WT tomogram. Inverting just the first arrival traveltimes is insufficient for reconstructing detailed features of the velocity model.

Figure 9 shows the RMS traveltime and RMS waveform residuals vs iteration number for the Exxon data reconstructions. Relative to the waveform residual at the 10th iterate, the waveform residual at the 46th iterate decreases by about 30%. For each WTW iteration, about 50 CPU minutes were required on a 2.5-Gigaflop Fujitsu VPX240/10, or about
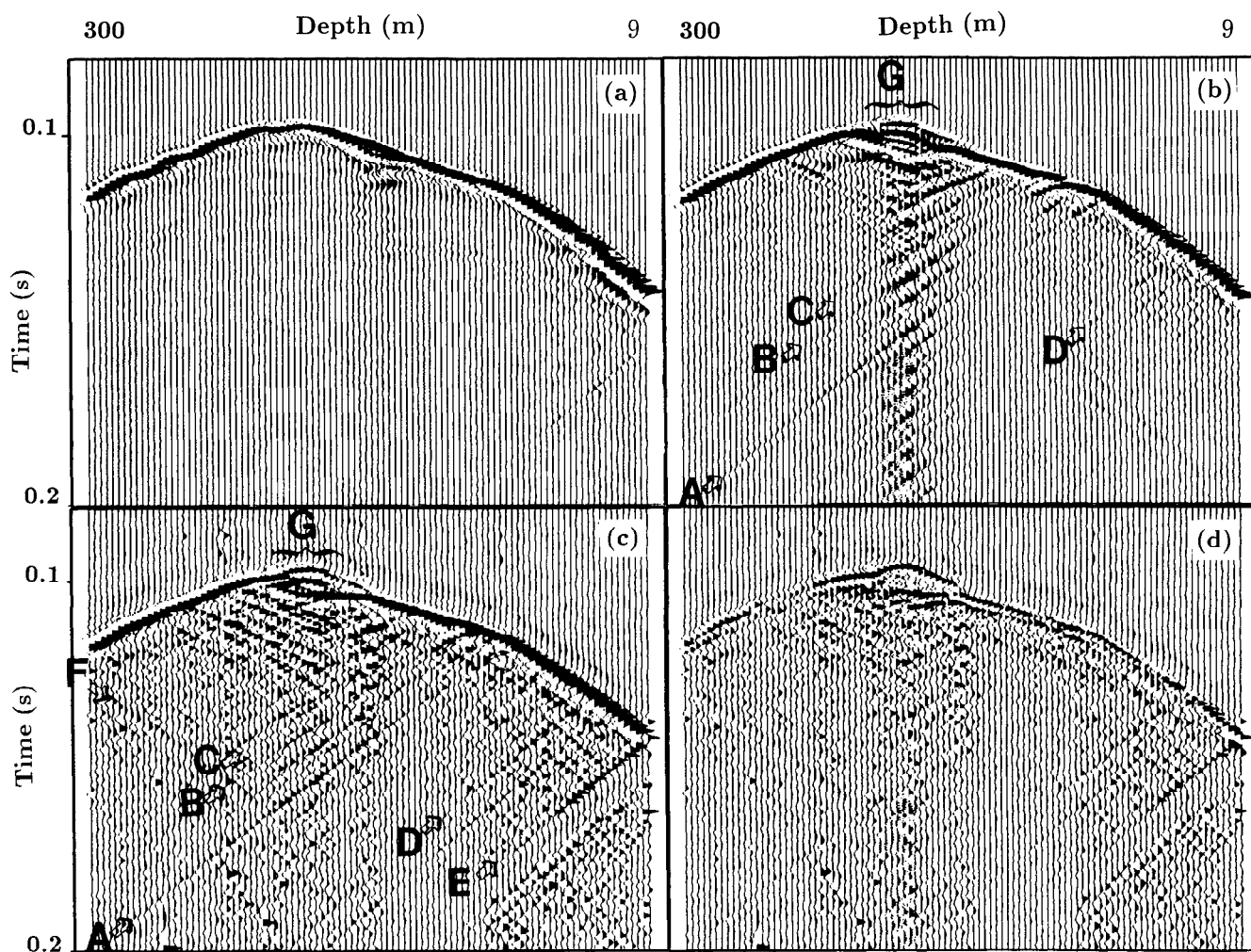


**FIG. 8.** Synthetic acoustic common shot gathers associated with the (a) WT tomogram in Figure 6(a), and (b) the WTW tomogram in Figure 6(c). Figure 8(c) depicts the corresponding observed shot gather, and Figure 8(d) depicts the seismogram residuals. The source location for these common shot gathers is at the depth 156 m.

30 CPU hours on an 80-Mflop Stardent 2000 (roughly equivalent in speed to an IBM RS6000 560 computer for finite-difference computations).

Figure 10(a) is a zoom view of the lower part of the Figure 6(c) WTW tomogram and shows highly resolved dipping layers (Figure 10) that are absent in the WT tomogram in Figure 6a. The layers appear to be about 3.0-4.5 m thick and are imaged because the WTW inversion accounts

for the secondary events (primarily reflection events) in the data.

Figure 11 shows the migrated crosshole section using the Friendswood crosshole data and a smoothed representation of the Figure 6(c) velocity model. In this case a constrained Kirchhoff-migration method (Qin and Schuster, 1993) was used and the migration velocity below the well was extrapolated from the bottom part of the Figure 6(c) tomogram. On the left is a 1-D synthetic seismogram computed from the sonic log in Figure 7 and the source wavelet in Figure 5(b). The migrated section and sonic log correlate reasonably well to provide complementary information to the WTW tomogram.

## CONCLUSION

We presented a hybrid velocity inversion method, i.e., WTW, that incorporates the advantages of both traveltime (i.e., robust convergence) and full wave inversion (i.e., high-resolution) methods. The wave equation traveltime inversion (WT) method is used to first reconstruct the long wavelength features of the velocity model, and then full waveform inversion is used to reconstruct the fine details of the model. This strategy avoids the problem of getting stuck in local minima when the starting model is far from the actual
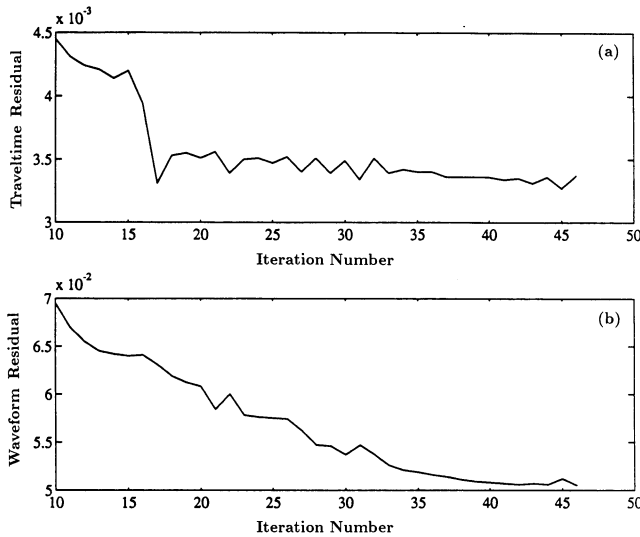


FIG. 9. RMS (a) traveltime residuals and (b) seismogram residuals vs iteration number for the WTW inversion of the Friendswood data.
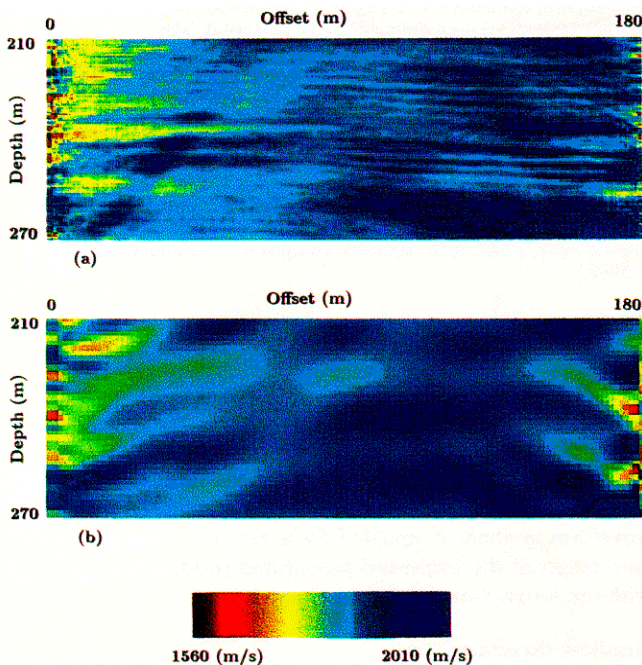


FIG. 10. Enlarged WTW tomogram showing structural detail. (a) WTW tomogram after 46 iterations. (b) WTW tomogram after 10 iterations (using only the first arrival traveltime information).
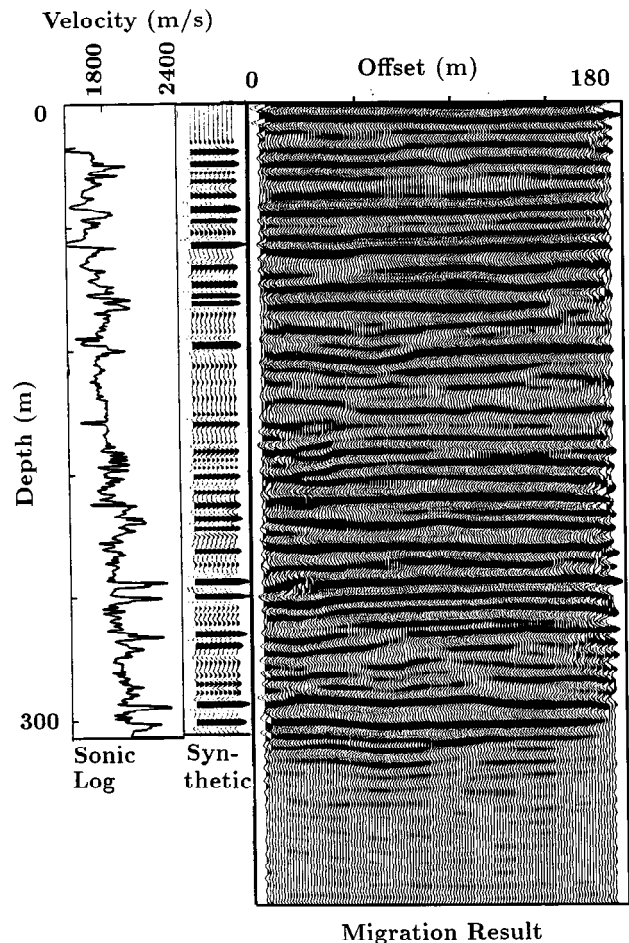


Migration Result

FIG. 11. Migrated Friendswood data using the WTW tomogram model in Figure 6(c) as the migration velocity model.

model. To partially alleviate amplitude problems caused by source radiation patterns or inconsistent geophone/source coupling, we normalize the traces of both the observed and synthetic seismograms.

Inversion of both synthetic and crosshole seismic data shows that the WTW method provides a significantly better model resolution than given by traveltime tomography. WTW tomograms show a very fine resolution (1 S-3.0 m) of the interface boundaries compared to the indistinct or smeared interfaces in the traveltime tomograms. Except for a DC shift at the 200-300 m interval, there was good agreement between the sonic log velocity and the tomogram's velocity profile near the source well. Subtle structural features completely absent in the Friendswood traveltime tomogram are clearly defined in the WTW tomogram. Such details can spell the difference between success and failure in enhanced oil recovery operations. This suggests that intermediate frequency data of high quality (i.e., distinct reflection arrivals) may be more useful than high frequency data of intermediate quality (i.e., good first breaks but indistinct reflection arrivals).

Future work on the WTW method should include increasing computation speed by the use of asymptotics, inverting for source radiation patterns and subwell velocities (Zhou and Schuster, 1993), and the use of the anisotropic and viscoelastic wave equation for P-velocity, S-velocity, and density inversion (Zhou et al., 1994).

## ACKNOWLEDGMENTS

## REFERENCES

Beydoun, W., and Mendes, M., 1989, Elastic ray Born & migration/inversion: Geophys. J., 97, 151-160.

Bilbro, G., Snyder, W., Garmer, S., and Gault, J., 1992, Mean field annealing: A formalism for constructing GNC-like algorithms: IEEE Transactions on Neural Networks, 3, 131-138.

Bishop, T., Bube, IS., Cutler, R., Langan, R., Love, P., Resnick, J., Shuey, R., Spindler, D., and Wyld, H., 1985, Tomographic determination of velocity and depth in laterally varying media: Geophysics, 50, 903-923.

Cai, W., and Schuster, G. T., 1993, Processing Friendswood crosswell seismic data for reflection imaging: 63th Ann. Internat. Mtg. Soc. Explor. Geophys., Expanded Abstracts, 92-94.

Chen, S. T., Zimmerman, L. J., and Tugnait, J. K., 1990, Subsurface imaging using reversed vertical seismic profiling and crosshole tomographic methods: Geophysics, 55, 1478-1487.

Crase, E., Wideman, C., Noble, M., and Tarantola, A., 1992, Nonlinear elastic waveform inversion of land seismic reflection data: J. Geophys. Res., v. 97-B4, 4685-4704.

Dines, K., and Lytle, R., 1979, Computerized geophysical tomography: Proc. IEEE, 67, 1065-1072.

Gauthier, O.) Virieux, J., and Tarantola, A., 1986, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results: Geophysics, 51, 1387-1403.

Gardner, G. H. F., Gardner, L. W., and Gregory, A. R., 1974, Formation velocity and density-The diagnostic basis of stratigraphic traps: Geophysics, 39, 770-780.

Ivansson, S., 1985, A study of methods for tomographic velocity estimation in the presence of low-velocity zones: Geophysics, 50, 969-988.

Levander, A., 1988, Fourth-order finite-difference P-SV seismograms: Geophysics, 53, 1425-1437.

Lines, L., 1988, Inversion of geophysical data, Geophysics Reprints No. 9: Soc. Expl. Geophys.

Luo, Y., and Schuster, G., 1990, Wave equation traveltime + waveform inversion: 60th Ann. Internat. Mtg., Soc. Explor. Geophys., Expanded Abstracts, 1223-1225.

Luo, Y., and Schuster, G., 1991a, Wave equation traveltime inversion: Geophysics, 56, 645-653.

Luo, Y., and Schuster, G., 1991 b, Wave equation inversion of skeletalized geophysical data: Geophys. J., 105, 289-294.

Mora, P., 1987, Nonlinear two-dimensional elastic inversion of seismic data: Geophysics, 52, 1211-1228.

Paulsson, B., Cook, N., and McEvilly, T., 1985, Elastic wave velocities and attenuation in an underground repository of nuclear waste: Geophysics, 50, 551-570.

Qin, F., and Schuster, G. T., 1993, Constrained Kirchhoff migration of cross-well seismic data: 63th Ann. Internat. Mtg., Soc. Explor. Geophys., Expanded Abstracts, 99-102.

Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data: Geophysics, 51, 1893-1903.

Tarantola, A., 1987, Inverse problem theory: Elsevier.

Zhou, C., and Schuster, G. T., 1993, Waveform inversion of subwell velocity structure: 63th Ann. Mtg., Soc. Explor. Geophys., Expanded Abstracts, 106-109.

Zhou, C., Schuster, G. T., Hassanzadeh, S., and Harris, J. M., 1994, Crosshole elastic WTW inversion of the McElroy data: 64th Ann. Mtg., Soc. Explor. Geophys., Expanded Abstracts, 50-53.

## APPENDIX A

We now heuristically show that the misfit gradient in equation (5) can be decomposed into both short wavelength and long wavelength components, where the short wavelength components are associated with reflection events and the long wavelength components are primarily associated with the transmitted arrivals.

For a single impulsive source, the waveform gradient $\gamma(x)$ in equation (5) can be rewritten as

$$\gamma(x) = G(x|x_s) \otimes g(x|x_s), \qquad (11)$$

where $g(x|x_s) = 2 \sum_r g(\mathbf{x}, t|\mathbf{x}_r, 0) * \delta\dot{p}_{rs}(t)/E$. and $G(x|x_s) = \dot{p}(\mathbf{x}, t|\mathbf{x}_s, 0)/c^3(\mathbf{x})$, and the notation for the time variables has been suppressed. The interpretation of equation (11) is that the waveform gradient is updated by a zero-lag temporal cross-correlation of the backward-propagated residual field $g(x|x_s)$ with the forward-modeled field $G(x|x_s)$.

### Gradient decomposition

We will now decompose the gradient field into a sum of transmission and reflection components, where the reflection component can be further divided into forward- and back-propagated parts. The decomposition formula shows

that the magnitude of the misfit gradient associated with the transmitted arrivals (henceforth called the transmission gradient) dominates over the gradient contributions from reflected arrivals (henceforth called the reflection gradient).

We will employ the notation $G_{ij}$ which describes the wavefield along the raypaths that connect the source at $x_i$ with the receiver at $x_j$. For the example of the single interface model in Figure 12(a), the Green's function $G(x_0|x_2)$ can be decomposed into a composite wavefield with support along the transmitted ray $G_{02}$, the downgoing ray $G_{01}$, and an upcoming ray $G_{12}$, i.e.,

$$G(x_2|x_0) = G_{02} + G_{01} + rG_{12}, \qquad (12)$$

where $r$ represents the reflection coefficient at the interface.

For a transmitted + reflected wavefield that is back-propagated from the receiver array, the Green's function $g(x_0|x_2)$ for a source at $x_2$ and a receiver at $x_0$ can be decomposed into a sum of back-propagated transmitted $g_{20}$, back-propagated downgoing $g_{21}$ and back-propagated up-coming $g_{10}$ parts, i.e.,

$$g(x_0|x_2) = g_{20} + rg_{21} + r^2g_{10}, \qquad (13)$$

where the associated raypaths are shown in Figure 12(b). Note that the wave downgoing from the receiver well is scaled by $r$ because we assume that this back-propagated reflection started out at the receiver well with an amplitude of $r$, After reflecting at the interface the amplitude is again scaled by a factor $r$ to yield the $r^2g_{10}$ term.

### Transmission traveltime tomography

Plugging equations (13) and (12) into equation (11) and neglecting all terms in $r$, we get the transmission gradient $\gamma_t(x)$

$$\gamma_t(x) = G_{02} \otimes g_{20}, \qquad (14)$$

which is, to zeroth-order in **r,** the gradient of the misfit function. For small **r,** this means that the dominant contribution to the velocity update is from the transmitted fields. Moreover, the correlation is nonzero along the transmitted raypath so that the velocity update for many source-receiver pairs is evenly, i.e., globally, distributed throughout the model. The velocity update is also evenly distributed along the ray since equation (14) applies nearly equal weighting to the velocity update along the transmitted raypath. This
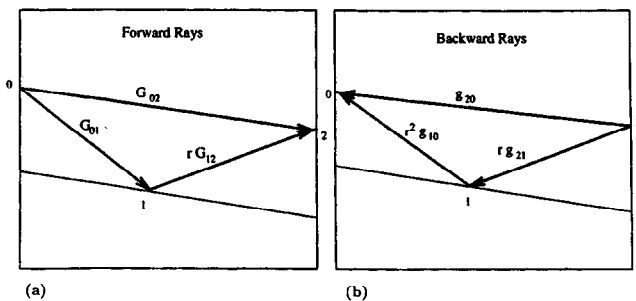
assumes that the geometrical spreading factors have been eliminated by a preconditioning method.

The idea that the transmission gradient updates the smooth components of velocity forms the basis of the inversion strategy for the WTW method. That is, the transmitted arrivals are weighted by the traveltime residuals and then back-projected into the medium to reconstruct the long wavelength components of the velocity field. After five or so iterations, the direct waves are muted out and waveform inversion is applied to the reflection events to reconstruct the shorter wavelength components of the velocity field. As an example, Figure 8(a) is devoid of prominent reflections because the traveltime tomogram is too smooth to generate reflections. In contrast, prominent reflections are extant in Figure 8(b) because the waveform tomogram contained the sharp boundaries of the reflector interfaces.

### Reflection waveform tomography

The previous section derived the transmission gradient formula and showed how it performs a global and smooth update of the velocity distribution. This section will derive the reflection gradient formula, which mainly performs a local and detailed update of the velocity model.

Eliminating the transmitted terms $g_{20}$ and $G_{02}$ in equations (12) and (13), and plugging the resulting formulae into equation (11), we get the reflection gradient $\gamma_r(x)$, i.e.,

$$\gamma_r(x) = (G_{01} + rG_{12}) \otimes (rg_{21} + r^2g_{10})$$

$$= r^2G_{01} \otimes g_{10} + rG_{01} \otimes g_{21}$$

$$+ r^2G_{12} \otimes g_{21} + r^3G_{12} \otimes g_{10}. \qquad (15)$$

Note that the crosscorrelations turn on only when the forward-propagated rays coincide in both space and time with the back-propagated rays. This statement implies that the rays labeled $G_{01}$ and $g_{21}$ in Figure 12 coincide only at $x_1$ on the reflector boundary so that $G_{01} \otimes rg_{21} \approx r\delta(x - x_1)$, where geometrical spreading terms have been ignored. Likewise, $G_{12} \otimes r^3g_{10} \approx r^3\delta(x - x_1)$. Therefore, the above equation reduces to:

$$\gamma_r(x) = r\delta(x - x_1) + r^2G_{12} \otimes g_{21}$$

$$+ r^2G_{01} \otimes g_{10} + r^3\delta(x - x_1); \qquad (16)$$

and neglecting third-order terms in $r$ this becomes

$$\gamma_r(x) = r\delta(x - x_1) + r^2G_{12} \otimes g_{21} + r^2G_{01} \otimes g_{10}. \qquad (17)$$

For small $r$, we see that the reflection gradient is dominated by the part that turns on at the boundary, namely the point $x_1$ where the migrated reflection correlates with the transmitted arrival. Complementary to the transmission gradient, the reflection gradient to first order in $r$ performs a local and detailed velocity update along the reflecting boundary, and no first-order corrections take place along the raypaths to update the smooth parts of the velocity field. This is one of the reasons for the slow convergence of iterative reflection waveform tomography. However, the second-order terms in equation (17) do contribute along the raypaths and so the waveform gradient can weakly update smooth parts of the velocity field.