Commun. Comput. Phys.
doi: 10.4208/cicp.2020.swsii.review

Vol. **28**, No. 1, pp. 1-40
July 2020

REVIEW ARTICLE

Seismic-Wave Scattering, Imaging, and Inversion

Lianjie Huang[1,*], Michael Fehler[2], Yingcai Zheng[3] and Xiao-Bi Xie[4]

[1] *Los Alamos National Laboratory, MS D452, Los Alamos, NM 87545, USA.*
[2] *Massachusetts Institute of Technology, Department of Earth, Atmospheric, & Planetary Sciences, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.*
[3] *University of Houston, Department of Earth & Atmospheric Sciences, Houston, Texas 77004, USA.*
[4] *University of California, Santa Cruz, CA 95064, USA.*

Received 27 April 2020; Accepted 6 May 2020

Abstract. Ru-Shan Wu has made seminal contributions in many research areas in geophysics, such as seismic-wave propagation, scattering, imaging, and inversion. We highlight some of his research in holography imaging, diffraction tomography, seismic-wave scattering and its applications to studying Earth's heterogeneity, one-way wave propagation and one-return wave modeling, beamlet and dreamlet applications, strong non-linear full-waveform inversion, and direct envelop inversion.

AMS subject classifications: 74J05, 74J10, 74J15, 74J20, 74J25, 35C07, 35L05

Key words: Beamlet, diffraction tomography, dreamlet, full-waveform inversion, holography, imaging, one-way, one-return, seismic wave propagation, seismic scattering.

1 Subsurface holographic imaging (1970s–)

In 1970s, Ru-Shan conducted translational research — applying optical holography to subsurface holographic imaging. In optical holography, a hologram is an interference pattern between the scattered light from an object and the reference light. In image reconstruction, the object is re-illuminated by the reference light to reproduce a three-dimensional light field, resulting in an image of the object. Ru-Shan and his colleagues first studied monochromatic (single-frequency) holography for imaging, and found that the main challenge in using elastic or electromagnetic waves for subsurface holographic imaging is its poor longitudinal resolution. To address this limitation, they developed

*Corresponding author. *Email addresses:* ljh@lanl.gov (L. Huang), fehler@mit.edu (M. Fehler), yzheng24@central.uh.edu (Y. Zheng), xxie@ucsc.edu (X.-B. Xie)

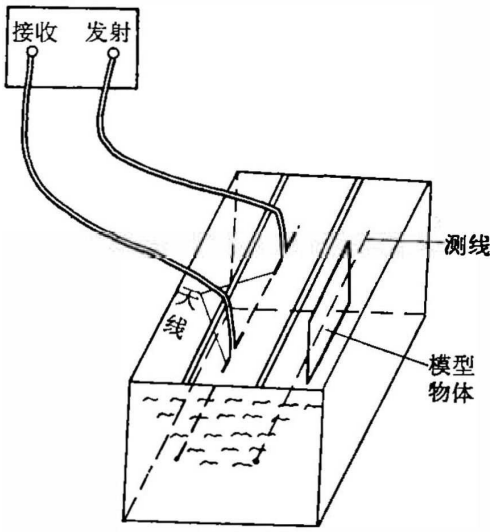


图 8 实验装置示意图

Figure 1: Setup of a water tank experiment for testing multifrequency synthetic detecting holography. (From [60])

the multifrequency synthetic detecting holography to improve the longitudinal image resolution, and published their research results in Chinese with an abstract in English in 1977 [60].

In the multifrequency synthetic detecting holography technique, after obtaining a hologram $H(\omega,\xi,\eta)$ (i.e. data acquisition) for each frequency ω within a given range from ω_1 to ω_2, the multifrequency holography image $M(x,y,z)$ is generated using (see Eq. 8 in [60])

$$M(x,y,z)=\int_{\omega_1}^{\omega_2}\int_{\xi}\int_{\eta}W(\omega)H(\omega,\xi,\eta)e^{-i\frac{\omega}{v}r(\xi,\eta,x,y,z)}d\xi d\eta d\omega, \tag{1.1}$$

where $W(\omega)$ is a weight for a monochromatic hologram at each ω, v is the velocity, and r is the distance from a source to a position (ξ,η) on a hologram via an imaging point (x,y,z).

In Ru-Shan's research on subsurface holographic imaging, he bridged many concepts in different fields: communication and antenna array, correlation functions and matched filters, and phased arrays. In addition, he studied the relationship between multifrequency synthetic detecting holography and "impulse holography," and derived the formulas for longitudinal resolution. He and his colleagues verified the feasibility of their technique using both two-dimensional (2D) synthetic data and laboratory data. They conducted the microwave multifrequency synthetic detecting holography using laboratory data acquired in a water tank (Fig. 1), demonstrating that their method can produce high-resolution and high-quality images. Fig. 2 depicts an image of two aluminum blocks reconstructed using their method with six frequencies.

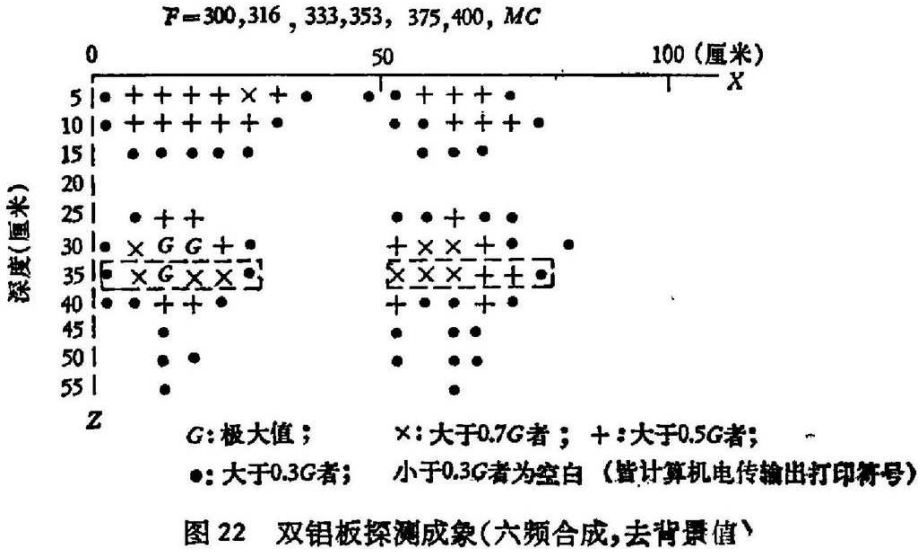


图 22 双铝板探测成象(六频合成,去背景值)

Figure 2: Image of two aluminum blocks reconstructed using six frequencies. (From [60])

After Ru-Shan came to the U.S. as a visiting scholar in 1979, the Nobel Physics Laureate Samuel C.C. Ting introduced him to seismologists at Massachusetts of Institute of Technology (MIT). Because of Ru-Shan's holography work, Professor Keiiti Aki invited Ru-Shan to work in his group, and later took him as a PhD student. Ru-Shan's PhD thesis work was mainly on seismic scattering that is highlighted in Section 3. His postdoctoral work at MIT on diffraction tomography (Section 2) is in line with his holography work.

2 Diffraction tomography (1980s–)

Wu and Toksöz (1987) [57] published an influential paper on diffraction tomography and multisource holography, and their relationship through theoretical analyses and numerical experiments. Their theory for diffraction tomography is based on the Born or Rytov approximation. They studied the performances of diffraction tomography and multisource holography by examining the information coverage in the spatial frequency domain and using numerical examples.

Wu and Toksöz (1987) [57] found that multisource holography is similar to Kirchhoff-type migration, and often produces distorted images because of spectral coverage issues. They found that the filtering operation of diffraction tomography helps correct the nonuniform coverage (including duplication) of the object spectrum in the reconstruction process, resulting in improved images. They also found that, multisource holography is better suited for imaging sharp boundaries with large acoustic impedance contrasts compared with diffraction tomography that is based on an approximation of weak inhomogeneities.

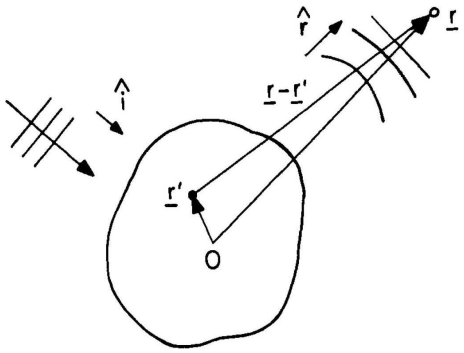


Figure 3: The geometry of the basic scattering experiment used to demonstrate the basic principle of diffraction tomography (from [57]).

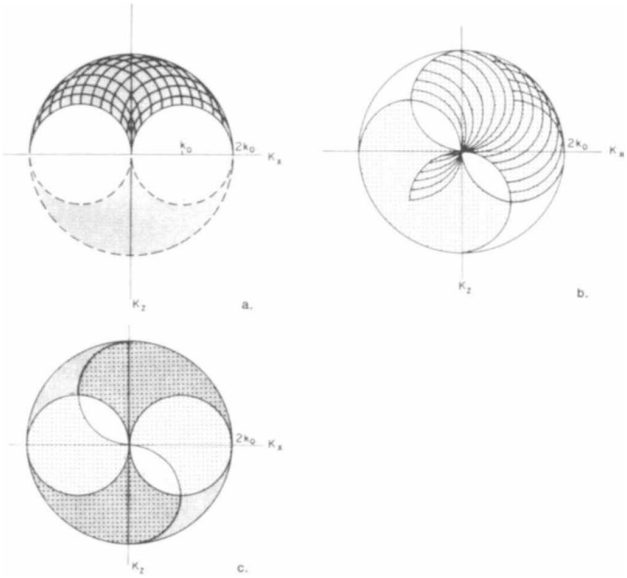


Figure 4: The spectral coverages of (a) surface reflection profiling data (the densely dotted area), (b) vertical seismic profiling data (the sparsely dotted area), (c) the combination of both datasets, where k_0 is the wavenumber of the wavefield, $K=(K_x,K_z)$ is the 2-D spatial frequency vector of the object function. (From [57])

Wu and Toksöz (1987) [57] demonstrated the basic principle of diffraction tomography using a plane wave incident on an object in a homogeneous, infinite medium, as displayed in Fig. 3. When a receiver is far from the object, the scattered wave from the object can be treated as a plane wave at the receiving point. They derived the relationship between the scattered field and the object spectrum for diffraction tomography.

They evaluated the performance of diffraction tomography for different seismic data acquisition geometries, including surface reflection profiling (SRP), vertical seismic profiling (VSP), and cross-hole measurements. Fig. 4(a) shows the poor spectral coverage of single-frequency surface reflection measurements. Fig. 4(b) depicts that the VSP ge-

ometry provides better single-frequency spectral coverage than surface reflection data. Fig. 4(c) displays the spectral coverage of the combined data set of VSP and SRP, leading to an improved spectral coverage.

3 Seismic scattering and Earth's heterogeneity (1970s–)

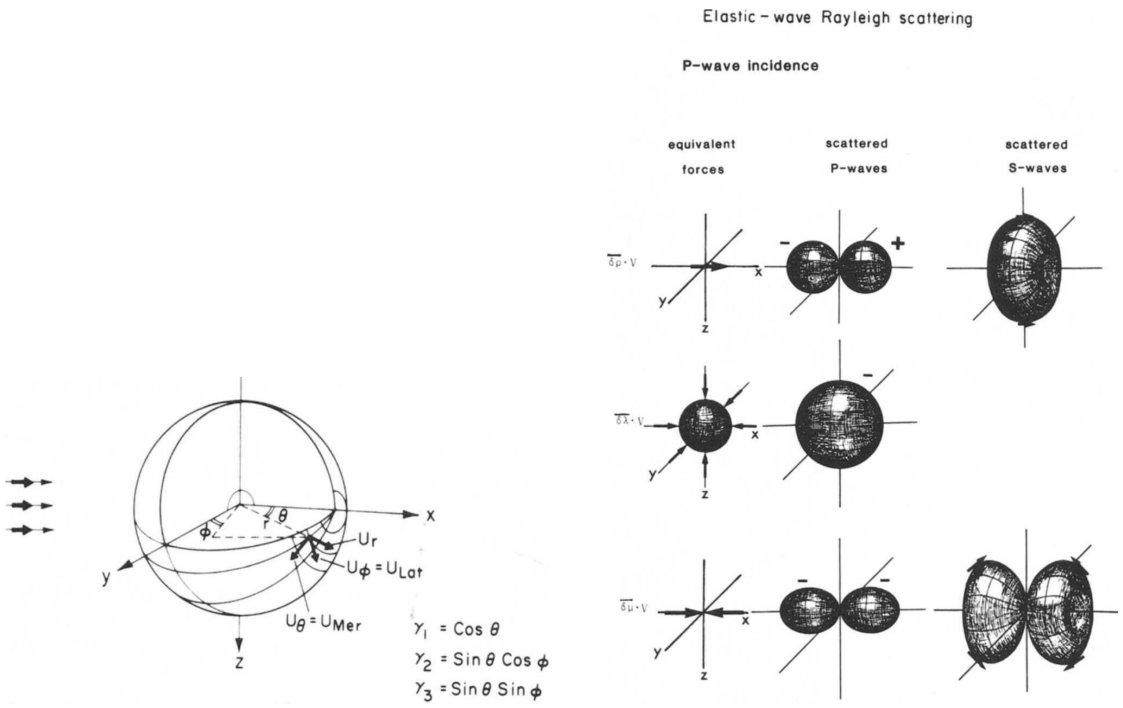
When Ru-Shan arrived at MIT in 1979, there was a great deal of activity at MIT around the topic of seismic scattering. This work was being undertaken by Keiiti Aki (Kei) and M. Nafi Toksöz along with their students, postdocs, and researchers. This topic was an excellent fit for the work that Ru-Shan had undertaken in China on holography. Scattering has been the foundation for a great deal of Ru-Shan's work whether addressing issues directly focused on characterizing heterogeneity in the Earth, or using scattering concepts in other areas of his research, such as imaging. Most of the work discussed in this section was done early in Ru-Shan's career. This work has had a deep and lasting impact and has been the foundation for countless papers written by investigators throughout the world. Ru-Shan addressed important theoretical topics and introduced several novel concepts into the seismic scattering community. He frequently used observations of others to motivate the topics that he addressed in his theoretical work. In addition, Ru-Shan frequently analyzed field data to support his theoretical developments.

Throughout this section, we refer to common parameters that are used to characterize random media. They are the velocity fractional fluctuation γ given by (Eq. 14 in [34])

$$\gamma = \left\langle \left(\frac{\delta C}{C_0} \right)^2 \right\rangle^{1/2}, \tag{3.1}$$

where C_0 is the average velocity of the medium and δC is the velocity variation about C_0. The spatial distribution of heterogeneities, δC, may be defined using one of various types of statistical functions. An autocorrelation function of the medium is often used. Autocorrelation functions can have a characteristic length, a, which dominates the wavelength-dependence of scattering during wave propagation. Typical autocorrelation functions used to characterize heterogeneity in the Earth include Gaussian, exponential, and von Kármán [30]. The exponential autocorrelation function is a special case of a von Kármán autocorrelation function.

Ru-Shan published several highly significant papers in the early 1980s on seismic scattering. The first paper that Ru-Shan published with his Ph.D. supervisor, Professor Keiiti (Kei) Aki, was an analysis of Born (e.g. single, weak) seismic scattering by an elastic heterogeneity [42]. In this paper, they showed how scattering could be analyzed as an equivalent source. They separately analyzed the cases of low-frequency scattering where the seismic wavelength is long compared to a, and the cases of high-frequency scattering where the seismic wavelength is small compared to a. They derived the scattering radiation patterns for P and S waves for both low- and high-frequency cases. Wu and Aki



(a) (b)

Figure 5: (a) Spherical coordinate system for P-wave incidence, and (b) the scattering patterns for different equivalent forces. (From [42])

(1985) [42] also showed how the radiation patterns differ depending on the character of the elastic heterogeneity, e.g. velocity difference, impedance contrast, or density contrast.

Fig. 5 (from [42]) shows an example of how scattering radiation patterns depend on the type of contrasts. Fig. 5(a) shows the spherical coordinate system used in the analysis of Rayleigh scattering and Fig. 5(b) displays the equivalent forces and scattering patterns for a plane P-wave incident on a density contrast (top), contrast in Lamé parameter λ (middle) and contrast in rigidity μ (bottom). Note that, depending on the type of contrast, the incident P wave may generate both scattered P and S waves. In the paper, Wu and Aki (1985) [42] also showed plots for the scattering from various combinations of contrasts in elastic parameters. This work provided a basis on which to improve our understanding of seismic scattering from different types of heterogeneities in the Earth and later for modeling seismic attenuation caused by scattering.

At the same time that Ru-Shan published his paper with Kei on elastic wave scattering from isolated heterogeneities, Ru-Shan and Kei looked at the issue of wave propagation through a medium with a distribution of random heterogeneities [41]. Aki (1969) [1] postulated that the coda of local earthquake waveforms, basically the long tail that follows the S-wave, is composed of S waves backscattered from medium heterogeneities. One important question that follows from this conjecture is what does this tell us about the

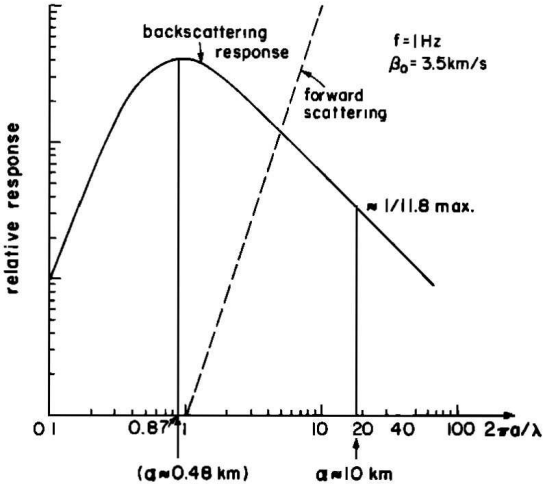


Figure 6: Relative backscattering strength for S-S scattering for media relation between medium correlation length, a, and incident S wave wavelength, λ. Values of a corresponding to 1 Hz S waves in a medium with S wave velocity of 3.5 km/s are listed below the x-axis label. (From [41])

character of heterogeneity in the Earth's crust. Wu and Aki (1985) [41] addressed this issue using several approaches. First they showed that a scalar wave equation gives similar results to the elastic wave equation for forward scattered high-frequency waves but that it does not give reliable results for low-frequency back scattered waves. This finding means that if they want to model coda waves as back scattered waves, an elastic wave equation must be used. They also looked at the relative amount of backscattered S waves that would result for a heterogeneous medium having a given characteristic scale.

Fig. 6 shows the relative amount of backscattering that occurs in a random medium as a function of the dimensionless parameter $2\pi a/\lambda = ka$, where λ is the wavelength of the incident S waves and k is the associated wavenumber. The characteristic length a is given on the abscissa for a 1-second period S wave in a medium with an S-wave velocity of 3.5 km/s. For this case, the maximum backscattering occurs when the characteristic length of the medium heterogeneity is 0.48 km. The wavelength for 1 Hz waves is 3.5 km so the scale of heterogeneity is small compared to the wavelength of the seismic waves when the backscattering is strongest.

3.1 Random heterogeneities: propagation and scattering in random media

The common practice in seismology for dealing with multiscale heterogeneities in the Earth is to smooth both the observed wavefield and the heterogeneity model. The objective of this smoothing is to allow one to ignore information about small-scale heterogeneities and obtain a deterministic model for the slowly varying, large-scale heterogeneities. Conversely, stochastic methods can obtain some statistical information about the small-scale heterogeneities from the statistics of the wavefield fluctuations. Statistical

parameters of the medium include information about the velocity and density perturbations in the medium: RMS perturbation, characteristic scale length, and the power spectrum, or correlation function. In this sense, deterministic and stochastic methods complement each other and allow one to explore multiscale complex media. The scale where deterministic and multiscale methods overlap varies depending on the medium, the wavelength of recorded waves, and the density of seismic recordings.

The study of stochastic characteristics of random media began by using forward-scattered waves. Early studies focused on the coherence of amplitude and phase of long-period seismic waves across arrays like the Large Amplitude Seismic Array (LASA) and the Norwegian Seismic Array (NORSAR) that were set up for the detection of nuclear explosions but whose data were ultimately used for many other purposes (e.g., [2]). Limited by the amount of information contained in these coherence functions, the medium model description was restricted to being a single layer of a uniform, isotropic random medium. Through the use of the scattering theory of Chrnov (1960) [11], several statistical parameters, such as the RMS velocity perturbation, the average scale length, and the total thickness of the layer, were inferred from the observed data.

Unfortunately, the single-layered model for interpreting amplitude and phase fluctuations did not fit the data well [50]. There was need to investigate the use of a model having multiple layers and different scale of heterogeneity in each layer. At the end of the 1980s, Ru-Shan worked with Stanley Flatté at the University of California, Santa Cruz, and introduced a new statistical observable, the angular coherence function (ACF) that defines the coherence in the seismic wavefield as a function of the angle of incidence. This new measure significantly increased the statistical information that could be obtained from data, and allowed the authors to derive a simple model of a layered, multiscale random medium [16] and apply it to data from NORSAR. Subsequently, Wu and Flatté (1990) [50] and Chen and Aki (1991) [10] introduced a new joint coherence function (JCF) or joint transverse coherence function (JTACF), and used the Rytov approximation to derive the theoretical relation between the joint coherence functions of array data and the heterogeneity spectrum of a layered random medium. All these theoretical work was based on a homogeneous background medium. Zheng and Wu (2008) [69] published a comprehensive JTACF theory for a depth-dependent background medium using the WKBJ Green's function. These developments in theory and methods greatly increased the amount of information obtained from the dataset.

Most researchers making array measurements of phase fluctuations in array data used the first arrival times of P waves. Line et al. (1998) [26] and Hong et al. (2005) [20] evaluated the relation between the JTACF and medium heterogeneity using numerical simulations. They found differences between the coherence functions derived from the simulated data and what was expected from the prior theory. Zheng and Wu (2005) [68] pointed out that the nature of the scattering in a random medium meant that the phase fluctuations were more reliably measured from unwrapped phase spectra rather than first arrival measurements. They were then able to resolve the differences between the numerical simulations and the previous theory.

3.2 Scattering and attenuation

With the realization that the heterogeneous Earth crust scatters waves came the idea that this scattering, which redistributes energy, attenuates the amplitudes of direct waves propagating between a source and a receiver. It was already well established that anelastic attenuation mechanisms remove energy from the wavefield. Laboratory measurements were made on rocks to better understand the mechanism of this anelastic mechanism (see e.g. Toksöz and Johnson, 1981 [28]). A compelling question was how big a role does scattering play in the measured attenuation of direct waves. Ru-Shan was one of the leaders in investigating this question. Here we discuss the use of single scattering models to calculate the amount and frequency dependence of the contribution of seismic scattering to attenuation. In the following section, we discuss the use of a multiple scattering approach that leads to a method for estimating the importance of scattering to seismic wave attenuation.

Ru-Shan published two papers on the contribution of single scattering to the attenuation of seismic waves in 1982 [34, 35]. In Wu (1982b) [35], Ru-Shan looked at the attenuation of the mean field for waves propagating in a random medium and the attenuation resulting from using the single scattering approximation. In this paper, he showed that attenuation predicted by both models is identical and does not depend on the statistics of the random medium. It does depend on γ and a. The predicted attenuation increases with increasing frequency. The predicted attenuation (Q^{-1} is shown as the solid line denoted with triangles in Fig. 7 for a selected values of γ and a that Ru-Shan chose as being appropriate from previous measurements of heterogeneity. Attenuation increases with increasing frequency. Fig. 7 also shows measurements of shear wave attenuation as a function of frequency made by Aki (1980) [3] in three regions of Kanto, Japan. The estimated attenuation of S waves at very low frequencies obtained using surface waves appears on the left portion of the plot. The field measurements show that there is a clear decrease in attenuation with increasing frequency above about 1 Hz while the value of attenuation at low frequency appears to be smaller than that for body waves at higher frequencies. Thus, one might infer that there is some frequency between about 0.1 Hz and 1 Hz where there is a peak in attenuation. The observations of attenuation are clearly in disagreement with the predictions made by the mean field theory and the single scattering approximation if scattering is a dominant contributor to overall seismic attenuation.

One issue that Ru-Shan pointed out in 1982 [35] is that measuring mean field attenuation in the field is not straightforward. He gives a detailed description of how one might measure it using seismic arrays. He argued that such measurements can be made using data from arrays such as LASA. He used amplitude and phase fluctuations reported by Aki in 1973 [2] to estimate mean field attenuation and compared it with predictions of his model. His results showed quite a good comparison.

Ru-Shan was not content to leave the subject of scattering attenuation with predictions made for a hard-to-achieve measurement condition. He wanted to model the attenuation in the direct wave that propagates between source and receiver. To accomplish

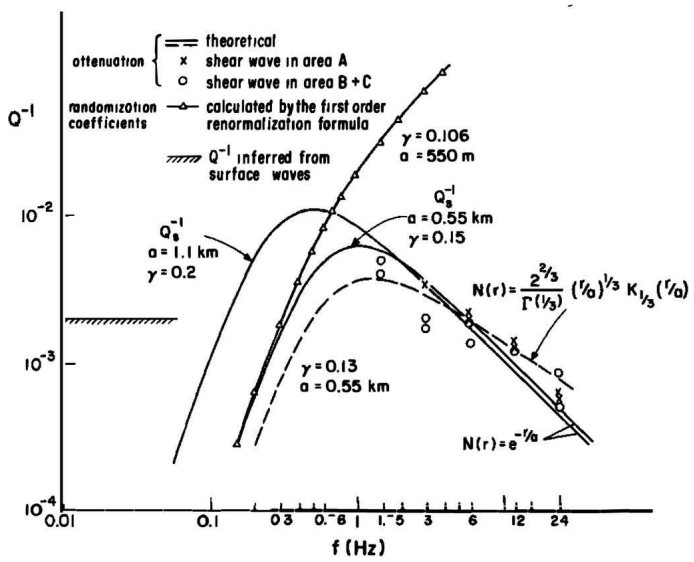


Figure 7: Attenuation of S waves as a function of frequency. Open circles and crosses show measurements of attenuation made by Aki (1980) [3] in Kanto, Japan. Solid curve with triangles shows attenuation predicted by mean field approach. Solid curves show attenuation predicted using a single scattering approach with attenuation being measured from the backscattered waves. Two solid curves show predictions using an exponential autocorrelation function to describe the medium heterogeneity with two values of γ and a. Dashed line shows attenuation predicted as for the solid curves except for a von Kármán autocorrelation function. (From [34])

this task, he began with the Born approximation to calculate the scattering caused by single scattering [34]. To make up for the use of the single scattering approximation in his model, he assumed that all energy that is back scattered is energy that is lost from the direct pulse. In this case, the backscattering is frequency dependent and depends on the statistics of the random medium. With this model, Ru-Shan found that there is a peak in attenuation and that the location of the peak depends on the approach he used to characterize the random medium (he used exponential and von Kármán autocorrelation functions). By appropriate choice of parameters, he was able to fit the field data shown in Fig. 7. As Ru-Shan pointed out, the good fits to the data ignore the expectation that some of the attenuation is likely caused by intrinsic attenuation.

3.3 Radiative transfer of seismic scattering

While Ru-Shan and others showed that a single backscattering model could predict a frequency dependent attenuation that is consistent with observations, the problem remained on how to determine the importance of scattering to overall attenuation. In 1985 and 1988, Ru-Shan published two important papers to address this question. One developed a theoretical model and the other applied the model to data. He summarized the importance of the work well in the first few sentences of the body of the theoretical paper [36] as follows:

> "Are the measured apparent attenuations for short-period seismic waves caused by anelasticity of the media or by scattering of the heterogeneities in the media? Is the single backscattering model a good approximation to the coda envelope decay or do we need a multiple scattering model which will have significant differences in describing the coda behavior from the single backscattering theory? These are long-standing problems."

Ru-Shan discussed various approaches for accounting for multiple scattering in models used to predict waveform characteristics [36]. He then proposed to borrow an approach from astrophysical optics to model the distribution of seismic energy in space and time, thus avoiding the complicated issue of phase interference that burdens many multiple scattering approaches. The approach is called the Radiative Transfer Theory and its roots are quite old (Schuster, 1905 [31]; Chandrasekhar, 1960 [5]). The theory includes multiple scattering. In 1985, Ru-Shan [36] translated the terminology of astrophysical optics into one appropriate for seismology. He introduced an important parameter that he called "Seismic Albedo" given as

$$B_0 = \frac{\eta_s}{\eta_s + \eta_a} = \frac{Q_s^{-1}}{Q_s^{-1} + Q_i^{-1}}, \tag{3.2}$$

where B_0 is albedo, η_s is scattering coefficient, η_a is anelastic absorption coefficient, Q_s^{-1} is the seismic quality describing attenuation caused by scattering, and Q_i^{-1} is the seismic quality describing attenuation caused by intrinsic mechanisms, such as those convert seismic energy into heat. Thus, albedo gives the ratio of scattering loss to total loss. In 1985 [36], Ru-Shan developed the radiative transfer theory for isotropic radiation from a seismic source and solved the resulting integral equation to obtain the distribution of normalized energy, e.g. energy normalized by source radiated energy, over space for different values of seismic albedo. These distributions can be thought of as measures of the integrated energy that passes through a given distance from the receiver.

Fig. 8 shows examples of the energy distributions. The middle panel shows total energy for different albedos. One interesting result that Ru-Shan noted and explained is that at some distances from the source and depending on B_0, the integrated energy is actually greater than the source radiated energy. As Ru-Shan explained, this phenomenon results from multiple scattering that causes a given packet of energy to pass by a given position multiple times and thus be counted in the integrated energy multiple times.

Ru-Shan followed his 1985 paper with one in 1988 [43] where he analyzed digital data collected in the Hindu Kush region of Northeastern Afghanistan. To evaluate data as a function of the source-receiver distance, Ru-Shan chose to evaluate several earthquakes that occurred at various distances from one station. This allowed him to ignore effects of local site amplification if he used data from multiple stations. He only needed to find a method to normalize for differences in radiated source energy among the earthquakes. He used the well-accepted coda normalization method to normalize seismograms for

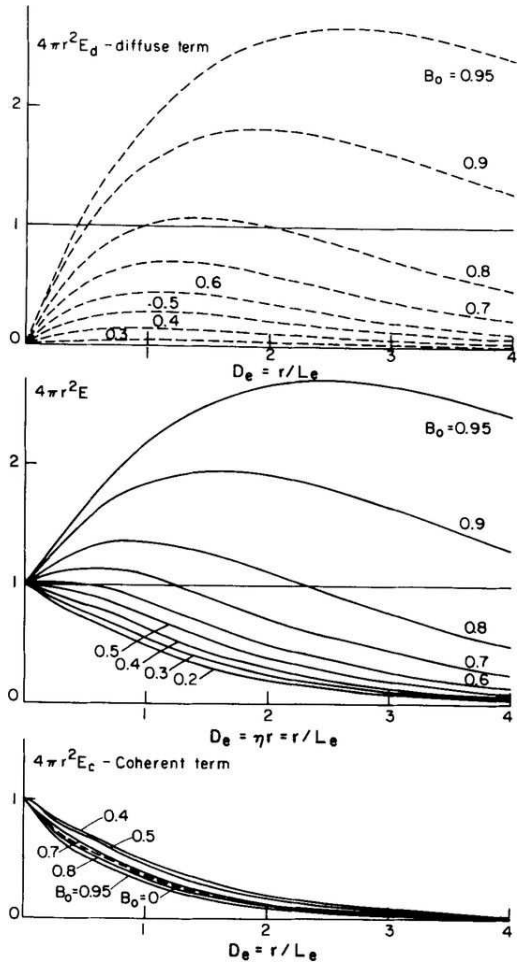


Figure 8: The normalized energy density distribution curves $4\pi r^2 E(r)$, where r is the propagation distance from the point source. At the top are the curves of the diffuse term, at the bottom are those of the coherent term; in the middle are the curves of the sum of the two terms. Here D_e is the numerical extinction distance, $L_e = 1/\eta_e$ is the extinction length of the medium, $\eta_e = \eta_s + \eta_a$ is the extinction coefficient, where η_s and η_a are the scattering coefficient and the absorption coefficient, respectively. $B_0 = \eta_s/(\eta_s + \eta_a)$ is the medium seismic albedo. (From [36])

variations in radiated source energy. He then used a Fourier Transform to find the energy as a function of distance and frequency. He found that at frequencies less than 1 Hz, the data appeared to predict very high albedo. He suspected that these low frequencies were contaminated with surface wave energy thus making direct application of the radiative transfer theory difficult. Normalized energy at high-frequency fell off very fast as a function of distance and indicated that the albedo is less than 0.5. Fig. 9 shows the plot of normalized energy versus distance for 1.5 and 2 Hz [43]. The figure also shows the distributions of energy calculated using the radiative transfer theory for several values of seismic albedo. While the authors were unable to reliably estimate the seismic

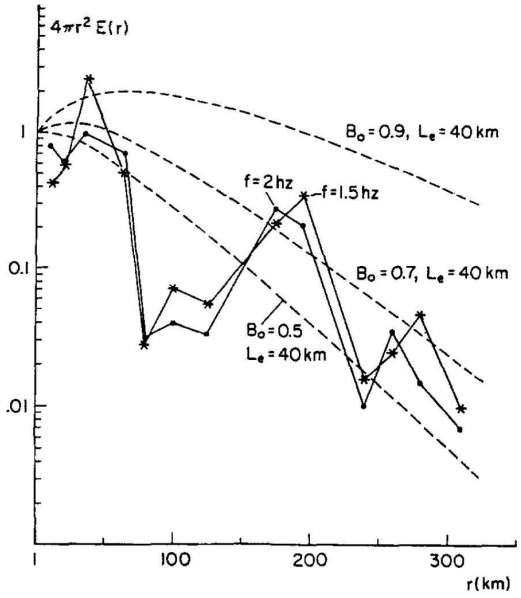


Figure 9: The comparison between the observed $4\pi r^2 E(r)$ for $f=1.5$ and 2 Hz (*EW* components) at station PEN in Hindu Kush and the theoretical predictions of different B_0's. The curve of $B_0=0.9$ is the prediction from the constant Q_i $(=2500)$ model, which does not match with the observation. (From [43])

albedo, they estimated that it is in the range of 0.5-0.7. This conclusion means that at high frequencies the importance of scattering is less than it is at low frequencies.

The fit to the radiative transfer theory model is done in the frequency domain and has the potential to ignore energy that arrives later than the end of the time window used to compute the spectra. Wu and Aki (1988) [43] thus evaluated whether the single scattering model or a diffusion model could fit the time history of the envelopes of the seismograms. The single scattering model is appropriate for a weak scattering regime and the diffusion model is appropriate for a strong scattering regime. Wu and Aki (1988) [43] pointed out that the diffusion model had been shown to work well to fit the envelopes of seismograms recorded on the moon. They first found that seismogram envelopes from the Hindu Kush could not be fit by one predicted by the diffusion model. They then showed that it can be well fit by an envelope predicted by the single scattering model. Combining the frequency domain analysis conducted using the radiative transfer theory with the successful fits to envelope traces in the time domain using the single scattering model, Wu and Aki (1988) [36] concluded that scattering is not a dominant mechanism for attenuation of high-frequency seismic waves. This conclusion has been later confirmed by studies conducted by many authors. There is some indication that at frequencies below about 2 Hz, scattering is the dominant attenuation mechanism (see e.g. [30]).

The study using the radiative transfer theory has led to a large number of papers that extended the theory to more complicated cases and a solution of the radiative transfer equation in the time domain was found (e.g. Zeng, 1991 [66]). A discussion of studies

conducted using the radiative transfer theory over many years following the pioneering work of Wu (1985) [36] and Wu and Aki (1988) [43] can be found in Sato et al. (2012) [30].

4 One-way wave propagation and one-return wave modeling (1990s–)

Ru-Shan made remarkable contributions in one-way wave propagation and one-return wave modeling. One-way methods use the scattering formulation in which the medium is considered as a background model with a perturbation. One main issue of the one-way methods is the wide-angle accuracy of the one-way wave propagator with respect to a pre-defined propagation direction. The accuracy depends on how to properly handle scattering by strong medium perturbations. To address this issue, Ru-Shan developed one-way methods using local propagators (e.g., beamlets and dreamlets; see Section 5) to reduce the perturbation. For strong scattering, Ru-Shan's key contribution helped understand the divergence issue of summing up the Born scattering series. He then developed the multiple-forward-single-backward renormalization method to solve the problem. Ru-shan also developed the thin-slab method and the super-wide angle screens to improve the wide-angle accuracy. He used his one-way methods to model long-distance high-frequency elastic Lg wave propagation excited by nuclear tests in the 1990s, a feat that is extremely difficult to achieve using the full-wave finite-difference method. Note that, although currently the reverse-time migration (RTM) is most commonly used in seismic imaging, one-way imaging methods dominated the field for more than 30 years since the 1970s. The many novel developments in one-way methods can provide fruitful ideas for future work in seismic inversion.

4.1 Acoustic and elastic one-way and one-return propagators: Generalized Screen Propagator (GSP) and De Wolf approximation

Ru-Shan pioneered the development of a suite of methods for efficient modeling of one-way wave propagation and one-return wave modeling in complex acoustic and elastic models. The methods are based on the multiple forward-scattering and single back-scattering approximations. These algorithms simulate primary reflections from heterogeneities and neglect the internal reverberations while accounting for all forward scattering phenomena, such as focusing/defocusing, diffraction, refraction, and interference. Ru-Shan and his colleagues implemented the one-return method using an iterative marching algorithm shuttling between space- and wavenumber-domains using FFTs. The one-way and one-return methods are accurate and highly efficient for seismic-wave modeling, particularly for long-distance propagation of high-frequency seismic waves. Ru-Shan and his colleagues have successfully used these methods for seismic modeling, imaging, computing finite-frequency sensitivity kernels for seismic tomography and migration velocity analysis, and modeling crustal guided waves.

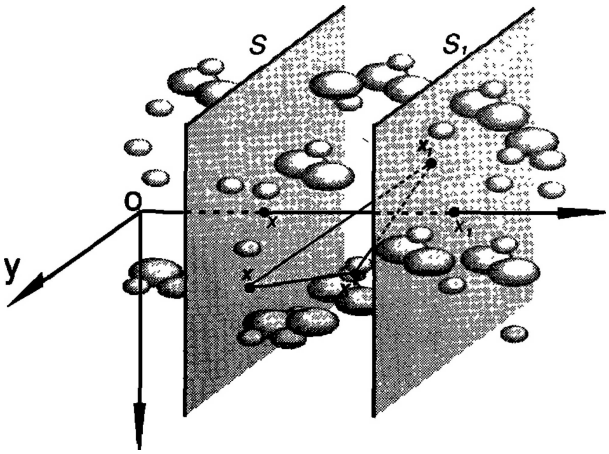


Figure 10: Geometry for derivation of the wide-angle complex-screen method. The medium is sliced into thinslabs. For scattering and propagation effects of the thinslab between plane S and S_1, when the wavefield is known on S, the wavefield on S_1 is obtained using the forward propagation approximation. (From [37])

In 1994, Ru-Shan [37] developed a wide-angle complex-screen method for an elastic one-way propagator. Fig. 10 depicts the basic concept of a one-way wave propagator. A 3D inhomogeneous medium is sliced into a series of thin slabs perpendicular to the main wave-propagation direction. With known wavefield incident at plane S, the wavefield at plane S_1 is calculated using the incident waves and all forward scattered waves from heterogeneities between S and S_1. Ru-Shan [37] derived a system of equations for one-way wave propagation in arbitrarily heterogeneous elastic media using elastic Rayleigh integrals and the local Born theory. Compared with the scalar phase-screen, the elastic complex-screen has the following features. (1) For P-P scattering and S-S in-plane scattering, the elastic complex-screen acts similar to scalar phase screens while both are 3-component vector fields. (2) For P-S and S-P conversions, the screen is no longer a pure phase-screen and becomes complex (with both phase and amplitude terms); both conversions are dominated by the shear modulus perturbation. The numerical results obtained using the complex-screen method agree well with 3D finite-difference wave-equation modeling and analytical solutions.

In 2003, Ru-Shan [39] studied the theory for one-wave wave propagation using the Born, Rytov, and De Wolf approximations [14,15]. He developed three dual-domain generalized screen propagators, including the hybrid pseudo-screen, the wide-angle Padé-screen, and the higher order generalized screen propagators. Compared with conventional one-way propagators, these methods can properly handle large-angle wave propagation fairly accurately in media with strong velocity contrasts. The renormalized forward propagator is given by

$$G_f = \sum_{m=0}^{\infty} \left[G_0 \varepsilon_f \right]^m G_0, \tag{4.1}$$

and the renormalized incident field is

$$u_f = \sum_{n=0}^{\infty} \left[G_0 \varepsilon_f\right]^n u^0. \tag{4.2}$$

The De Wolf approximation becomes

$$u = u_f + G_f \varepsilon_b u_f. \tag{4.3}$$

In these equations, ε_f and ε_b are forescattering and backscattering potentials, respectively, G_0 is the Green's operator for the background medium, $G_f \varepsilon_b u_f$ is the primary backscattered field under the De Wolf approximation.

The above approximation is a multiple-forescattering, single-backscattering approximation. To model full wavefield including multiple backscattering using renormalized forward Green's function, Ru-Shan introduced De Wolf series [39, 62],

$$\begin{aligned} u &= u_f + G_f \varepsilon_b u_f + G_f \varepsilon_b G_f \varepsilon_b u_f + \cdots \\ &= \sum_{m=0}^{M} \left[G_f \varepsilon_b\right]^m u_f. \end{aligned} \tag{4.4}$$

The De Wolf approximation is the first order approximation. The difference between the De Wolf series and the Born series is in the Green's functions. De Wolf series uses the forward-scattering renormalized Green's function, while Born series uses the background Green's function. Consequently, the De Wolf series avoids the serious divergence problem of Born series, since the backscattering coefficient is always smaller than one.

Fig. 11 shows an example of 3D migration using the generalized screen propagator (GSP) that is based on the De Wolf approximation. It compares 2D slices of 3D images obtained using the C3 subset of 3D SEG/EAGE salt model. The left column is for vertical slices at Line 90 and the right column is for horizontal slices at depth 126. Panels (a), (b) and (c) are velocity models and images from the split-step Fourier method and the GSP method, respectively.

Wu et al. (2006) [62] developed two versions of the one-return method based on the De Wolf approximation: one is the wide-angle, dual-domain formulation, and the other is the screen formulation with small-angle approximation. Their results demonstrated that the forward scattered waves are mainly controlled by velocity perturbations, while the backscattered waves are primarily dominated by impedance perturbations. Fig. 12 schematically illustrates the flowchart of the thin-slab method for implementing the one-return approximation, where (a) shows the original medium sliced into thin-slabs, and (b) depicts the iterative procedure of transmitted and reflected wave calculations using the one-return approximation. They numerically demonstrated the effectiveness of the thin-slab and screen methods, and the wide-angle capability of the dual-domain implementation [61]. Fig. 13 is the 3D French model for numerical test. Fig. 14 compares synthetic seismograms calculated using finite-difference (solid lines) and thin-slab methods

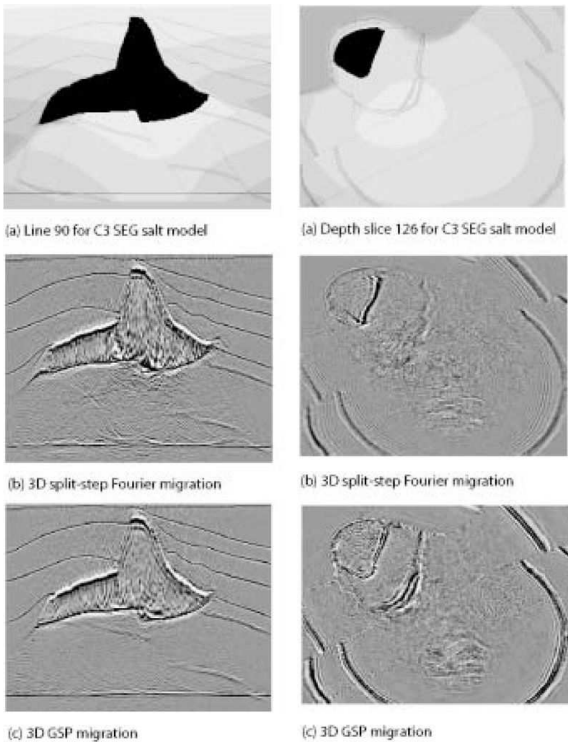


Figure 11: Comparison between 2D slices of 3D images obtained using different methods for the C3 subset of the 3D SEG/EAGE salt model. The left panels are vertical slices at Line 90, and the right panels are horizontal slices at depth 126. Panels (a), (b) and (c) are the velocity models, images produced using the split-step Fourier method and images generated using the GSP migration, respectively. (From [39])

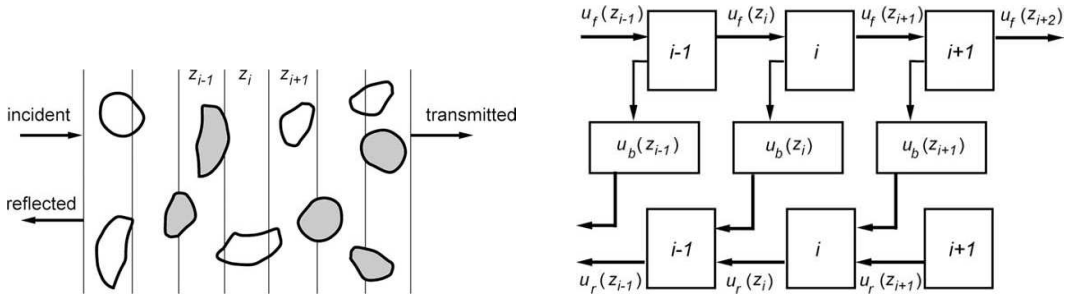


Figure 12: Schematic illustration of the thin-slab method for implementing the one-return approximation: (a) the original medium is sliced into thin-slabs; (b) the iterative procedure of transmitted and reflected wave calculations by the one-return approximation. (From [62])

(dashed lines) methods. The results show generally good agreement between the two methods. However, the one-return approach is approximately 2-3 orders of magnitude faster than an elastic FD algorithm.

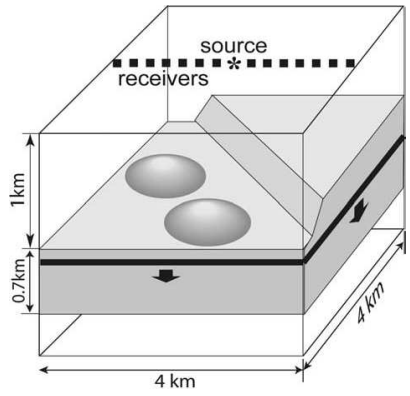


Figure 13: 3D French model. The fat solid lines indicate the horizontal surface at which the thin-slab method is connected to the reflectivity method. The structure in grey has a perturbation of -10% in both P and S velocities. (From [62])

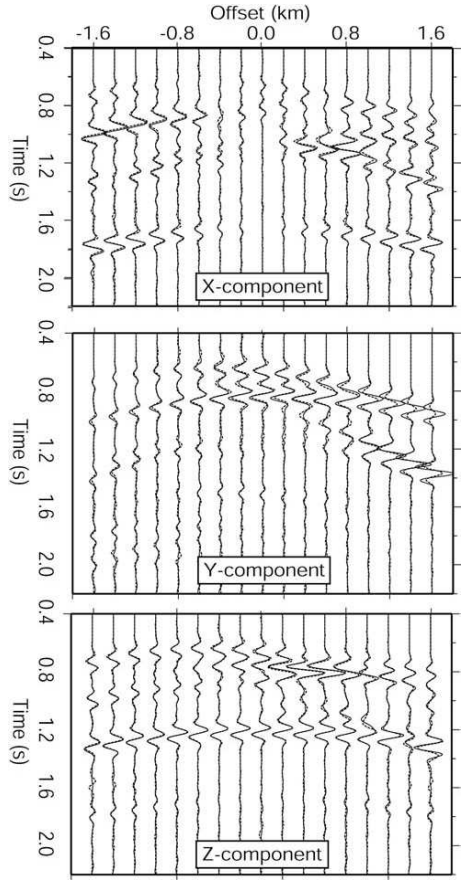


Figure 14: Comparison of synthetic seismograms calculated using finite difference (solid lines) and thin-slab methods (dashed lines). From top to the bottom are X-, Y- and Z-components of displacement, respectively. (From [62])

Ru-Shan also developed one-way propagators with an alternative geometry, i.e., with a horizontal main propagation direction and vertical screens. This geometry is suitable for calculating turning waves for velocity tomography, and for simulating crustal guided waves. Wu et al. (2000) [54] developed the half-space one-way propagator for simulating SH-wave propagation in the crustal waveguide. They used the method to simulate the arriving regional wave Lg in different types of complex crustal waveguides including those with small-scale random heterogeneities. They studied the influence of these heterogeneities on Lg amplitude attenuation and Lg coda formation. Wu et al. (2000) [53] used the one-way propagator to study the energy partition and attenuation of Lg-waves in complex crustal waveguides with both large-scale structures and small-scale random heterogeneities. Their studies specifically addressed leakage attenuation caused by large-angle forward scattering from random heterogeneities, removing the guided wave energy out of the trapped mode and those leaking into the mantle. The equivalent Q for leakage attenuation as a function of character scale of the random heterogeneities agrees well with the scattering theory.

4.2 Super-wide-angle GSP

To overcome the angle limitation of conventional one-way propagator, Ru-Shan developed the super-wide-angle one-way propagator [24, 25, 52]. By combining two orthogonally propagating one-way wavefields using wavefront reconstruction, the method is accurate for super-wide angle (greater than 90°) propagation and can model turning waves. The method could be used for imaging steep subsalt reflectors and overhanging salt flanks. They numerically verified the efficacy of the method. Fig. 15 displays the modeling result of turning waves in a linear gradient velocity model obtained using the super-wide angle one-way operator.

5 Beamlet and dreamlet (1990s–)

5.1 Concept of beamlets and frames

Mosher, Foster, and Wu (1996) [29] first introduced the idea of beamlet. They implemented the beamlet in the frequency-space-wavenumber domain. The beamlet, like a flashlight beam localized in space and direction, uses local reference velocities. The use of the local velocities decreases the medium perturbations in the scattering formulation of one-way propagators (Fig. 16). Therefore, the beamlet provides an approach to improving propagation accuracy. In contrast to the beamlet method, global one-way methods use a global reference velocity, leading to large perturbations and large errors in the propagators. Another advantage of using the beamlet is that a beamlet is automatically equipped with the local propagation angles (see Fig. 17), which is important in correcting the image amplitudes for the finite acquisition aperture. Physically, the properties of the beamlets are attractive. In fact, a similar idea was introduced in 1946 by Gabor [17]

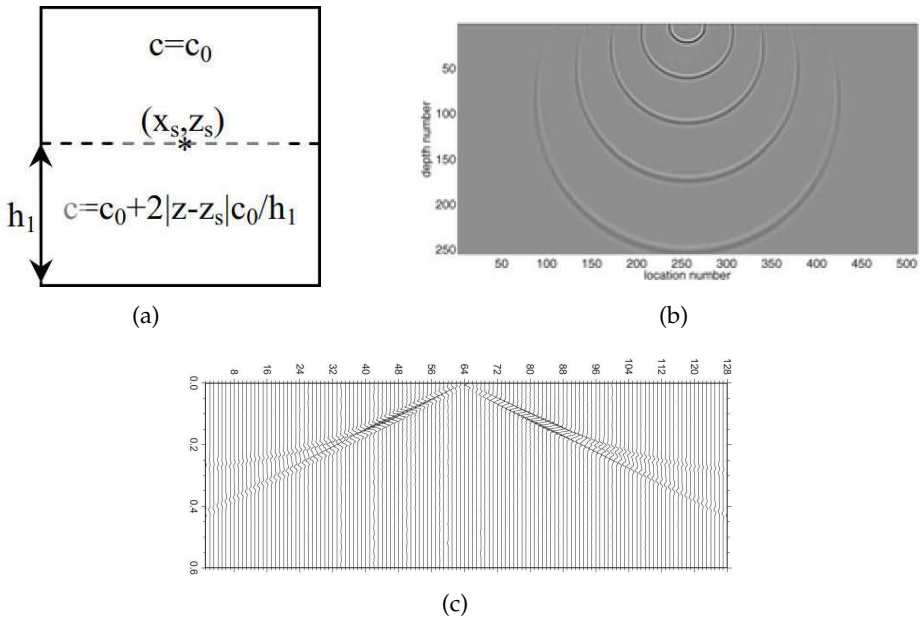


Figure 15: Modeling turning waves using a wavefield reconstruction method. (a) velocity model; (b) snapshots obtained using the wavefield reconstruction method; (c) synthetic seismograms observed on the surface. The vertical coordinate is reduced traveltime and the horizontal axis is trace number on the surface (the point source is in the middle). The direct arrivals are weakened to render the signals of turning waves more clearly. (From [52])

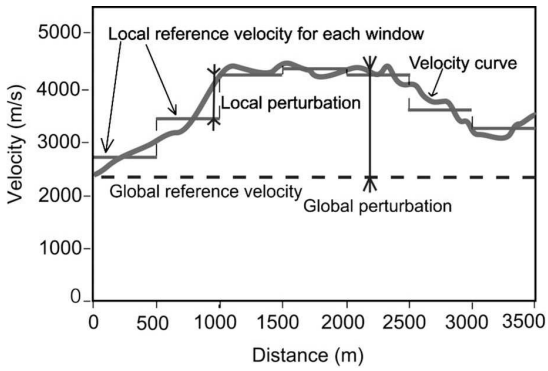


Figure 16: Background velocity profile (using local reference velocities) and local perturbations in the local perturbation theory, compared with the global reference velocity (dashed line) and global perturbations in standard perturbation methods. (From [58])

who used overlapping Gaussian window modulated sinusoids of different frequencies to represent communication signals. Daubechies [12, 13] later showed that Gabor's reconstruction was not stable and redundancy (i.e., smaller distance between neighboring Gaussian windows) is needed to achieve a stable signal reconstruction. In Gabor's original decomposition/reconstruction, there is no redundancy (or redundancy factor is one).

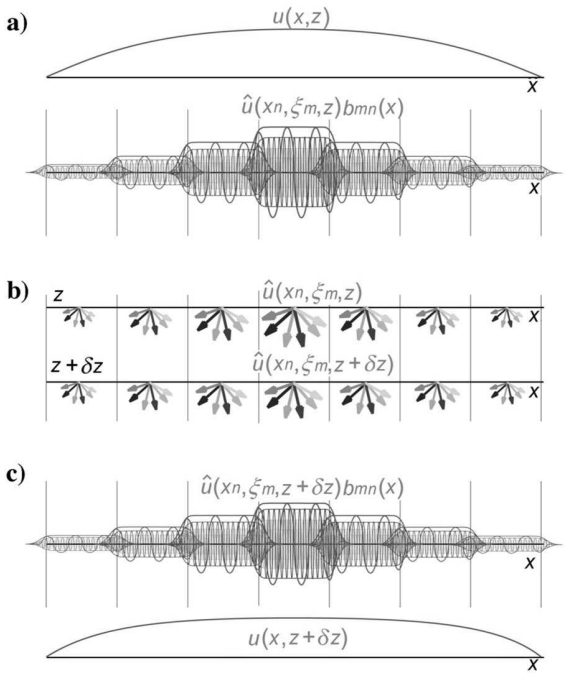


Figure 17: One-way wave propagation in the beamlet domain. (a) Space-domain wavefield is decomposed into the summation of beamlets. (b) Wave propagation in the beamlet domain. (c) Space-domain wavefield at depth is reconstructed using the new beamlets. (From [58])

Daubechies [12,13] used frame vectors in signal analysis of the windowed Fourier transform. Loosely speaking, frame vectors are similar to basis vectors. Both frame vectors and basis vectors span the same space. However, the number of frame vectors is more than the number of the basis vectors. Therefore, the use of the frame vectors to decompose and reconstruct signals is redundant.

5.2 Gabor-Daubechies beamlets

Ru-Shan pioneered the use of *frames* and the *dual frames* to construct new beamlet propagators [46] in 2001, to create new imaging tools [47] in 2002, and target-oriented beamlet migration based on the Gabor-Daubechies (G-D) frame decomposition [9]. The G-D beamlet imaging shows a much better image than the Kirchhoff migration (Fig. 18).

5.3 Local cosine beamlets

Ru-Shan introduced the *local cosine beamlets* (LCB) [58] to overcome some computational efficiency drawback in the G-D beamlets. The G-D frames vectors are redundant and computationally inefficient. To reconstruct a signal using the frame coefficients, one needs to use the dual frame vectors [47]. On the other hand, the LCB is a set of basis

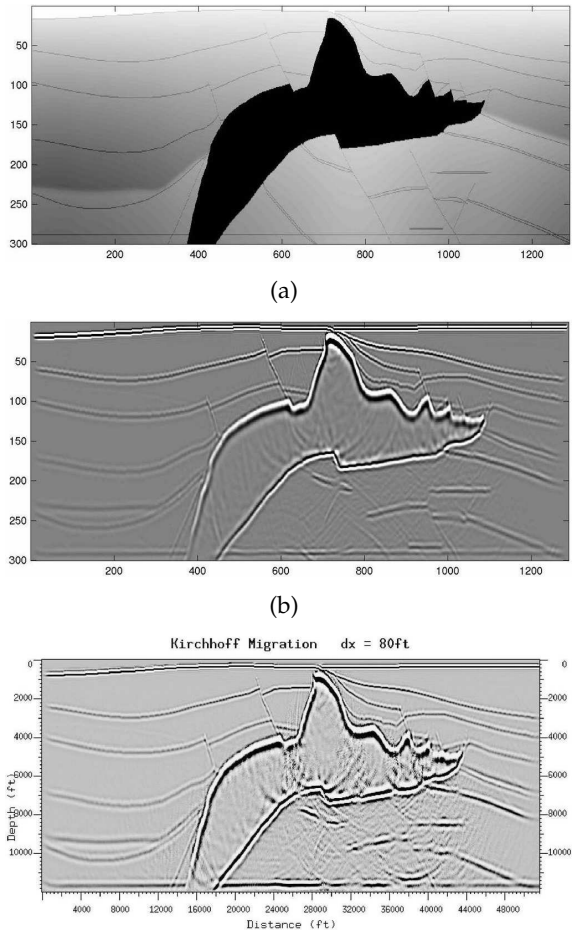


Figure 18: (a) Velocity model of the 2D cross-section AA' of the 3D SEG/EAGE salt model; (b) Image obtained using beamlet migration; (c) Image obtained using Kirchhoff migration. (From [47])

vectors without redundancy and there is a fast LCB transform algorithm available [58]. LCB can also handle large-contrast media and produce high-quality images (Fig. 19).

5.4 Beamlet transform applications: true-amplitude imaging and directional illumination

The amplitude of a seismic migration image is usually not proportional to the local reflectivity. The true-amplitude imaging is to preserve the local reflectivity image amplitude. Ru-Shan contributed greatly in this area. He found that acquisition system plays a bigger role than the propagator amplitude accuracy in true-amplitude imaging. He further showed that it is necessary to correct the image amplitude using the directional illumina-

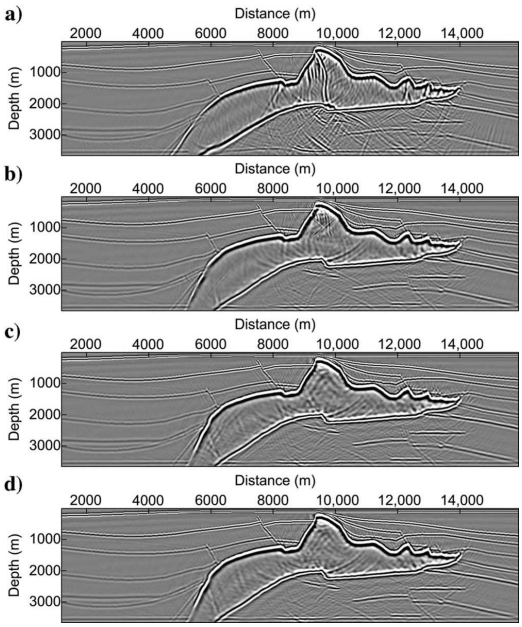


Figure 19: Comparison of image quality and cost of poststack migration for the 2D cross-section AA' of the 3D SEG/EAGE salt model using different methods. (a) Split-step Fourier method. Computing cost is set to 1.00. (b) GSP method. Computing cost is approximately 1.53. (c) LCB method. Computing cost is approximately 1.55. (d) LCB method. Computing cost is approximately 1.66. (From [58])

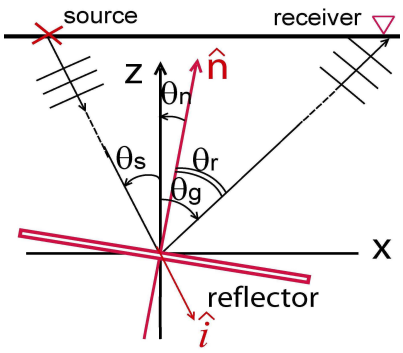


Figure 20: The definition of reflection (θ_r) and dip (θ_n)angles. For the same receiver array aperture, the received energy depends on the dip of the reflector. (From [56])

tion in local dip/reflection angle domain. His true-amplitude imaging methods improve amplitude-balanced images and reduce artifacts.

Wu and Chen in 2006 [48] developed a method to evaluate directional illumination and acquisition aperture efficacy using wave theory-based beamlet decomposition. They introduced the image conditions in the local angle domain (Fig. 20), defined the local image matrix, and calculated the direction illumination in the image space.

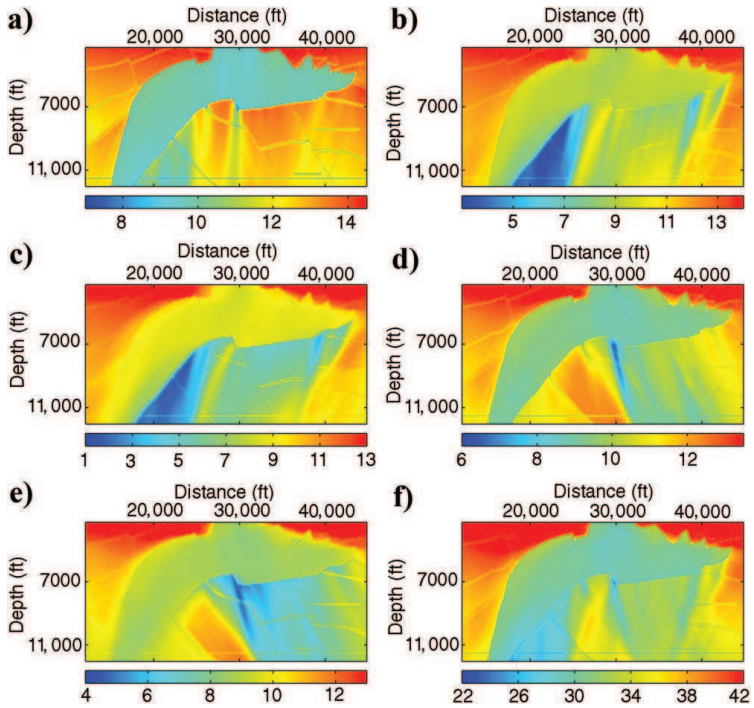


Figure 21: Directional illumination album for the dominant frequency (15 Hz) from all the 325 shots for a 2D cross-section of the 3D SEG/EAGE salt model: (a) vertical direction; (b) -30° (from vertical to left); (c) -45° (from vertical to left); (d) $+30^\circ$ (from vertical to right); (e) $+45^\circ$ (from vertical to right); and (f) total illumination intensity. (From [48])

The acquisition-aperture efficacy matrix and acquisition dip-response vector can be used to quantify the efficacy of an acquisition configuration at a given subsurface point. As numerical examples, Wu and Chen (2006) [56] calculated the direction illumination maps and acquisition dip response maps in different velocity models. Fig. 21 depicts the 15-Hz directional illumination album for a 2D cross-section of the 3D SEG/EAGE salt model, where (a) to (f) are illumination intensities for different angles and for total intensity. Fig. 22 shows the acquisition dip response album of the 2D cross-section of the 3D SEG/EAGE salt model, where (a) to (f) are response intensities at different dipping angles and for total response intensity.

Wu et al. [49, 56] developed an angle-domain amplitude correction method for true-amplitude wave-equation migration. The correction formulation relates the local image matrix to the local scattering matrix. The local image matrix bears the footprints of the acquisition aperture and propagation path effects, and the local scattering matrix is directly related to the medium property. They obtained the local image matrix using beamlet decomposition. They carried out the acquisition aperture correction using the acquisition aperture efficacy matrix or acquisition dip response, which includes effects from both acquisition configuration and complex overburden structures. Fig. 23 shows that the

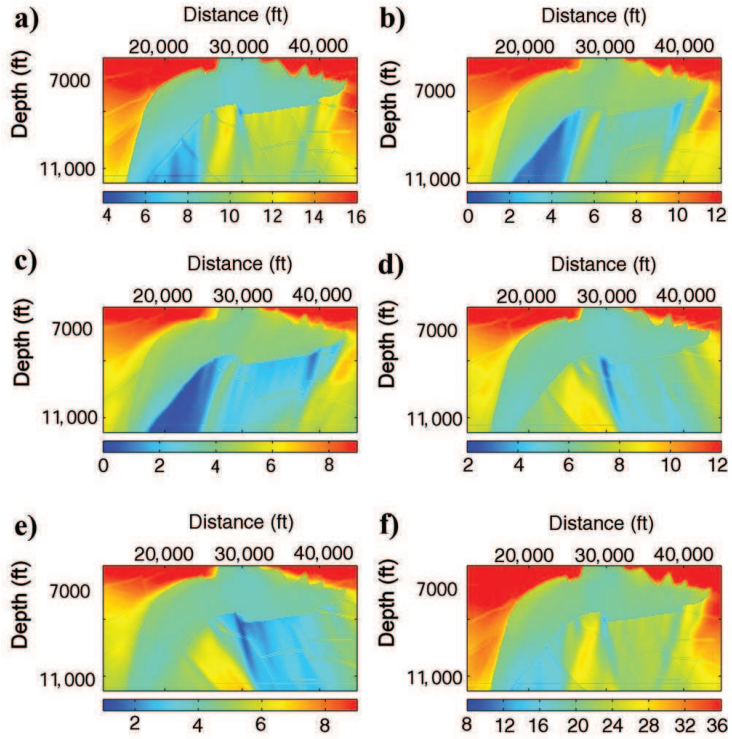
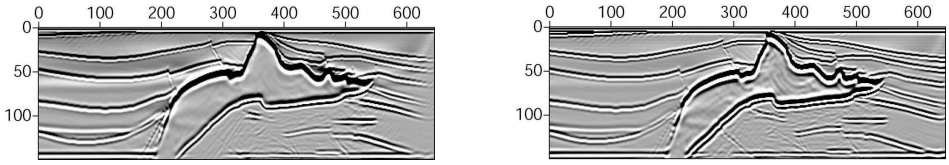


Figure 22: ADR album of a 2D cross-section of the 3D SEG/EAGE salt model from all the 325 shots with 176 left-hand-side receivers for each shot: (a) horizontal dip; (b) $+30^\circ$ (down from horizontal); (c) $+45^\circ$ (down from horizontal); (d) -30° (up from horizontal); (e) -45° (up from horizontal); (f) total response intensity. (From [48])



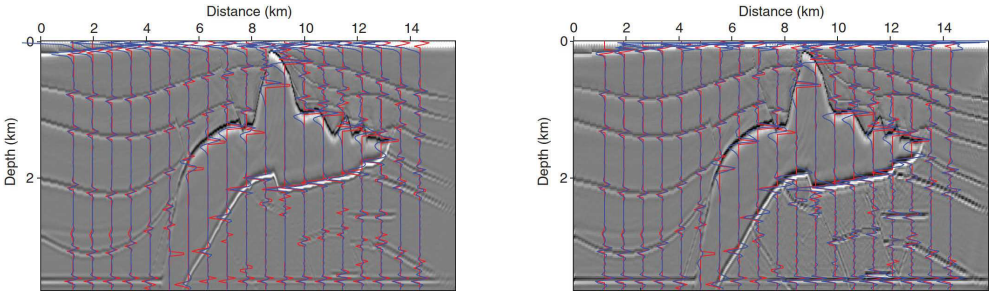
(a) Image without correction (b) Image after correction

Figure 23: Comparison of the images for a 2D cross-section of the 3D SEG/EAGE model before and after the amplitude correction. (From [56])

angle-domain amplitude correction greatly reduces the distortion of image amplitudes caused by the acquisition aperture effect.

Wu et al. (2004) [49, 56] developed two types of correction methods: one for common reflection-angle images and the other for total strength images. These methods significantly improve the amplitude fidelity and image quality.

The concept of the acquisition dip response (ADR) correction not only is applicable for one-way propagators but also for full-wave reverse-time migration [65] to achieve true-



(a) (b)

Figure 24: The migration image compensated by (a) source illumination and (b) ADR. Superimposed traces are the theoretical normal reflectivity (red) and the image amplitudes (blue) [65].

reflection imaging. Local beams (beamlet), local angle-domain, and local phase-space are all closely linked. Xie et al. (2006) [64] developed a full-wave-equation-based method for seismic illumination analysis. They used a local plane-wave analysis method to calculate localized directional energy fluxes for both source and receiver waves. They obtained an illumination matrix and applied it to target-oriented or volumetric illumination analyses. This method is flexible and practical for illumination analysis in complex 2D and 3D velocity models with nontrivial acquisition and target geometries.

Yan et al. (2014) [65] developed an amplitude correction method using the acquisition dip response (ADR). In the method, the mono-frequency (the dominant frequency) Green's functions from shot/geophone locations are generated and decomposed locally into incident/scattered plane waves to construct ADRs of different dip angles. The conventional RTM image is decomposed into a stack of common dip images, which are compensated individually using the corresponding ADRs. Then, the corrected images are stacked again to form the final image. They numerically demonstrated the efficacy of the method. Fig. 24 shows RTM images for a 2D cross-section of the 3D SEG/EAGE salt model with the source illumination correction in (a) and the ADR correction in (b). The theoretical normal reflectivity (red) and the image amplitude (blue) are superimposed on the images in Fig. 24. The ADR corrected image reproduces true impedance contrast of subsurface structures, particularly for steeply dipping structures and in subsalt regions. In true-reflection imaging, the image amplitude is proportional to the local reflectivity.

5.4.1 Dreamlet concept

Imaging in the compressed domain is a dream for geophysicists. Dreamlet = drumbeat (time-frequency, $t-\omega$, localization) + beamlet (space-wavenumber, $x-k$, localization). Beamlets represent space localization using a windowed $x-k$ transform. Because seismic events are sparse in time, it is better to also have time localization using a windowed $t-\omega$ transform. In fact, seismic events are sparse in both time and space, most of the dreamlet coefficients are close to zero, and this property can be used to significantly compress seismic waveform datasets.

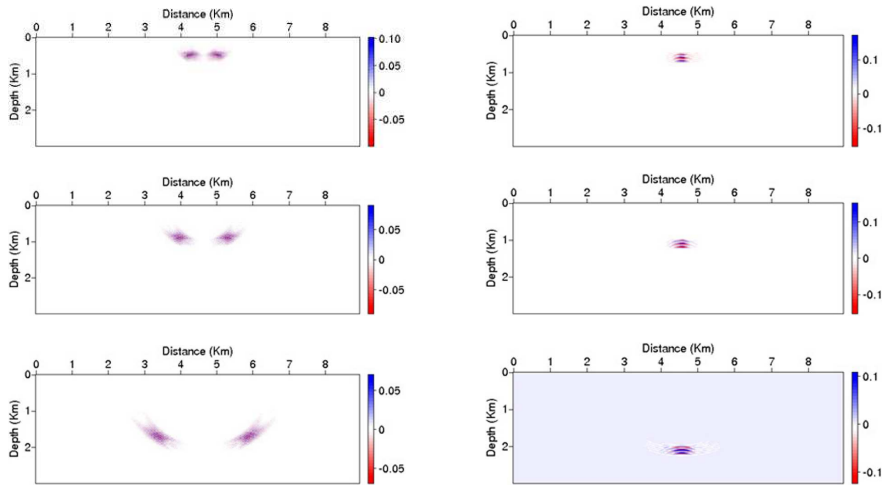


Figure 25: Snapshots of a single dreamlet propagation. Because of the symmetrical property of the local-cosine basis, the dreamlet has two symmetric lobes of packets. On the left is the dreamlet of $\bar{\xi}=8\Delta\bar{\xi}$, and on the right $\bar{\xi}=0$. ξ is the horizontal wavenumber and $\Delta\xi$ is the sampling interval of ξ. (From [59])

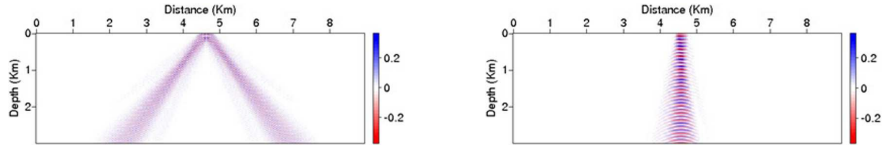


Figure 26: Wave spreading of beamlet propagation. On the left is the beamlet of $\bar{\xi}=8\Delta\bar{\xi}$, and on the right, $\bar{\xi}=0$. (From [59])

Wu et al. (2008) [59] employed dreamlets for wave propagation and imaging (Fig. 25). Compared with the dreamlet propagation that is localized in time-space, the beamlet spreads its energy along the entire wave path (i.e. not localized in time) (Fig. 26) [59].

5.5 Dreamlet wavefield compression and denoising

Because of the complete localization of the dreamlet (time-space-frequency-wavenumber localization), using dreamlets to decompose seismic data can achieve significantly better compression compared with curvelets [4, 51] and beamlets (Fig. 27). Dreamlets have the highest compression ratio. Because of this property, dreamlets can also be used for denoising [22].

5.6 Dreamlet imaging and migration

Imaging using dreamlets with selected significant coefficients improves the computational efficiency because the number of significant coefficients are small. Wu and his colleagues successfully employed the dreamlet concepts in prestack depth migration

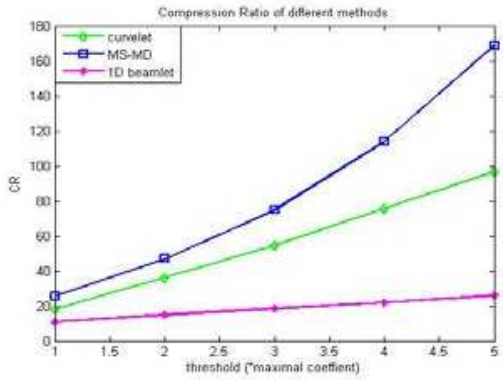


Figure 27: Seismic-wavefield compression ratios (CR) of different methods. As shown in the figure, the threshold varies from 1% to 5% of the maximal coefficients in each decomposition method, and the blue-square, green-diamond, magenta-asterisk lines stand for the CR's of multi-scale, multi-dimensional local harmonic (self-similar dreamlet), curvelet, and beamlet methods, respectively. (From [59])

with source-receiver survey sinking [33], in Gabor-frame-based Gaussian packet migration [18], and in migration with Gaussian wave packets based on Poincare wavelets [19].

6 Strong-nonlinear FWI and Direct Envelope Inversion (DEI) (2010s–)

6.1 Background

In full-waveform inversion (FWI), the observation is recorded seismic traces $d_{obs}(r,s,t)$, where r, s are the receiver, source locations, respectively, and t is time. The goal of FWI is to find *the* true subsurface model m_{true} such that the forward modeled seismic traces in m_{true} match the observation

$$A(m_{true}) = d_{obs}(r,s,t), \tag{6.1}$$

where A is a modeling operator solving the wave equation in model m_{true} for the source and receiver (s,r). The model is usually a 3D model such as the P-wave velocity.

How to find the true model m_{true}? In FWI, one guesses an initial model m_0 that is hopefully *not far* from m_{true}, such that we can use a perturbation expansion

$$A(m_0+\delta m) \approx A(m_0) + A'(m_0)\delta m + \mathcal{O}(\delta m) = d_{obs}(r,s,t) = A(m_{true}), \tag{6.2}$$

or

$$\delta d = d_{obs}(r,s,t) - A(m_0) \approx A'(m_0)\delta m, \tag{6.3}$$

from which δm can be solved if the terms in $\mathcal{O}(\delta m)$ can be ignored. $A'(m_0)$ is called the Fréchet derivative that expresses a linear relationship between the wavefield perturbation δd and the model perturbation δm. However, it is possible that the initial model

m_0 may not be close to the true model. In this case, one can update the model from $m_0 \to m_1 = (m_0 + \delta m)$ using δm solved from Eq. (6.3). We then use m_1 as the new initial model to repeat the process until some convergence criterion can be achieved. Unfortunately, the convergence is usually slow or not guaranteed at all, which depends on the choice of the initial model m_0.

Two problems are outstanding in FWI:

- Can we ignore $\mathcal{O}(\delta m)$ in (6.2)? If we do so, what physics is ignored?
- Can FWI come up with its own initial model m_0?

Ru-Shan has contributed significantly in addressing these problems.

6.2 High-order Fréchet derivatives and multiple scattering

Like the Taylor series, we can use higher-order functional Fréchet derivatives that are multi-linear mapping operators with properly defined norms (for details, see [6, 32]) for the perturbation expansion in (6.2)

$$A(m_0+\delta m) = A(m_0) + A'(m_0)\delta m + \frac{1}{2}A''(m_0)\delta m^2 + \cdots + \frac{1}{n!}A^{(n)}(m_0)\delta m^n + \cdots, \qquad (6.4)$$

where A', A'', and $A^{(n)}$ are the first, second, and n^{th} order Fréchet derivatives. The series (6.4) is closely related to the Born series. In 2014, Ru-Shan [63] produced an important insight that

> "There is a one-to-one correspondence between the mathematical n-th order Fréchet derivative and the physical n-th order multiple scattering term in the Born scattering series."

Therefore, ignoring high-order terms in (6.2) is the same as ignoring multiple scattering. However, multiple scattering cannot be ignored when the heterogeneity size is large compared to the wavelength or when the velocity contrast/perturbation is strong.

Owing to this reason, FWI using only the first-order Fréchet derivative works effectively when the initial model is close to the true model, which implies two things: the medium contrast $||\delta m||$ is small and the spatial size of $||\delta m||$ is small (i.e., the large-scale structure in the true model is well captured in the initial model). If the FWI initial model is just a homogeneous model, to invert for the large-scale inhomogeneities, we need to have low-frequency seismic data corresponding to long wavelengths to justify the use of only the Fréchet derivative and make the first-order Born approximation valid. If we do not have low-frequency data, we must include higher-order derivatives and use multiple scattering.

Ru-Shan's work in waveform inversion focuses on strong scattering and on how to obtain the low-wavenumber initial model without using low-frequency data.

6.3 Strong scattering modeling

Seismic inversion and modeling are tightly related to each other. The ability to invert for large-scale strong-contrast scatterers depends on our understanding of modeling wave interaction with such scatterers. Multiple scattering must be included for modeling strong scattering. Historically, multiple scattering was described in the Born scattering series, with each term describing a scattering order. When the background field encounters a scatterer, it produces a first-order scattering. The first-order scattered wave encounters another scatterer and produces a second-order scattering, and so on. Seismic modeling with a finite number of terms in the Born series is called the naive Born approximation. Unfortunately, the Born series is usually not convergent. To handle the divergence issue, Ru-Shan has made significant contribution by recognizing that

> "by re-ordering and renormalizing the Born series into the multiple-forward-single-backward De Wolf series, one can get a convergent field."

Ru-Shan's work on seismic modeling of strong scattering started in the 1990s [23, 38, 62]. For a more recent work, we refer readers to [23]. In particular, he used the De Wolf series (DWS) renormalization. In DWS, rather than dealing with the omni-directional scattering, he re-ordered the scattering series into a multiple-forward and single-backward scattering series. Therefore, the divergent parts can be *progressively canceled* as more terms are added. Fig. 28 shows a comparison between the naive Born approximation and the DWS using a resampled 2D cross-section of the 3D SEG/EAGE salt model. The naive Born calculation diverges as more and more terms/orders are used. The renormalized DWS converges.

6.4 Nonlinear sensitivity kernel

In waveform inversion, a sensitivity kernel maps model perturbations (e.g., velocity anomaly) to seismic data perturbations. For example, the commonly used Fréchet derivative is a linear mapping kernel. This linear kernel is not sufficient when the heterogeneity scale a is large in size compared with the wavelength λ, or when the medium contrast is strong, or both. Using the DWS, Ru-Shan derived the nonlinear sensitivity kernel [63] using the multiple forward scattering renormalization (Fig. 29). Figs. 29(b)-(d) show that, when the heterogeneity size is small compared with the wavelength ($a \sim 0.1\lambda$), the linear Fréchet kernel (Fig. 29(b) works well compared to the exact result (Fig. 29(c). However, when the heterogeneity size is large ($a \sim 5\lambda$), the linear kernel (Fig. 29(e) cannot capture the sensitivity (Fig. 29(f) whereas the nonlinear DWS kernel can (Fig. 29(g). The implication of this observation in waveform inversion is obvious. If we rely on the Fréchet derivative, we have to find conditions that make the linear kernel valid. This means that the availability of the low-frequency data is crucial as low-frequency waves have long wavelengths, resulting in a small value of a/λ. If we do not have low-frequency data, we need to have a better initial model so that the heterogeneities are of small sizes or weak perturbations.

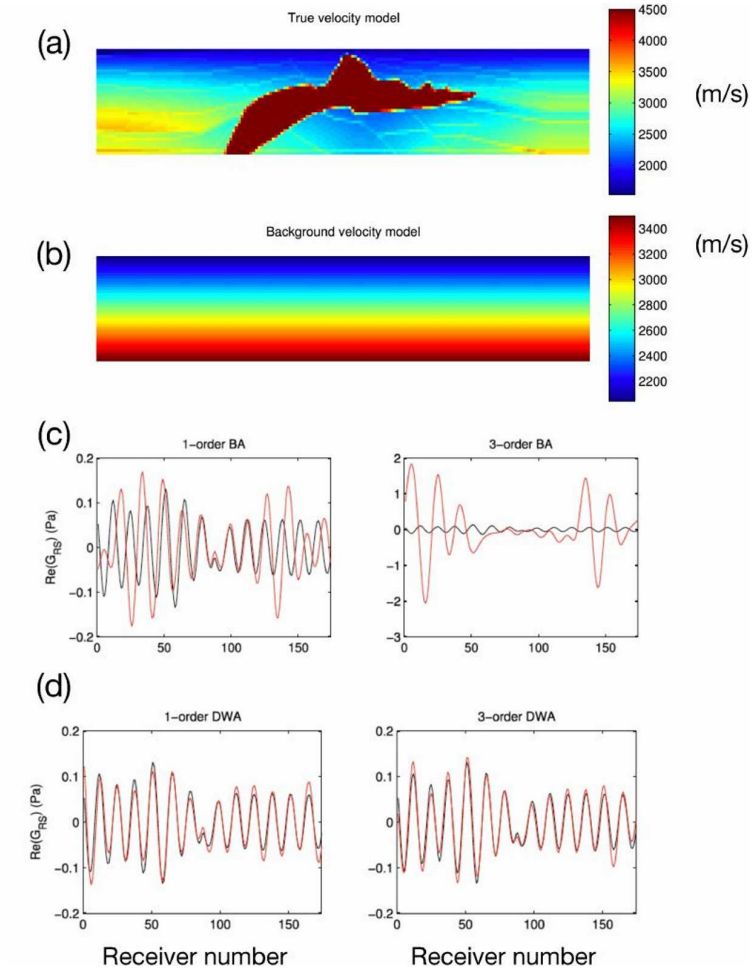


Figure 28: Comparison of modeling strong scattering using the naive Born series and the De Wolf series. (a) A resampled 2D cross-section of the 3D SEG/EAGE salt model. The source is at the top center of the model and 174 receivers are evenly distributed on top of the model. The source wavelet is a 15 Hz Ricker. (b) Background model; (c) Real part of the recorded wavefield at 10 Hz, computed by summing the naive Born series (first-order Born: BA-1 and the 3rd order Born: BA-3); (d) Real part of the wavefield at 10 Hz, computed by summing the renormalized de Wolf series (DWS) (first-order DWS: DWA-1 and the 3rd order DWS: DWA-3). (Modified from [23])

The nonlinear kernel states that

> "If multiple scattering nonlinear kernel is used in the waveform inversion, the need for low-frequency data is reduced."

We emphasize here that the 'single' or 'multiple' scattering is referred to the interaction between the background field and the unknown heterogeneities yet to be inverted. The wavefield (i.e. Green's function) in the background model can itself contain multiple scattering.

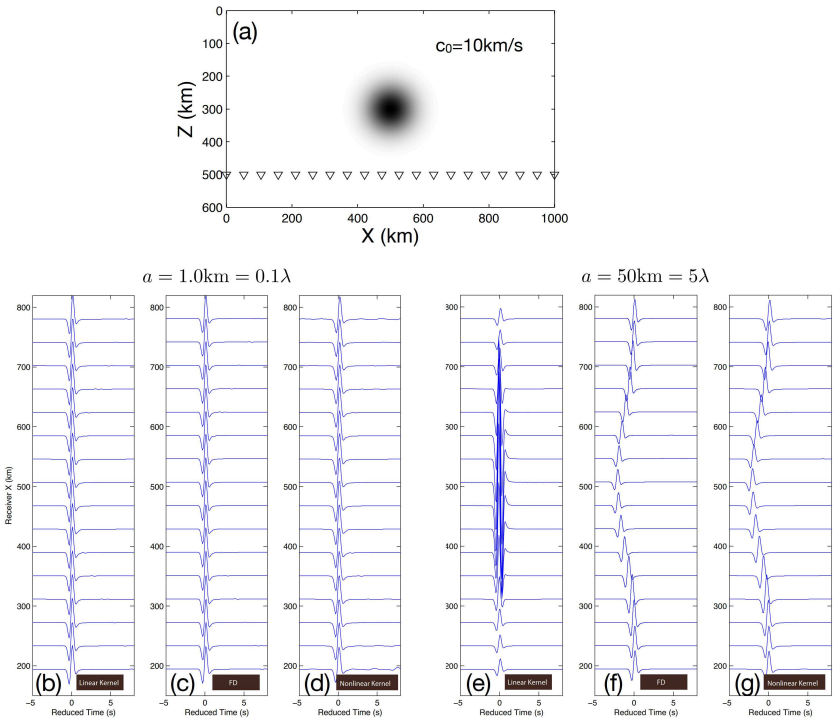


Figure 29: Comparison of sensitivity kernels computed using different modeling methods for two different-sized heterogeneities. (a) A Gaussian anomaly in a homogeneous background medium. At the center of the Gaussian anomaly, the velocity is 20% higher than the background and the velocity anomaly decays away from the center in a Gaussian fashion. The standard deviation of the Gaussian profile are $a=1$ km (b-d) and $a=50$ km (e-g). The source is located at the top center of the model and the receivers are at $Z=500$ km. The parameter λ is the P-wavelength in the background medium. Modeled seismic waveforms are calculated using (b and e) the first-order Born approximation or the linear Frechet derivative kernel, (c and f) the full-wave finite difference (FD), and (d and g) the De Wolf series using multiple forward scattering (i.e., the non-linear kernel). In (b-g), the time is reduced. The reduced traveltime is defined as the true traveltime minus the background traveltime. The FD result is assumed to be accurate. (Modified from Figs. 3-4 in Wu and Zheng [63])

6.5 Envelope and signed envelope inversion

Conventional FWI works well if the initial model is close to the true model. In fact, the initial model only needs to be approximately correct at long-wavelength scales. For an arbitrary starting model, it is necessary for FWI to use the low-frequency content in the data because FWI is based on the single-scattering (Fréchet) inversion engine. However, the problem is that real active seismic data may not contain low frequencies (lower than 5 Hz). Within the framework of conventional FWI, the lack of low-frequency is a main issue for FWI convergence. For the purpose of model building, one can find a model to just fit certain aspects of the data. For example, traveltime tomography can generate a good smooth large-scale velocity model using only the first-arrival times.

Ru-Shan has conducted a series of systematic works to find an effective approach for inversion of seismic reflection waveform envelopes. The inversion focuses on using

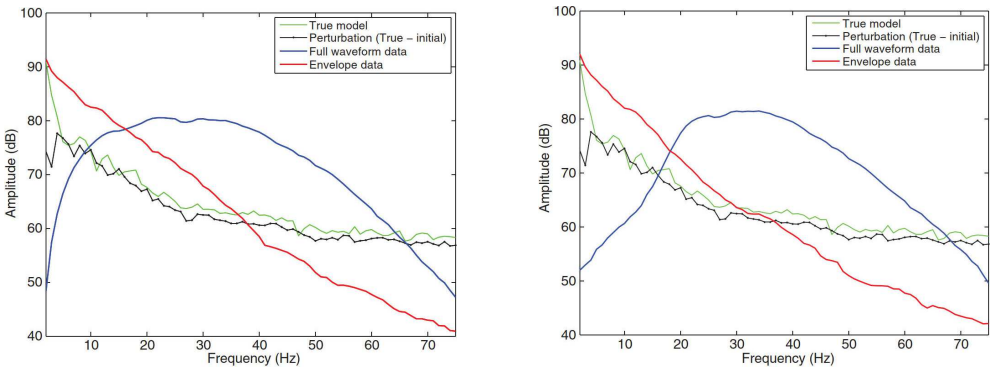


Figure 30: (a) Comparison of waveform spectrum and envelope spectrum for the Marmousi data with a Ricker source excitation and (b) same as (a) but with a low-cut source excitation. The low-cut source is obtained by a 5-Hz low-frequency taper. "True-model" means the spectrum of the 1D vertical profile of the true velocity after depth to time conversion. The "Perturbation" spectrum is the spectrum of the 1D profile of the velocity model difference after depth to time conversion. (From [55])

the seismic envelope and the modulation signal model (instead of the commonly-used convolution) to invert for large-scale velocity structure [8, 27, 55]. The advantage of this method is that the workflow is the same as that in FWI but the data misfit is the envelope difference. Ru-Shan initially developed an indirect envelope inversion (EI) method [55]. "Indirect" is meant to contrast with the newly developed "direct envelope inversion" (DEI). In the indirect EI, the gradient operator for updating the model is derived through the chain rule of differentiation with respect to the data functional, which is a double linearization. Hence, the waveform sensitivity operator (Fréchet derivative) can be used in EI but the effective data residual (adjoint source) is changed to the one related to the envelope. The envelope has low frequency (often called ultra-low-frequency, namely frequencies lower than the lowest frequency of the source frequency band) information (not to be confused with the low frequency in the waveform wiggles) (Figs. 30(a)-(b). Therefore, fitting the envelope (long-period signal) can avoid the cycle skipping problem commonly encountered in FWI. Note that the envelope inversion is not meant to replace FWI, but to produce the initial model for FWI. Numerical examples in Figs. 31 and 32(a)-(d) show that using EI followed by FWI is better than using FWI alone (Figs. 32(e)-(f)).

Ru-Shan then developed the multi-scale direct envelope inversion (MSDEI) with new Fréchet derivatives [44, 45]. As we mentioned above, the indirect envelope inversion (EI) applies the chain rule of differentiation to the data functional to use the waveform Fréchet derivative,

$$\frac{\delta \mathbf{e}}{\delta \mathbf{v}} = \frac{\delta \mathbf{e}}{\delta \mathbf{p}} \frac{\delta \mathbf{p}}{\delta \mathbf{v}}, \tag{6.5}$$

where $\delta \mathbf{e}/\delta \mathbf{v}$ is the envelope Fréchet derivative and $\delta \mathbf{p}/\delta \mathbf{v}$ is the waveform Fréchet derivative. In strong-scattering cases, both the envelope-to-waveform derivative and the waveform Fréchet derivative on the right-hand side of the above equation are strongly

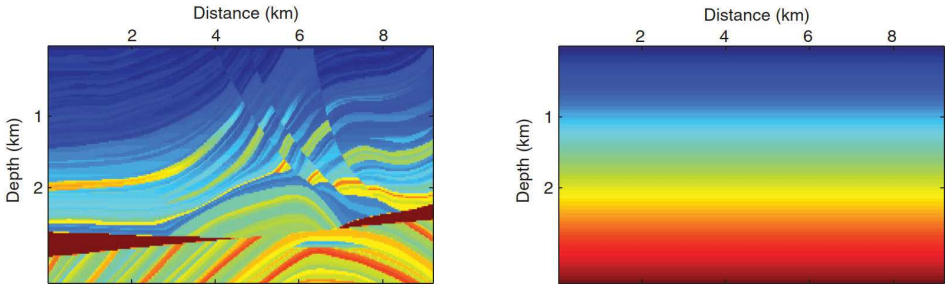


Figure 31: (a) True Marmousi model and (b) linear gradient initial model. (From [55])

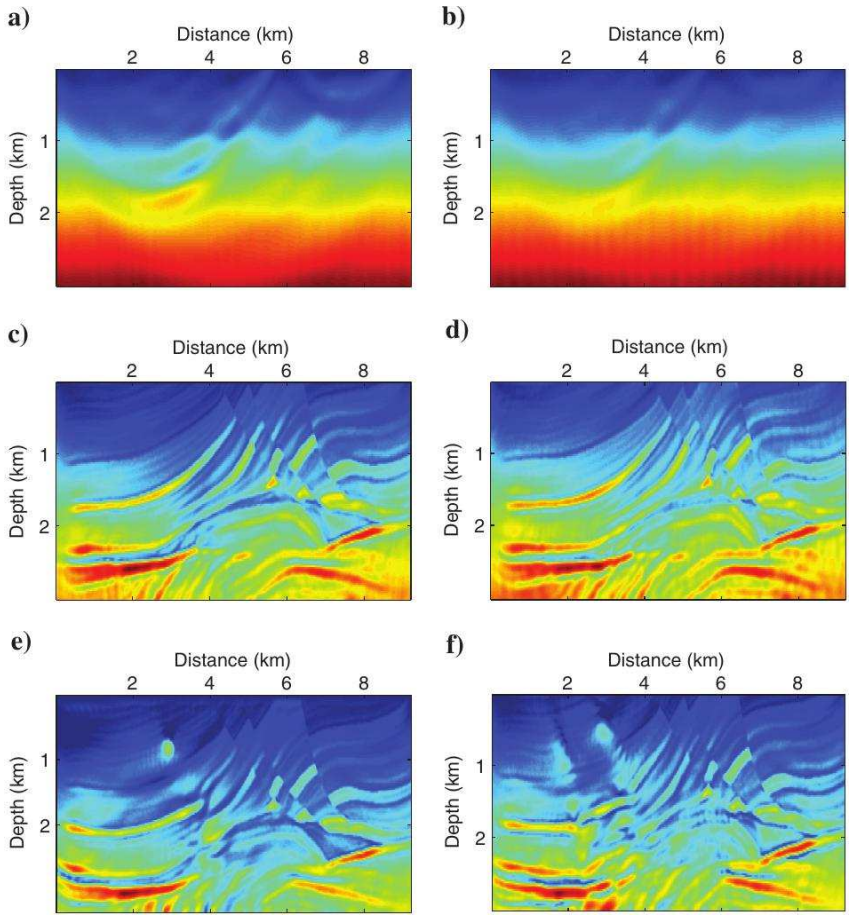


Figure 32: Seismic envelope inversion tests on the Marmousi model. (a) Smooth background obtained from envelope inversion (EI) using the full-band source, (b) same as (a) but with the low-cut source, (c) combined EI + WI inversion result (waveform inversion using the smooth background from envelope inversion) using the full-band source, (d) combined EI + WI inversion result using the low-cut source, (e) direct waveform inversion (FWI) from the linear initial model using the full-band source, and (f) same as (e) but using the low-cut source. (From [55])

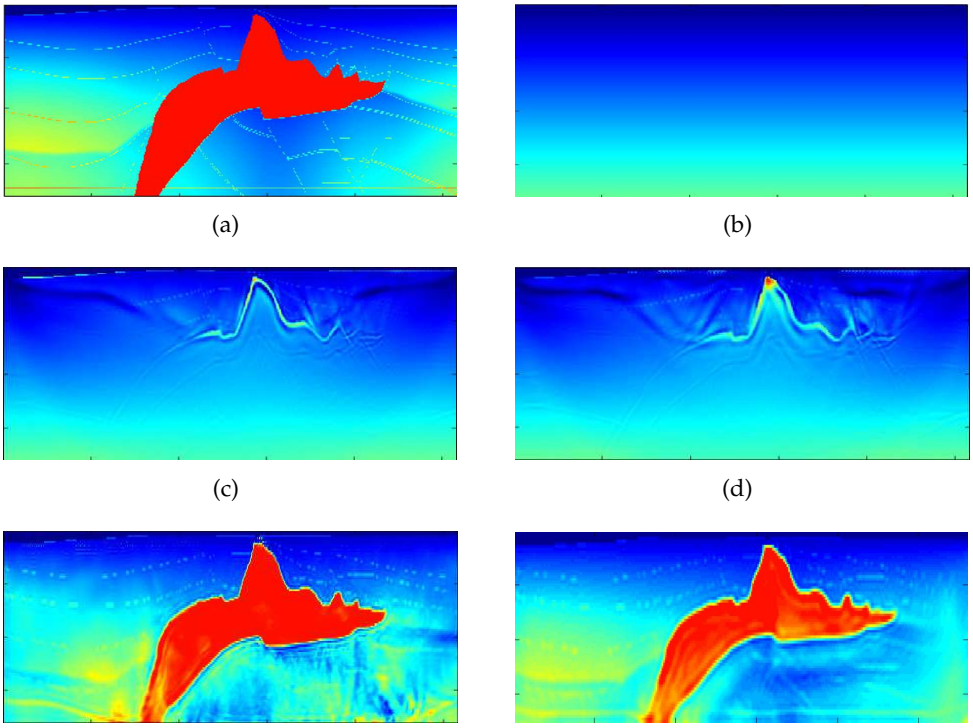


Figure 33: MSDEI results obtained using envelope data with polarity (envelope by signed demodulation) for a 2D SEG/EAGE salt model: (a) True model; (b) Initial model; (c) Result of conventional FWI; (d) Result of indirect envelope inversion (using waveform Fréchet derivative) + FWI; (e) Result of MSDEI + FWI; (f) Result of MSDEI + FWI using envelope phase-data in the local time-frequency domain. (Modified from [40])

nonlinear, so the chain rule of differentiation is a double linearization that renders the failure of the indirect EI when applied to strong-scattering cases. To overcome this limitation, Ru-Shan introduced the direct envelope Fréchet derivative and the calculation formulas [44, 45]. Hence, not only the limitation of the waveform Fréchet derivative $\delta\mathbf{p}/\delta\mathbf{v}$, which is based on the weak-scattering theory, is avoided, but also the severe distortion by the data filter $\delta\mathbf{e}/\delta\mathbf{p}$ is removed. This feature tremendously expanded the application range of DEI (direct envelope inversion). Fig. 33 gives a summary of the results from MSDEI applied to the 2D SEG/EAGE salt model and the comparison to the results from other inversion methods.

Ru-Shan used a 2D SEG/EAGE salt model with strong scattering to verify the capability of his MSDEI method. Fig. 33(a) is the true model and Fig. 33(b) is the initial model for inversion. The model in Fig. 33(a) contains the large-scale salt dome and the strong velocity perturbation (from the background model). The source is a Ricker wavelet with a dominant frequency of 9 Hz (cut off signals below 4 Hz). Fig. 33(c) is the result of conventional FWI, and Fig. 33(d) is the result of conventional envelope-inversion (indi-

rect EI). As expected, both conventional FWI and indirect EI fail in this strong-scattering case. The last two figures (Figs. 33(e),(f)) are the results obtained using direct envelope inversion (DEI). Fig. 33e is the result of MSDEI with signed envelopes [7, 8]. Fig. 33(f) is the result of MSDEI using envelope data through time-frequency analysis [21]. DEI with the signed envelope gives excellent recovery of the salt structure, since signed envelopes keep the polarity information of reflections in the data, and the inversion process can better recover the positive or negative velocity jumps. The MSDEI method with the envelope phase in the local time-frequency domain and the normalized envelope amplitude can enhance the signal of reflections from deep structures, so the inversion has better recovery of deep structures (Fig. 33(f)). Conventional FWI can add the high-wavenumber component to the inversion. Therefore, MSDEI + FWI gives a better recovery of a broad wavenumber spectrum of the model. Some results have also been obtained for source-independent MSDEI [67].

7 Summary

We have highlighted only some of Ru-Shan Wu's research. We marvel at the numerous remarkable innovative ideas and methods that he has contributed to three major fields in seismology: seismic scattering, imaging, and inversion. He has successfully discovered deep connections among seemingly unrelated concepts across different physical sciences. He has pioneered research particularly in diffraction tomography, seismic scattering in random media, scattering and attenuation, radiative transfer of seismic scattering, acoustic and elastic one-way wave propagation and imaging, beamlet and dreamlet for seismic-wave modeling and imaging, strong-nonlinear full-waveform inversion, and direct envelop inversion. Ru-Shan constantly looks for scientific challenges and consistently identifies important ones. He can easily initiate a new research topic and make impactful contributions. We look forward to his continued contributions in addressing many challenging problems in geophysics.

Acknowledgments

We thank Ru-Shan Wu for his help in providing most figures in the paper and for his very careful review of the paper.

References

[1] K. Aki. Analysis of the seismic coda of local earthquakes as scattered waves. *Journal of Geophysical Research*, 74:615–631, 1969.

[2] K. Aki. Scattering of P waves under the Montana Lasa. *Journal of Geophysical Research*, 78:1334–1346, 1973.

[3] K. Aki. Attenuation of shear waves in the lithosphere for frequencies from 0.05 to 25 Hz. *Phys. Earth Planet. Int.*, 21:50–60, 1980.

[4] E. Candès and D. Donoho. Curvelets: a surprisingly effective nonadaptive representation for objects with edges. In A. Cohen, C. Rabut, and L. Schumaker, editors, *Curves and Surfaces Fitting*, pages 105–120. Venderbilt University Press, Nashville, 2000.
[5] S. Chandrasekhar. *Radiative Transfer*. Dover Publications, Inc, New York, 1960.
[6] K. Chang. *Methods in Nonlinear Analysis*. Springer-Verlag, 2005.
[7] G. Chen, R.-S. Wu, Y. Wang, and S. Chen. Multi-scale signed envelope inversion. *Journal of Applied Geophysics*, 153:113–126, 2018.
[8] G. X. Chen, R.-S. Wu, and S. C. Chen. Multiscale direct envelope inversion: Algorithm and methodology for application to the salt structure inversion. *Earth and Space Science*, 6(1):174–190, 2019.
[9] L. Chen, R.-S. Wu, and Y. Chen. Target-oriented beamlet migration based on gabor-daubechies frame decomposition. *Geophysics*, 71(2):S37–S52, 2006.
[10] X. Chen and K. Aki. General coherence functions for amplitude and phase fluctuations in a randomly heterogeneous medium. *Geophysical Journal International*, 105:155–162, 1991.
[11] L. Chernov. *Wave Propagation in a Random Medium (English translation by R. A. Silverman)*. McGraw-Hill, New York, 1960.
[12] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Info. Theory*, 36:961–1005, 1990.
[13] I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, Pennsylvania, 1992.
[14] D. De Wolf. Electromagnetic reflection from an extended turbulent medium: Cumulative forward-scatter single-backscatter approximation. *IEEE Transactions on Antennas and Propagation*, AP-19:254–262, 1971.
[15] D. De Wolf. Renormalization of em fields in application to large-angle scattering from randomly continuous media and sparse particle distributions. *IEEE Transactions on Antennas and Propagation*, AP-33:608–615, 1985.
[16] S. M. Flatté and R.-S. Wu. Small-scale structure in the lithosphere and asthenosphere deduced from arrival time and amplitude fluctuations at norsar. *Journal of Geophysical Research*, 93(B6):6601–6614, 1988.
[17] D. Gabor. Theory of communication. *Journal of the Institution of Electrical Engineers*, 93:429–457, 1946.
[18] Y. Geng, R.-S. Wu, and J.-H. Gao. Gabor frame based fast Gaussian packet migration. *Geophysical Prospecting*, 62:1432–1452, 2014.
[19] E. Gorodnitskiy, M. Perel, Y. Geng, and R.-S. Wu. Depth migration with Gaussian wave packets based on poincare wavelets. *Geophysical Journal International*, 205(1):314–331, 2016.
[20] T. Hong, R.-S. Wu, and B. Kennett. Stochastic feature in scattering. *Phys. Earth Planet. Inter.*, 148:131–148, 2005.
[21] Y. Hu, R.-S. Wu, L. Han, and P. Zhang. Joint multiscale direct envelope inversion of phase and amplitude in the time-frequency domain. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7):5108–5120, 2019.
[22] W. Huang, R.-S. Wu, and R. Wang. Damped dreamlet representation for exploration seismic data interpolation and denoising. *IEEE Transactions on Geoscience and Remote Sensing*, PP(99):1–14, 2018.
[23] M. Jakobsen and R.-S. Wu. Renormalized scattering series for frequency-domain waveform modelling of strong velocity contrasts. *Geophysical Journal International*, 206(2):880–899, 2016.
[24] X. Jia and R.-S. Wu. Calculation of the wave propagation angle in complex media: application to turning wave simulations. *Geophysical Journal International*, 178(3):1565–1573, 2009.
[25] X. Jia and R.-S. Wu. Superwide-angle one-way wave propagator and its application in imag-

ing steep salt flanks. *Geophysics*, 74(4):S75–S83, 2009.
[26] C. Line, R. Hobbs, J. Hudson, and D. Synder. Statistical inversion of controlled-source seismic data using parabolic wave scattering theory. *Geophys. J. Int.*, 132:61–78, 1998.
[27] J. Luo and R.-S. Wu. Seismic envelope inversion: reduction of local minima and noise resistance. *Geophysical Prospecting*, 63(3):597–614, 2015.
[28] N. M. N. Toksż and D. H. Johnston, editors. *Seismic Wave Attenuation*. Soc. Expl. Geophys., Tulsa, Oklahoma, 1981.
[29] C. C. Mosher, D. J. Foster, and R.-S. Wu. Phase shift migration with wave packet algorithms. In *Mathematical methods in Geophysical Imaging IV*, volume 2822, pages 2-16. SPIE, 1996.
[30] H. Sato, M. C. Fehler, and T. Maeda. *Seismic Wave Propagation and Scattering in the Heterogeneous Earth: Second Edition*. Springer, 2012.
[31] A. Schuster. Radiation through a foggy atmosphere. *Astrophys. J.*, 2:1–22, 1905.
[32] G. Teschl. *Nonlinear Functional Analysis*. University of Vienna, 1998.
[33] B. Wu, R.-S. Wu, and J. Gao. Dreamlet source-receiver survey sinking prestack depth migration. *Geophysical Prospecting*, 61(1):63–74, 2013.
[34] R.-S. Wu. Attenuation of short period seismic waves due to scattering. *Geophysical Research Letters*, 9(1):9–12, 1982.
[35] R.-S. Wu. Mean field attenuation and amplitude attenuation due to wave scattering. *Wave Motion*, 4(3):305–316, 1982.
[36] R.-S. Wu. Multiple scattering and energy transfer of seismic waves — separation of scattering effect from intrinsic attenuation — I. Theoretical modelling. *Geophysical Journal of the Royal Astronomical Society*, 82(1):57–80, 1985.
[37] R.-S. Wu. Wide-angle elastic wave one-way propagation in heterogeneous media and an elastic wave complex-screen method. *Journal of Geophysical Research*, 99(B1):751–766, 1994.
[38] R.-S. Wu. Synthetic seismograms in heterogeneous media by one-return approximation. *Pure and Applied Geophysics*, 148(1-2):155–173, 1996.
[39] R.-S. Wu. Wave propagation, scattering and imaging using dual-domain one-way and one-return propagators. *Pure and Applied Geophysics*, 160(3-4):509–539, 2003.
[40] R.-S. Wu. Towards a theoretical background for strong-scattering inversion – direct envelope inversion and Gel'fand-Levitan-Marchenko theory. *Commun. Comput. Phys.*, 28(1):41–73, 2020.
[41] R.-S. Wu and K. Aki. Elastic wave scattering by a random medium and the small-scale inhomogeneities in the lithosphere. *Journal of Geophysical Research-Solid Earth and Planets*, 90(Nb12):261–273, 1985.
[42] R.-S. Wu and K. Aki. Scattering characteristics of elastic-waves by an elastic heterogeneity. *Geophysics*, 50(4):582–595, 1985.
[43] R.-S. Wu and K. Aki. Multiple scattering and energy transfer of seismic waves — separation of scattering effect from intrinsic attenuation. II. Application of the theory to Hindu Kush region. In Special Issue "Seismic Wave Scattering and Attenuation," Ed. by R.-S. Wu and K. Aki. *Pure and Applied Geophys.*, 128:49–80, 1988.
[44] R.-S. Wu and G. Chen. Multi-scale seismic envelope inversion for salt structures using a new direct envelope Fréchet derivative. In *Expanded Abstracts, 87th SEG Annual Meeting*, pages 1523–1525, 2017.
[45] R.-S. Wu and G. Chen. Multi-scale seismic envelope inversion using a new envelope Fréchet derivative for strong-nonlinear full waveform inversion. *arXiv: http://arxiv.org/abs/1808.05275*, 2018.
[46] R.-S. Wu and L. Chen. Beamlet migration using Gabor-Daubechies frame propagator. In

Extended Abstracts, EAGE 63rd Annual Meeting, P074, 2001.
[47] R.-S. Wu and L. Chen. Wave propagation and imaging using Gabor-Daubechies beamlets. *Theoretical and Computational Acoustics*, World Scientific, New Jersey, 661-670, 2002.
[48] R.-S. Wu and L. Chen. Mapping directional illumination and acquisition dip-response using beamlet propagators. *Geophysics*, 71:S147–159, 2006.
[49] R.-S. Wu, S. Chen, and M. Luo. Migration amplitude correction in angle domain using beamlet decomposition. In *Expanded abstracts, EAGE 66th Annual Meeting, G029*, 2004.
[50] R.-S. Wu and S. M. Flatté. Transmission fluctuations across an array and heterogeneities in the crust and upper mantle. In Special Issue "Seismic Wave Scattering and Attenuation," Ed. by R.-S. Wu and K. Aki. *Pure and Applied Geophysics*, 132(1-2):175–196, 1990.
[51] R.-S. Wu and J. Gao. Beamlets and curvelets in wavefield decomposition, propagation and imaging: A review on the application of wavelet to wave propagation and imaging. *WTOPI Technical Report*, 14:3–40, 2007.
[52] R.-S. Wu and X. Jia. Accuracy improvement for super-wide angle one-way waves by wavefront reconstruction. *76th Ann. Mtg., Soc. Expl. Geophys., Expanded Abstracts*, pages 2976–2980, 2006.
[53] R.-S. Wu, S. Jin, and X.-B. Xie. Energy partition and attenuation of Lg waves by numerical simulations using screen propagators. *Physics of the Earth and Planetary Interiors*, 120:227–243, 2000.
[54] R.-S. Wu, S. Jin, and X.-B. Xie. Seismic wave propagation and scattering in heterogeneous crustal waveguides using screen propagators: I. SH waves. *Bulletin of the Seismological Society of America*, 90(2):401–413, 2000.
[55] R.-S. Wu, J. Luo, and B. Wu. Seismic envelope inversion and modulation signal model. *Geophysics*, 79(3):WA13–WA24, 2014.
[56] R.-S. Wu, M. Luo, S. Chen, and X.-B. Xie. Acquisition aperture correction in angle-domain and true-amplitude imaging for wave equation migration. In *Expanded abstracts, SEG 74th Annual Meeting*, pages 937–940, 2004.
[57] R.-S. Wu and M. N. Toksoz. Diffraction tomography and multisource holography applied to seismic imaging. *Geophysics*, 52(1):11–25, 1987.
[58] R.-S. Wu, Y. Wang, and M. Luo. Beamlet migration using local cosine basis. *Geophysics*, 73(5):S207–S217, 2008.
[59] R.-S. Wu, B. Wu, and Y. Geng. Seismic wave propagation and imaging using time-space wavelets. In *Expanded abstracts, SEG 78th Annual Meeting*, pages 2983–2987, 2008.
[60] R.-S. Wu, Y. Wu, C. Liu, C. Peng, and M. Wang. Multi-frequency synthetic detecting holography. *Acta Geophysica Sinica*, 20:33–38, 1977.
[61] R.-S. Wu and X.-B. Xie. Modeling primaries of acoustic/elastic waves by one-return approximation. *Leading Edge [Tulsa, OK]*, 28(5):576–581, 2009.
[62] R.-S. Wu, X.-B. Xie, and X.-Y. Wu. One-way and one-return approximations (De Wolf approximation) for fast elastic wave modeling in complex media, Chapter 5. In R.-S. Wu and V. Maupin, editors, *Advances in Wave Propagation in Heterogeneous Earth*, Advances in Geophysics, volume 48, pages 265–322. Elsevier, 2007.
[63] R.-S. Wu and Y. Zheng. Non-linear partial derivative and its De Wolf approximation for non-linear seismic inversion. *Geophysical Journal International*, 196(3):1827–1843, 2014.
[64] X.-B. Xie, S. Jin, and R.-S. Wu. Wave-equation-based seismic illumination analysis. *Geophysics*, 71(5):S169–S177, 2006.
[65] R. Yan, H. Guan, X.-B. Xie, and R.-S. Wu. Acquisition aperture correction in the angle domain toward true-reflection reverse time migration. *Geophysics*, 79(6):S241–s250, 2014.

[66] Y. Zeng. Compact solutions for multiple scattered wave energy in time domain. *Bull. Seismol. Soc. Am.*, 81:1022–1029, 1991.
[67] P. Zhang, R.-S. Wu, and L. Han. Source-independent seismic envelope inversion based on the direct envelope Fréchet derivative. *Geophysics*, 83(6):R581–R595, 2018.
[68] Y. Zheng and R.-S. Wu. Measurement of phase fluctuations for transmitted waves in random media. *Geophysical Research Letters*, 32(14):L14314, 2005.
[69] Y. Zheng and R. S. Wu. Theory of transmission fluctuations in a depth-dependent background medium. In H. Sato and M. Fehler, editors, *Earth Heterogeneity and Scattering Effects on Seismic Waves*, volume 50 of *Advances in Geophysics*, pages 21–41. Elsevier, 2008.