VITERBI SPARSE SPIKE DETECTION AND A COMPOSITIONAL ORIGIN TO ULTRALOW-VELOCITY ZONES

by

Samuel Paul Brown

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Geophysics

Department of Geology and Geophysics

The University of Utah

December 2015

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of Samuel Paul Brown

has been approved by the following supervisory committee members:

Michael S. Thorne	Chair	3/22/2013 Date Approved
Michael S. Zhdanov	Member	3/22/2013 Date Approved
Richard D. Jarrard	Member	3/22/2013 Date Approved
Kristine L. Pankow	Member	3/22/2013 Date Approved
Paul R. Williamson	Member	4/3/2013 Date Approved

The dissertation has also been approved by John M. Bartley

Chair of the Department of Geology and Geophysics

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

Accurate interpretation of seismic travel times and amplitudes in both the exploration and global scales is complicated by the band-limited nature of seismic data. We present a stochastic method, Viterbi sparse spike detection (VSSD), to reduce a seismic waveform into a most probable constituent spike train. Model waveforms are constructed from a set of candidate spike trains convolved with a source wavelet estimate. For each model waveform, a profile hidden Markov model (HMM) is constructed to represent the waveform as a stochastic generative model with a linear topology corresponding to a sequence of samples. The Viterbi algorithm is employed to simultaneously find the optimal nonlinear alignment between a model waveform and the seismic data, and to assign a score to each candidate spike train. The most probable travel times and amplitudes are inferred from the alignments of the highest scoring models. Our analyses show that the method can resolve closely spaced arrivals below traditional resolution limits and that travel time estimates are robust in the presence of random noise and source wavelet errors.

We applied the VSSD method to constrain the elastic properties of a ultralow-velocity zone (ULVZ) at the core-mantle boundary beneath the Coral Sea. We analyzed vertical component short period *ScP* waveforms for 16 earthquakes occurring in the Tonga-Fiji trench recorded at the Alice Springs Array (ASAR) in central Australia. These waveforms show strong pre and postcursory seismic arrivals consistent with

ULVZ layering. We used the VSSD method to measure differential travel-times and amplitudes of the post-cursor arrival *ScSP* and the precursor arrival *SPcP* relative to *ScP*. We compare our measurements to a database of approximately 340,000 synthetic seismograms finding that these data are best fit by a ULVZ model with an *S*-wave velocity reduction of 24%, a *P*-wave velocity reduction of 23%, a thickness of 8.5 km, and a density increase of 6%. We simultaneously constrain both *P*- and *S*-wave velocity reductions as a 1:1 ratio inside this ULVZ. This 1:1 ratio is not consistent with a partial melt origin to ULVZs. Rather, we demonstrate that a compositional origin is more likely.

For my daughter, Cadence.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my gratitude to Sebastian Rost and Lowell Miyagi for their invaluable contributions to Chapter 3. I would also like to express my deep appreciation for Michael Thorne's mentorship and support throughout this process. I would like to thank my daughter for being a constant source of inspiration. I gratefully acknowledge the Rocky Mountain Oilfield Testing Center (RMOTC) and the U.S. Department of Energy for providing the Teapot Dome seismic data set, and the University of Utah Center for High Performance Computing (CHPC) for providing computer resources and support. I would also like to acknowledge the contributions of five anonymous reviewers from Geophysics, whose comments and suggestions greatly improved Chapter 2. Michael Thorne and Samuel Brown were partially supported by NSF grant EAR-1014749, and Samuel Brown was partially supported by a grant from Total E&P. Sebastian Rost was partially supported by NERC grant NE/H022473/1. Lowell Miyagi was partially supported by NSF grant EAR-1344579. Lowell Miyagi would like to thank Zhixue Du for the helpful discussions. Figures were drawn using the Generic Mapping Tools (http://gmt.soest.hawaii.edu/; Wessel and Smith, 1998).

CHAPTER 1

INTRODUCTION

The bandlimited nature of seismic data obscures details of individual arrivals. Above the resolution limit, constructive and destructive interference between closely spaced arrivals distorts apparent travel times and amplitudes. Below the resolution limit, the reflection response of individual layers merge into an effective media response. In order to more accurately estimate travel times and amplitudes, and to push existing resolution limits, the effects of interference between closely spaced arrivals must be unraveled. Traditional deconvolution can enhance resolution by shaping the underlying wavelet, but cannot add frequency content (which was not recorded). Sparse spike deconvolution methods, which are reviewed in Chapter 2, attempt to decompose a bandlimited seismic record into a constituent spike train representing the differential travel times and relative amplitudes of all resolvable arrivals. In this dissertation I present a stochastic method, Viterbi sparse spike detection (VSSD), which predicts the most likely spike train and explains a bandlimited seismic record given an approximate source time function estimate.

1.1 Technical Contributions

The VSSD method is introduced in Chapter 2. The theoretical framework, which borrows ideas from biological sequence analysis, and the algorithmic details are presented. Seismic wedge models are used to demonstrate that the method can resolve closely spaced arrivals below traditional resolution limits. An analysis of the method's sensitivity to noise and errors in the source wavelet estimate is also presented. The results demonstrate that travel time estimates are robust in the presence of random noise and source wavelet errors, and reasonable expectations for the fidelity of relative amplitude picks are established. The method is particularly well suited to fine-scale interpretation problems such as thin bed interpretation, tying seismic images to well logs, and the analysis of anomalous waveforms in global seismology. As an example, the method is applied to interpret thin layers in the Teapot Dome oilfield.

In Chapter 3, the VSSD method is applied to constrain the elastic properties of a ultralow-velocity zone (ULVZ) at the core-mantle boundary beneath the Coral Sea. The dataset analyzed consists of vertical component short period *ScP* waveforms for 16 earthquakes occurring in the Tonga-Fiji trench recorded at the Alice Springs Array (ASAR) in central Australia, which show strong pre and postcursory seismic arrivals consistent with ULVZ layering. The direct *P*-wave arrival is windowed and tapered to provide a source time function estimate. We used the VSSD method to measure differential travel-times and amplitudes of the postcursor arrival *ScSP* and the precursor arrival *SPcP* relative to *ScP*. The nonlinear nature of the VSSD method, which allows the underlying source time function to dilate and contract, is able to account for the increased attenuation in the *ScP* waveform versus the *P*-wave based source estimate. The

measurements were compared to a database of approximately 340,000 synthetic seismograms. The measured differential travel times and relative amplitudes of *SPcP* and *ScSP* were used to define Gaussian distributions with which the likelihood of synthetic models was determined. Our results show that these data are best fit by a ULVZ model with an *S*-wave velocity reduction of 24%, a *P*-wave velocity reduction of 23%, a thickness of 8.5 km, and a density increase of 6%. We simultaneously constrain both *P*- and *S*-wave velocity reductions as a 1:1 ratio inside this ULVZ, which is consistent with a compositional origin to ULVZs.

CHAPTER 2

VITERBI SPARSE SPIKE DETECTION

Reprinted with permission from Society of Exploration Geophysicists. Brown, S. P., and M. S. Thorne (2013), Viterbi sparse spike detection, Geophysics, 78(4), V157–V169, doi:10.1190/GEO2012-0209.1.

GEOPHYSICS, VOL. 78, NO. 4 (JULY-AUGUST 2013); P. V157–V169, 13 FIGS., 1 TABLE.
10.1190/GEO2012-0209.1

Viterbi sparse spike detection

Samuel P. Brown[1] and Michael S. Thorne[2]

ABSTRACT

Accurate interpretation of seismic traveltimes and amplitudes in the exploration and global scales is complicated by the band-limited nature of seismic data. We discovered a stochastic method to reduce a seismic waveform into a most probable constituent spike train. Model waveforms were constructed from a set of candidate spike trains convolved with a source wavelet estimate. For each model waveform, a profile hidden Markov model (HMM) was constructed to represent the waveform as a stochastic generative model with a linear topology corresponding to a sequence of samples. Each match state in the HMM represented a sample in the model waveform, in which the amplitude was represented by a Gaussian distribution. Insert and delete states allowed the underlying source wavelet to dilate or contract, accounting for nonstationarity in the seismic data and errors in the source wavelet estimate. The Gaussian distribution characterizing each sample's amplitude accounted for random noise. The Viterbi algorithm was employed to simultaneously find the optimal nonlinear alignment between a model waveform and the seismic data and to assign a score to each candidate spike train. The most probable traveltimes and amplitudes were inferred from the alignments of the highest scoring models. The method required no implicit assumptions regarding the distribution of traveltimes and amplitudes; however, in practice, the solution set may be limited to mitigate the nonuniqueness of solutions and to reduce the computational effort. Our analyses found that the method can resolve closely spaced arrivals below traditional resolution limits and that traveltime estimates are robust in the presence of random noise and source wavelet errors. The method was particularly well suited to fine-scale interpretation problems such as thin bed interpretation, tying seismic images to well logs, and the analysis of anomalous waveforms in global seismology.

INTRODUCTION

A problem of fundamental importance in all branches of seismology is accurate traveltime and amplitude picking. However, the problem is generally complicated by a low signal-to-noise ratio (S/N) or constructive/destructive interference effects from band-limited wavelets. Figure 1 provides a simple illustration of this problem. The top three traces show individual arrivals consisting of 25-Hz Ricker wavelets scaled in amplitude and shifted in time. The lowermost trace shows the composite waveform obtained by summing the individual arrivals. Because these are zero phase wavelets, traveltimes and amplitudes can be estimated from picks at the locations of local amplitude maxima or minima at peaks or troughs. The black spikes in Figure 1 depict manually picked traveltimes and amplitudes of the individual arrivals comprising the composite waveform, whereas the gray spikes depict the actual traveltimes and amplitudes. The traveltime is correctly picked for the first arrival, but destructive interference from the second arrival has led to an erroneously low amplitude pick. The second arrival pick has significant time and amplitude errors, and the third arrival pick has slight traveltime and amplitude errors. This simple noise-free example demonstrates how constructive and destructive interference from overlapping wavelets complicates the picking process and can lead to erroneous traveltime and amplitude picks.

Accurate picking requires consideration of constructive and destructive interference from neighboring arrivals and noise. As the separation between individual arrivals decreases, the problem shifts from being one of picking accurate traveltimes and amplitudes to resolving the two separate arrivals. Figure 2 demonstrates the resolution criteria proposed by Rayleigh and Ricker (see Ricker,

Manuscript received by the Editor 13 June 2012; revised manuscript received 27 March 2013; published online 24 June 2013.
[1]Petroleum Geo-Services, Houston, Texas, USA and University of Utah, Department of Geology and Geophysics, Salt Lake City, Utah, USA. E-mail: samuel.brown@pgs.com.
[2]University of Utah, Department of Geology and Geophysics, Salt Lake City, Utah, USA. E-mail: michael.thorne@utah.edu.

1953). Figure 2a defines two metrics for a Ricker wavelet, which determine both resolution criteria. The wavelet breadth b is given as the distance between the maximum absolute amplitudes of the symmetric side lobes, while the temporal resolution limit (TR) (Kallweit and Wood, 1982) is the distance between inflection points on the central lobe. With sufficient separation, individual main-lobe peaks are visible and separate arrivals are easily resolved, as shown in Figure 2b. For two uniform-polarity events of equal amplitude, Rayleigh's limit of resolution is reached when two wavelets are separated by $b/2$, or the peak-to-trough time. As Figure 2c illustrates, this is the minimum separation at which two distinct peaks are still visible in the composite waveform. The resolution limit was extended by Ricker to the smaller distance TR. At this separation, the composite wavelet no longer has two distinct peaks; however, the curvature at the central maximum is zero, resulting in a flat spot, as can be seen in Figure 2d.

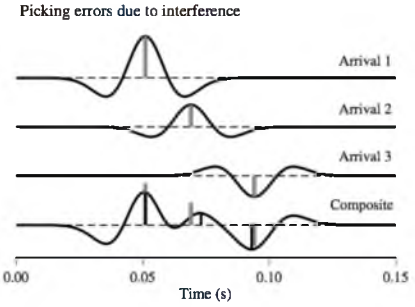

Figure 1. The top three traces show individual 25-Hz Ricker wavelets. The lowermost trace shows a composite waveform consisting of the sum of the individual Ricker wavelets from the upper traces. The gray spikes represent actual arrival times and amplitudes, while the black spikes represent manually picked arrival times and amplitudes.

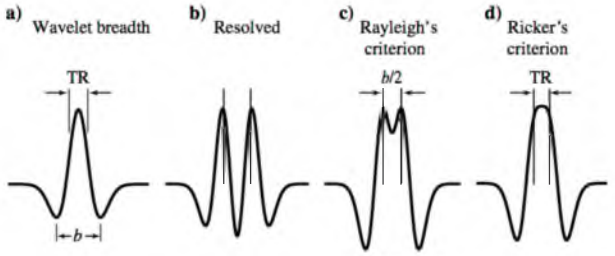

Figure 2. Resolution criteria with decreasing wavelet separation for two uniform-polarity Ricker wavelets. (a) Definition of wavelet breadth b and TR (Kallweit and Wood, 1982) for a Ricker wavelet. The wavelet breadth is the distance between side-lobe troughs, while TR is the distance between inflection points on the central lobe. The separation between arrivals is sufficient that the events in (b) are clearly resolved. (c) Rayleigh's criteria $b/2$ represents the limit of resolvability in which two distinct peaks remain. (d) A zero curvature-induced flat spot allows for event resolution at Ricker's limit TR.

Resolution limits, as well as the effects of constructive and destructive interference on waveform character, can be visualized with a wedge model, as shown in Figure 3. The underlying earth model contains three layers, with impedance varying discretely with depth, such that the absolute value of the zero-offset reflection coefficients at each interface is equal. The thickness of the middle layer decreases from left to right, forming the wedge. The reflectivity caused by impedance contrasts is represented as a function of two-way traveltime by the spikes in Figure 3. Synthetic seismic traces representing an idealized zero-offset acquisition were created by convolving a 25-Hz Ricker wavelet with the reflectivity series. Figure 3a shows a case in which the reflection coefficient is positive at both interfaces, representing an increase in impedance with depth. Figure 3b presents a mixed-polarity wedge model. In this case, a high-impedance wedge is inserted into a whole space, such that the polarity of the top and bottom reflection coefficients is reversed. This type of model would be characteristic of either a low-impedance bed surrounded by identical high-impedance beds, or vice versa, and it is of particular value because this phenomenon is likely to occur in a shale-sand-shale sequence.

In Figure 3a, the transition from well-resolved peaks through Rayleigh's ($b/2$) and Ricker's (TR) resolution limits can clearly be seen from the thickest to thinnest regions of the wedge (from left to right in Figure 3). However, in the mixed-polarity case (Figure 3b) Ricker's (TR) limit has no meaning. Widess (1973) proposes an amplitude-based methodology for resolving thin beds in the mixed-polarity case. Widess notes that at an approximate distance of $\lambda_b/8$, where λ_b is the dominant wavelength, the composite waveform stabilizes into a good approximation of the derivative of the original zero-phase wavelet. Thus, the waveform shape is virtually indistinguishable for beds less than λ_b, but the bed thickness may be inferred from the peak amplitude of the composite wavelet. At thicknesses below the established resolution criteria, the composite waveform for the uniform-polarity and opposite-polarity cases can be visually interpreted as a single arrival; however, the composite wavelet in the opposite-polarity case will approximate the derivative of the original zero-phase pulse with a strong amplitude response dependent on bed thickness (e.g., Figure 3a and 3b). This amplitude information is routinely used by seismic interpreters to construct tuning curves (Bacon et al., 2003) for thin-bed interpretation. In practice, this requires calibration of amplitudes to a known bed thickness. This method is further complicated by the presence of noise and uncertainty in the source wavelet estimate (Widess, 1973).

Figure 4a shows apparent thickness in two-way traveltime and apparent amplitude as a function of true thickness for the synthetic traces pictured in Figure 3a. The solid black line at 45° shows the ideal case where the apparent thickness determined matches the true thickness of the layer. The blue line shows the actual recovered thickness. If one could accurately determine thickness in all cases, the solid blue line should perfectly track the solid black line. However, the apparent thickness increases slightly before Rayleigh's criterion ($b/2$) is reached because of interference between the wavelets. Below the TR limit, the separate arrivals are unresolved

and the composite arrival appears to be an individual arrival. The solid horizontal black line shows the true amplitude of both arrivals, while the red line shows the measured amplitude. For traveltimes at or above Rayleigh's limit, the amplitudes are measured from two separate peaks. Below Rayleigh's limit, only a single peak remains. Destructive interference lowers amplitudes near the resolution criteria, and constructive interference increases amplitudes as the differential traveltime goes to zero.

In Figure 4b, the apparent thickness and amplitude as a function of true thickness for the mixed-polarity wedge model (Figure 3b) are shown. Here, the composite waveforms are comprised of two equal-amplitude, opposite-polarity 25-Hz Ricker wavelets. In this case, it is possible to resolve the two arrivals for all thicknesses; however, the apparent separation is limited to TR. Constructive interference increases amplitudes near the resolution limits, while destructive interference diminishes amplitudes as the differential traveltime goes to zero.

The process of detecting traveltimes and amplitudes of each arrival in a band-limited seismic trace is a form of sparse spike deconvolution. In recent decades, numerous techniques have been developed in this area. Kormylo and Mendel (1983) introduce a maximum-likelihood sparse spike deconvolution algorithm based on state-variable technology, which is capable of estimating a Bernoulli-Gaussian sparse spike train and the correct phase of the seismic wavelet. Wiggins (1978) and Sacchi et al. (1994) propose minimum entropy sparse spike deconvolution methods. Velis (2008) proposes a stochastic sparse spike deconvolution method that incorporates impedance constraints into simulated annealing iterations to find a solution consisting of the least number of spikes, which explains the observed data when convolved with a model wavelet. Phase errors in the wavelet estimate are handled to some extent by computing an optimal phase shift to match the model wavelet to an effective data wavelet. Kaaresen and Taxt (1998) use a Bayesian framework to derive a maximum a posteriori multichannel deconvolution estimate. Their method alternates between steps of wavelet and reflectivity estimation, and it handles continuity between traces by modeling local dependencies. Heimer et al. (2007) use dynamic programming to constrain multichannel blind seismic deconvolution such that reflections must form continuous paths across consecutive traces, representing consistent layers in the earth model. This method is improved upon by Heimer et al. (2009) by incorporating the Viterbi algorithm (Viterbi, 1967) for Markov-Bernoulli random field modeling in place of the previous dynamic programming algorithm, to allow layers to split, merge, and terminate across traces. Dynamic programming methods are also used by Liner and Clapp (2004) to nonlinearly align seismic traces, and the Viterbi algorithm is used by Clapp (2008) to autopick seismic horizons.

In this paper, we propose a stochastic method, using profile hidden Markov models (HMMs) and the Viterbi algorithm (Durbin et al., 1998), to resolve composite waveforms into their constituent spike trains. This effectively poses the sparse spike deconvolution problem as a pattern recognition problem. This method is adapted from the application of the Viterbi algorithm to profile HMMs by Eddy (1995) to construct alignments between, and aid in the identification of, evolutionarily related protein sequences. The Viterbi algorithm is used in the single-channel deconvolution context of picking individual arrival traveltimes and amplitudes. Unlike the other deconvolution algorithms outlined above, the method presented here makes no implicit assumptions about the distributions of traveltimes and amplitudes nor the phase of the wavelet. However, in practice, the method is generally employed within a constrained solution space to mitigate the problem of nonuniqueness and to bound the computation time. While the algorithm does not provide updated estimates of the wavelet, it is relatively insensitive to the types of errors expected in the practice of source wavelet estimation, and it can be applied in cases in which data are insufficient to generate meaningful statistics, such as a single seismic trace. Unlike tuning curves, this method implicitly accounts for noise and errors in the wavelet estimate, and its accuracy degrades gracefully as noise levels and wavelet errors increase.

THEORY

Our goal is to analyze a composite waveform in a data trace specifically to determine the differential traveltimes and relative amplitudes of the individual arrivals. Our ultimate goal is to deconstruct the composite waveform into a spike train, where each spike is aligned on the true arrival time with the correct amplitude of the individual seismic arrivals that went into constructing the composite waveform. To do this, we first define a solution space that consists of a set of unique spike trains, where each spike train is an element of the set containing all possible spike trains in the solution space.

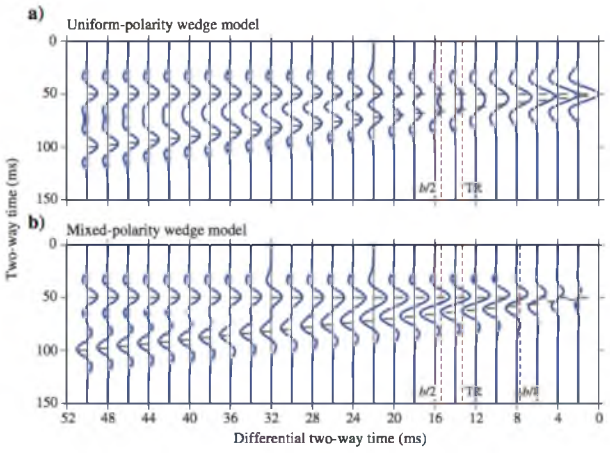

Figure 3. Seismic wedge models that demonstrate the effects of constructive and destructive interference in band-limited seismic traces as the temporal separation of individual arrivals decrease. (a) A seismic wedge model with uniform-polarity reflection coefficients, representing three layers in which acoustic impedance increases with depth. (b) A seismic wedge model with opposite-polarity reflection coefficients, representing a high-impedance layer between two low-impedance layers.

That is, each element consists of a unique collection of differential traveltimes and relative amplitudes that could describe the composite trace. From here on, we refer to one element in our set of spike trains as a model. The goal is to determine which model best describes the composite trace. One approach to solving this problem is to (1) build a reference trace for each point in the solution space by convolving it with a source wavelet, (2) find the best alignment between each reference trace and the data trace using crosscorrelation, and (3) pick the optimal solution through minimizing a residual error norm on the difference between the data trace and the reference traces.

The above algorithm may be a useful method to solve this problem, but it is not without challenges. For example, if an L2-norm is used in the optimization step, the algorithm will not be strongly affected by the presence of Gaussian-distributed random noise, but it does not implicitly handle errors in the wavelet estimate. The approach we propose here is similar to the above algorithm, except that it uses stochastic models to represent the points in our solution space. The proposed algorithm is: (1) Build a profile HMM for each model in which a noise model and a wavelet estimate are assumed at the onset. (2) For each profile HMM, use the Viterbi algorithm to simultaneously find the optimal alignment to the composite waveform and assign a log-likelihood score. Here, the alignment may be nonlinear. (3) Extract traveltimes and amplitudes from the alignment of the highest-scoring HMM.

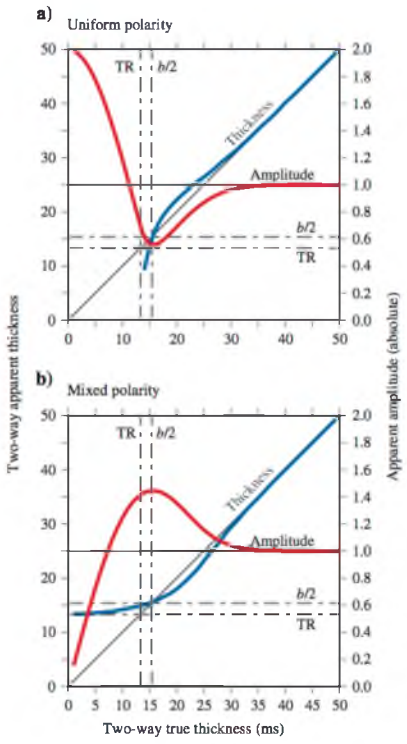

Figure 4. (a) Apparent thickness and amplitude as a function of event separation for two equal-amplitude uniform-polarity 25-Hz Ricker wavelets. The two arrivals are not resolved below the TR. (b) Apparent thickness and amplitude as a function of event separation for two equal-amplitude opposite-polarity 25-Hz Ricker wavelets. In this case, it is possible to detect both arrivals for all separations; however, the apparent thickness is limited to TR.

The first step is to construct a profile HMM for each model, which will be based on a synthetic trace obtained by convolving the model spike train with the assumed wavelet estimate. A profile HMM can be pictured as a finite state automaton such as the one pictured in Figure 5. This is a simplified version of the Plan 7 HMM architecture that Eddy (1995) introduces for analyzing protein sequences. Each of the nodes (diamonds and circles in Figure 5) represents a state, and each of the edges (arrows in Figure 5) represents a possible state transition. Markov models represent a stochastic process for generating sequences, which in this case, will be a sequence of samples constituting a seismic trace. In a Markov model, each state emits a single value or token and the stochastic process is represented by the random set of transitions between states. The model is characterized by its states and by transition probabilities between the states. The underlying stochastic process exhibits the Markov property, which states that the conditional probability distribution of future states depends only on the current state, resulting in a memoryless stochastic process. In an HMM, each state may generate many different tokens based on a state-specific emission probability distribution. The name is derived from the fact that the actual state sequence that produced a given output is hidden and can only be described in terms of probabilities, because many different state sequences can produce the same output. The Viterbi algorithm takes advantage of the Markov property, and it employs dynamic programming to efficiently determine the most likely state sequence for generating a specific output sequence with a specific HMM.

Our goal in using the HMM is to find the best nonlinear alignment between the composite waveform (our data) and a model in our solution space. An alignment can be thought of as a unique path traced through the state machine. The B and E states are for starting and ending alignments; M1, M2, M3, M4, and M5 are match states, and each one is associated with a sample from the model trace. Each match state is characterized by a Gaussian distribution with the amplitude of the associated model trace sample defining the mean. This Gaussian distribution over sample amplitudes represents the implicit random noise model. I1, I2, I3, and I4 are insert states, which allow multiple samples from the data trace to be aligned to a single sample from the model trace. This, in effect, allows the embedded wavelet to dilate. Insert states are also characterized by a Gaussian distribution whose mean is the average of the two surrounding match state means. D2, D3, and D4 are delete states, which allow model samples to be skipped, effectively allowing the embedded wavelet to contract. The dilation and contraction facilitated by the insert and delete states introduce a nonlinearity in alignment between the model and data that is not present in deterministic convolutional models.

To demonstrate the alignment procedure, we refer to Figure 6. Figure 6a shows a composite waveform derived from a candidate model from our set of all possible models that may explain the data trace shown in Figure 6b. One possible way to align the model and data trace is shown in Figure 6c. Consulting Figure 5, we start at the begin state (B) and proceed through the state machine as follows:

For this HMM, the first state is always a match state (M1), so we match the first sample from the model trace to the first sample from the data trace. From the M1 state, transitions exist to the I1, M2, and D2 states. In this example, the transition goes to the M2 state, so we match the second sample from the model trace to the second sample of the data trace. In a similar manner, we next transition through the M3, M4, and M5 states, matching the corresponding samples from the model trace to the data trace. From the M5 state, we can only transition to the end state (E) and we are finished. The end result shown in Figure 6c just matches each sample from the model trace to the data trace.

Another possible alignment is shown in Figure 6d. We start at the begin state, and as before, we transition to the M1 state matching the first sample from the data trace to the first sample of the model trace. Then we transition to the M2 state matching the second sample from the data trace to the second sample of the model trace. From the M2 state, we then transition to the D3 state, which means that the third sample of the model trace is discarded. The M4 state follows, which matches the third data sample to the fourth model sample. The fourth data sample is also aligned to the fourth model sample through the I4 state. Finally, the M5 state is traversed, matching the fifth data and model samples, and the end state (E) once again concludes the alignment.

The alignment in Figure 6e is similar to the one in Figure 6d. In this case, the D2 state is traversed rather than the D3 state. From visual inspection, the alignment of Figure 6d does a better job matching amplitudes than the alignment in Figure 6e. When comparing the alignments from Figure 6c and 6d, it is clear that the amplitudes are more closely matched for the four match states traversed in Figure 6d than they are in Figure 6c. For this reason, the alignment of Figure 6d may be considered the best qualitatively. The quantitative selection of the best alignment, however, will depend on the HMM parameterization as described below.

The next step in our procedure is to use the Viterbi algorithm to find the optimal alignment between a composite waveform in a data trace and a candidate model, which can be described by a sequence of states traced through the HMM. The Viterbi algorithm uses dynamic programming to determine the solution in $\mathcal{O}(N \times M)$ operations, where M is the number of states in the HMM (for example, in Figure 5, this HMM has 14 states) and N is the number of samples in a data trace.

The optimal alignment is defined as the alignment that maximizes the following probability:

$$P^{\pi}(\text{data}|\text{model}) = \Pi_i a^{\pi}_{kl} g^{\pi}_l(d_i), \tag{1}$$

where i is a sample index for the data trace, d_i is a sample of the data trace, a_{kl} is a transition probability associated with transitioning from state k to l, and g_l is a state-dependent probability distribution that determines how likely it is to observe the value of d_i in a given state. The state integer π represents a specific state sequence that determines the values of k and l. The optimal state sequence π^* maximizes the probability of fit between the data and the model and determines the optimal alignment.

The matrix a_{kl} determines the probability of transitioning from any given state k to any other state l. For a profile HMM, most values of this matrix are zero, and the nonzero components correspond to the edges in the state machine graph. Typically, transitions to match states are assigned higher probabilities than transitions to insert and delete states, which have the effect of penalizing wavelet

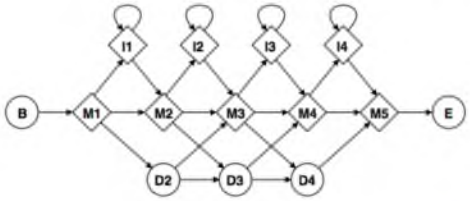

Figure 5. The profile HMM used for waveform analysis pictured as a finite state automaton.

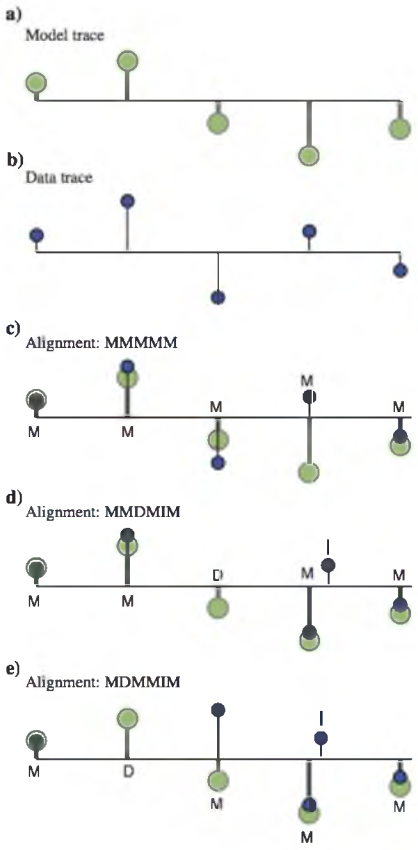

Figure 6. (a) A model trace used to build an HMM represented as a stem plot. (b) A data trace that is to be aligned to the model trace. (c, d, e) Three examples of alignments between the data and model traces. In each panel, the green and blue stems correspond to samples from the model and data trace, respectively.

dilation and contraction. Formal training algorithms exist that can optimize the values of a_{kl} given sufficient training data; however, the existence of adequate training data and overspecialization of the HMM to the training data are problematic. We have found empirically that match state transition probabilities of approximately 0.5 and insert/delete transition probabilities of approximately 0.25 work well in most cases. These probabilities can be fine-tuned if more or less contraction/dilation is desired. There is an interplay between transition and emission probabilities, but having match transitions approximately twice as likely as insert/delete transitions will provide for one or two samples of dilation or contraction where needed, without distorting the underlying wavelet beyond recognition. Each match and insert state has associated with it an emission probability distribution g_l. The mean of each state-dependent emission distribution is set to the amplitude of the associated model trace sample. Representing model trace sample amplitudes with Gaussian distributions accounts for random noise and errors in the wavelet estimate. The standard deviation of the amplitude distributions is set as a run-time parameter. The standard deviations should be inversely proportional to the S/N of the data trace because amplitude differences are penalized more harshly as the standard deviation shrinks.

For computational efficiency, the model and data traces are converted to an integer representation, and discrete probability mass functions are used in the HMM to limit the solution space. Also, computation of equation 1 is subject to numeric underflow because it involves the multiplication of many small probabilities. For this reason, the computations are carried out in a logarithmic space. More specifically, log-odds scores are calculated as a proxy for the probability in equation 1, and are defined as

$$S(\text{data}|\text{model}) = \log\left(\frac{P(\text{data}|\text{model})}{P(\text{data}|W)}\right), \qquad (2)$$

where W is a white-noise model. In this logarithmic computation space, the optimal alignment represents the path through the

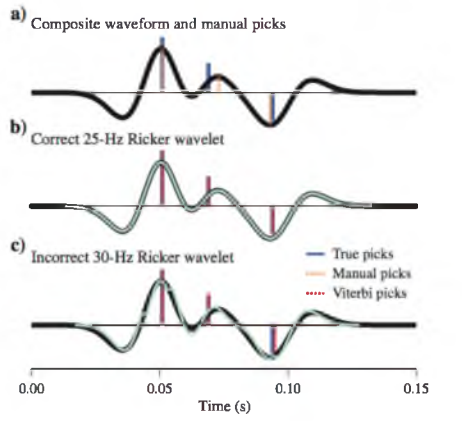

Figure 7. In each panel, the black trace is an example composite waveform (Figure 1) used to demonstrate the VSSD technique. (a) Manual picks for the three arrivals are shown as dashed orange lines, whereas true arrival times and amplitudes are represented by blue lines. (b) Viterbi picks (red lines) using the correct 25-Hz Ricker wavelet. (c) Viterbi picks using a 30-Hz Ricker wavelet. (b and c) Green waveforms represent the means of the emission distributions for all match and insert states traversed in the optimal alignments.

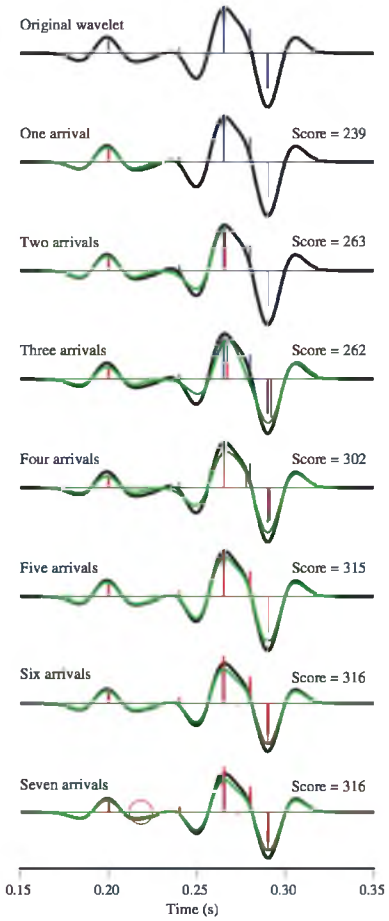

Figure 8. The behavior of the VSSD algorithm for solution spaces characterized by different numbers of arrivals is demonstrated by matching a composite waveform with five arrivals with models comprising one to seven arrivals. The black trace is the composite waveform consisting of five arrivals whose traveltimes and amplitudes are indicated by the blue spikes. Each of the lower panels displays the VSSD results and their log-odds scores for an increasing number of arrivals. In each panel, the red spike indicates the best fit arrival(s) and the green trace represents the modeled trace. Note that the relative amplitudes of the sixth and seventh (red circle) arrivals are very close to zero and that their inclusion does not significantly improve the waveform fit.

HMM that generates the data trace with the highest possible log-odds score. The log-odds score is a ratio of the likelihood of a given model producing a given data trace over the likelihood of a random model producing the data trace.

Referring back to Figure 6, the alignment in Figure 6c may be optimal if the HMM parameterization penalizes dilation and contraction (insert and delete states) more than amplitude errors (amplitude distributions). Conversely, the alignment in Figure 6d may be optimal if amplitude errors are penalized more than dilation and contraction.

Once the optimal log-odds score has been calculated for each model, the model with the highest score is chosen. This brings us to the final step in our procedure, which is to determine the amplitudes and differential traveltimes of our arrivals. The amplitudes are immediately available from the model parameters, and the arrival times of each event can be calculated by analyzing the optimal state path to determine the relative position of match states corresponding to each model spike.

The nonlinear alignment provided by the Viterbi algorithm is the key to this method's ability to work in the presence of noise and with imperfect knowledge of the source wavelet. Through the insert and delete states, the constituent wavelets are allowed to dilate and contract, which compensates for errors in the wavelet phase and amplitude spectra in tandem with random noise when combined with the Gaussian noise model. This ability is crucial to waveform analysis because wavelet estimates are never exact and wavelets in recorded seismograms are inherently nonstationary. This capability also enables interpretation of thin beds in depth-migrated images, in which the dominant wavelength of the embedded wavelet will change as a function of velocity. The construction of a profile HMM using a wavelet estimate and a Gaussian noise model is discussed below.

As an example, we analyze the composite waveform introduced in Figure 1. The composite waveform (black trace, Figure 7a) was created by convolving a 25-Hz Ricker wavelet with the spike train shown in blue. In this case, guided by our prior knowledge, we select an initial solution space consisting of three spikes. That is, for the composite waveform, we find the model that best reproduces the composite waveform using three arrivals. For reference, Figure 7a displays possible manual picks (dashed orange lines) for this composite trace. Figure 7b shows the result obtained through our proposed technique with the correct source wavelet. The dashed red lines show the spike train that the Viterbi process has determined to be the best model. As expected, in this case, we recover the amplitude and relative timing information exactly because we knew exactly how many arrivals we should search for and had perfect knowledge of the source. The green line in Figure 7b represents the aligned model waveform, which comprises the means of all match and insert states traversed in the optimal state path. It is notable that the technique is capable of properly determining the amplitude and timing of these arrivals in which there is strong constructive/destructive interference. However, precise prior knowledge of the source wavelet is usually not available. Figure 7c shows an example using our method with a 30-Hz source wavelet. Even though we use the incorrect wavelet, the traveltime and amplitudes recovered are still better than those picked manually. Note that the green waveform in Figure 7c contains discontinuities due to traversing insert states. This simple example demonstrates the utility of our proposed technique in determining arrival time and amplitude information for a composite wavelet with overlapping arrivals.

The number of arrivals modeled is a key parameter of the candidate solution space. Figure 8 demonstrates the behavior of the VSSD algorithm for different numbers of modeled arrivals. The first trace in Figure 8 illustrates a synthetic waveform composed of five arrivals. The true amplitudes and traveltimes are depicted by the blue lines. The subsequent traces show VSSD results for model spaces containing one to seven arrivals, along with the log-odds scores. When only one arrival is allowed, only a portion of the waveform is matched (the score only reflects the portion of the waveform matched). Allowing for two arrivals increases the portion of the waveform that is matched, but three arrivals are needed to cover the entire duration of the waveform. Note, however, that the log-odds score actually decreases when going from two arrivals to three arrivals. Despite the fact that more match states are traversed, the amplitude discrepancies are larger in the three-arrival result for this example. The four-arrival result increases the fidelity of the fit as well as the log-odds score. The five-arrival result provides a near-perfect fit. Increasing the number of modeled arrivals beyond five in this example does not change the result in a significant way, as the bottom two traces in Figure 8 demonstrate. The sixth and seventh picked arrivals represent spurious low-amplitude events,

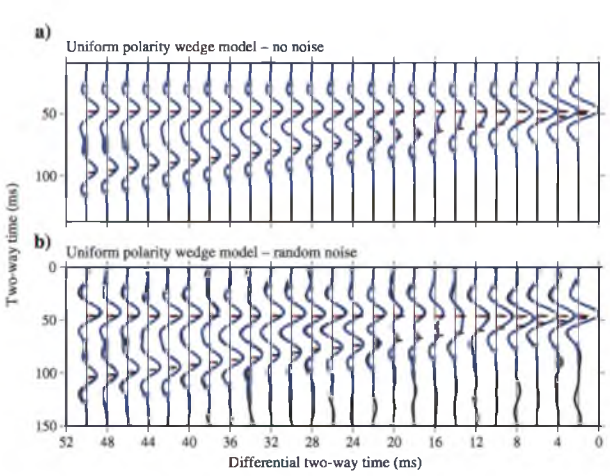

Figure 9. (a) Uniform-polarity wedge model (black curves and spikes) with Viterbi waveforms (blue traces) and picks (red spikes) overlain. The blue curves represent the Viterbi-aligned waveforms corresponding to the means of amplitude distributions for traversed match and insert states. The red spikes represent the amplitudes and traveltimes of the Viterbi-aligned spike train. (b) The same as (a) with the addition of band-limited random noise.

which are essentially fitting quantization errors from the conversion from the floating point to integer representations of the input trace. The amplitude of the seventh event (red circle) is so low that it is not visible in the figure.

Within a small window, an arbitrary number of arrivals may be modeled if the amplitudes for any given arrival are allowed to be zero. As the number of arrivals grows, the solution space grows exponentially, and it may become computationally unfeasible. To mitigate this problem, the method is generally applied in a series of small, overlapping windows in which only a small number of arrivals are modeled. Regularization may be applied by enforcing consistency of the results within the window overlaps. The problem of nonuniqueness can either be mitigated with a priori knowledge about the distribution of expected arrivals (e.g., by using impedance logs) or by penalizing results that do not show lateral consistency as interference patterns evolve across traces. The VSSD method may be used to tie well logs to seismic data in time or in depth because the nonlinear alignment can accommodate velocity dependent dominant wavelengths in the source function. The VSSD method may also be used to extend interpretations away from a well. In what follows, we show how this technique can be applied to determining bed thickness and arrival amplitude for wedge models similar to those in the Introduction.

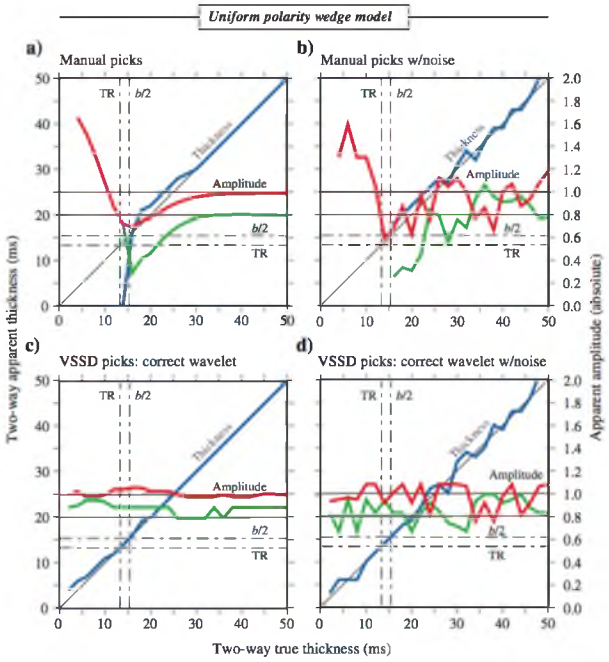

Figure 10. (a) Manual thickness and amplitude picks for the uniform-polarity wedge model. The blue line represents picked bed thickness, the red line represents picked amplitudes for the top interface, and the green line represents picked amplitudes for the lower interface. (b) The same as (a) except with the addition of band-limited random noise. (c) Viterbi thickness and amplitude picks using the correct 25-Hz, 15° phase-rotated wavelet. (d) The same as (c) except with the addition of band-limited random noise.

RESOLUTION TESTS

To examine the effectiveness of the Viterbi technique in determining arrival time and amplitude information for overlapping waveforms, we apply it to synthetic seismograms created for wedge models similar to those introduced in the Introduction. The advantage of analyzing the wedge models is that we can systematically examine amplitudes and differential traveltimes of arrivals through a steady variation in overlap of the arrival wavelets. First, we examine uniform-polarity wedge models, and then we examine the mixed-polarity case.

Uniform-polarity wedge models

A uniform-polarity wedge model is generated with zero-offset reflection coefficients for the upper and lower interfaces set at a constant ratio of 0.8 (see the "Introduction" section for further description on the basic design of the wedge models). That is, if we normalize the reflection coefficient of the upper interface to 1.0, then the lower interface has a normalized reflection coefficient of 0.8. The synthetic traces are constructed through convolution with a 25-Hz Ricker wavelet with a 15° phase rotation. A second set of synthetic traces is constructed with the addition of band-limited random noise with an S/N of 10. The nonuniform reflection coefficients, deviations from zero-phase, and the addition of random noise are intended to simulate real-world conditions more realistically.

Figure 9a shows the noise-free uniform-polarity wedge model waveforms (black traces) and spikes representing the true arrival times and amplitudes (black spikes), overlain by VSSD waveforms (blue traces) and VSSD spikes (red spikes). A common problem in deconvolution is that accurate information about the source time function is not always known. In this example, we use a 25-Hz, 15° phase-rotated wavelet, which is the correct source time function. When using the correct source, the Viterbi method provides almost-perfect traveltime variation all of the way down to the thinnest portion of the wedge. The largest deviations from the true thicknesses are 2 ms (5% of the dominant wavelength) at the smallest separation of the wedge model (at 2 and 4 ms differential two-way times; see Figure 9a). Figure 9b shows the results for the same uniform wedge model with the addition of band-limited random noise. The addition of the random noise increases the misfit between true arrival times (black spikes) and VSSD-picked arrival times (red spikes), but the maximum error has only increased to 3 ms or 7.5% of the dominant wavelength.

In Figure 10a, we show thicknesses and amplitudes from manually picked arrivals for the noise-free case. The blue line represents two-way thickness, while the red and green lines represent amplitudes for the upper and lower interfaces, respectively. Because two separate arrivals are not resolved below TR, the amplitude of the upper interface arrival increases significantly. Figure 10b shows the manual picks for the noisy uniform-polarity wedge model. In this case, the two separate arrivals are only resolved down to $b/2$ and the amplitude of the upper-layer arrival is not accurately recovered even at the largest thicknesses of the wedge model.

In Figure 10c, we summarize results from the VSSD-picked thicknesses and amplitudes, corresponding to the red spikes in Figure 8a. The two events are clearly resolved for all thicknesses, and amplitudes are also reasonably well estimated for all thicknesses. The quality of the thickness estimate degrades slightly in the neighborhood of $b/2$ and TR, but it is still much better than that obtained from manual picks. Figure 10d shows the VSSD picks for the noisy wedge model. Both events remain resolved for all thicknesses. The thickness is also recovered in the thinnest region of the wedge (2-ms two-way thickness). There is more scatter in the amplitude measurements, but there is no spike in amplitude for the upper-layer amplitude as is seen in the manually picked case.

Mixed-polarity wedge models

A mixed-polarity wedge model is generated with zero-offset reflection coefficients for the upper and lower interfaces set at a constant ratio of -0.8. That is, if we normalize the reflection coefficient of the upper interface to 1.0, then the lower interface has a normalized reflection coefficient of -0.8. Once again, the synthetic traces are constructed through convolution with a 25-Hz Ricker wavelet with a 15° phase rotation, and a second set of synthetic traces is constructed with the addition of band-limited random noise.

In Figure 11a, we show thicknesses and amplitudes from manually picked arrivals for the noise-free case. Once again, the blue line represents two-way thickness, while the red and green lines represent amplitudes for the upper and lower interfaces, respectively. Unlike the uniform-polarity case, the two separate arrivals are resolved below TR, but the thickness is overestimated as the two wavelets merge into an approximation of the first derivative of the wavelet (Widess, 1973). Also, the amplitudes of the two events become indistinguishable before $b/2$ is reached. Figure 11b shows the manual picks for the noisy uniform-polarity wedge model. The effect of the noise is primarily manifested in additional scatter in the amplitude picks.

In Figure 11c, we summarize results from the VSSD-picked thicknesses and amplitudes. The two events are clearly resolved for all thicknesses with much greater fidelity below TR than is the case using manual picks. The amplitude picks are nearly perfect except for a small region around Widess' $b/8$ criteria (Widess, 1973), where the combined wavelet approximates the first derivative. Figure 11d shows the VSSD picks for the noisy wedge model. These picks have the same general characteristics as the noise-free picks; however, noise has introduced scatter into the amplitude and thickness estimates.

Sensitivity to errors in the wavelet phase, wavelet frequency, and random noise

Because we would typically not expect to have an exact wavelet with which to perform our analysis, the sensitivity of the VSSD technique to errors in the wavelet estimate is explored with a suite of different Ricker wavelets. To test the method's sensitivity to frequency content, 15 wavelets with peak frequencies equally spaced between 18 and 32 Hz, representing errors of ±7 Hz, and with the correct 15° phase rotation were used. This range of frequency content represents a 39% increase in the dominant wavelength on the low end and a 22% decrease in the dominant wavelength on the high end. Twenty-one wavelets with the correct peak frequency and phase rotations equally spaced between −35° and 65°, representing phase errors of ±50°, were used to test the method's sensitivity to

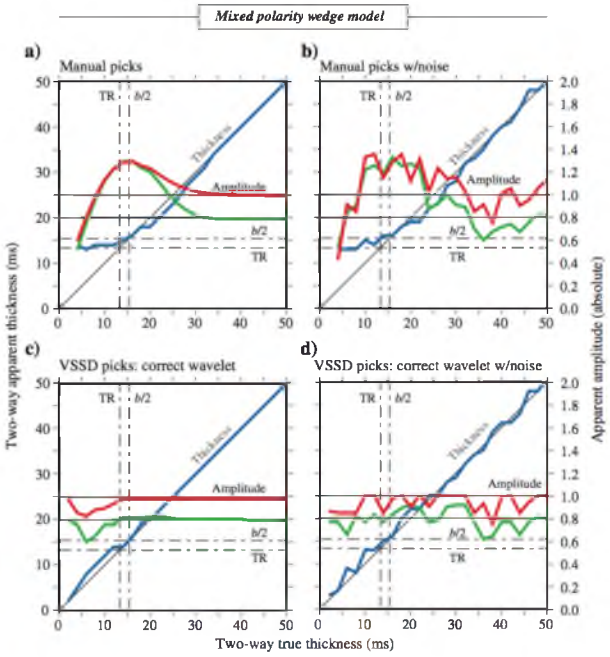

Figure 11. (a) Manual thickness and amplitude picks for the mixed-polarity wedge model. The blue line represents picked bed thickness, the red line represents picked amplitudes for the top interface, and the green line represents picked amplitudes for the lower interface. (b) The same as (a) except with the addition of band-limited random noise. (c) Viterbi thickness and amplitude picks using the correct 25-Hz, 15° phase-shifted wavelet. (d) The same as (c) except with the addition of band-limited random noise.

phase. The VSSD method begins to degrade for errors in phase and frequency content outside of these selected ranges, so that the two events are not consistently resolved for all thicknesses. However, the selected range of parameters provides an ample test suite spanning a larger range of errors than would typically be expected in a wavelet estimate. VSSD analysis was carried out separately for each of the 36 candidate wavelets for the noise-free and noisy uniform-polarity wedge models in Figure 9.

Figure 12 shows the VSSD-determined thicknesses for each of our test cases varying frequency, phase, and additive noise. Results for the 15 test wavelets with varying frequency content are shown in Figure 12a in the absence of noise. The color bar indicates the departure from the correct peak frequency of 25 Hz. The lower-frequency wavelets tend to underestimate the thickness slightly, while the higher-frequency wavelets tend to overestimate the thickness slightly, with a reversal of this general trend in the vicinity of TR and $b/2$ (i.e., between two-way thicknesses 13.33 and 15.38 ms). Wavelet contraction (delete states) tends to pull the center of the individual arrivals together when the frequency content is erroneously low, whereas wavelet dilation (insert states) tends to push the arrivals apart when the frequency content is erroneously high. The reversal of the trend between 12 and 22 ms is due to the effects of destructive interference between main and side lobes. The largest measured errors are 6 ms, with the errors being within 4 ms for all thickness greater than 8 ms. The addition of noise shows the same general trends as displayed in the noise-free model. However, the departure from true thickness is somewhat larger than in the noise-free model, showing that the addition of noise has a larger effect than the frequency content of the wavelet on thickness estimates.

The sensitivity to wavelet phase on thickness measurements is demonstrated in Figure 12b. This shows that thickness estimates are essentially insensitive to phase errors within the $\pm 50°$ range. One outlier, an 8-ms error corresponding to a 35° phase rotation, indicates the beginning of a departure from acceptable phase errors. For phase rotations larger than $\pm 50°$, the wavelet shape has departed significantly enough from the true wavelet such that the algorithm

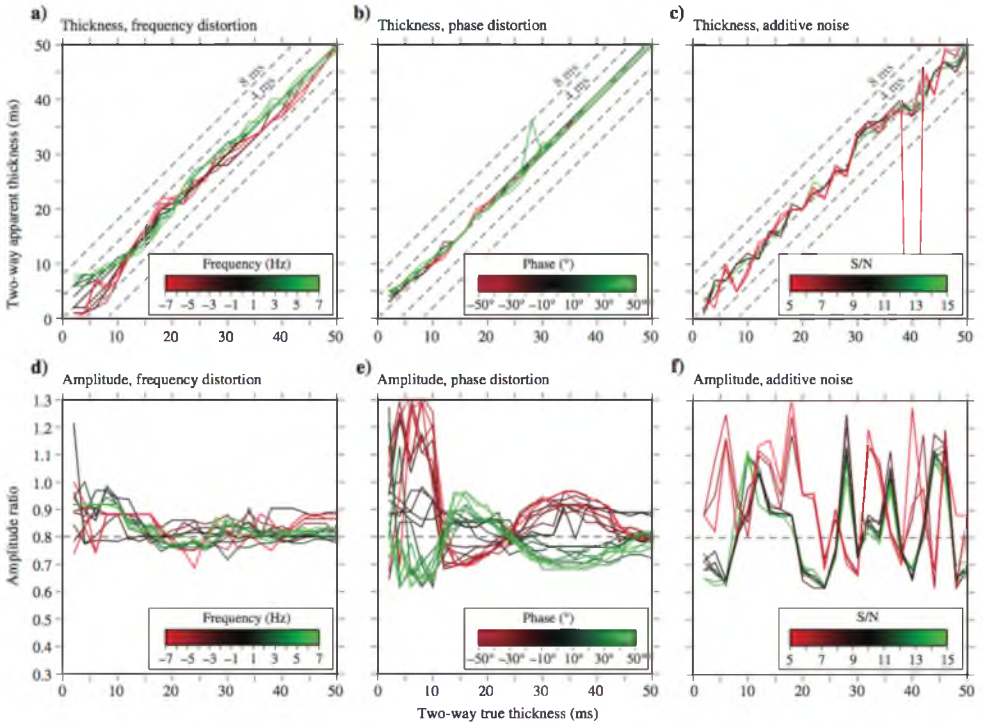

Figure 12. Apparent thickness curves for variations in the dominant frequency or phase of the source wavelet for the uniform-polarity wedge model. Results shown here in (a, b, d, e) are for the noise-free model. (a) Results for wavelets with dominant frequencies ranging from 18 to 32 Hz, representing an error range of ±7 Hz. (b) Results for wavelets with the correct dominant frequency, but with phase errors up to ±50°. (c) Results for wavelets with the correct dominant frequency and phase, but with additive noise with S/Ns from 5 to 15. (d) The same as in (a) except amplitude ratio measurements are displayed. (e) The same as in (b) except amplitude ratio measurements are displayed. (f) The same as in (c) except amplitude ratio measurements are displayed.

cannot accurately account for the effects of constructive and destructive interference across the entire range of differential traveltimes. The addition of noise slightly increases the error in thickness estimates, but it does not introduce a large departure from the noise-free case.

Figure 12d and 12e demonstrates the sensitivity of relative amplitude estimates to frequency content and phase rotation for the noise-free models. The value plotted for each wavelet is the ratio of the two picked amplitudes where the correct amplitude ratio is 0.8. Here, the sensitivities are reversed from those of thickness estimates. That is, when examining the thickness estimates, we observed that the measurements were less sensitive to differences in the wavelet phase than they were to differences in the frequency content. No clear dependence on frequency content can be inferred. That is, errors in amplitude ratio measurements for the noise-free and noisy case show deviations with no systematic trend with respect to wavelet frequency. Yet, phase errors demonstrate clear trends. Negative phase rotations (Figure 12e) tend to overestimate the amplitude ratio, while positive rotations tend to underestimate the ratio. This trend reverses in the 12–22-ms region, as was also noted in the frequency-dependent thickness estimates. The phase dependence of the amplitude estimates can be understood intuitively by considering the manner in which the shape of a wavelet changes as successively larger phase rotations are applied. Negative phase rotations tend to front load the energy in the wavelet, while positive phase rotations tend to back load the energy. For positive phase rotations at thicknesses below 12 ms, the back loading of energy gives more weight to the first arrival as the two wavelets constructively interfere. In the anomalous region between 12 and 22 ms, the opposite is true, as the back-loaded energy in the first arrival undergoes destructive interference. The addition of noise has more devastating effects on amplitude ratio measurements than on thickness estimates.

Figure 12c demonstrates the sensitivity of traveltime picks to different levels of random noise. In this case, the correct wavelet was used and the S/N varied from 5 to 15. The differential traveltime errors are within 4 ms, showing that the traveltimes are generally insensitive to the inclusion of random noise. However, at an S/N of five, a couple of the events are not correctly resolved, as indicated by the red lines that leave the area of the plot. In this case, a visual analysis of the waveforms would also give an erroneous interpretation. Figure 12f shows the sensitivity of amplitude picks to random noise. This result shows that the amplitude estimates are more sensitive to random noise than they are to wavelet errors in either phase or frequency, for all thicknesses. From these observations, one can infer that traveltime estimates are much more robust than amplitude estimates, especially in the presence of noise.

APPLICATION OF VSSD METHOD

We have presented the VSSD method for picking seismic arrival times and amplitudes. In certain cases, the VSSD method can resolve these arrivals below the standard resolution limits. In the previous section, we analyzed the sensitivity of this technique using a variable-thickness wedge model and synthetic seismograms. Here, we show the utility of this method with field data.

Our example demonstrates the application of the method at the exploration scale and comes from the Teapot Dome Oilfield. Teapot Dome is located ~35 miles north of Casper, Wyoming, on Naval Petroleum Reserve No. 3. The field has a long history of production dating back to the early 1900s, and it is currently used as a testing center for emerging technologies. Figure 13a shows an extraction of crossline 123 from a time-migrated 3D seismic cube provided by the Rocky Mountain Oilfield Testing Center (http://www.rmotc.doe.gov/datasets.html). Reflections in the analysis window are bounded by large, consistent arrivals (indicated by gray curves in Figure 13a). However, the central reflections exhibit substantial lateral variation due to subtle variations in bed thickness. The primary observations are as follows: (1) a pair of lower amplitude peaks visible just after the first large peak between inlines 75 and 88 that merges into a single peak as the inline number increases (feature is highlighted with dashed red lines), (2) increasing complexity of the preceding trough as the inline number increases (highlighted with a dashed green line), and (3) the third central peak is clearly visible in the lower inlines, but it diminishes by inline 108 (indicated with a dashed blue line). Here, we use the VSSD method to explore the subtle variations in these arrivals.

To use the VSSD technique, we must first choose an appropriate source time function. A wavelet estimate was obtained by aligning and stacking multiple traces along a set of consistent regional horizons. The results of our analysis are shown in Figure 13b for every

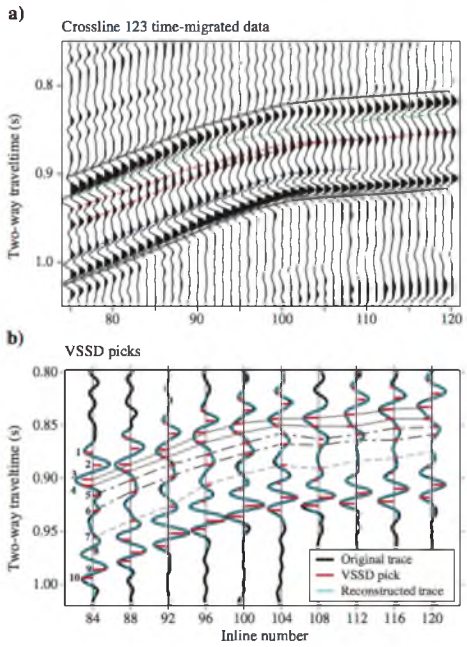

Figure 13. (a) A windowed portion of crossline 123 from the Teapot Dome 3D time-migrated data set. The gray lines indicate the VSSD analysis window. The dashed green, red, and blue lines indicate lateral variations in waveform character of particular interest. (b) VSSD picks (red lines) and reconstructed waveforms (blue curves) for every fourth trace. The VSSD method identified 10 arrivals (numbered on inline 84) with good lateral consistency across the section.

fourth trace from inlines 84–120. The method identifies 10 consistent reflectors that are numbered from 1 to 10 starting from the top and labeled for inline 84 in Figure 13b. One exception is the failure to resolve events 1 and 2 on inline 108. This means that the highest scoring alignment for that trace skipped several match states corresponding to these two arrivals. At either end of the analysis window, there remain constructive and destructive interference from neighboring events that are not considered. It is likely that the preceding events would need to be considered by expanding the analysis window to properly resolve events 1 and 2 on inline 108. The results corresponding to the primary observations of lateral inhomogeneity are discussed below.

- The two peaks highlighted by red dashed lines in Figure 13a correspond to arrivals 5 and 6 in Figure 13b (dashed-dotted line). The two peaks, clearly visible at inline 84, merge into a single arrival as the inline number increases, much like the wedge models from the Introduction. The VSSD picks show a small thinning trend as the inline number increases, but not as much as would be expected from a visual inspection of the waveform.
- Note, however, that the trough that precedes arrivals 5 and 6 (green dashed line in Figure 13a) corresponding to arrivals 3 and 4 in Figure 13b (solid gray lines) shows a thickening trend over this same lateral window. The result implies that the interference of all these arrivals produces a waveform that visually exaggerates the thinning of the bed bounded by arrivals 5 and 6. The combination of these four arrivals consistently matches the data across the range of inlines, indicating sedimentary beds with slight lateral thickness variations.
- The disappearing peak (blue dashed line in Figure 13a) corresponds to the zone of waveform interference between arrivals 7 (dashed gray line in Figure 13b) and 8. On the left side of the section, the peak correlates more strongly with arrival 7. By inline 96, the peak occurs between arrivals 7 and 8. As the inline number increases, the peak itself disappears, yet arrivals 7 and 8 remain with consistent polarities and relatively consistent amplitudes. One exception is the amplitude of arrival 8 on inline 88, which appears to be overestimated. This is likely due to strong constructive interference from arrivals 9 and 10. This amplitude may be better resolved by lengthening the analysis window.

Overall, the VSSD method, applied separately to each individual trace, has produced spike trains that match the data well while demonstrating a good degree of lateral consistency, despite the fact that there were no constraints promoting lateral consistency between traces. This improves our confidence in the VSSD method as a tool for thin-bed interpretation.

DISCUSSION

We have analyzed the sensitivity of the VSSD technique to variations in bed thickness, measuring apparent thickness and the amplitude ratio of arrivals reflecting off the layers. In addition, we added noise and source wavelet frequency and phase distortions. Table 1 summarizes the results from these tests. The apparent trends from Table 1 are

- Thickness estimates are relatively insensitive to errors in wavelet phase.
- When only random noise and frequency content are considered, amplitude estimates tend to be slightly more robust for mixed-polarity events.
- Amplitude estimates tend to be equally sensitive to random noise and large phase errors.
- Thickness estimates are much more robust than amplitude estimates.

Table 1. Results of the sensitivity of the VSSD technique to wavelet errors and random noise.

Uniform-polarity wedge models				Mixed-polarity wedge models			
Noise	Δf (Hz)	$\Delta\varphi$ (°)	Max error (%)	Noise	Δf (Hz)	$\Delta\varphi$ (°)	Max error (%)
Thickness measurements as a percentage of dominant wavelength							
No	0	0	5	No	0	0	5
Yes	0	0	5	Yes	0	0	5
No	±7	0	10	No	±7	0	12.5
Yes	±7	0	15	Yes	±7	0	15
No	0	±50	5	No	0	±50	5
Yes	0	±50	10	Yes	0	±50	7.5
Amplitude ratio measurements							
No	0	0	21	No	0	0	9
Yes	0	0	64	Yes	0	0	27
No	±7	0	25	No	±7	0	15
Yes	±7	0	64	Yes	±7	0	45
No	0	±50	62	No	0	±50	64
Yes	0	±50	64	Yes	0	±50	64

The VSSD technique is a versatile tool that provides accurate timing and amplitude information for seismic arrivals when these arrivals are nearly overlapping. However, this technique is not a blind-deconvolution method. That is, the method requires a priori knowledge about (1) the source wavelet and (2) the number of expected arrivals within the time window of interest.

With respect to the first point, the analyses in the section Resolution Tests primarily show the sensitivity of this technique with respect to errors in our knowledge of the seismic source. In exploration seismology applications, the source function is generally known well enough to be adequate for the VSSD method. For applications in earthquake seismology, we often do not have detailed knowledge of the source time function. However, empirical source time functions can be generated to alleviate this problem. For example, it is possible to stack reference seismic phases to build an empirical source time function (e.g., Thorne and Garnero, 2004).

With respect to the second point, precise knowledge of the number of arrivals within our time window is not necessary. In practice, we know the approximate number of expected arrivals. One can either test different numbers of arrivals, comparing log-odds scores and the quality of the waveform match, or one can allow the amplitudes of arrivals to be zero, which will explicitly search over multiple numbers of arrivals in a single experiment. Blocked impedance models derived from well logs present an excellent source of information to constrain the solution space. This method could be used to perform nonlinear seismic-well ties in either time or depth.

Figures in this paper were drawn using the Generic Mapping Tools (http://gmt.soest.hawaii.edu/; Wessel and Smith, 1998).

CONCLUSIONS

We have presented a stochastic method to decompose individual seismic traces into sparse spike trains describing the traveltimes and relative amplitudes of individual arrivals comprising band-limited signals. The method has been shown to have resolving power below Rayleigh's and Ricker's criteria under appropriate conditions. It has also been shown that estimated traveltimes are relatively insensitive to random noise and errors in the source time function estimate. No assumptions are made regarding the distribution of traveltimes and amplitudes; however, interpreted information can be used to limit the solution space. Application to field data has shown that single-channel sparse spike detection produces results with good lateral consistency. This technique shows promise in targeted exploration scale interpretation, and it may provide an alternative method of tying seismic images to well logs. Further applications may include decomposing overlapping pre- and postcursor teleseismic phases into individual arrivals or improving arrival time measurements in earthquake location or seismic tomography problems.

ACKNOWLEDGMENTS

We thank the Rocky Mountain Oilfield Testing Center (RMOTC) and the United States Department of Energy for providing the Teapot Dome seismic data set. We gratefully acknowledge the University of Utah Center for High Performance Computing (CHPC) for computer resources and support. We would also like to thank five anonymous reviewers for their constructive suggestions and comments. MT and SB were partially supported by National Science Foundation (NSF) grant EAR-0952187, and SB was partially supported by a grant from Total E&P.

REFERENCES

Bacon, M., R. Simm, and T. Redshaw, 2003, 3-D seismic interpretation: Cambridge University Press.

Clapp, R. G., 2008, Lloyd and Viterbi for QC and auto-picking: Stanford Exploration Project, SEP-134.

Durbin, R., S. R. Eddy, A. Krogh, and G. Mitchison, 1998, Biological sequence analysis: Probabilistic models of proteins and nucleic acids: Cambridge University Press.

Eddy, S. R., 1995, Multiple alignment using hidden Markov models: *in* C. J. Rawlings, ed., Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology: AAAI Press, 114–120.

Heimer, A., I. Cohen, and A. A. Vassiliou, 2007, Dynamic programming for multichannel blind seismic deconvolution: 77th Annual International Meeting, SEG, Expanded Abstracts, 1845–1849.

Heimer, A., I. Cohen, and A. A. Vassiliou, 2009, Multichannel seismic modeling and inversion based on Markov-Bernoulli random field: 79th Annual International Meeting, SEG, Expanded Abstracts, 2322–2326.

Kaaresen, K. F., and T. Taxt, 1998, Multichannel blind deconvolution of seismic signals: Geophysics, **63**, 2093–2107, doi: 10.1190/1.1444503.

Kallweit, R. S., and L. C. Wood, 1982, The limits of resolution of zero-phase wavelets: Geophysics, **47**, 1035–1046, doi: 10.1190/1.1441367.

Kormylo, J. J., and J. M. Mendel, 1983, Maximum-likelihood seismic deconvolution: IEEE Transactions on Geoscience and Remote Sensing, **GE-21**, 72–82, doi: 10.1109/TGRS.1983.350532.

Liner, C. L., and R. G. Clapp, 2004, Nonlinear pairwise alignment of seismic traces: Geophysics, **69**, 1552–1559, doi: 10.1190/1.1836828.

Ricker, N., 1953, Wavelet contraction, wavelet expansion, and the control of seismic resolution: Geophysics, **18**, 769–792, doi: 10.1190/1.1437927.

Sacchi, M. D., D. R. Velis, and A. H. Cominguez, 1994, Minimum entropy deconvolution with frequency-domain constraints: Geophysics, **59**, 938–945, doi: 10.1190/1.1443653.

Thorne, M. S., and E. J. Garnero, 2004, Inferences on ultralow-velocity zone structure from a global analysis of SPdKS waves: Journal of Geophysical Research, **109**, 1978–2012, doi: 10.1029/2004JB003010.

Velis, D. R., 2008, Stochastic sparse-spike deconvolution: Geophysics, **73**, no. 1, R1–R9, doi: 10.1190/1.2790584.

Viterbi, A., 1967, Error bounds for convolutional codes and an asymptotically optimum decoding algorithm: IEEE Transactions on Information Theory, **13**, 260–269, doi: 10.1109/TIT.1967.1054010.

Wessel, P., and W.H.F. Smith, 1998, New, improved version of generic mapping tools released: EOS — Transactions of the American Geophysical Union, **79**, 579, doi: 10.1029/98EO00426.

Widess, M. B., 1973, How thin is a thin bed?: Geophysics, **38**, 1176–1180, doi: 10.1190/1.1440403.

Wiggins, R. A., 1978, Minimum entropy deconvolution: Geoexploration: International Journal of Mining and Technical Geophysics and Related Subjects, **16**, 21–35, doi: 10.1016/0016-7142(78)90005-4.

CHAPTER 3

A COMPOSITIONAL ORIGIN TO ULTRALOW-VELOCITY ZONES

Reprinted with permission from John Wiley & Sons, Inc. Brown, S. P., M. S. Thorne, L. Miyagi, and S. Rost (2015), A compositional origin to ultralow-velocity zones, Geophys. Res. Lett., 42, 1039–1045, doi:10.1002/ 2014GL062097.

AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2014GL062097

Key Points:
- New constraints on a ULVZ provide a 1:1 V_S:V_P ratio
- The 1:1 ratio indicates a compositional origin to the ULVZ
- ULVZ composition may be Fe-enriched ferropericlase

Supporting Information:
- Figures S1–S7 and equations (S1)–(S7)

Correspondence to:
M. S. Thorne,
michael.thorne@utah.edu

Citation:
Brown, S. P., M. S. Thorne, L. Miyagi, and S. Rost (2015), A compositional origin to ultralow-velocity zones, *Geophys. Res. Lett.*, *42*, 1039–1045, doi:10.1002/2014GL062097.

Received 3 OCT 2014
Accepted 9 JAN 2015
Accepted article online 14 JAN 2015
Published online 18 FEB 2015

A compositional origin to ultralow-velocity zones

Samuel P. Brown[1], Michael S. Thorne[1], Lowell Miyagi[1], and Sebastian Rost[2]

[1]Department of Geology and Geophysics, University of Utah, Salt Lake City, Utah, USA, [2]School of Earth and Environment, University of Leeds, Leeds, UK

Abstract We analyzed vertical component short-period *ScP* waveforms for 26 earthquakes occurring in the Tonga-Fiji trench recorded at the Alice Springs Array in central Australia. These waveforms show strong precursory and postcursory seismic arrivals consistent with ultralow-velocity zone (ULVZ) layering beneath the Coral Sea. We used the Viterbi sparse spike detection method to measure differential travel times and amplitudes of the postcursor arrival *ScSP* and the precursor arrival *SPcP* relative to *ScP*. We compare our measurements to a database of 340,000 synthetic seismograms finding that these data are best fit by a ULVZ model with an *S* wave velocity reduction of 24%, a *P* wave velocity reduction of 23%, a thickness of 8.5 km, and a density increase of 6%. This 1:1 V_S:V_P velocity decrease is commensurate with a ULVZ compositional origin and is most consistent with highly iron enriched ferropericlase.

1. Introduction

A salient feature of the core-mantle boundary (CMB) region is the existence of ultralow-velocity zones (ULVZs). ULVZs have been detected using many seismic phases, including *SPdKS* [e.g., *Thorne and Garnero*, 2004], *PcP* [e.g., *Mori and Helmberger*, 1995], *ScP* [e.g., *Garnero and Vidale*, 1999], *ScS* [*Avants et al.*, 2006], *PKP* precursors [e.g., *Vidale and Hedlin*, 1998], and anomalies in travel time or slowness of a variety of different phases [e.g., *Xu and Koper*, 2009]. These studies show that ULVZs are characterized by a wide range of elastic parameters. For example, *S* wave velocity reductions (δV_S) have been reported as large as 45% (all percentages reported with respect to the Preliminary Reference Earth Model (PREM) [*Dziewonski and Anderson*, 1981]), *P* wave velocity reductions (δV_P) as large as 20%, density increases ($\delta\rho$) of up to 10%, and thicknesses of up to 40 km (see *Thorne and Garnero* [2004] for a review). Nevertheless, strong trade-offs typically exist in the model space [e.g., *Garnero and Helmberger*, 1998], and many of these parameters are uncertain.

Of the elastic parameters, the *P* wave velocity reduction is the most well constrained parameter. The seismic phase *SPdKS* is primarily sensitive to the *P* wave velocity reduction at the base of the mantle [see, e.g., *Rondenay et al.*, 2010]. In the southwest Pacific Ocean region, *P* wave velocity reductions of 10 to 15% are well constrained [*Thorne et al.*, 2013; *Zhang et al.*, 2009] in at least one ULVZ. Density contrast is less certain, with one study providing a constraint of 10 ± 5% in a ULVZ beneath the Coral Sea [*Rost et al.*, 2005] and another study providing density constraints in ULVZs beneath the Philippine Islands from 5–10% and 20–25% [*Idehara*, 2011]. The *S* wave velocity reduction appears quite variable, with one study demonstrating an *S* wave velocity reduction of roughly 7% beneath the central Pacific using postcursors to the *ScS* phase [*Avants et al.*, 2006], while *Idehara et al.* [2007] examined *ScP* postcursors beneath the Philippine Sea to show that the *S* wave velocity decrease must be at least 20%. Additional evidence beneath the Coral Sea from *ScP* arrivals shows an *S* wave velocity decrease of 24% [*Rost et al.*, 2006]. Due to the trade-off between *P* wave velocity, *S* wave velocity, and thickness, it is difficult to determine the ULVZ elastic parameters precisely. Nonetheless, constraining these parameters is paramount in determining what ULVZs physically represent.

The seismic phase *ScP* is of utmost importance in studying ULVZs as it is sensitive to all elastic parameters [*Garnero and Vidale*, 1999]. *ScP* is an *S* wave that converts to a *P* wave at its reflection on the CMB. Figure 1a shows the *ScP* raypath through the mantle, with the direct *P* wave path for comparison. The reason *ScP* is well suited for ULVZ studies is because of the existence of at least two precursors and one postcursor that are predicted if the *ScP* wavefield interacts with a ULVZ. These additional phases (Figure 1b) are (1) *SdP*—a precursor occurring from the reflection off the top of the ULVZ, (2) *SPcP*—a precursor occurring when the downgoing *S* wave converts to a *P* wave at the top of the ULVZ, and (3) *ScSP*—a postcursor that occurs when the upgoing *ScS* wave converts to a *P* wave at the top of the ULVZ. A synthetic seismogram showing the predicted *ScP* waveform for the PREM model is compared to the predicted waveform for a model including a

AGU Geophysical Research Letters 10.1002/2014GL062097

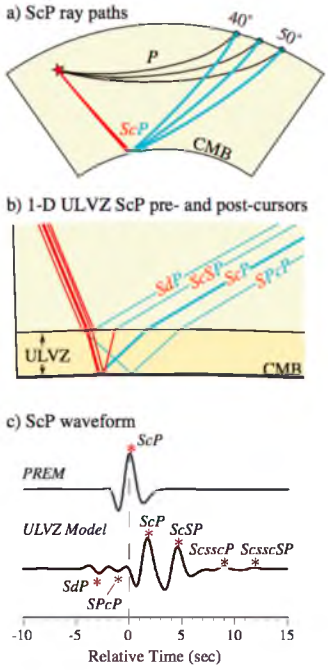

Figure 1. (a) Direct *P* wave (black) and *ScP* rays are shown at epicentral distances of 40°, 45°, and 50°. (b) Detail of raypaths near the core-mantle boundary for precursory and postcursory rays associated with a ULVZ. Raypaths are drawn for a ULVZ model with thickness = 20 km, $\delta V_S = -30\%$, and $\delta V_P = -10\%$. Rays are color coded red and blue for the *S* and *P* legs of the raypaths, respectively. (c) Vertical component displacement synthetic seismograms calculated for the PREM (Figure 1c, top) and ULVZ (Figure 1c, bottom) models. The ULVZ model is the same as for which rays are drawn in Figure 1b. Seismograms are aligned at zero time on the PREM *ScP* arrival. In addition to the arrivals drawn in Figure 1b, additional reverberations inside the ULVZ are also observed—phases labeled *ScsscP* and *ScsscSP*. Raypaths and synthetics in Figures 1b and 1c are calculated for an epicentral distance of 45°.

1-D ULVZ in Figure 1c. Additional reverberations within the ULVZ, phases labeled *ScsscP* and *ScsscSP* in Figure 1c, can be observed in synthetics yet are typically too low amplitude to be observed in the data.

The value of *ScP* in determining the *S* wave velocity reduction is that the amplitudes of the *SdP* and *ScSP* arrivals are most strongly dependent on the *S* wave velocity reduction [*Rost and Revenaugh*, 2003]. However, precise determination of the amplitudes of the *ScSP* postcursor is complicated by the presence of the *ScP* coda, which is most pronounced in the short-period data. Additionally, both of the precursor arrivals are predicted to have relatively low amplitudes with respect to *ScP* (see Figure 1c), and the majority of studies have only identified a single precursor due to the interference of the two low-amplitude phases [e.g., *Rost et al.*, 2005]. In this paper we examine *ScP* waveforms interacting with a known ULVZ in the southwestern Pacific Ocean region [*Rost et al.*, 2005, 2006]. We use the Viterbi sparse spike detection (VSSD) method [*Brown and Thorne*, 2013] to accurately determine differential travel times and amplitude ratios for *ScP* precursor and postcursor arrivals. We compare this set of measurements to synthetic predictions constraining *S* and *P* wave velocity reductions in a single ULVZ.

2. *ScP* Data and Viterbi Sparse Spike Detection Method

We analyze a data set of *ScP* records from 26 earthquakes in the Fiji-Tonga area recorded at the Alice Springs Array located in central Australia. Individual records were band-pass filtered between 1 and 4 Hz and beamformed on the *ScP* arrival to provide a single high signal-to-noise ratio trace for each event (see *Rost et al.* [2006] for the details on event beams, event locations, and bouncepoint location). These events provide a tight cluster of *ScP* bounce points (roughly 30 km × 30 km) on the CMB centered at 167.5°E longitude and 25°S latitude, spanning an epicentral distance range from 40.9° to 42.8°. *Rost et al.* [2006] identified a precursor (arriving 1.9 s prior to the *ScP* arrival) identified as the *SdP* arrival. An *SPcP* precursor and *ScSP* postcursor are also apparent in these data; yet these arrivals are partially masked by destructive interference from the *ScP* arrival.

We use the Viterbi sparse spike detection (VSSD) technique [*Brown and Thorne*, 2013] to search for *ScP* precursor and postcursor directly in the recorded traces while simultaneously obtaining estimates of their relative amplitudes and differential travel times. Given an approximate source time function, the VSSD method provides a statistically robust way to search a predefined set of spike trains for the best nonlinear fit to a composite waveform, accounting for Gaussian distributed random noise, errors in the source time function, and constructive/destructive interference between closely spaced arrivals. A synthetic trace is built for each candidate spike train through convolution with the source time function. A profile hidden Markov model (HMM) [*Eddy*, 1995] is built for each synthetic trace. The profile HMM represents a stochastic model for generating synthetic waveforms, the most likely of

AGU Geophysical Research Letters 10.1002/2014GL062097

which will be the synthetic upon which the model is based. The model consists of a linear array of match states, where each match state corresponds to a sample in the synthetic seismogram, and is characterized by a Gaussian distribution centered on the synthetic sample amplitude. The linear topology is augmented with insert and delete states between match states, which allow the waveforms to contract and dilate, facilitating nonlinearity in waveform alignments. The Viterbi algorithm [*Viterbi*, 1967] is applied to find the optimal nonlinear alignment between a data trace and the profile HMM. This optimal alignment represents the path through the HMM, which generates the data trace with the highest probability. The optimal path, π^*, is the state sequence which optimizes the following equation:

$$P^{\pi}(\text{Data}|\text{Model}) = \prod_i a^{\pi}_{kl}\, g^{\pi}_l(d_i), \tag{1}$$

where i is a sample index for the data trace, d_i is a sample of the data trace, a_{kl} is a transition probability associated with transitioning from state k to l, and g_l is a Gaussian distribution associated with each sample amplitude in the synthetic model. The state integer π represents a specific state sequence that determines the values of k and l. A log-odds score S is assigned to each profile HMM, based on the optimal alignment to a data trace:

$$S(\text{Data}|\text{Model}) = \log\left(\frac{P(\text{Data}|\text{Model})}{P(\text{Data}|W)}\right), \tag{2}$$

where W is a white noise model. The log-odds score is a ratio of the likelihood of a given model producing a given data trace over the likelihood of a white noise model producing the data trace. Travel times and amplitudes for each individual arrival in a complex waveform are derived from the highest scoring model. Relative amplitudes are taken directly from the spike train from which the model was built, and differential travel times are adjusted based on the optimal alignment.

We create an empirical source, between 2 and 3 s in length, composed of a windowed and tapered direct *P* wave arrival for each event (Figure 2a). The direct *P* wave arrival was too noisy to provide a suitable source time function estimate for 10 of the 26 events, which led to their exclusion from this analysis. The *ScP* arrival travels a longer path through the mantle and is more attenuated than the *P* wave arrival. Thus, the frequency content of our empirical source is slightly higher than for the *ScP* arrival, but this variation in frequency content is accommodated for by the VSSD method [*Brown and Thorne*, 2013].

3. Results

Figure 2a shows an example where we search for multiple (1 to 6) arrivals in a beamed trace from a single event recorded at Alice Springs Seismic Array using the VSSD method. When only one arrival is searched for, the algorithm finds the *ScP* arrival. When we search for two arrivals, we recover the *ScSP* arrival, the next highest-amplitude arrival after *ScP*. The third arrival finds an additional postcursor, and the fourth arrival finds the *SPcP* precursor. Looking for five and six arrivals does not reveal a consistent *SdP* arrival, and the increase in log likelihood score starts to flatten. Additionally, the amplitude of arrivals greater than four becomes increasingly lower and within the noise level; thus, we expect that no more than four arrivals can consistently be determined in these data. The overall data fit increases when searching for five or six arrivals but does not significantly influence the travel time or amplitude of the arrivals already determined, and we hence limit our search to the five largest amplitude arrivals in the *ScP* wavefield for the events in our data set.

Figure 2b shows VSSD alignments and picks for the 16 events analyzed. The inclusion of the postcursor beyond *ScSP* is necessary to get a good waveform fit for *ScSP*. This additional postcursor is possibly a diffracted arrival off the far edge of the ULVZ and is too large amplitude to be consistent with one of the ULVZ reverberations shown in Figure 1c. Figure 3a shows the differential travel times versus relative amplitudes for all 16 events. The average *ScP*-*SPcP* differential travel time is 0.9094 ± 0.0455 s, and the average *ScSP*-*ScP* time is 0.8625 ± 0.0806 s. Average amplitude ratios are *SPcP*/*ScP* = −0.2555 ± 0.0671 and *ScSP*/*ScP* = 0.5073 ± 0.1448.

Synthetic seismograms are computed using the generalized ray method (GRM) [*Helmberger*, 1983]. We use the PREM background model with a 1-D ULVZ model embedded at the base of the mantle. Synthetic seismograms are computed for a 500 km deep event at an epicentral distance of 42°. We used the GRM in order to model individual arrivals obtaining a spike train similar to those obtained using the VSSD technique. We compute synthetic seismograms where we allow the following ULVZ parameters to vary: (1) δV_S decreases from 0% to 50% in 1% increments, (2) δV_P decreases from 0% to 30% in 1% increments, (3) $\delta\rho$ variations from

AGU Geophysical Research Letters 10.1002/2014GL062097

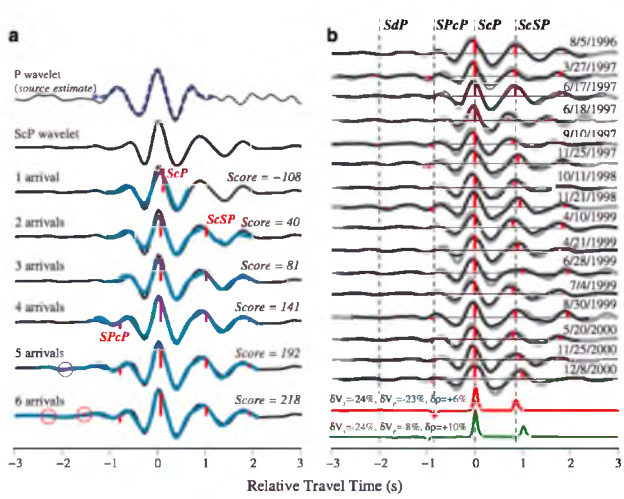

Figure 2. (a) Analysis of a single event occurring on 25 November 2000. The top trace (black line) shows the direct *P* wave. The dashed blue line shows the windowed *P* wavelet used as the source time function estimate. The second trace from the top (heavy black line) shows the raw *ScP* wavelet which is repeated for each trace below. Each subsequent trace shows the best fit VSSD solution for this event using 1 through 6 arrivals. In each case, the red spike shows the best fit amplitude and travel time for the number of arrivals being fit. The light blue trace is an overlay of the source time function convolved with the best fit spike train. The log-odds score is shown to the right for each fit. Because the amplitudes for arrivals 5 and 6 are low, we have highlighted their positions with red circles. (b) Travel time and amplitude picks (red spikes) computed by the VSSD method for the 16 events analyzed in this study are shown. The beam-formed *ScP* trace for each event is shown as the gray trace. The VSSD-derived waveforms are overlain (black traces). The lowermost two traces are GRM synthetic predictions convolved with a 0.25 s dominant period source time function at an epicentral distance of 42°. The red trace is for a ULVZ model with $\delta V_S = -24\%$, $\delta V_P = -23\%$, thickness = 8.5 km, and density = +6%, and the green trace is for a ULVZ model with $\delta V_S = -24\%$, $\delta V_P = -8\%$, thickness = 8.5 km, and density = +10%.

−10% to 20% in 1% increments, and (4) ULVZ thickness (h) from 1 to 20 km in 1 km increments. In total, we have a database of synthetic seismograms for nearly 340,000 ULVZ models.

In order to characterize the likelihood of models from the suite of synthetics, four Gaussian probability mass functions were defined from the means and variances of the measured differential travel times and relative amplitudes. Figure 3b shows P_1 ($\partial V_S, \partial V_P$), the likelihood of models considering only differential travel times and amplitudes of *ScSP* relative to *ScP*, which we define as

$$P_1(\partial V_S, \partial V_P) = \text{argmax}_h(< P_{aScSP}(\partial V_S, \partial V_P, \delta\rho, h) * P_{tScSP}(\partial V_S, \partial V_P, \delta\rho, h) >), \quad (3)$$

where h denotes ULVZ thickness, $P_{aScSP}(\partial V_S, \partial V_P, \delta\rho, h)$ is the probability assigned to the *ScSP*/*ScP* differential amplitude for a given synthetic model, $P_{tScSP}(\partial V_S, \partial V_P, \delta\rho, h)$ is the probability assigned to the *ScSP*-*ScP* differential travel time, and <> denotes an ensemble average over density variations. Examination of Figure 3b demonstrates that the *ScSP* postcursor provides a strong constraint on δV_S, giving an approximate value of −25%, which is consistent with the *Rost et al.*'s [2006] estimate of δV_S.

Figure 3c shows P_2 ($\partial V_S, \partial V_P$), the likelihood of models considering differential travel times and amplitudes of both *ScSP* and *SPcP* relative to *ScP*, which we define as

$$P_2(\partial V_S, \partial V_P) = \text{argmax}_h \begin{pmatrix} < P_{aScSP}(\partial V_S, \partial V_P, \delta\rho, h) * P_{tScSP}(\partial V_S, \partial V_P, \delta\rho, h)^* \\ P_{aSPcP}(\partial V_S, \partial V_P, \delta\rho, h) * P_{tSPcP}(\partial V_S, \partial V_P, \delta\rho, h) > \end{pmatrix}, \quad (4)$$

where $P_{aSPcP}(\partial V_S, \partial V_P, \delta\rho, h)$ is the probability assigned to the *SPcP*/*ScP* differential amplitude for a given synthetic model, $P_{tSPcP}(\partial V_S, \partial V_P, \delta\rho, h)$ is the probability assigned to the *ScP*-*SPcP* differential travel time pick.

AGU Geophysical Research Letters 10.1002/2014GL062097

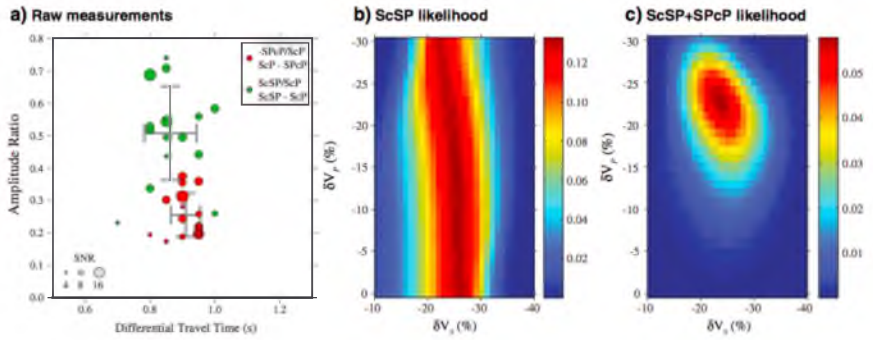

Figure 3. (a) VSSD measured differential travel times versus relative amplitudes. The red circles show *ScP-SPcP* and −*SPcP/ScP* differentials, whereas the green circles show the *ScSP-ScP* and *ScSP/ScP* differentials. The gray crosses show the average and 1σ error bars for the measurements. The size of the circles is scaled by the signal-to-noise ratio of the *ScP* beam. (b) Velocity variation likelihood, constrained only by *ScSP* travel times and amplitudes. (c) Velocity variation likelihood, constrained by *ScSP* and *SPcP* travel times and amplitudes.

We see a clear peak in the likelihood plot (Figure 3c) located at $\delta V_P = -23\%$ and $\delta V_S = -24\%$. The most likely individual model is given by

$$P_3 = \max\left(\frac{P_{aScSP}(\partial V_S, \partial V_P, \delta\rho, h) * P_{tScSP}(\partial V_S, \partial V_P, \delta\rho, h)^*}{P_{aSPcP}(\partial V_S, \partial V_P, \delta\rho, h) * P_{tSPcP}(\partial V_S, \partial V_P, \delta\rho, h)} \right) \quad (5)$$

This provides a best fit model with the following parameters: $\delta V_S = -24\%$, $\delta V_P = -23\%$, $\delta\rho = +6\%$, and $h = 8.5$ km. The synthetic seismogram for this best fit model is shown in Figure 2b. The additional information from *SPcP* resolves the ambiguity between δV_P and h. Density variation within the ULVZ is the least well constrained of the elastic parameters and does not lead to a significant change in model probability over the range of density variations tested.

4. Discussion and Conclusions

The results of *Rost et al.* [2006] fit synthetic waveforms and determined a best fit model of $\delta V_S = -24\%$, $\delta V_P = -8\%$, $\delta\rho = +10\%$, and $h = 8.5$ km, which we will refer to as the *partial-melt model* due to its 3:1 V_S:V_P ratio. A synthetic for this model is also shown in Figure 2b. There is not a large amount of difference in amplitude or travel time for the *SPcP* and *ScSP* arrivals for these different predictions. Yet the partial-melt model predicts a negative polarity *SdP* arrival, whereas our best fit model predicts a positive polarity *SdP* arrival. We are not able to constrain amplitude or travel time of the *SdP* arrival in our individual event beams. We further apply the VSSD technique to a trace created by forming a double beam of all events in order to further enhance an *SdP* arrival. However, the *SdP* arrival is within the noise level, and the polarity of the *SdP* arrival cannot be constrained (see supporting information); hence, the partial-melt model cannot be entirely ruled out.

The origins of ULVZs have been suggested to be due to a variety of causes including a metal-bearing layer [*Manga and Jeanloz*, 1996], iron-enriched postperovskite [*Mao et al.*, 2006], iron enrichment of ferropericlase (Fp) [*Wicks et al.*, 2010], and partial melt [*Berryman*, 2000; *Labrosse et al.*, 2007; *Williams and Garnero*, 1996]. Of these scenarios, partial melt has received the most attention. This partial-melt origin for ULVZs predicts a 3:1 ratio for reduction of V_S:V_P wave speeds [*Williams and Garnero*, 1996], which is incompatible with our most likely model which has a 1:1 ratio. This reduction is consistent with compositional or mineralogical variations, and thus, we explore these scenarios.

Using elastic properties and densities available in the literature, we calculate changes in velocity and density for several mineralogical models at conditions of 136 GPa and 3500 K. Since the detailed mineralogical and chemical composition at the base of the mantle is uncertain, rather than attempt to model a complete mineralogical mixture, we perform a computation where the mineral phase of interest is mixed with PREM

velocities and densities [*Wicks et al.*, 2010]. We explore possible enrichment of Fp with 35% Fe (Fp35) [*Chen et al.*, 2012], Fp with 84% Fe (Fp84) [*Wicks et al.*, 2010], hexagonal close packed Fe [*Lin et al.*, 2005], $FeSiO_3$ postperovskite [*Stackhouse and Brodholt*, 2008], Fe_2O_3 postperovskite [*Stackhouse and Brodholt*, 2008], $FeSiO_3$ perovskite [*Stackhouse et al.*, 2006], and B2 structured FeSi [*Caracas and Wentzcovitch*, 2004; *Ono*, 2013; *Vočadlo et al.*, 2002]. We find that 25–30% enrichment of Fp84 is the only model tested that provides a decrement ratio of 1:1 with a 23% decrease in velocities and a 6% increase in density. Notably, Fp35 does not provide a 1:1 decrease at these velocity reductions, which implies that very high Fe content is required (i.e., in excess of 35% substitution). Thus, our seismic data are most consistent with a solid state ULVZ enriched in high Fe content Fp, perhaps even approaching pure wüstite (FeO) end-member composition.

Since it is controversial if temperatures at the base of the mantle are hot enough to induce significant melting of mantle material [*Andrault et al.*, 2011], and because our observations are incompatible with partial melt, we propose a mechanism that does not require melting of mantle material. In this mechanism, ULVZs are formed through entrainment of core material followed by chemical reaction and gravitational settling. This mechanism can explain observations for the formation of solid state (1:1) and partially molten (3:1) ULVZs. When Fp is in contact with molten Fe, a morphological instability occurs, allowing liquid iron droplets to penetrate deep into Fp, providing a mechanism for entrainment of core material into the mantle [*Otsuka and Karato*, 2012]. Liquid Fe has a very high affinity for oxygen, so it is expected that liquid Fe should "strip" oxygen away from the surrounding mantle material [*Asahara et al.*, 2007]. As this Fe-enriched material percolates upward from the core-mantle boundary, the steep geotherm in this region is likely to cause precipitation of solid material. We expect that since the amount of Fe drawn out of the core is relatively small compared to the volume of the surrounding mantle [*Otsuka and Karato*, 2012], the Fe droplets should become highly enriched in oxygen. If there is sufficient oxygen enrichment, the material lies on the oxygen-rich side of the Fe-FeO eutectic, pure solid FeO will precipitate [*Seagle et al.*, 2008]. Reactions between liquid Fe and mantle silicates to form FeO and FeSi have been documented in high pressure-temperature experiments [e.g., *Knittle and Jeanloz*, 1989]. If FeSi is produced by such a reaction [*Knittle and Jeanloz*, 1989], and it has a lower melting temperature than Fe [e.g., *Santamaría-Pérez and Boehler*, 2008], it should remain in the molten state, likely alloyed with the molten Fe droplets. This residual molten FeSi alloy will be denser than mantle silicate [*Caracas and Wentzcovitch*, 2004] and may eventually drain back to the core in the liquid state but may temporarily form partially molten ULVZs. Percolation of Fe droplets into the mantle is strongly temperature controlled [*Otsuka and Karato*, 2012], and we expect that this process is broadly occurring across the CMB. However, in hotter regions (e.g., near Large Low Shear Velocity Provinces (LLSVPs)), this process will be more efficient. Localized regions of Fe percolation could also give rise to temporary partially molten ULVZs. However, as the Fe reacts with silicate mantle, FeO collection at the base of the mantle will occur. *McNamara et al.* [2010] have shown that a dense ULVZ-like layer will get swept toward the edges of LLSVPs. Hence, FeO-rich ULVZs may pile up into ULVZs observed at LLSVP boundaries.

Acknowledgments

We gratefully acknowledge the University of Utah Center for High Performance Computing for the computer resources and support. M.T. and S.B. were partially supported by NSF grant EAR-1014749. S.R. was partially supported by NERC grant NE/H022473/1. L.M. was partially supported by NSF grant EAR-1344579. L.M. would like to thank Zhixue Du for the helpful discussions. Figures were drawn using the Generic Mapping Tools [*Wessel and Smith*, 1998]. Data used in this study may be downloaded from http://web.utah.edu/thorne/pubs.html.

The Editor thanks an anonymous reviewer for assisting in the evaluation of this paper.

References

Andrault, D., N. Bolfan-Casanova, G. Lo Nigro, M. A. Bouhifd, G. Garbarino, and M. Mezouar (2011), Solidus and liquidus profiles of chondritic mantle: Implication for melting of the Earth across its history, *Earth Planet. Sci. Lett.*, *304*, 251–259, doi:10.1016/j.epsl.2011.02.006.

Asahara, Y., D. J. Frost, and D. C. Rubie (2007), Partitioning of FeO between magnesiowüstite and liquid iron at high pressures and temperatures: Implications for the composition of the Earth's outer core, *Earth Planet. Sci. Lett.*, *257*, 435–449, doi:10.1016/j.epsl.2007.03.006.

Avants, M., T. Lay, and E. J. Garnero (2006), A new probe of ULVZ *S* wave velocity structure: Array stacking of *ScS* waveforms, *Geophys. Res. Lett.*, *33*, L07314, doi:10.1029/2005GL024989.

Berryman, J. G. (2000), Seismic velocity decrement ratios for regions of partial melt in the lower mantle, *Geophys. Res. Lett.*, *27*(3), 421–424, doi:10.1029/1999GL008402.

Brown, S. P., and M. S. Thorne (2013), Viterbi sparse spike detection, *Geophysics*, *78*(4), V157–V169, doi:10.1190/GEO2012-0209.1.

Caracas, R., and R. M. Wentzcovitch (2004), Equation of state and elasticity of FeSi, *Geophys. Res. Lett.*, *31*, L20603, doi:10.1029/2004GL020601.

Chen, B., J. M. Jackson, W. Sturhahn, D. Zhang, J. Zhao, J. K. Wicks, and C. A. Murphy (2012), Spin crossover equation of state and sound velocities of $(Mg_{0.65}Fe_{0.35})O$ ferropericlase to 140 GPa, *J. Geophys. Res.*, *117*, B08208, doi:10.1029/2012JB009162.

Dziewonski, A. M., and D. L. Anderson (1981), Preliminary reference Earth model, *Phys. Earth Planet. Inter.*, *25*, 297–356.

Eddy, S. R. (1995), Multiple alignment using hidden Markov models, in *Proceedings of the Third International Conference Intelligent Systems for Molecular Biology*, pp. 114–120.

Garnero, E. J., and D. V. Helmberger (1998), Further structural constraints and uncertainties of a thin laterally varying ultralow velocity layer at the base of the mantle, *J. Geophys. Res.*, *103*(B6), 12,495–12,509, doi:10.1029/98JB00700.

Garnero, E. J., and J. E. Vidale (1999), *ScP*: A probe of ultralow velocity zones at the base of the mantle, *Geophys. Res. Lett.*, *26*(3), 377–380, doi:10.1029/1998GL900319.

Helmberger, D. V. (1983), Theory and application of synthetic seismograms, in *Earthquakes: Observation, Theory and Interpretation*, edited by H. Kanamori, pp. 173–222, Soc. Ital. di Fis, Bologna.

Idehara, K. (2011), Structural heterogeneity of an ultralow velocity zone beneath the Philippine Islands: Implications for core-mantle chemical interactions induced by massive partial melting at the bottom of the mantle, *Phys. Earth Planet. Inter.*, *184*, 80–90, doi:10.1016/j.pepi.2010.10.014.

Idehara, K., A. Yamada, and D. Zhao (2007), Seismological constraints on the ultralow velocity zones in the lowermost mantle from core-reflected waves, *Phys. Earth Planet. Inter.*, *165*, 25–46, doi:10.1016/j.pepi.2007.07.005.

Knittle, E., and R. Jeanloz (1989), Simulating the core-mantle boundary: An experimental study of high-pressure reactions between silicates and liquid iron, *Geophys. Res. Lett.*, *16*(7), 609–612, doi:10.1029/GL016i007p00609.

Labrosse, S., J. W. Hernlund, and N. Coltice (2007), A crystallizing dense magma ocean at the base of the Earth's mantle, *Nature*, *450*, 866–869, doi:10.1038/nature06355.

Lin, J.-F., W. Sturhahn, J. Zhao, G. Shen, H.-K. Mao, and R. J. Hemley (2005), Sound velocities of hot dense iron: Birch's law revisited, *Science*, *308*, 1892–1894, doi:10.1126/science.1111724.

Manga, M., and R. Jeanloz (1996), Implications of a metal-bearing chemical boundary layer in D" for mantle dynamics, *Geophys. Res. Lett.*, *23*(22), 3091–3094, doi:10.1029/96GL03021.

Mao, W. L., H.-K. Mao, W. Sturhahn, J. Zhao, V. B. Prakapenka, Y. Meng, J. Shu, Y. Fei, and R. J. Hemley (2006), Iron-rich postperovskite and the origin of ultralow velocity zones, *Science*, *312*, 564–565.

McNamara, A. K., E. J. Garnero, and S. Rost (2010), Tracking deep mantle reservoirs with ultralow velocity zones, *Earth Planet. Sci. Lett.*, *299*, 1–9, doi:10.1016/j.epsl.2010.07.042.

Mori, J., and D. V. Helmberger (1995), Localized boundary layer below the mid-Pacific velocity anomaly identified from a PcP precursor, *J. Geophys. Res.*, *100*(B10), 20,359–20,365, doi:10.1029/95JB02243.

Ono, S. (2013), Equation of state and elasticity of B2-type FeSi: Implications for silicon in the inner core, *Phys. Earth Planet. Inter.*, *224*, 32–37, doi:10.1016/j.pepi.2013.08.009.

Otsuka, K., and S.-I. Karato (2012), Deep penetration of molten iron into the mantle caused by a morphological instability, *Nature*, *429*, 243–246, doi:10.1038/nature11663.

Rondenay, S., V. F. Cormier, and E. M. Van Ark (2010), SKS and SPdKS sensitivity to two-dimensional ultralow velocity zones, *J. Geophys. Res.*, *115*, B04311, doi:10.1029/2009JB006733.

Rost, S., and J. Revenaugh (2003), Small-scale ultralow velocity zone structure imaged by *ScP*, *J. Geophys. Res.*, *108*(B1), 2056, doi:10.1029/2001JB001627.

Rost, S., E. J. Garnero, Q. Williams, and M. Manga (2005), Seismological constraints on a possible plume root at the core-mantle boundary, *Nature*, *435*, 666–669, doi:10.1038/nature03620.

Rost, S., E. J. Garnero, and Q. Williams (2006), Fine-scale ultralow velocity zone structure from high-frequency seismic array data, *J. Geophys. Res.*, *111*, B09310, doi:10.1029/2005JB004088.

Santamaría-Pérez, D., and R. Boehler (2008), FeSi melting curve up to 70 GPa, *Earth Planet. Sci. Lett.*, *265*, 743–747, doi:10.1016/j.epsl.2007.11.008.

Seagle, C. T., D. L. Heinz, A. J. Campbell, V. B. Prakapenka, and S. T. Wanless (2008), Melting and thermal expansion in the Fe-FeO system at high pressure, *Earth Planet. Sci. Lett.*, *265*, 655–665, doi:10.1016/j.epsl.2007.11.004.

Stackhouse, S., and J. P. Brodholt (2008), Elastic properties of the postperovskite phase of Fe_2O_3 and implications for ultralow velocity zones, *Phys. Earth Planet. Inter.*, *170*, 260–266, doi:10.1016/j.pepi.2008.07.010.

Stackhouse, S., J. P. Brodholt, and G. D. Price (2006), Elastic anisotropy of $FeSiO_3$ end-members of the perovskite and postperovskite phases, *Geophys. Res. Lett.*, *33*, L01304, doi:10.1029/2005GL023887.

Thorne, M. S., and E. J. Garnero (2004), Inferences on ultralow velocity zone structure from a global analysis of *SPdKS* waves, *J. Geophys. Res.*, *109*, B08301, doi:10.1029/2004JB003010.

Thorne, M. S., E. J. Garnero, G. Jahnke, H. Igel, and A. K. McNamara (2013), Mega ultralow velocity zone and mantle flow, *Earth Planet. Sci. Lett.*, *364*, 59–67, doi:10.1016/j.epsl.2012.12.034.

Vidale, J. E., and M. A. H. Hedlin (1998), Evidence for partial melt at the core-mantle boundary north of Tonga from the strong scattering of seismic waves, *Nature*, *391*, 682–685.

Viterbi, A. (1967), Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, *IEEE Trans. Inf. Theory*, *13*(2), 260–269.

Vočadlo, L., K. S. Knight, G. D. Price, and I. G. Wood (2002), Thermal expansion and crystal structure of FeSi between 4 and 1173 K determined by time-of-flight neutron powder diffraction, *Phys. Chem. Miner.*, *29*(2), 132–139, doi:10.1007/s002690100202.

Wessel, P., and W. H. F. Smith (1998), New, improved version of Generic Mapping Tools released, *Eos Trans. AGU*, *79*(47), 579.

Wicks, J. K., J. M. Jackson, and W. Sturhahn (2010), Very low sound velocities in iron-rich (Mg,Fe)O: Implications for the core-mantle boundary region, *Geophys. Res. Lett.*, *37*, L15304, doi:10.1029/2010GL043689.

Williams, Q., and E. J. Garnero (1996), Seismic evidence for partial melt at the base of Earth's mantle, *Science*, *273*, 1528–1530.

Xu, Y., and K. D. Koper (2009), Detection of a ULVZ at the base of the mantle beneath the northwest Pacific, *Geophys. Res. Lett.*, *36*, L17301, doi:10.1029/2009GL039387.

Zhang, Y., J. Ritsema, and M. S. Thorne (2009), Modeling the ratios of SKKS and SKS amplitudes with ultralow velocity zones at the core-mantle boundary, *Geophys. Res. Lett.*, *36*, L19303, doi:10.1029/2009GL040030.

APPENDIX

PROFILE HIDDEN MARKOV MODELS AND THE VITERBI ALGORITHM

A.1 Markov Chains

A Markov chain is a statistical model that can be used to generate discrete time sequences through the process of stochastic modeling. We can visualize a Markov chain as a finite state automata (FSA) like the one pictured in Figure A.1 In this diagram, the cirlces and diamonds represent states, and the arrows represent state transitions. The set of states, ϕ, can be defined as:

$$\phi = \{B, 0, 1, E\}. \tag{A.1}$$

We may refer to states either by their label, or by an index, s, which will represent their respective positions in the definition of ϕ. The Markov chain will be used to generate a discrete time sequence, indexed by the variable t. At each time step, t, of a stochastic simulation, the model will be in one particular state, represented by the random variable S_t, which may take on any of the values of ϕ.

The state B is a special state, representing the beginning of a sequence. A stochastic simulation will start in state B, or in other words $S_0 = B$. At each time step, $t=1,2,3,...T$, the value of S_t will be updated. The simulation will continue until the terminal time, T,

defined as the time at which the E state is entered. Upon entry, each of the diamond-shaped states will emit a value corresponding to the state label. For this reason, these states are referred to as emitting states. B and E are nonemitting states, which model the beginning and ending of sequences. Assuming that the E state is entered at time T, the Markov chain in Figure A.1 will generate a sequence of $T-1$ samples, whose values are either 0 or 1.

A stochastic simulation is governed by a matrix of state transition probabilities, $\boldsymbol{A}$, which can be defined as:

$$\boldsymbol{A} = \{a_{ij} \mid a_{ij} = P(S_t = j \mid S_{t-1} = i)\}. \tag{A.2}$$

This states that the probability of entering state j at time t, $P(S_t=j)$, is a conditional probability, dependent on the state S_{t-1}. Each entry in $\boldsymbol{A}$ corresponds to one of the arrows in Figure A.1. By definition:

$$\forall j \sum_i a_{ij} = 1, \tag{A.3}$$

which simply states that the probabilities associated with transitions from any given state sum to 1. The probability of generating a given sequence, S, can be written as:

$$P(S) = P(S_T, S_{T-1}, \ldots, S_1, S_0). \tag{A.4}$$

By definition, $S_0 = B$, so $P(S_0) = 1$. We will define each random experiment which produces an element of a sequence, S_t, to be an independent random experiment, which allows Equation A.4 to be rewritten as:

$$P(S) = P(S_T|S_{T-1},S_{T-2},...,S_1)P(S_{T-1}|S_{T-2},S_{T-3},...,S_1)... P(S_1). \quad (A.5)$$

From the FSA diagram, it is easy to see that $P(S_t)$ depends only on the previous state, as $P(S_t = j)$ is simply one of the entries of $\boldsymbol{A}$, with i determined by S_{t-1}. This is a statement of the Markov property, which allows us to rewrite Equation A.5 as:

$$P(S) = P(S_T|S_{T-1})P(S_{T-1}|S_{T-2})... P(S_1), \quad (A.6)$$

or equivalently as:

$$P(S) = \prod_{t=1}^{T} a_{S_{t-1}S_t}. \quad (A.7)$$

Stated simply, the Markov property indicates that the probability of entering any given state is dependent solely on the previous state, and not on the entire sequence of preceding states. It should be noted that the sum of the probabilities of all possible sequences sums to 1, or:

$$\sum_{S} P(S) = 1. \quad (A.8)$$

Given values for the matrix $\boldsymbol{A}$, it is a simple matter to carry out stochastic simulations with a computer program. In the examples that follow, superscripts will be used to differentiate sequences output from individual simulations. For example, using the transition probabilities defined in Table A.1, one stochastic simulation produces the following sequence, S^1:

$$S^1 = [01111001000],$$

with the production probability $P(S^1) = 4.89039 \times 10^{-3}$. A subsequent simulation produces the sequence:

$$S^2 = [10001000010010100101111000010001101101 10],$$

with the production probability $P(S^2) = 4.93535 \times 10^{-07}$.

So far the discussion has concentrated on the generation of time sequences, but Markov chains can be used to generate any kind of discrete sequence. For example, the Markov chain in Figure A.2, in which the states represent nucleic acids, can be used to generate DNA sequences. In this case the state sequence is indexed by position rather than time. One could also generate protein sequences by creating a larger Markov chain with 20 emitting states. In this case, each emitting state would correspond to one of the 20 amino acids that make up proteins. The analysis of protein sequences has played an important role in the development of profile hidden Markov models, which will be discussed later.

A.2 Hidden Markov Models

In the Markov chains previously discussed, each state emits a unique symbol that is also used as a state label. We can create a class of more general models by separating the state labels from the emission symbols. Consider the FSA pictured in Figure A.3, with the state set, $\phi = \{B, P1, P2, E\}$. We will now introduce a set of emission symbols, $\varphi = \{0,1\}$. Each emitting state, *P1* and *P2*, will be allowed to emit any symbol from φ based on a state-specific emission probability distribution, $\boldsymbol{G}$, defined as:

$$\boldsymbol{G} = \{g_{\mathrm{mn}} | g_{\mathrm{mn}} = P(x_t = n | S_t = m)\}, \tag{A.9}$$

where x_t is the symbol emitted at time t, m is the state index, and n is an index into the set of emission symbols, φ. The random variable D_t will represent the symbol emitted at time t. The value of D_t will be null during time steps that occupy nonemitting states. For reasons which will soon become clear, this new type of model is referred to as a hidden Markov model.

The separation of state labels and symbol emissions effectively adds a second stochastic stage to simulations. The first stage produces state transitions, and the the second stage emits symbols from φ using state-specific emission probabilities. In a two-stage stochastic simulation, the probability of a sequence S becomes:

$$P(S) = P(D_T|S_T)\, P(S_T|S_{T-1})\, P(D_{T-1}|S_{T-1}) P(S_{T-1}|S_{T-2}) \ldots P(D_1|S_1) P(S_1), \tag{A.10}$$

which can be written more compactly as:

$$P(S) = \prod_{t=1}^{T} g_{S_t x_t} a_{S_{t-1} S_t} \tag{A.11}$$

If we were to assign the transition probabilities from Table A.2 and the emission probabilities from Table A.3 to the HMM in Figure A.3, we would produce an HMM which is equivalent to the Markov chain in Figure A.1 with the transition probabilities defined in Table A.1. This is because the emission probabilities in Table A.3 only allow exclusive emissions of 0 for *P1* and 1 for *P2*. The emission probabilities in Table A.4, however, allow both *P1* and *P2* to emit either 0 or 1. Using the emission probabilities in

Table A.4, and the transition probabilities in Table A.2, one stochastic simulation produces the following sequence:

$$S^3 = [111101010000101011111111010010100101111111000000011101],$$

with a production probability of $P(S^3) = 2.1667 \times 10^{-16}$, and a subsequent simulation produces this sequence:

$$S^4 = [01101110],$$

with a production probability of $P(S^4) = 1.08714 \times 10^{-3}$.

When we considered stochastic simulations involving Markov chains, the state sequence traversed during the simulation could be directly inferred from the output sequence itself. This is not the case with these two hidden Markov model simulations. Many state sequences exist that are capable of generating these output sequences. Each possible state sequence has its own production probability, which is not necessarily unique. Given only the output sequence, the state sequence which produced the output sequence is hidden, which is why these types of models are called hidden Markov models. Even given the production probability and the output sequence, an inversion for the exact state sequence may be ill-posed as the solution generally exhibits non-uniqueness.

While it may not be possible to recover the exact state sequence that produced a given observed sequence, it is possible to recover the state sequence which produced an observed sequence with the highest production probability, $P^*(S)$. The optimal state sequence, π^*, can be thought of as an optimal path traced through the HMM state

machine. A naive approach to discovering the optimal state sequence would be to explicitly calculate and compare the production probabilities for each possible path through the HMM that generates the desired output sequence. Since the number of possible paths grows geometrically with the number of elements in the output sequence, a more computationally-efficient approach is desired. Fortunately a better solution exists, and is known as the Viterbi algorithm (Viterbi 1967). The Viterbi algorithm uses dynamic programming to discover the optimal path with a computational cost that is proportional to the number of states times the number of elements in the output sequence. Should there exist multiple state sequences which generate a given sequence with probability, $P^*(S)$, we shall be satisfied with whichever one the Viterbi algorithm produces, and refer to this sequence as the optimal state sequence, π^*. This can be justified by the statistical nature of this approach.

The Viterbi algorithm is enumerated in Figure A.4. The algorithm utilizes a dynamic programming approach to find the optimal state sequence. The sequence elements are essentially streamed through the model, and for each state the path which has that particular state emitting the current sequence element is discovered. Due to the Markov property, the maximal probability of any given state emitting any sequence element can be uniquely determined by examining solely the probabilities associated with the previous sequence element. The optimal path for the entire sequence is found by tracing backwards through all the state transitions leading to the highest probability for the terminal sequence element.

A.3 Profile Hidden Markov Models

Often times when working with HMMs, the goal is not to produce sequences through stochastic modeling, but to use the HMM to gain insight into data that has been collected through field recordings or by some other means. When we use an HMM to analyze data, we are making the assumption that the data itself was generated by some stochastic process, and that the HMM in question models that stochastic process with a sufficient level of accuracy. Typically the Viterbi algorithm is used to calculate the optimal state sequence corresponding to the data to be analyzed, and the optimal state sequence is then used to make inferences with regards to the data. Profile HMMs are a particularly effective tool for this type of analysis. Profile HMMs exhibit an overall linear topology, which is created by only allowing specific state transitions. Figure A.5 shows the Plan 7 HMM architecture (Eddy 1995), which is an example of a profile HMM. Due to the linear topology of profile HMMs, we can think of the optimal state sequence as representing the best nonlinear alignment between the data and the model representing a stochastic process.

Profile HMMs were originally introduced by Krogh et al. (1994) to align biological sequences. The Plan 7 HMM architecture, pictured in Figure A.5, was introduced by Eddy (1995) for searching databases of multiple protein alignments to obtain functionally-related matches to a target protein sequence. A multiple alignment is a probabilistic model obtained by aligning several protein sequences which are related through evolutionary processes. The protein sequences making up a multiple alignment represent a family of proteins that have similar functionality. Central to the Plan 7 HMM is a set of (M)atch, (I)nsert and (D)elete states. The Plan 7 HMM takes its name from the

seven state transitions between (M)atch, (I)nsert and (D)elete states.

In the protein sequence search application of Eddy (1995) , each set of (M)atch, (I)nsert and (D)elete states corresponds to a single column in a multiple alignment. The (M)atch and (I)nsert states can emit any of the 20 amino acids according to state-specific probability distributions. The emission distribution for a (M)atch state would typically be generated from the frequency with which each amino acid occurs in the corresponding column of a multiple alignment. The (I)nsert state allows for evolutionarily plausible insertions of amino acids into a protein sequence. The (D)elete state is a nonemitting state, which allows for deletions of amino acids.

Typically, a Plan 7 HMM would be built from many multiple alignments in which the (M)atch, (I)nsert, and (D)elete states from each alignment are concatenated to form a long, linear HMM. New states, S and T are introduced to begin and terminate state sequences. The B and E states serve similar functions as before, except now they model the beginning and ending of subsequences. Multiple subsequences can be combined by traversing the J state. The N, C, and J states have self-transitions to consume elements of the data sequence that do not match the main model well. The optimal state sequence will be analyzed to identify multiple alignments that match the observed data well. This is generally indicated by the traversal of several (M)atch states corresponding to a particular multiple alignment in a subsequence of the optimal path, and may indicate that the corresponding multiple alignment is functionally related to the observed data.

In Chapter 2, we adapt the Plan 7 HMM to model seismic data. Instead of multiple protein alignments, the model will be built on discretely sampled and scaled wavelets. Assuming a Gaussian-distributed noise model, the (M)atch state emission probabilities

will be Gaussian distributions, with the mean corresponding to the amplitude of a particular sample. The optimal state path will illuminate the most probable instances of wavelets contained in a trace, and will allow us to directly infer their travel times and amplitudes.

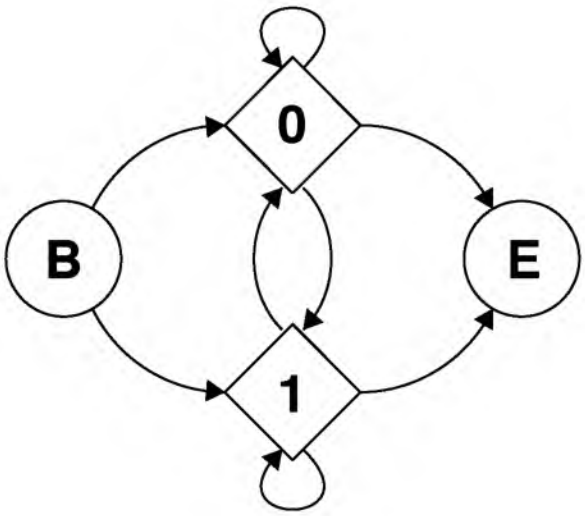

Figure A.1. A Markov chain for generating binary sequences, visualized as a finite state automata. The circles and diamonds represent states, and the arrows represent state transitions.

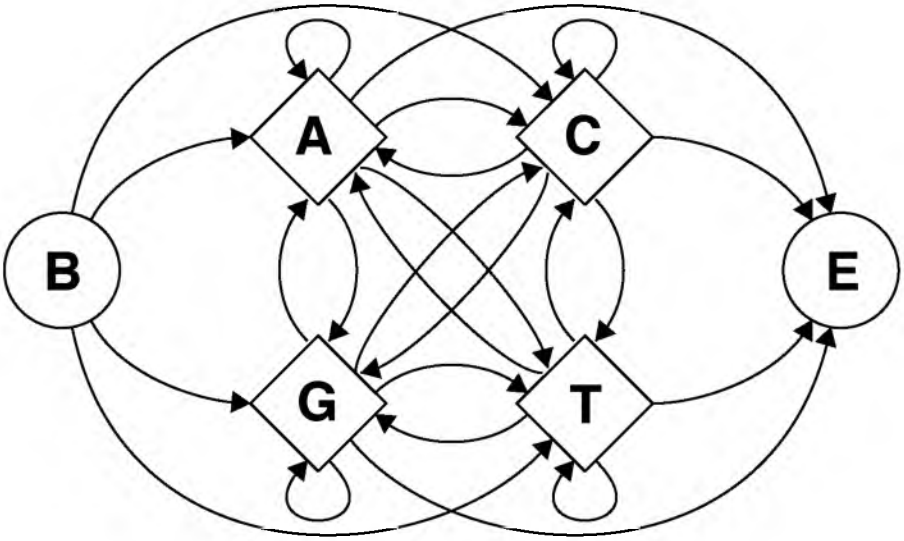

Figure A.2. A Markov chain for generating DNA sequences.

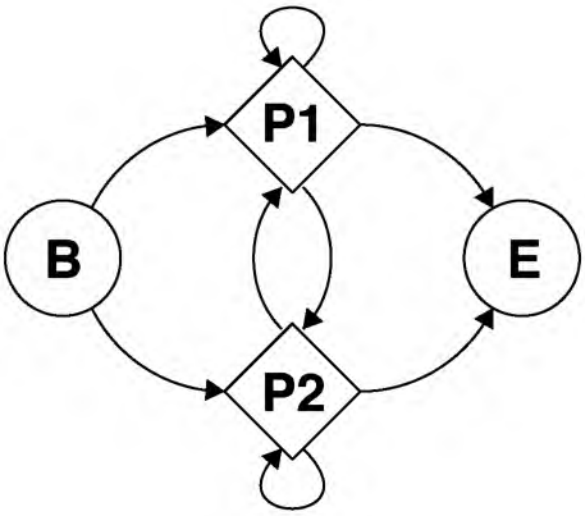

Figure A.3. A hidden Markov model for generating binary sequences. Both the P1 and P2 states may emit either 0 or 1, based on an emission probability mass function.

1. Initialization $P_1(j) = g_{1,x_1}$	$\gamma_1(j) = 0$
2. Recursion $for\ all\ times\ t, t = 1 \ldots T-1$: $P_{t+1}(j) = \max_i P_t(i) a_{ij} g_{j,x+1}$	$\gamma_{t+1}(j) = \arg\max_i P_t(i) a_{ij}$
3. Termination $P^*(S) = \max_i P_T(i)$ $\pi_T^* = \arg\max_i P_T(i)$	
4. Optimal Path Back-Tracking $for\ all\ times\ t, t = T-1 \ldots 1$: $\pi_T^* = \gamma_{t+1}(\pi_{t+1}^*) = 1$	

Figure A.4. The Viterbi algorithm for calculating the optimal state sequence π^*. Adapted from Viterbi (1967).

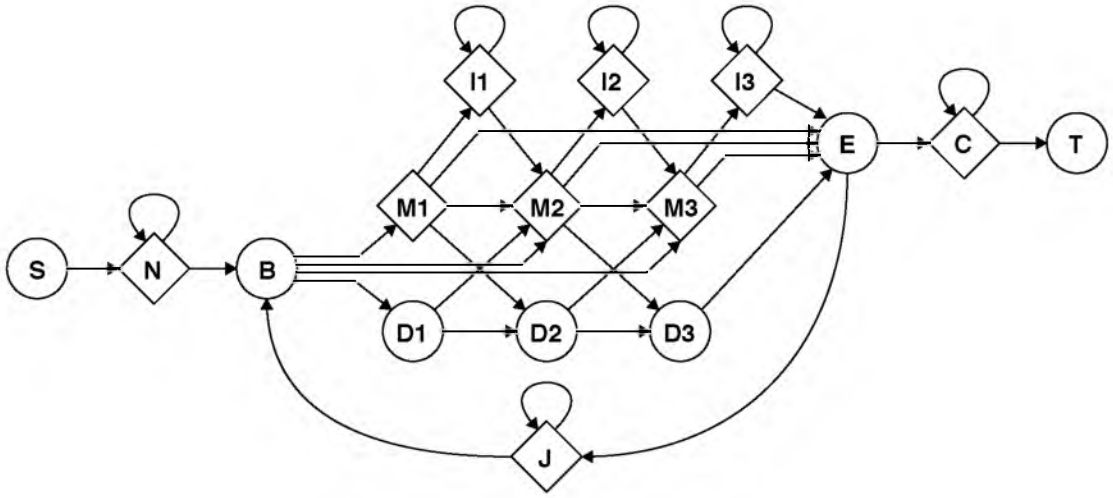

Figure A.5. Plan 7 HMM architecture. The allowed state transitions give the HMM a linear topology, and the name is derived from the seven state transitions between (M)atch, (I)nsert, and (D)elete states. Adapted from Durbin et al. (1998).

Table A.1. Example transition matrix for the Markov chain in Figure A.1. Each row indicates the transition probabilities from a given state to all other states. Zeros indicate the absence of state transitions.

State	B	0	1	E
B	0	.5	.5	0
0	0	.5	.45	.05
1	0	.52	.41	.07

Table A.2. Example transition matrix for the hidden Markov model in Figure A.3. The transition probabilities are the same as in Table A.1, however the state names have been changed to emphasize the fact that state labels no longer imply symbol emissions.

State	B	P1	P2	E
B	0	.5	.5	0
P1	0	.5	.45	.05
P2	0	.52	.41	.07

Table A.3. Example emission probabilities for the hidden Markov model pictured in Figure A.3. The probability mass functions defined by this table make the hidden Markov model in Figure A.3 equivalent to the Markov chain in Figure A.1

State	0	1
P1	1	0
P2	0	1

Table A.4 Example emission probabilities for the hidden Markov model in Figure A.3.

State	0	1
P1	.55	.45
P2	.4	.6

A.4 References

Durbin, R., S.R. Eddy, A. Krogh and G.J. Mitchison, 1998, Biological Sequence Analysis: Probabilistic Models of Proteins of Nucleic Acids: Cambridge University Press.

Eddy, S.R., 1995, Multiple Alignment Using Hidden Markov Models: Proceedings of the Third International Conference Intelligent Systems for Molecular Biology**,** 114–120.

Krogh, A., M. Brown, I. S. Mian, K. Sjolander and D. Haussler, 1994, Hidden Markov Models in Computational Biology: Applications to Protein Modeling: J. Mol. Biol. **235**, 1501-1531.

Viterbi, A., 1967, Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm: IEEE Transactions on Information Theory, **13**, no. 2**,** 260–269.