Application of the Hilbert−Huang Transform to the Analysis of Molecular Dynamics Simulations

Stephen C. Phillips, Robert J. Gledhill, and Jonathan W. Essex*

Department of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, U.K.

Colin M. Edge

GlaxoSmithKline, New Frontiers Science Park (North), Coldharbour Road, Harlow CM19 5AD, U.K.

Received: May 24, 2002; In Final Form: April 15, 2003

The Hilbert−Huang transform (HHT) is a new method for the analysis of nonstationary signals that allows a signal's frequency and amplitude to be evaluated with excellent time resolution. In this paper, the HHT method is described, and its performance is compared with the Fourier methods of spectral analysis. The HHT is then applied to the analysis of molecular dynamics simulation trajectories, including enhanced sampling trajectories produced by reversible digitally filtered molecular dynamics. Amplitude-time, amplitude-frequency, and amplitude-frequency-time spectra are all produced with the method and compared to equivalent results obtained using wavelet analysis. The wavelet and HHT analysis yield qualitatively similar results, but the HHT provides a better match to physical intuition than the wavelet transform. Moreover the HHT method is able to show the flow of energy into low-frequency vibrations during conformational change events and is able to identify frequencies appropriate for amplification by digital filters including the observation of a 10 cm^{-1} shift in target frequency.

1. Introduction

The analysis of the frequencies of motion present in a molecular dynamics (MD) simulation helps our understanding of the dynamics and has particular relevance to the study of conformational change. It is commonly thought that conformational changes occur through large amplitude low frequency motions. This assumption is the basis for the digitally filtered molecular dynamics (DFMD) technique[1] and the newer reversible digitally filtered molecular dynamics (RDFMD) technique,[2] both of which seek to enhance the rate of conformational change by introducing energy specifically into the low-frequency vibrations.

The Fourier transform is the most commonly used method of spectral analysis, so much so that the term "spectrum" is often taken to mean the Fourier spectrum. However, great care must be taken in the physical interpretation of Fourier spectra (and derived techniques), and uncritical use must be avoided. All analysis methods based on the Fourier transform have similar associated problems due to the way that the Fourier transform acts on the data: the Fourier transform takes a data set, replicates it to infinity both forward and backward in time, and calculates the linear combination of sine waves necessary to reproduce the replicated data set.[3] The Fourier transform is a mathematically exact method but its physical interpretation is difficult. Discontinuities in the infinite data set (arising either from the original data or from the replication process) cause leakage away from the true frequencies. The same problem occurs if the data set contains time-localized events.[3] The frequency leakage is an unavoidable aspect of the Fourier transform because it must fit invariant sine waves to a time-localized signal. A nonperiodic signal, or one with time-localized events, is said to be a "nonstationary" data set. Molecular dynamics simulations in the condensed phase inevitably produce nonstationary data.

Derivatives of the Fourier method include the spectral density,[1,4] the spectrogram[5] (also called a sliding frequency distribution or SFD[6]) and wavelet analysis.[7] To calculate a spectrogram, the Fourier transform is applied using a window that cuts out all of the data apart from a localized time section. By sliding the window along the time axis and repeatedly calculating the Fourier transform, a time-frequency distribution can be obtained. For instance, the Fourier transform of the first 5 ps of a simulation may be compared with the Fourier transform of the final 5 ps of data. Since the spectrogram method relies on the Fourier transform, the same problem of nonstationarity in the data occurs. With the spectrogram method, the data in each time window must be stationary, or the same spectral leakage effects of the basic Fourier transform will return. Even if the data is stationary within a time window, the spectrogram method has an additional problem of limited frequency resolution. To obtain precise time information, a narrow time window must be employed. The frequency resolution is inversely proportional to the length of data analyzed, so the Fourier transform of a small time window gives a low-frequency resolution. One cannot obtain both time-localized and frequency-localized information with this method.

Wavelet analysis is a generalized spectrogram method. Whereas the spectrogram essentially fits sine waves to portions of the data set, the wavelet method fits a function chosen by the user. The wavelet transform of a function, $f(t)$, is defined as

$$w(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) g^*\left(\frac{t-b}{a}\right) \mathrm{d}t \qquad (1)$$

* To whom correspondence should be addressed. E-mail: J.W.Essex@soton.ac.uk.

10.1021/jp0261758 CCC: $25.00
Published on Web 05/23/2003

where a is the scale dilation parameter, b is the translation parameter, and the function $g(t)$ is the mother wavelet (with * representing the complex conjugate). The mother wavelet used in this work is the Morlet wavelet which is the product of a complex sine wave with a localizing Gaussian envelope

$$g(t) = \exp(i\omega_0 t)\exp\left(-\frac{t^2}{2}\right) \tag{2}$$

where ω_0 is a constant. A large ω_0 gives good frequency resolution at the expense of poor time resolution. A compromise value of 6 was used in this work.

By varying the dilation parameter, a, the frequency analyzed is varied. At each frequency, the translation parameter, b, may take a range of values corresponding to the time position along the data sample. A complete wavelet spectrum is obtained by taking all appropriate values of a and b. Wavelet transforms where large parts of the scaled and translated mother wavelet lie outside the data set are unreliable. This means that at low frequencies (where the dilation parameter is large) there are few points that may be analyzed. In practice, the computation of the wavelet spectrum is usually performed using a series of Fourier convolutions.

The use of wavelets in the analysis of MD trajectories is limited by its poor time definition at low frequency. However, of the three techniques discussed, the wavelet transform is the best method for analyzing nonstationary data and has become extremely popular in the field of image analysis and compression.[8]

Neither the spectrogram nor the wavelet method can simultaneously provide both good frequency and time resolution. The recently described Hilbert−Huang transform (HHT)[9] is a new method for analyzing nonstationary signals in the frequency and time domains. An examination of the application of this technique to the analysis of molecular dynamics computer simulations is reported in this paper.

2. Theory

2.1. Hilbert Transform. Whereas the Fourier transform takes a time-domain signal, $x(t)$, and moves it into the frequency domain, the Hilbert transform[10] of $x(t)$ produces another time-domain signal. Specifically, the Hilbert transform of a real-valued function $x(t)$ over the range $-\infty < t < \infty$ is another real-valued function $\tilde{x}(t)$ defined by

$$\tilde{x}(t) = \int_{-\infty}^{\infty} \frac{x(u)}{\pi(t-u)}\,\mathrm{d}u \tag{3}$$

This is the convolution of $x(t)$ with $1/\pi t$. As a simple example, the Hilbert transform of a cosine wave is a sine wave of the same frequency and amplitude.

More generally, it may be shown[10] that the Hilbert transform of a signal, $x(t)$, leaves the magnitude of $x(t)$ unchanged but changes the phase by $\pi/2$ (though no calculation of the phase is performed). The rapid diminution of the $1/(t-u)$ function means that its product with the signal is heavily biased to points close to t. The Hilbert transform thus acts on time-localized data.

The original signal, $x(t)$, and its Hilbert transform, $\tilde{x}(t)$, may be considered to be part of a complex signal $z(t)$

$$z(t) = x(t) + i\tilde{x}(t) \tag{4}$$

This can also be written in terms of amplitude, A, and phase, ϕ

$$z(t) = A(t)\mathrm{e}^{i\phi(t)} \tag{5}$$

where

$$A(t) = \sqrt{x^2(t) + \tilde{x}^2(t)} \tag{6}$$

$$\phi(t) = \tan^{-1}\left(\frac{\tilde{x}(t)}{x(t)}\right)$$

Knowing the phase of the signal at time t is useful because we can then consider the rate of change of the phase angle with time. The phase angle of a high frequency motion will change more quickly than that of a lower frequency motion. The relationship between frequency and rate of change of phase is linear and the instantaneous frequency, $f(t)$, is defined as

$$f(t) = \left(\frac{1}{2\pi}\right)\frac{\mathrm{d}\phi(t)}{\mathrm{d}t} \tag{7}$$

To summarize, if we have a signal, $x(t)$, we can calculate the signal's Hilbert transform, $\tilde{x}(t)$, and by combining $x(t)$ with $\tilde{x}(t)$ we can obtain the signal's phase. The derivative of the phase with respect to time gives the signal's instantaneous frequency, $f(t)$. Unfortunately, to obtain meaningful and well-behaved instantaneous frequencies, the wave to be analyzed must have no riding waves and be locally symmetrical about its mean point as defined by the envelopes of local maxima and local minima. This type of wave is not common in real-world situations and so the Hilbert transform has, until recently, not been widely applied to real-world data.

In practice, the Hilbert transform may be computed by taking the Fourier transform of the data, setting all of the negative-frequency components to zero, doubling the positive-frequency components, and back-transforming.[10] This method has been used by Huang et al.[9] but can cause some problems. The Fourier transform replicates the data set to infinity both forward and backward in time and transforms this infinite data set. By using the Fourier transform to compute the Hilbert transform, the Hilbert transform is also working on this infinite data set. This can cause ripples in the Hilbert spectrum at the ends of the (unreplicated) data set because of the discontinuity in the data produced by replication. These ripples create errors in the instantaneous frequencies and amplitudes calculated from the affected regions. The effect of this is to create regions at both ends of the spectrum that are unreliable. The regions where these artifacts were significant were excluded from our analysis. Other methods for calculating the Hilbert transform are available, but for computational ease and efficiency the Fourier method has been used in this work. Alternative methods would likely have very similar end-effect problems to the Fourier transform.

The minimum frequency that can be extracted from a dataset with the HHT method is stated as $1/T$ by Huang et al.[9] where T is the length in time of the dataset. These limits placed on the HHT process by the long-ranged nature of the Hilbert transform have been taken into account in the spectra included within this paper; a fuller discussion of the resolving capabilities of the technique will follow in a later publication.

2.2. Empirical Mode Decomposition. Empirical Mode Decomposition (EMD) is a new method of signal analysis described in detail elsewhere.[9] The aim of EMD is to decompose a signal (which may be nonstationary) into a set of "intrinsic mode functions" (IMFs), where the characteristics of each IMF are such that they may be Hilbert transformed. Thus, through

the Hilbert transform, the instantaneous frequency of each IMF at any point in time may be calculated.

An IMF is defined as a wave in which (a) the number of extrema and the number of zero-crossings differ by at most one and (b) at any point the mean of the envelope defined by the local maxima and the envelope defined by the local minima is zero. This definition satisfies the conditions necessary for the Hilbert transform to work. The original signal may be recovered by summing the IMF components. This also suggests the algorithm to use for decomposing the input data into these IMFs.

The algorithm proceeds by subtracting each recovered IMF from the original data set until either the recovered IMF or the residual data is too small, in the sense of the integrals of their absolute values or the residual data has no turning points. The final residual component once all of the wavelike IMFs are subtracted from the data is similar to the "DC component" obtained by Fourier transform procedures but also contains the overall trend. To find an IMF, the local mean is repeatedly determined and subtracted from the data set until the number of extrema and zero-crossings in the residual data differ by at most one (this process is termed "sifting"). The local mean is found by taking the mean of a curve through all of the maxima and a curve through all of the minima. The maxima and minima curves are, in the current work, defined by cubic spline functions. Once a data set has been decomposed into its IMFs, each IMF is separately Hilbert transformed. The output of the Hilbert transform yields the instantaneous frequency and amplitude at every point in time along each IMF. If the IMF component is considered to be a harmonic oscillator of variable amplitude and frequency, then its signal energy at time t can be written in terms of the signal amplitude, A, and frequency ν

$$E(t) \propto \nu(t)^2 A(t)^2 \tag{8}$$

It is this energy that will be displayed in the HHT spectra that follow.

2.3. Hilbert−Huang Transform. The combination of the EMD method with the Hilbert transform provides a potentially powerful analysis tool, recently named the HHT. The method has been applied in various fields, including biology,[9,11] geophysics,[12] and solar physics.[13] In this section, the method will be tested on some simple signals; in the next, it will be applied to data taken from MD simulations.

The HHT method was tested on three simple data sets to verify our implementation of the method and to build confidence in its use. The first was a sine wave with an abrupt change of frequency around its midpoint. As this already met the criteria for being an IMF, it was processed directly with the Hilbert transform. The resulting graph showed the transition point and distinguished the two frequencies with precision. The energy/frequency/time data produced were then integrated along their time axis to produce a frequency/energy distribution which we compared with a Fourier spectral density plot.[4] The two plots were very similar, but the HHT plot had sharper peaks and lacked the ripple artifacts present in the Fourier plot. The second test data set was constructed by summing two constant frequency, varying amplitude sine waves. This signal did not meet the IMF criteria and required processing with the EMD algorithm. Although EMD did not exactly reproduce the original sine waves, the resulting IMFs showed excellent frequency definition and gave the correct trend in amplitude values of the signal components.

In the final test data set, the sum of two sine waves was once again taken. On this occasion, the amplitudes of both waves were fixed at the same value and the frequency of one wave was set to 100 units. The frequency of the other wave was varied linearly from 25 to 666 units but the sampling rate was chosen to give a maximum sampled frequency of 500 units. Figure 1 plots the HHT of this wave. There is clearly a problem in the ringed part of the spectrum. The physical interpretation of the very low frequency (and low energy) data points is also unclear. For the majority of the data set, however, the two clearest signals are very sharply defined to the correct values. Also, the rebounding of the high frequency signal when it reaches approximately 500 units (the Nyquist limit) reminds us of the necessity with any analysis method to sample the data sufficiently.

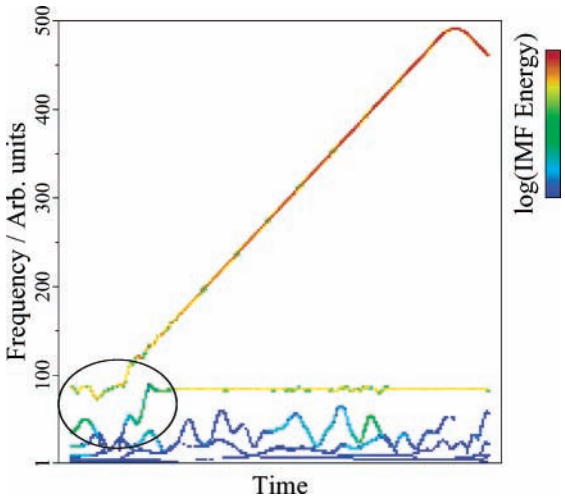

Figure 1. HHT of two fixed-amplitude sine waves. One wave has a frequency of 100 units, whereas the frequency of the other sine wave varies linearly from 25 to 666 units. The energy is plotted on a logarithmic scale. The region where the EMD algorithm has failed is ringed.

The problems in Figure 1 are caused by the EMD method, not the Hilbert transform. The ringed area of Figure 1 underlines the need for all of the IMF components to be interpreted together if the data being investigated do not possess a clear, physically meaningful separation of scales—in such circumstances it is impossible to interpret an isolated IMF.[9] A related problem has also been found by Huang[9] and attributed to inaccuracies introduced by the spline fitting procedure used in the algorithm. The Akima spline[14] function, which fits a curve to a small local set of points was tried as a possible alternative to the conventional cubic spline, which requires the solution of a set of equations involving all datapoints. We found that this made little difference to the final output. Even though the HHT method seems to have problems in some circumstances, it will be demonstrated below that useful and consistent data can be extracted from MD trajectories using the procedure.

3. Applications

In this section, the HHT is applied to investigate conformational change in Brownian dynamics (BD) and molecular dynamics simulations. Many of the MD trajectories were generated using the reversible digitally filtered molecular dynamics (RDFMD) method[2] where the amplitudes of vibrations in a chosen frequency range are amplified by repeated application of a digital filter.

3.1. Brownian Dynamics. The combination of BD and a simple analytic potential energy surface provides a simple test case to examine the performance of the HHT when applied to

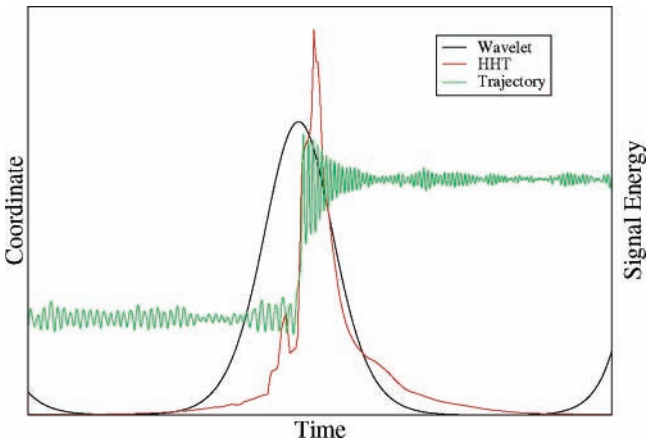

Figure 2. Trajectory taken by the BD simulation. The particle moves from the higher potential well to the lower potential well. Also plotted are the wavelet and HHT spectra integrated over the low-frequency region defined in Figure 3.

conformational transitions. A one-dimensional energy surface with analytic gradients was created from a short Fourier series

$$V(x) = 5\cos 2\pi x + 2\cos 4\pi x + \frac{3}{2}\cos \pi x \qquad (9)$$

The surface was designed to contain two potential wells, one higher in energy than the other, with different characteristic frequencies separated by a low barrier, and to rise steeply to a high value on either side of these wells. The BD algorithm employed is a variation[15] of Ermak and Buckholtz[16] consisting of a standard velocity Verlet integrator, damped by friction and driven by random forces with $k_BT/m = 0.0025$ (m is the particle mass), the friction constant, ξ, set to 0.0001 and a time-step, δt, of 0.02. Multiple simulation runs were performed with different random number seeds until a trajectory containing a transition event between the potential wells was observed. The trajectory is shown in Figure 2. All of the graphs for the BD simulation and analysis have arbitrary units.

Analysis of the trajectory using the HHT gave 10 IMFs; the Hilbert transforms of these IMFs are plotted in Figure 3, along with a Morlet wavelet transform of the same data set for comparison. The mean signal energy of each IMF was calculated, and only points above 40% of mean are plotted for clarity in the HHT figure. The two frequencies of motion associated with vibrations in each of the two minima are clearly visible and also an increase in low-frequency motion around the transition point (circled in the Figure). The wavelet spectrum of the data shows the same frequency and energy trends as the HHT spectrum, but the behavior at low frequencies is less clear. Also, the frequency definition in the wavelet spectrum is much lower. Integrating the IMFs in the time dimension gives a plot of amplitude versus frequency, shown in Figure 4. The figure comprises two clear peaks to which Gaussian curves have been fitted, and also shows large amplitude low-frequency motion. The ratio of the two frequencies, taken as the centers of the Gaussian peaks is 1.72. This compares well with the frequency ratio of 1.62 obtained analytically by applying the normal mode approximation. The discrepancy in the two frequency ratios is a result of the motion in the wells being anharmonic, and therefore the analytical derivation (using the harmonic approximation) is different to the numerical (real life) result.

The HHT analysis reveals a marked increase in the amplitude of the low frequency motions around the transition point. This is further clarified by Figure 2 which plots the signal energy of the lowest frequencies as a function of time. An integrated wavelet spectrum is also presented for comparison. Although both methods are clearly capable of picking up the conformational change event, each provides a different picture of this event. The wavelet transform shows a broad peak centered on the point of transition, whereas the HHT gives a peak that rises sharply at the transition point and rolls off less rapidly after it. It seems that the low frequency components being detected by the HHT are associated with the gradually damped high-amplitude motions of the particle losing energy to friction after transition to the lower energy state. This is different to the result obtained from the integrated wavelet spectrum but has a clear interpretation.

In conclusion, the HHT analysis of the BD trajectory has shown that the HHT method is able to isolate characteristic frequencies of motion (comparable to the analytical result) and has revealed an increase in the energy of low frequency motion just after the transition point in accordance with intuition. The Morlet wavelet transform of the data is similar to the HHT result, but with apparently poorer time resolution at lower frequency.

3.2. Molecular Dynamics. Unless otherwise stated, all molecular dynamics simulations were performed using DL-PROTEIN,[17] modified to include the RDFMD algorithm.[2] The first MD simulation to be analyzed is a conformational transition in gas-phase united-atom pentane induced by the RDFMD method using a $0-150\ cm^{-1}$ amplifying filter to add energy to the low-frequency modes very gradually.[2] The force field parameters were derived from the OPLS parameters,[18] SHAKE[19]

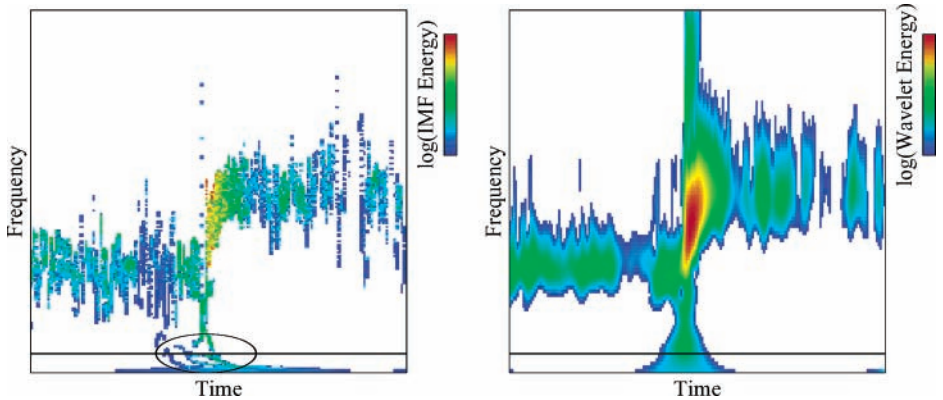

Figure 3. Left: HHT plot of the BD trajectory with an energy cutoff at 40% of the mean applied for clarity. The circled region highlights an increase in low-frequency motion around the transition point. The horizontal line shows the high-frequency limit of the low-frequency region integrated over in Figure 2. Right: Morlet wavelet transformation of the BD trajectory with low energy regions clipped to white for ease of comparison.

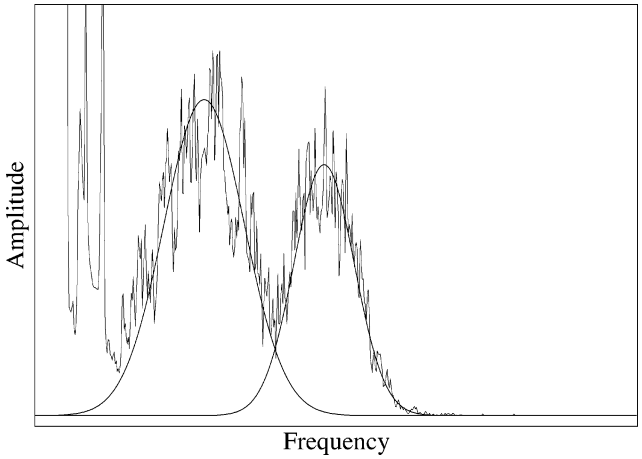

Figure 4. Amplitude versus frequency for the BD trajectory. A Gaussian curve has been fitted to each peak in order to calculate the peak centers.

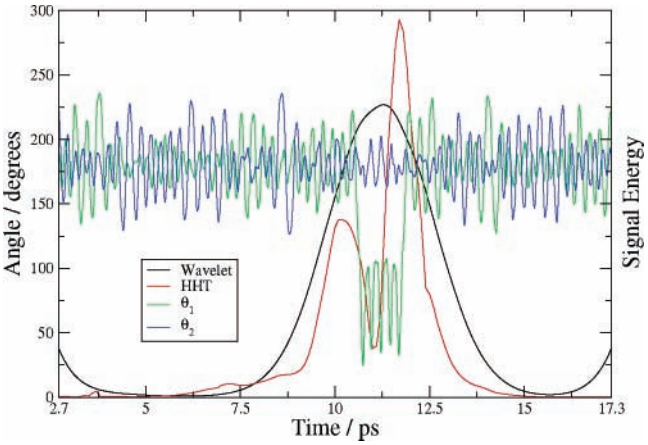

Figure 5. Two dihedral angles of pentane, along with the integrated HHT signal energy in the 2−25 cm^{-1} region and the integrated wavelet signal energy in the 16−25 cm^{-1} region (the lowest possible frequencies were used in both cases).

was applied to all bonds, a 2 fs time-step was employed and the NVE ensemble was used with an initial temperature of 300 K. Figure 5 shows the trajectories of the two dihedral angles in pentane (labeled θ_1 and θ_2). The HHT plot of the dihedral angles and the HHT plot of just the frequency range amplified by the filter are shown in Figure 6, along with wavelet transforms of the same data for comparison. Not all of the points in each IMF are displayed. The mean signal energy of each IMF is calculated, and only points above half that mean are displayed. Figure 6 shows that while the pentane molecule is in the alternate conformation (10.5−12 ps) the dihedrals oscillate at a higher frequency. The low frequency spectra of Figure 6 show an increase in the low frequency energy as each transition occurs. The Morlet wavelet spectra in Figure 6 show the same basic trends at higher frequencies, but become increasingly difficult to interpret toward the lower end of the spectrum. In particular, the very low frequency region targetted by the filter is quite poorly resolved.

As has been noted earlier, the HHT spectrum can be integrated in time to produce a spectral density equivalent. This is not useful when the data set is nonstationary as in this case. What is useful is to integrate the HHT spectrum in the frequency axis to provide a measure of the variation with time of the signal energy of any chosen frequency band. The integral of Figure 6 over the region below 25 cm^{-1} is shown in Figure 5. The integration captures just the lowest frequency IMFs and clearly shows the large increase in low-frequency energy during both transitions; the first transition is directly induced by RDFMD, and the second transition is spontaneous. An integrated wavelet spectrum is also included for comparison; it shows a single broad peak centered between the two transition points. The HHT marginal spectrum has clearly resolved the two conformational transitions and is in better agreement with physical intuition than the Morlet wavelet.

The second simulation to be analyzed is gas-phase alanine dipeptide, simulated using the united-atom OPLS force field.[20] All bonds were constrained using SHAKE,[19] a 2 fs time-step was employed and the simulations were run in the NVE ensemble with an initial temperature of 293 K. RDFMD was used to enhance the low-frequency motion present in the ϕ and ψ angles. Conformational change was caused by applying an amplifying filter centered around 50 cm^{-1} five times, followed by four applications of a similar filter centered around 40 cm^{-1}. It was necessary to use both a 50 cm^{-1} filter and a 40 cm^{-1} filter because it was found that after five applications the 50 cm^{-1} filter was no longer effective. It was postulated that the predominant motion may have shifted in frequency from 50 to 40 cm^{-1} as energy was introduced into the oscillation.[2] The HHT method allows further investigation of this phenomenon and shows that a frequency shift does indeed occur.

Figure 7 shows the two $\phi-\psi$ trajectories of the alanine dipeptide molecule in the initial potential well, one before any RDFMD application and one after five 50 cm^{-1} amplifying filters. Figure 8 shows the time-integrated frequency-energy plots of the HHT spectrum for the 8−70 cm^{-1} segment of the two trajectories. It is clear that not only has the majority of the energy shifted to lower frequency but that the original filter ideally should have targeted 40 cm^{-1} and the following filter 33 cm^{-1}. In this case therefore, the HHT method is able to show a shift in the characteristic vibrational frequency as energy is added to the system. Owing to the short 4 ps time scale of this experiment, Morlet wavelet visualization of the results is impossible. A single Morlet wavelet contains just under 10 whole Fourier wavelengths, and as a result, the lowest frequency obtainable with it on a data set of this length is just over 79 cm^{-1}, rendering the entire region shown in Figure 8 inaccessible. A wavelet with fewer Fourier wavelengths would facilitate the analysis of lower frequency regions, but such a wavelet would have a reduced frequency resolution.

Thus, in this system, the HHT method is able to identify the frequencies appropriate for amplification by RDMFD, with the accuracy of these frequencies being independently confirmed based on the system's response to RDFMD amplification.[2] An equivalent analysis using the Morlet wavelet is impossible.

The HHT analysis method presents an opportunity to analyze the frequency and energy characteristics of a spontaneous transition and to compare the results with an RDFMD-induced transition. A spontaneous conformational transition occurred during an attempt to equilibrate chloroform-solvated alanine dipeptide.[2] The dipeptide was solvated in a box of side 33 Å containing 310 chloroform molecules. The united-atom OPLS force field[20] was employed and the simulation was run in the NPT ensemble at 1 atm and 293 K using the Melchionna[21] adaptation of the Hoover barostat and thermostat. SHAKE[19] was applied to all bonds and a 2 fs time-step employed. The region around the transition point was analyzed using the HHT method and the time-energy graph for 8−25 cm^{-1} is presented in Figure 9 along with the same analysis of a transition induced by RDFMD using multiple applications of a 0−100 cm^{-1} amplifying filter.[2] Both transition paths take the conformation from the

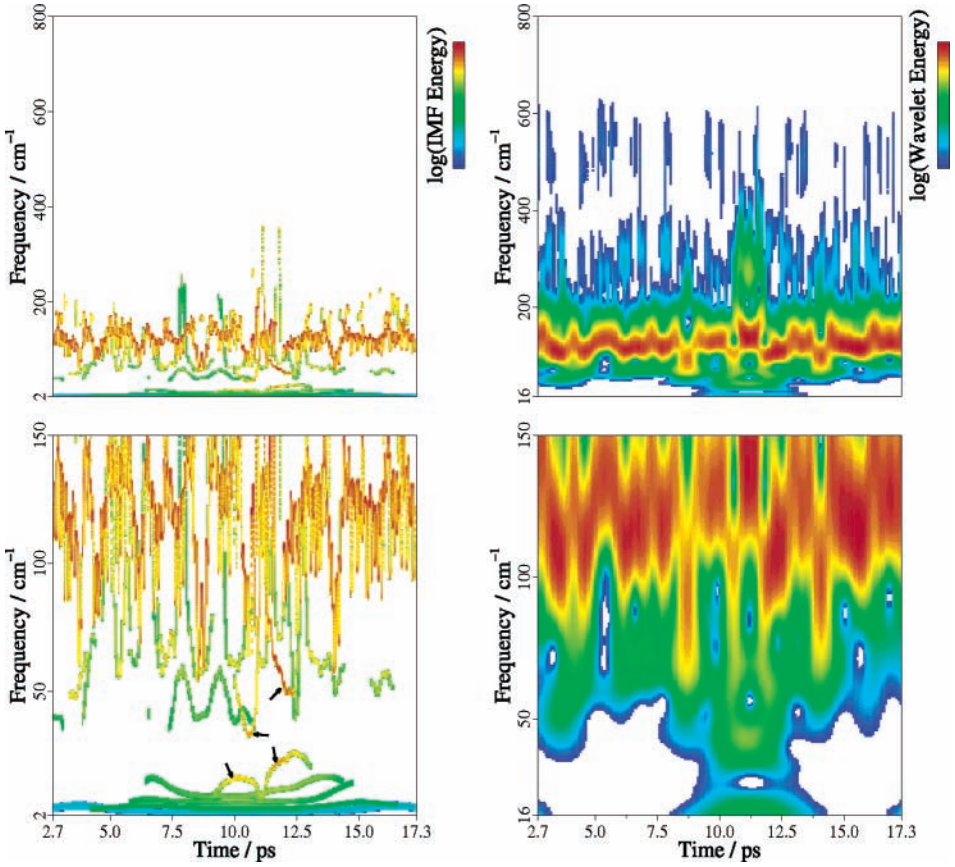

Figure 6. Top left: the HHT plot of the dihedral angles of pentane. Bottom left: the low-frequency region of the HHT plot. Both HHT plots use a logarithmic energy scale and a cutoff of half the mean. The regions where the low-frequency IMFs increase in energy during the transitions are marked by arrows. Top right: a Morlet wavelet plot of the dihedral angles. Bottom right: the low-frequency region of the wavelet plot. Both wavelet plots use a logarithmic energy scale, and have low values clipped to white for ease of comparison.

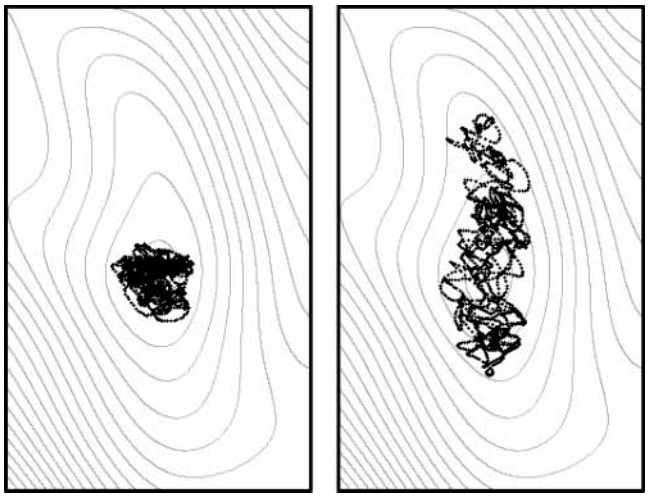

Figure 7. Alanine dipeptide $\phi-\psi$ trajectories around the initial potential well. Left: trajectory before any filter applications. Right: trajectory after four applications of the 50 cm^{-1} amplifying filter.

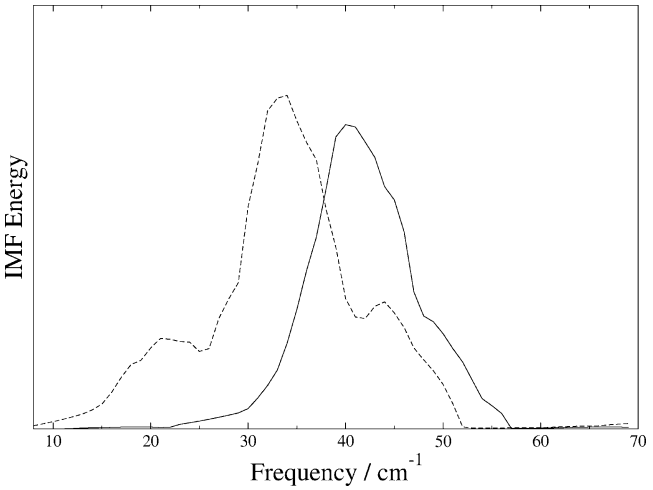

Figure 8. Time integrals of the HHT transforms of the two alanine dipeptide trajectories for the frequency range 8−70 cm^{-1}. The solid line is for the initial trajectory and the dashed line is for the trajectory resulting from applying four successive 50 cm^{-1} amplifying filters.

same initial minimum to the same final minimum via the same transition point. The induced transition occurs approximately five times faster than the natural transition (according to the peak widths in Figure 9) but during both transitions energy moves into the low-frequency modes and this low-frequency energy is not present in the rest of the simulation.

Finally, a conformational transition in the pentapeptide YPGDV solvated in TIP3P water[22] has been analyzed by HHT. The pentapeptide was solvated in a cubic box of side 28.9Å containing 805 water molecules and a Na^+ counter-ion. The all-atom CHARMM[23] force field was used along with the PME method[24] for electrostatics, SHAKE[19] was applied to all bonds containing hydrogen atoms, and a 2 fs time-step was employed. The simulations were run primarily in the NVE ensemble using a Langevin thermostat set at 300 K with a damping parameter of 10 ps^{-1}. YPGDV was simulated with the NAMD program[25] modified for RDFMD. NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana−Champaign. Conformational change was

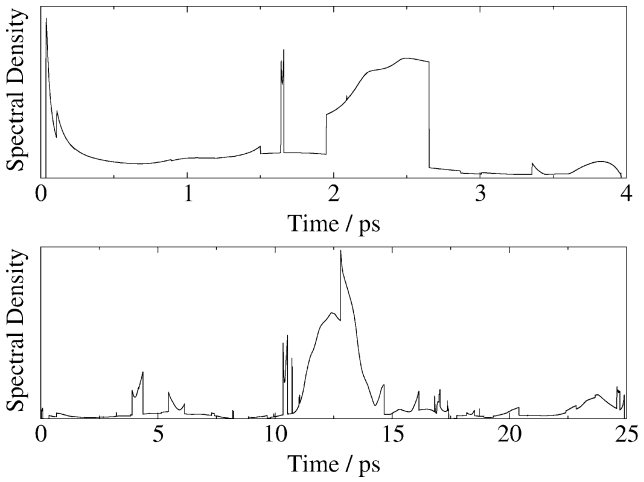

Figure 9. Time-energy plots of the 8−25 cm^{-1} frequency band for two conformational transitions in chloroform-solvated alanine dipeptide. Top: an RDFMD trajectory. Bottom: a spontaneously occurring event.

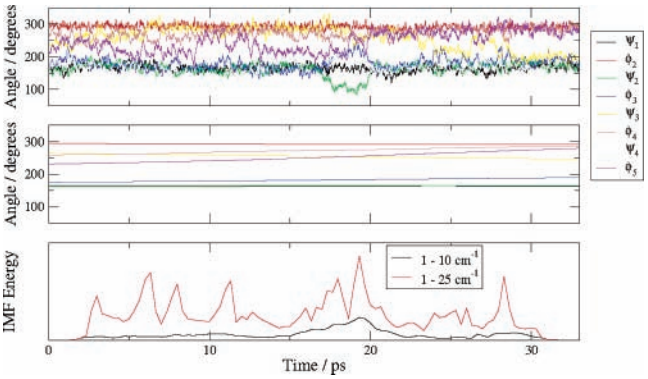

Figure 10. (a) ϕ and ψ trajectory of a 20 ps YPGDV simulation. (b) The residual data after the IMFs have been subtracted from a. (c) The Hilbert time-energy spectra for two frequency bands.

enhanced by applying a 0−25 cm^{-1} amplifying filter. For further details of this simulation, see Phillips et al.[2] Spontaneous conformational change was also observed and a plot of the eight backbone dihedral angle trajectories during one of these spontaneous conformational changes is shown in Figure 10a.

The eight ϕ and ψ trajectories were processed through the EMD algorithm to produce 61 IMFs. The residual data after the IMFs were extracted are shown in Figure 10b. The residual data effectively show the trends in the dihedral angles as explained in section 2.2. The HHT time-energy plots for the 1−25 and 1−10 cm^{-1} frequency bands are shown in Figure 10c. Large peaks in the 1−25 cm^{-1} frequency band are present during the conformational change events (in angles ψ_2, ϕ_3, ψ_4, and ϕ_5), vindicating the choice of a 0−25 cm^{-1} amplifying filter for the RDFMD application on this system.

4. Conclusion

In this paper, the application of the HHT method to the analysis of the nonstationary signals associated with a molecular dynamics trajectory has been presented. The HHT method offers an alternative to the wavelet and advantages over the spectrogram and Fourier transform methods. The Fourier transform and spectral density have limited use in analyzing the frequency characteristics of molecular dynamics trajectories. Care must be taken not to over-interpret the Fourier spectra, and only data sets where the system is in equilibrium (and the data set is approximately stationary) may be analyzed. The spectrogram has some use, but it is limited by the suboptimal way of trading-off frequency and time resolution. The Morlet wavelet and HHT methods yield qualitatively similar information; either can be used to analyze important features of the amplitude-frequency-time distribution of a signal. However, frequency or time integration of HHT yields results more consistent with physical intuition for the phenomena under investigation. The expected observation by HHT of reduced oscillation frequency after the addition of energy by the RDFMD[2] technique offers corroboration that the spectra of low-frequency bands have physical meaning.

A test on the HHT method indicates that artifacts in the analysis are possible both near the ends of the dataset, because of the properties of the Hilbert transform, and at time-points where components of the signal cross in frequency. Huang noted such artifacts and suggested that using an alternative spline fitting procedure in the EMD algorithm may yield improvement. Our application of the localized, cubic Akima spline did not enhance the algorithm, although higher order splines may still yield improvements. It is important to remember that where there is no clear separation of scales in the data, all IMF components must be considered together for meaning to be extracted from the derived spectrum; all data presented in this paper are consistent with this requirement.

Despite these artifacts, useful and consistent data have been obtained from the analysis of conformational transitions. The integration of the frequency axis to provide a time-energy spectrum during a transition is particularly useful. The analysis of five different conformational transitions provides a consistent picture of an increase in the energy of low-frequency oscillations during conformational transition events, vindicating the choices made in designing the RDFMD amplifying filters for these systems. Furthermore, a shift in the target vibrational frequency on the addition of energy has also been observed and independently confirmed based on the system's response to RDFMD amplification.[2] Thus, the HHT may be used to select appropriate frequencies for amplification in DFMD[1] or RDFMD[2] experiments.

Finally, the similarities between the low-frequency energy profiles of induced and spontaneous transitions validates the DFMD[1] and RDFMD[2] method of promoting conformational change by introducing energy into low-frequency modes.

Acknowledgment. We should like to thank Dr. Harvey Lewis for drawing the HHT method to our attention. S.C.P. would like to thank GlaxoSmithKline for their generous support. J.W.E. is a Royal Society University Research Fellow. This work is supported by the EPSRC and the BBSRC.

Note Added after ASAP Posting. This article was released ASAP on 5/23/2003. The title running head was changed from "Application of the HHT to MS Simulations" to "Application of the HHT to MD Simulations". The correct version was posted on 5/30/2003.

References and Notes

(1) Phillips, S. C.; Essex, J. W.; Edge, C. M. *J. Chem. Phys.* **2000**, *112* (6), 2586−2597.

(2) Phillips, S. C.; Swain, M. T.; Wiley, A. P.; Essex, J. W.; Edge, C. M. *J. Phys. Chem. B* **2003**, *107*, 2098−2110.

(3) Brigham, E. O. *The Fast Fourier Transform*; Prentice Hall: Englewood Cliffs, NJ, 1980.

(4) Sessions, R. B.; Dauber-Osguthorpe, P.; Osguthorpe, D. J. *J. Mol. Biol.* **1989**, *210* (3), 617−633.

(5) Oppenheim, A. V.; Schafer, R. W. *Digital Signal Processing*; Prentice Hall: Englewood Cliffs, NJ, 1989.

(6) Sessions, R. B.; Osguthorpe, D. J.; Dauber-Osguthorpe, P. *J. Phys. Chem.* **1995**, *99* (22), 9034−9044.

(7) Torrence, C.; Compo, G. P. *Bull. Am. Meteorol. Soc.* **1998**, *79* (1), 61−78.

(8) Usevitch, B. E. *IEEE Signal Process. Magn.* **2001**, *18* (5), 22−35.

(9) Huang, N. E.; Shen, Z.; Long, S. R.; Wu, M. L. C.; Shih, H. H.; Zheng, Q. N.; Yen, N. C.; Tung, C. C.; Liu, H. H. *Proc. R. Soc. London Ser. A-Math. Phys. Eng. Sci.* **1998**, *454*, 903−995.

(10) Bendat, J. S. *The Hilbert Transform and Applications to Correlation Measurements*; Brüel & Kjær: Nærum, Denmark, 1985.

(11) Echeverria, J. C.; Crowe, J. A.; Woolfson, M. S.; Hayes-Gill, B. R. *Med. Biol. Eng. Comput.* **2001**, *39* (4), 471−479.

(12) Chen, K. Y.; Yeh, H. C.; Su, S. Y.; Liu, C. H.; Huang, N. E. *Geophys. Res. Lett.* **2001**, *28* (16), 3107−3110.

(13) Komm, R. W.; Hill, F.; Howe, R. *Astrophys. J.* **2001**, *558* (1), 428−441.

(14) Akima, H. *Commun. ACM* **1972**, *15*, 914−918.

(15) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, U.K., 1987.

(16) Ermak, D. L.; Buckholtz, H. *J. Comput. Phys.* **1980**, *35*, 169−182.

(17) Dlprotein user manual. Melchionna, S.; Cozzini, S.; University of Rome, Rome, Italy, 1998.

(18) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. *J. Am. Chem. Soc.* **1984**, *106* (22), 6638−6646.

(19) Ryckaert, J. P.; Ciccoti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

(20) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 0 (6), 1657−1666.

(21) Melchionna, S.; Ciccotti, G.; Holian, B. L. *Mol. Phys.* **1993**, *78* (3), 533−544.

(22) Jorgensen, W. L. *J. Phys. Chem.* **1983**, *87* (26), 5304−5314.

(23) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4* (2), 187−217.

(24) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98* (12), 10089−10092.

(25) Kalé, L.; Skeel, R.; Bandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283−312.