

# Wavelet Transform for Analysis of Molecular Dynamics

Attila Askar\* and A. Enis Cetin†

Department of Mathematics, Koç University, Istinye, 80860 Istanbul, Turkey

Herschel Rabitz

Department of Chemistry, Princeton University, Princeton, New Jersey 08544

Received: July 2, 1996; In Final Form: October 16, 1996<sup>®</sup>

The wavelet transform (WT) is suggested as an analysis tool in molecular dynamics by analogy with its use in signal and image processing. In particular, the WT excels in detecting rare events such as transitions that are localized in both time and frequency. Coupled with this property, the WT is also an efficient method for separating phenomena of different time scales. In this regard, the WT permits filtering out the high-frequency noise without completely omitting the high-frequency phenomena whose contribution is crucial in cases where the dynamics is localized in frequency and time. Two time series are studied, resulting respectively from the deterministic dynamics of a three-dimensional polymer chain and the stochastic dynamics of a one-dimensional polymer model subjected to random forces within the Langevin equation formalism. The WT is observed to excel in reconstructing the original signal by a subset of the basis used in the analysis and in identifying the occurrence of rare phenomena by examining the wavelet energies at high-resolution levels.

## 1. Introduction

Wavelets and the wavelet transform (WT) were first introduced to analyze the time series for seismic signals.<sup>1</sup> Subsequently they found applications in a wide spectrum of fields<sup>2,3</sup> ranging from digital signal and image processing<sup>4–9</sup> to, recently, chemical physics.<sup>10,11</sup> The WT is basically an orthogonal basis expansion procedure with the wavelet basis functions being a two-parameter family obtained from a single generating function. The two parameters in the present context relate to the time and frequency coordinates. The time and frequency domains are covered respectively by the shifts and dilations of this single generating function by varying the two independent variables. The frequency content of a time series provides useful information in many applications, and the Fourier transform (FT) is the most widely used method. A signal that is localized in the frequency domain corresponds to a near periodic time history that extends almost indefinitely. A serious drawback with the FT is its weakness for detecting local time features of a signal. This can be understood by noting that a signal that is localized in time necessitates a wide frequency spectrum of Fourier components. In practical terms, sudden changes and transitions in the signal cannot be captured in a readily observable way by the FT. The well-known Gibbs phenomenon is a concrete demonstration of this feature of the Fourier series as a global basis representation.<sup>2</sup>

The WT is gaining wide acceptance as a convenient time–frequency analysis method that excels in areas where the FT is a weak performer. Unlike the FT, the WT can capture phenomena that are simultaneously localized in time and frequency. Consequently, since signal features related to both time and frequency can be extracted, the WT can be used effectively in the analysis of nonstationary data. Finally, the WT is an attractive multiresolution data analysis method through its ability to produce subsignals at various scales in time and with different frequency contents. The subsignals may be

studied separately, with each level providing information about different aspects of the original signal. For example, in the field of signal waveform coding, speech and image signals can be compactly represented by their subsignals.<sup>4–7</sup> The WT decomposition of a sound signal or an image into various levels of resolution permits discarding the portion of the signal that does not contribute significantly. This can have dramatic consequences for practical applications. In most cases, a few percent of the WT data for the whole signal suffices, and these reduced data are transmitted to the receiver. Since the significant information content of both low- and high-frequency data is kept, the signal is reconstructed at the receiver end with a remarkable quality.

A strong similarity exists between the needs in processing image and sound signals and the time series produced in molecular dynamics: high-frequency “noise” is present in all of these cases, and the common challenge is to identify and discard the unimportant data. A major difference is, however, that the acoustic or visual signals are available and the WT post processes these. Conversely, in molecular dynamics these time series are to be generated by a numerical algorithm. The further challenge in this latter field is thus to transform the operator at the outset so that it does not generate nonconsequential high-frequency features. In the physical world, in an analog manner, the molecule itself minimizes the participation from high-frequency motion and averages out their contribution. An earlier attempt in this direction is found in the subphase method<sup>12–15</sup> through the basic philosophy of filtering out the higher frequency modes. In the crudest approximation, the high-frequency modes are completely ignored, while the higher order schemes can incorporate some of their contribution in an averaged manner. An added challenge is to retain large amplitude, infrequent rapid transitions, and the WT may be especially useful in this regard. This paper presents a wavelet analysis of two time series generated by numerical integrations in molecular dynamics and suggests the WT as a possible tool for constructing solutions. The two examples studied are (i) the deterministic dynamics of a three-dimensional polymer model<sup>13</sup> and (ii) the stochastic dynamics of a one-dimensional chain subjected to random forces

† Present address: Department of Electrical Engineering, Bilkent University, Ankara, 06533, Turkey.

<sup>®</sup> Abstract published in *Advance ACS Abstracts*, November 15, 1996.

within the Langevin equation formalism.<sup>12</sup> These studies indicate that the WT is a useful tool for analyzing the numerical data representing the dynamics of chain molecules as well as for reconstructing them systematically from a subset of the data. Specifically, it is shown that the bonds that undergo significant motions can be detected very efficiently and the time series can be reconstructed with much simplified data. The signal compaction property of the WT is compared with and observed to be superior to that of the FT for the time series such as dihedral angle transitions. The fundamentals of the WT are presented in the next section, followed by applications in section 3, and concluding remarks in section 4.

Although analysis of existing molecular dynamics trajectories is interesting in itself, the primary significance of this work is the potential it holds for wavelets as a basis for performing the dynamics directly. The potential gains with such an approach could be quite significant. In particular, a longstanding goal in macromolecular dynamics has been the desire to carry out computations at very long time scales. Perhaps the most notable of such challenges involves the study of protein-folding dynamics. Although it is too early to project the ultimate role of wavelets for this purpose, the special capability of simultaneously acting as filters and retaining local time information on the dynamics suggests that they may be very valuable for this purpose. The present paper takes a modest step in this direction.

## 2. Fundamentals of Wavelet Analysis

Wavelets are a set of basis functions for representing functions in the space  $L_2(R)$ .<sup>2,3</sup> Two equally important and related features distinguish them from other basis representations: a hierarchy of nested spaces and simultaneous scales for time and frequency for each level. This structure permits the representation of a temporal function at different resolution levels. In the presentation here, first the hierarchical multiresolution nature of the wavelets, which provides a representation of a function at various resolutions, is discussed and then the time and frequency characteristics of wavelets are presented at the end of the section.

Toward the goal of building the hierarchy of spaces at different resolution levels, first a reference space  $V_0$  is considered with a function within it denoted as  $f_0(t)$  and with the associated basis set  $\{\phi_{0,k}\}$ . A function  $f_0(t)$  in  $V_0$  is interpreted as the projection of the function  $f(t)$  in  $L_2(R)$  into the subspace  $V_0$ . The basis functions in wavelet analysis are generated by the translations of a single function  $\phi(t)$ , called the “scaling function”. For a convenient presentation, the independent variable  $t$  axis is scaled such that the subintervals are delimited at integer values. The basis functions of  $V_0$  are then obtained by the translations of  $\phi(t)$  as

$$\phi_{0,k} = \phi(t - k), \quad k \in Z \quad (1)$$

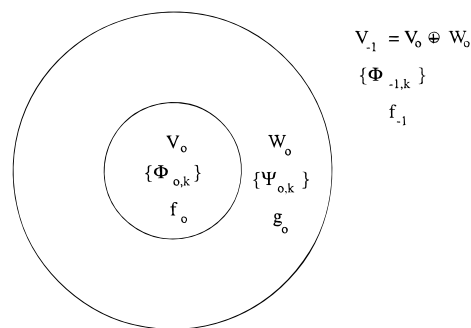
The scaling function  $\phi(t)$  can always be chosen such that the basis  $\{\phi_{0,k}\}$  is orthonormal. The basic relation that defines  $\phi(t)$  will be introduced below.

The next step in constructing a hierarchy of spaces is to introduce a larger space  $V_{-1}$ . For any function  $f_0(t) \in V_0$  define  $f_{-1}(t)$  by the isomorphism

$$f_{-1}(t) = f_0(2t) \quad (2)$$

The isomorphism implies that  $V_0$  is a proper subspace of  $V_{-1}$ :  $V_0 \subset V_{-1}$ . A direct implication of the manner in which  $V_{-1}$  is generated is to identify

$$\phi_{-1,k}(t) = 2^{1/2} \phi(2t - k), \quad k \in Z \quad (3)$$



**Figure 1.** Hierarchy of the spaces  $V_{-1}$ ,  $V_0$ , and  $W_0$ .

as a basis for the space  $V_{-1}$ . The new basis  $\{\phi_{-1,k}(t)\}$  is readily seen to be orthonormal as a consequence of the manner  $V_{-1}$  is defined.

A new subspace  $W_0$  arises as the complement of  $V_0$  with respect to

$$W_0 \oplus V_0 = V_{-1} \quad (4)$$

The subspace  $W_0$  can be chosen to be orthogonal to  $V_0$ . Since  $V_0$  is the subspace of  $V_{-1}$  at lower resolution,  $W_0$  is the subspace of  $V_{-1}$ , which contains the higher order structures. Equivalently, in the frequency domain,  $V_0$  and  $W_0$  correspond respectively to the subspaces containing the low- and high-frequency components of the functions in  $V_{-1}$ . A convenient basis for  $W_0$  can be generated in a manner analogous to those for  $V_0$  and  $V_{-1}$  by the translations of some appropriate function  $\psi(t) \in W_0$  as

$$\psi_{0,k} = \psi(t - k), \quad k \in Z \quad (5)$$

The function  $\psi(t)$  is called the “mother wavelet”.

As a consequence of the hierarchy of the spaces  $V_0$ ,  $W_0$ , and  $V_{-1}$ , a function  $f_{-1}(t)$  can always be decomposed into its low-resolution and high-structure components  $f_0(t)$  and  $g_0(t)$ , respectively, as

$$f_{-1}(t) = f_0(t) + g_0(t) \quad (6)$$

In view of the respective bases introduced, the above function has the representations

$$\begin{aligned} f_{-1}(t) &= \sum_k c_{-1,k} \phi_{-1,k}(t) \\ f_0(t) &= \sum_k c_{0,k} \phi_{0,k}(t) \\ g_0(t) &= \sum_k w_{0,k} \psi_{0,k}(t) \end{aligned} \quad (7)$$

Figure 1 illustrates the notation and the relative configuration of the spaces, functions, and bases introduced above.

The next stage of the presentation establishes the relation between the respective basis sets. In view of the preceding construction, the scaling function  $\phi(t) \in V_0$  and the mother wavelet function  $\psi(t) \in W_0$  are also contained in  $V_{-1}$ . Therefore, they can each be represented in terms of the basis set  $\{\phi_{-1,k}\}$  for the latter space as

$$\begin{aligned} \phi(t) &= \sum_k h_k \phi_{-1,k}(t) = \sum_k h_k \sqrt{2} \phi(2t - k) \\ \psi(t) &= \sum_k g_k \phi_{-1,k}(t) = \sum_k g_k \sqrt{2} \phi(2t - k) \end{aligned} \quad (8)$$

where

$$\begin{aligned} h_k &= \langle \phi, \phi_{-1,k} \rangle \\ g_k &= \langle \psi, \phi_{-1,k} \rangle \end{aligned} \quad (9)$$

The first relation in eq 8 together with the condition that  $\sum_h |h_k| < \infty$  determines the restriction on  $\phi(t)$  and is called the “dilation equation”. The sequences  $h_k$  and  $g_k$  are called the discrete-time filters or, more specifically, the “low-pass” and the “high-pass” filters in the signal processing literature.<sup>4</sup> The orthogonality of the subspaces  $V_0$  and  $W_0$  and the orthonormality of the basis  $\{\psi_{0,k}\}$  lead to the conditions

$$\begin{aligned} \sum_k \psi(t + n\pi k) \phi(t + 2\pi k) &= 0 \\ \sum_h |\psi(t + 2\pi k)|^2 &= 1 \end{aligned} \quad (10)$$

The above conditions are obtained from  $\int \psi(t) \psi(t - k) dt = \delta_{0,k}$  and  $\sum_n \int \psi(t) \phi(t - n) dt = 0$  and the isometry property of the Fourier transform.<sup>2</sup>

In view of the relations in eq 8, the expansion coefficients in eq 7 for the spaces  $V_0$  and  $W_0$  can now be shown to be related to those in  $V_{-1}$ . Indeed, considering the orthonormality of the bases, the inner products of eq 7 respectively with  $\phi_{0,n}$  and  $\psi_{0,n}$  yields

$$\begin{aligned} c_{0,n} &= \sum_k \langle \phi_{0,n}, \phi_{-1,k} \rangle c_{-1,k} \\ w_{0,n} &= \sum_k \langle \psi_{0,n}, \phi_{-1,k} \rangle c_{-1,k} \end{aligned} \quad (11)$$

In view of the definitions of eqs 1, 3, and 5, the following change of variables is introduced:

$$\begin{aligned} \phi_{0,n}(t) &= \phi(t - n) = \phi(t') \\ \psi_{0,n}(t) &= \psi(t - n) = \psi(t') \\ \phi_{-1,k}(t) &= \phi(2t - k) = \phi(2t' - (k - 2n)) = \phi_{-1,k-2n}(t') \end{aligned} \quad (12)$$

where  $t' = t - n$ . In terms of the variable  $t'$  and the definitions of the coefficients  $h_k$  and  $g_k$  in eq 8, the inner products in eq 11 read

$$\begin{aligned} \langle \phi_{0,n}, \phi_{-1,k} \rangle &= \langle \phi, \phi_{-1,k-2n} \rangle = h_{k-2n} \\ \langle \psi_{0,n}, \phi_{-1,k} \rangle &= \langle \psi, \phi_{-1,k-2n} \rangle = g_{k-2n} \end{aligned} \quad (13)$$

Finally, the substitution of the inner products as given in eqs 13 and eq 11 yields the desired relations between the expansion coefficients in eq 8 for the spaces  $V_0$ ,  $W_0$  and those in  $V_{-1}$  as

$$\begin{aligned} c_{0,n} &= \sum_k h_{k-2n} c_{-1,k} \\ w_{0,n} &= \sum_k g_{k-2n} c_{-1,k} \end{aligned} \quad (14)$$

The expressions above are called the “decomposition relations”, as they provide the coefficients for  $f_0(t)$  and  $g_0(t)$  in terms of those for  $f_{-1}(t)$ . The complementary relations for expressing  $c_{-1,n}$  in terms of  $c_{0,k}$  and  $w_{0,k}$  are derived by taking the inner

product of eq 6 with  $\phi_{-1,n}$ . This inner product, in view of eq 7 and the orthonormality of the basis, yields

$$c_{-1,n} = \sum_k \langle \phi_{-1,n}, \phi_{0,k} \rangle c_{0,k} + \sum_k \langle \phi_{-1,n}, \psi_{0,k} \rangle w_{0,k} \quad (15)$$

By use of the transformations of eq 13 in eq 15, the expansion coefficients in eq 8 for the space  $V_{-1}$  are related to those of  $V_0$  and  $W_0$  as

$$c_{-1,n} = \sum_k h_{n-2k} c_{0,k} + \sum_k g_{n-2k} w_{0,k} \quad (16)$$

The expression above is called the “synthesis relation”, as it provides the coefficients for  $f_{-1}(t)$  in terms of those for  $f_0(t)$  and  $g_0(t)$ . The decomposition relation in eq 14 and the synthesis relation in eq 16 are complementary for going between the lower and higher scales of resolution.

The preceding discussion was kept simple in terms of the number of spaces involved. A hierarchy of spaces can be defined in higher and lower resolutions starting with the reference space  $V_0$ . Thus we have an infinite sequence of nested subspaces as ...,  $V_1 \subset V_0 \subset V_{-1} \subset V_{-2}$ , ..., along with the subspaces  $W_j$ , which are the orthogonal complements of  $V_j$ :

$$V_{j-1} = W_j \oplus V_j, \quad j \in \mathbf{Z} \quad (17)$$

This construction implies that

$$V_{j-1} = W_j \in W_{j+1} \in W_{j+2} \in \dots \quad (18)$$

and

$$\bigoplus_{j=-\infty}^{\infty} W_j = L^2(R) \quad (19)$$

As  $j$  decreases or conversely increases, the subspace  $V_j$  approaches  $L_2(R)$  or  $\{0\}$ . This hierarchy allows the multiresolution representation of functions in  $L_2(R)$ . Hence the expressions in eqs 6 and 7 are generalized to read

$$f_j(t) = f_{j+1}(t) + g_{j+1}(t) \quad (20)$$

where

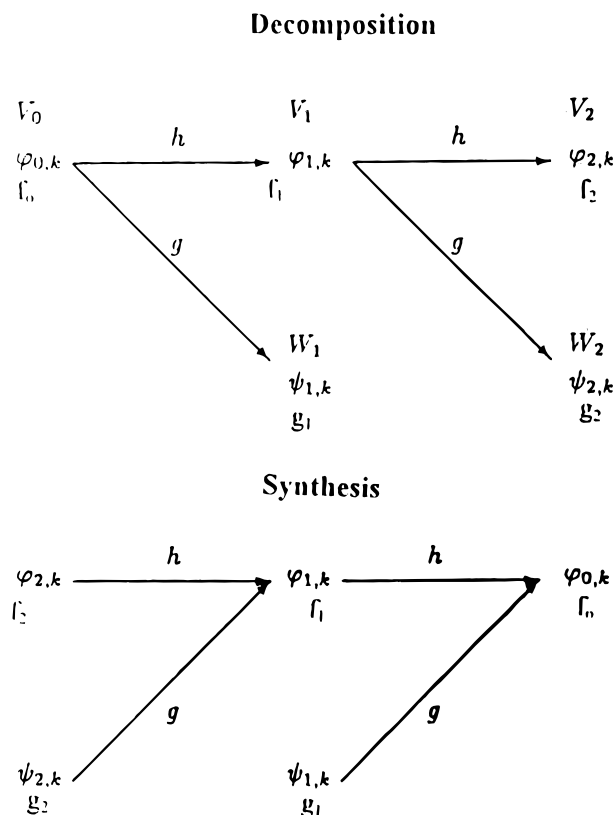
$$\begin{aligned} f_j(t) &= \sum_k c_{j,k} \phi_{j,k}(t) \\ f_{j+1}(t) &= \sum_k c_{j+1,k} \phi_{j+1,k}(t) \\ g_{j+1}(t) &= \sum_k w_{j+1,k} \psi_{j+1,k}(t) \end{aligned} \quad (21)$$

The elements of the discussion on the hierarchical representations along with the relevant components of basis sets are illustrated in Figure 2a for decomposition and in Figure 2b for synthesis.

The basis sets above are generated by the dilations and translations of the scaling and mother wavelet function as

$$\begin{aligned} \phi_{j,k}(t) &= 2^{-j/2} \phi(2^{-j}t - k) \\ \psi_{j,k}(t) &= 2^{-j/2} \psi(2^{-j}t - k) \end{aligned} \quad (22)$$

Moreover, it is evident from the derivations above that the dilation equation in (14) is a universal relation that is valid for



**Figure 2.** (a, top) Decomposition of the subspaces  $V_0, V_1, W_1, V_2, W_2, \dots$  and the related functions. (b, bottom) Synthesis of the subphase  $V_0$  from  $V_1$ , and  $W_1$ , etc. The filter coefficients  $h_k$  and  $g_k$  provide the link between the consecutive subspaces.

all orders of resolution spaces with the same coefficients  $h_k$  and  $g_k$ :

$$\begin{aligned} c_{j+1,k} &= \sum_n h_{n-2k} c_{j,n} \\ w_{j+1,k} &= \sum_n g_{n-2k} c_{j,n} \end{aligned} \quad (23)$$

and similarly for the generalization of eq 16,

$$c_{j,n} = \sum_k g_{n-2k} w_{j+1,k} + \sum_k h_{n-2k} c_{j+1,k} \quad (24)$$

The recursive application of the decomposition in (18) yields

$$f_j(t) = \sum_{l=j+1}^{\infty} \sum_{k=-\infty}^{\infty} w_{l,k} \psi_{l,k}(t) \quad (25)$$

and

$$f(t) = \lim_{j \rightarrow -\infty} f_j(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} w_{j,k} \psi_{j,k}(t) \quad (26)$$

With the representation of the limit function in  $L_2(R)$ ,  $f_j(t)$  is interpreted as the projection of  $f(t)$  into the subspace  $V_j$ . The above representation makes the set  $\{\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k) \mid j, k \in \mathbf{Z}\}$  an orthonormal basis for  $L_2(R)$ . The coefficients  $w_{j,k}$  are given by the inner product:

$$w_{j,k} = \int_{-\infty}^{\infty} f(t) \psi_{j,k}(t) dt \quad (27)$$

The coefficients  $w_{j,k}$  are defined as the WT of the function  $f(t)$ .

In the practical application of wavelet analysis, a given signal  $f(t)$  is first projected onto the subspace  $V_0$ ; that is, the coefficients  $c_{0,n}$  are computed. For the representations in the subsequent subspaces, the wavelet coefficients  $w_{j,k}$  ( $j = 1, 2, \dots, J$ ) and  $c_{J,k}$  are computed in a recursive manner from eq 23. In this way,  $f(k\Delta t)$  where  $\Delta t$  is a suitable sampling period, is represented by  $c_{J,k}$  and  $w_{j,k}$  ( $j = 1, 2, \dots, J$ ). This representation is called the finite wavelet transform of  $f(t)$ , and it uniquely specifies the set of samples  $f(k\Delta t)$ .<sup>6</sup> The coefficients  $c_{J,k}$  determine a coarse approximation to  $f(t)$  at the resolution scale  $J$ , and the wavelet coefficients  $w_{j,k}$  ( $j = 1, 2, \dots, J$ ) are the “details” at resolution scales  $j = 1, 2, \dots, J$ , respectively. Another common approach assumes that the components of the given signal  $f(t)$  in  $V_0$  are the discrete values  $f(k\Delta t)$ . This assumption is consistent with taking  $\phi_{0,k}(t)$  as highly localized about  $t = k\Delta t$ . Consequently, these components are expressed in the basis  $\phi_{0,k}$  as  $\langle f(t), \phi_{0,k} \rangle = f(k\Delta t)$ . This approximate starting strategy is computationally more efficient.

In the final part of this section, the time and frequency dependence of the wavelets is discussed. The familiar FT of a function

$$F(w) = \int_{-\infty}^{+\infty} f(t) \exp(-iwt) dt \quad (28)$$

as well as the Fourier series as its discrete analog

$$F_k = \frac{1}{T} \int_{-T}^T f(t) \exp(-kt/T) dt \quad (29)$$

provide information on the frequency content of  $f(t)$  over the whole space. However, this frequency content does not convey information about the local behavior of  $f(t)$ . That is, the FT cannot provide information on the frequencies that are operative during the evolution process. This is an important consideration for nonstationary processes. For example a sudden change in  $f(t)$  cannot be detected from the FT or Fourier series. Equally important, the FT is a very poor representation procedure for functions  $f(t)$  with sudden changes. This leads to a very broad Fourier spectrum manifested by the Gibbs phenomenon when the series, by necessity, is truncated at a finite number of frequencies. Another useful interpretation is provided by viewing the FT within the wavelet formalism. Within this framework, the FT is a one-parameter basis generated by the dilations of the “mother function”,  $\varphi(t) = \exp(it)$  into  $\varphi_w(t) = \exp(iwt)$ ,  $w \in \mathbf{R}$ . This structure of the FT makes the basis functions completely localized in frequency while having an infinite extent in time. In contrast, the wavelet  $\psi(t)$  is a “band-pass” function; that is, the Fourier transform  $\hat{\psi}(w)$  of  $\psi(t)$  is 0 at  $w = 0$  and  $w \rightarrow \infty$ . Moreover, with decreasing  $j$ ,  $\psi_{j,k}$  is increasingly localized in time while more spread out in frequency. Conversely, as  $j$  increases,  $\psi_{j,k}$  becomes more localized in frequency and broader in time. Therefore, low and high  $j$  values correspond respectively to high and low frequencies. The ability of the wavelets to describe the frequency content of a process during each phase of its evolution follows from their local nature. Indeed, since the wavelet  $\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k)$  is centered around  $t = 2^j k$ , the WT coefficient  $w_{j,k}$  captures the features of  $f(t)$  around  $t = 2^j k$  at the resolution scale  $j$ . Due to this feature, the WT is called a time-frequency analysis tool.

The simplest example of a wavelet basis is the well-known Haar basis, in which the mother wavelet is

$$\psi(t) = \begin{cases} 1/2, & \text{if } 0 \leq t \leq 1/2 \\ -1/2, & \text{if } 1/2 < t < 1 \\ 0, & \text{otherwise} \end{cases} \quad (30)$$

the corresponding scaling function is

$$\phi(t) = \begin{cases} 1, & \text{if } 0 \leq t < 1 \\ 0, & \text{otherwise} \end{cases} \quad (31)$$

The filter coefficients are  $h_k = 1/\sqrt{2}$  for  $k = 0, 1$ , and  $h_k = 0$  otherwise, and  $g_0 = 1/\sqrt{2}$ ,  $g_1 = -1/\sqrt{2}$ , and  $g_k = 0$  otherwise. It can easily be shown that the Haar wavelet and the scaling function satisfy eqs 8–10. From the behavior of  $\psi(t)$  one may expect that  $w_{j,k}$  as a function of  $k$  will be sensitive to sudden changes in  $f(t)$  to higher resolution for decreasing  $j$ . Similarly,  $\phi(t)$  will result in  $c_{j,k}$  reflecting the actual shape of the function  $f(t)$  at increasing resolution for decreasing  $j$ . This interpretation will be evident in the examples below.

The calculations in this paper utilized the Daubechies wavelet with eight nonvanishing coefficients.<sup>2</sup> The low-pass filter coefficients for this case are

$$\{h_k\} = (0.230378, 0.714847, 0.630881, -0.0279838, -0.187035, 0.030841, 0.032883, -0.0105974) \text{ for } k = 0, 1, 2, \dots, 7 \quad (32)$$

and  $h_k = 0$ , otherwise. The high-pass filter  $g_k$  is related to the low-pass filter  $h_k$  through the relation

$$g_k = -1^k h_{8-k} \quad (33)$$

The shapes of the corresponding wavelet and scaling functions may be found in many references, e.g., ref 2.

### 3. Applications

In this section, two applications are considered: a deterministic three-dimensional polymer<sup>13</sup> and a one-dimensional polymer chain subjected to random forces.<sup>12</sup> The dynamics are governed by the Newton and Langevin equations, respectively.

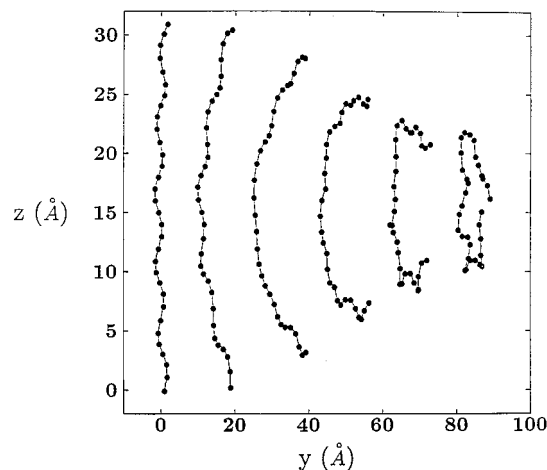
**(i) Three-Dimensional Polymer with Deterministic Dynamics.** The kinematics of the polymer chain studied here is described in terms of the bond lengths  $b_{jk}$ , bending angles  $\theta_{jkl}$ , and the dihedral (or torsion) angles  $\phi_{jklm}$ . The various long-range nonbonded interactions are not included. Here,  $jk$ ,  $jkl$ , and  $jklm$  indicate respectively two, three, and four atom interactions in the polymer with the indices labeling the particles. The kinematic variables are defined in terms of the Cartesian coordinates  $\mathbf{x}_j$  of the particles forming the polymer as<sup>13,15</sup>

$$b_{jk} = |\mathbf{x}_j - \mathbf{x}_k|, \quad \cos \theta_{jkl} = \mathbf{c}_{jk} \cdot \mathbf{c}_{kl}, \quad \sin \phi_{jklm} = \frac{\det[\mathbf{c}_{jk}, \mathbf{c}_{kl}, \mathbf{c}_{lm}]}{\sin \theta_{jkl} \sin \theta_{klm}} \quad (34)$$

Above,  $\det$  denotes the triple product of the vectors appearing as arguments which are calculated as a determinant, and  $\mathbf{c}_{jk}$  are the unit vectors along the bonds between the particles at  $\mathbf{x}_j$  and  $\mathbf{x}_k$ :  $\mathbf{c}_{jk} = (\mathbf{x}_k - \mathbf{x}_j)/b_{jk}$ . The potential energy is taken typically as

$$V = \sum K_{ij}(b_{ij} - b_{ij}^0)^2 + \sum K_{ijk}^B(\theta_{ijk} - \theta_{ijk}^0)^2 + \sum K_{ijkl}^T[1 - \cos 3(\phi_{ijkl} - \phi_{ijkl}^0)] \quad (35)$$

The values of the coefficients  $K_{ij}$ ,  $K_{ijk}^B$ , and  $K_{ijkl}^T$  are known for various bonds along with the equilibrium values  $b_{ij}^0$ ,  $\theta_{ijk}^0$ , and  $\phi_{ijkl}^0$ .<sup>12</sup> The dynamics of a system of 32 particles, which are initially arranged along a helix and eventually fold into a globular form, generates the time series to be analyzed.



**Figure 3.** Snapshots for the time history of the polymer configuration.

Associated with these particles are 6° coordinates for rigid body motion, 29 torsional angles, 30 bending angles, and 31 intra-atomic bonds corresponding to the 96 variables for the 32 particles. The torsional angles are the kinematic variables that undergo the most severe changes, including transitions between various local equilibrium configurations. Indeed, the torsional interaction potential seen in eq 35 has three local equilibrium points at the values  $\phi_{ijkl} - \phi_{ijkl}^0 = 0, 2\pi/3$ , and  $4\pi/3$ . The dynamical equations in Cartesian coordinates are given by the Newton equations and the initial conditions below:

$$\ddot{\mathbf{x}} = \mathbf{F}(\mathbf{x}), \quad \mathbf{x}(0) = 0, \quad \dot{\mathbf{x}}(0) = \mathbf{v}_0 \quad (36)$$

Above  $\mathbf{x}$  represents collectively all of the coordinates, for simplicity in the presentation the masses are taken as normalized, and  $F_j = -\partial V/\partial x_j$ . More details on this system can be found in ref 12. The snapshots in Figure 3 show the extreme nature of the dynamical deformations of the system. This level of deformation was made possible by choosing a large impulsive initial velocity  $\mathbf{v}_0$  solely in the eighth mode of the linearized system at the initial equilibrium configuration as a helix. This initial energy distribution corresponds to an extreme nonequilibrium state with the energy concentrated only in one mode. This energy tends to equidistribute itself as the system evolves through time.

All of the time histories  $f(t)$  are available at the discrete time instants  $t = k\Delta t$ , with  $k$  a positive integer. In this study, Daubechies wavelets are used, but similar results can be obtained using other wavelets as well.<sup>16</sup> As in many digital signal processing problems, it is assumed that the scaling coefficients  $c_{0,k}$  at the resolution scale  $j = 0$  are equal to the samples  $f(k\Delta t)$ , i.e.,

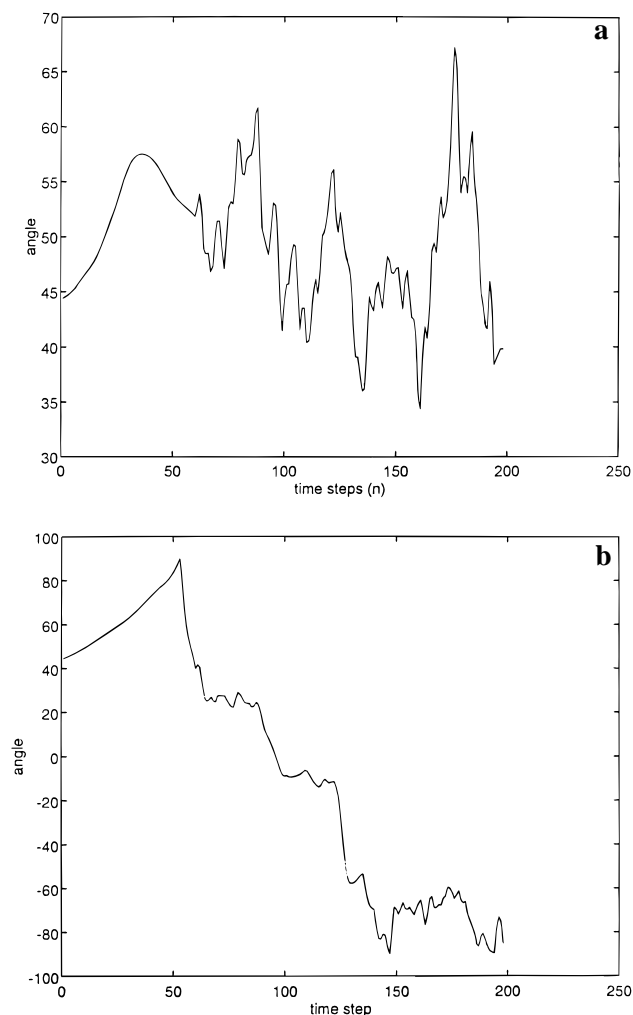
$$c_{0,k} = f(k\Delta t), \quad k = 0, 1, 2, \dots \quad (37)$$

This means that there is a function  $f_0(t) \in V_0$  such that

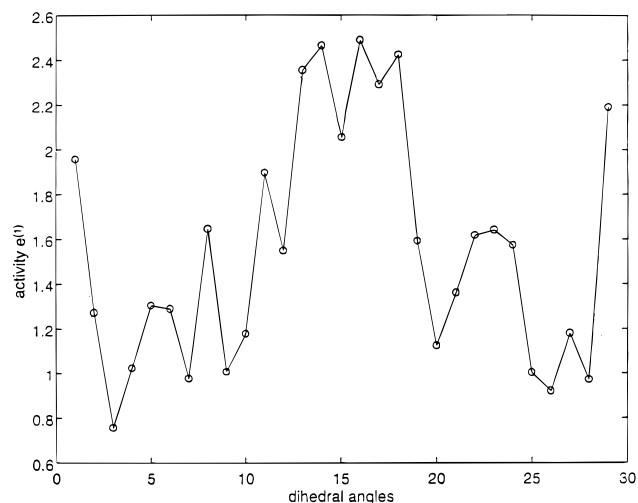
$$f_0(t) = \sum_k f(k\Delta t) \phi_{0,k}(t) \quad (38)$$

where  $\phi_{0,k}(t)$  is the scaling function (1) of the subspace  $V_0$ . The function  $f_0(t)$  is considered to be a close approximation<sup>2,6</sup> to the actual time function  $f(t)$ . The wavelet transform coefficients,  $w_{1,k}$  and the inner products  $c_{1,k}$  at the resolution scale  $j = 1$  are then obtained by eq 23. Likewise,  $c_{2,k}$  and  $w_{2,k}$ , are obtained from  $c_{1,k}$  by the recursive use of these latter equations. In the present applications, this iterative procedure is repeated until  $j = 3$ .

The plots in Figures 4–8 display some results based on the dynamics of the system described above. The time histories

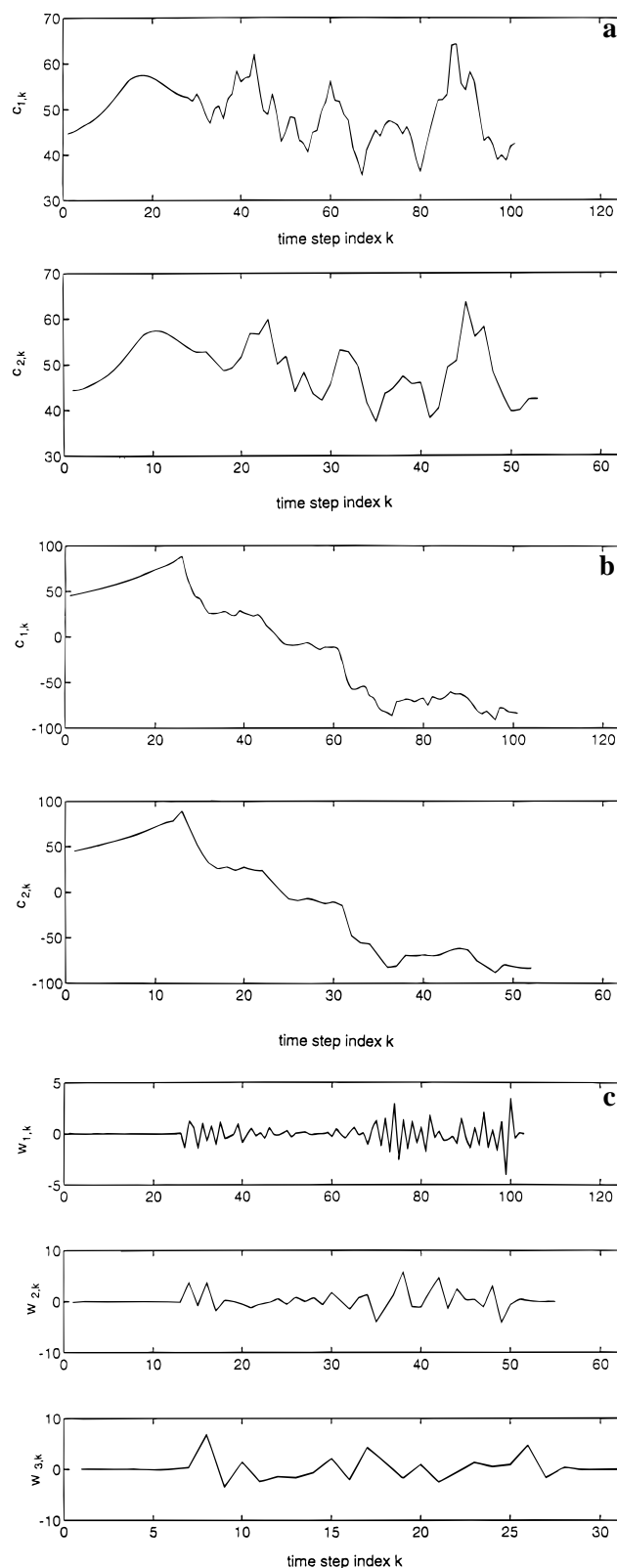


**Figure 4.** Time histories for representative internal coordinates. The data are sampled at 200 time steps  $\Delta t$ . (a) A torsional angle with small oscillations. (b) A torsional angle with large oscillations and a transition from local equilibrium near  $60^\circ$  to one near  $-60^\circ$ .



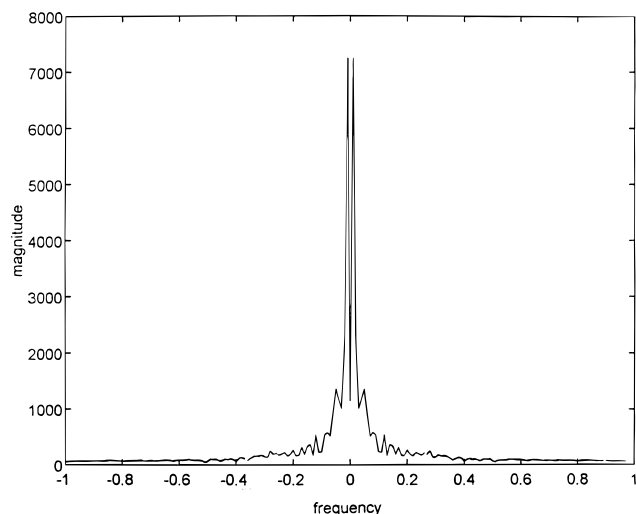
**Figure 5.** Wavelet activities,  $e^{(1)} = \sum_k |w_{1,k}|^2$ , for the 29 dihedral angles of the polymer.

of the torsional angles show the largest variability between the various bonds, and two of these are chosen for the analysis. The time series in Figure 4a,b for these two dihedral angles are the primary data to be analyzed. The first plot shows the dynamics for a dihedral angle that remains near its initial local equilibrium while undergoing small amplitude oscillations. The second dihedral angle undergoes a large change including a

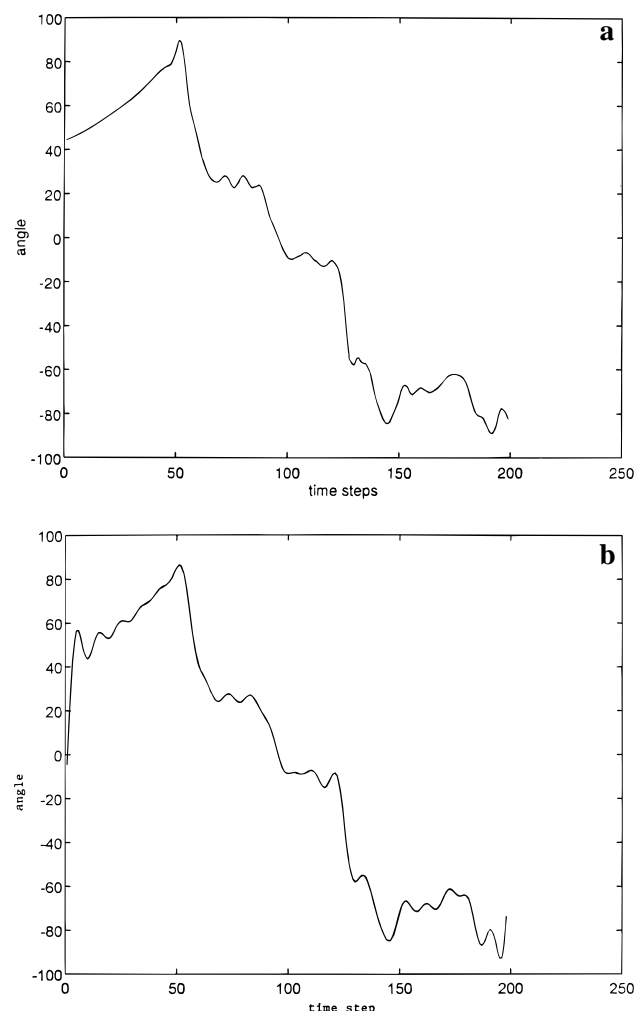


**Figure 6.** (a) Scaling coefficients  $c_{1,k}$  and  $c_{2,k}$  corresponding to the time histories of the torsional angles presented in Figure 4a. The time step index  $k$  corresponds respectively to  $2\Delta t$  and  $2^2\Delta t$  in the top and bottom plots (since in the function  $\psi_{j,k}(t) = \psi(t/2^j) - k$  the time step has to be expanded by  $2^j$  at resolution scales  $j = 1, 2$ ). (b) Scaling coefficients  $c_{1,k}$  and  $c_{2,k}$  corresponding to the time histories of the dihedral angles presented in Figure 4b. (c) Wavelet coefficients (the detailed subsignals),  $w_{j,k}$ ,  $j = 1, 2, 3$ , are plotted for the time history shown in Figure 4b. The time step index  $k$  corresponds to  $2\Delta t$ ,  $2^2\Delta t$ , and  $2^3\Delta t$  in the top, middle, and bottom plots.

transition from the local equilibrium position near  $\phi_{ijkl} - \phi_{ijkl}^0 = 0$  to the one near  $\phi_{ijkl} - \phi_{ijkl}^0 = 2\pi/3$ , as shown in Figure 4b.



**Figure 7.** Magnitude of the FT for the time history in Figure 4b. The frequency is expressed in units of  $1/2\Delta t$ .



**Figure 8.** Signal reconstructed by excluding some of the high-resolution and high-frequency components for the time history in Figure 4b: (a) WT; (b) FT. 41 coefficients are used in both cases.

The “activity” of the WT coefficients is a measure of the high-frequency activity in the data. The activity,  $e^{(1)}$ , of the WT coefficients at resolution scale 1 is defined as follows:

$$e^{(1)} = \sum_k |w_{1,k}|^2 \quad (39)$$

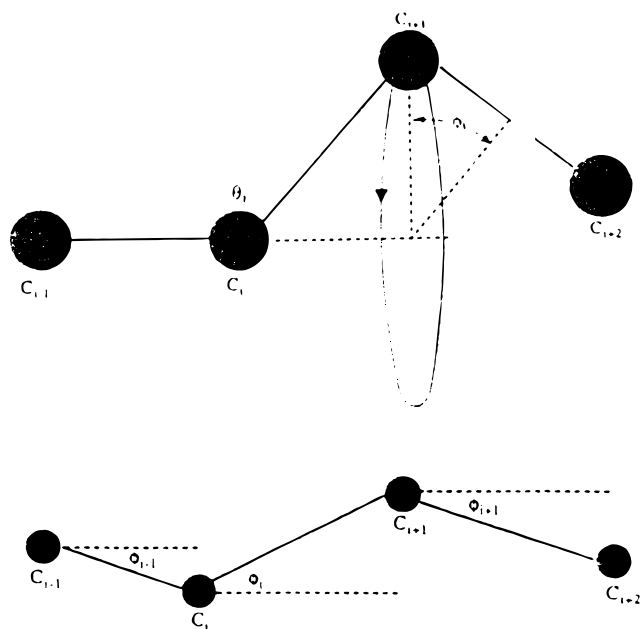
In Figure 5 the effectiveness of the WT as a diagnostics tool is

investigated by plotting  $e^{(1)}$  for the 29 torsional angles of the polymer. Indeed, the three large peaks at angles 14, 16, and 18 correspond precisely to the torsional angles that made transitions from one local equilibrium angle to the next in the interaction potential. The activity of the torsion angles 13, 15, and 17 is high because they are significantly disturbed neighbors of the transition events located at 14, 16, and 18. The plots in Figure 4a,b are respectively the time histories for the fourth and the 16th activity in Figure 5.

The low-pass filter  $h_k$  and the coefficients  $c_{j,k}$ ,  $j = 1, 2, \dots$ , via eq 23 define coarser approximations to the time history at the resolution scales  $j = 1, 2, \dots$ . Figures 6a,b shows the plots for  $c_{j,k}$ ,  $j = 1, 2$ , corresponding to the time histories of the dihedral angles presented in Figure 4a,b, respectively. It is observed that the significant characteristics of the signal are captured at both levels of coarseness. Note that in Figure 6a,b the scale of the time axis changes as  $j$  increases. This is due to the fact that  $c_{j,k} = \langle f(t), 2^{-j/2}\phi(2^{-j}t - k) \rangle$  describes the local activity of  $f(t)$  around  $t = 2^j k$  as the smoothing function  $\phi(2^{-j}t - k)$  is centered around  $t = 2^j k$ . Between  $c_{j,k}$  and  $c_{j,k+1}$  an intermediate value  $c_{j,k+1/2} = \sum_n h(n - 2k - 1)c_{j-1,n}$  can be computed. The coefficients  $c_{j,k}$  together with the intermediate values  $c_{j,k+1/2}$  are called the low-pass filtered version of the sequence  $c_{j-1,k}$  in the signal processing literature, and the sequences  $c_{j,k}$ ,  $j = 1, 2, \dots$ , capture the low-frequency activity or equivalently the main trends of the original time history.

Complementary to these results are the wavelet coefficients  $w_{j,k}$ ,  $j = 1, 2, 3$  (i.e., the detailed subsignals), plotted in Figure 6c for the time history shown in Figure 4b. The extrema of the WT are used to detect sudden changes in signals such as edge detection in images.<sup>8,9</sup> It can be observed from Figure 6c that the sequence of wavelet coefficients  $w_{1,k}$  reflects a significant extremum when the torsional angle data make a sudden transition at  $t = 53\Delta t$ . The wavelet coefficients  $w_{j,k}$  describe the projection of the time history onto the subspace  $W_j$ , which contains the “detailed” part of the subspace  $V_{j-1}$ . This is due to the fact that the wavelet  $\psi(t)$  is a “band-pass” function, that is, the Fourier transform  $\hat{\psi}(w)$  of  $\psi(t)$  is 0 at  $w = 0$  and  $w \rightarrow \infty$ . Therefore the inner products  $w_{j,k} = \langle f(t), 2^{-j/2}\psi(2^{-j}t - k) \rangle$  capture changes in the function  $f(t)$ . For example a relatively high valued wavelet coefficient  $w_{1,k_0}$  means that around  $t = 2^1 k_0$  a transition occurs in the original time history. In Figure 6c the wavelet coefficients  $w_{1,k}$  are very close to zero until  $k = 26$  or equivalently  $t = 2^1 \times 26$ , and their values significantly increase after that. Similarly, the wavelet coefficients  $w_{2,k}$  have also relatively small values up to  $k = 13$  or  $t = 2^2 \times 13$ . The wavelet sequence  $w_{1,k}$  together with the intermediate values  $w_{1,k+1/2} = \sum_n g(n - 2k - 1)c_{j-1,n}$  is called the high-pass filtered version of the time history in the signal processing literature. For comparison, the power spectrum of the time history in Figure 4b, shown in Figure 7, illustrates that it is not possible to detect the transition points from the FT, which is only the function of the frequency.

To investigate the signal compaction capability of the WT and FT, parts a and b of Figure 8 respectively display the signal reconstructed by excluding some of the high-resolution and high-frequency components in the two schemes. The signal in Figure 8a is reconstructed from first 24 wavelet and 17 scaling coefficients, which are significantly large in absolute value compared to the other coefficients. In a similar spirit of data reduction, 41 low-frequency components of the FT are used to reconstruct the signal shown in Figure 8b. Clearly the signal in Figure 8a is a better approximation to the original signal shown in Figure 4b. The ringing effects in Figure 8b are due to the Gibbs phenomenon, which does not occur in Figure 8a.



**Figure 9.** Section of a three-dimensional chain molecule and its two-dimensional projection.

Similar data compaction results can be obtained for the other time series as well. This compaction comparison becomes even more favorable with the WT for larger sets of data points.

**(ii) Polymer Chain Subjected to Random Forces.** The primary purpose of this case is to study the performance of the WT in a random force field as opposed to the deterministic example above. The linear chain model<sup>12</sup> is obtained by following the construction by Weiner and Pear.<sup>17</sup> A section of the three-dimensional chain and its projection onto a plane are shown in Figure 9. To construct a simplified model, the stretching of interatomic bonds is not permitted, and the rigid bond lengths are taken to all have the common value  $b^0$ . Secondly, the valence angles  $\theta_{j-1,j,j+1}$  between the atoms  $j-1$ ,  $j$ , and  $j+1$  are taken as fixed, while the torsional rotation of the bond between the atoms  $j$  and  $j+1$  is allowed. The unknowns in the model are therefore the torsional angles  $\phi_j$  through which the bond between the atoms  $j$  and  $j+1$  has rotated. The angles are measured relative to some fixed reference line, which will be termed the  $x$ -axis; consequently the rotation is now visualized as taking place about the axes perpendicular to the two planes formed by the bonds between the atoms  $j-1$ ,  $j$  and  $j, j+1$ , as seen in Figure 9. The projection of the position  $x_j$  of the atoms onto the  $x$ -axis leads to

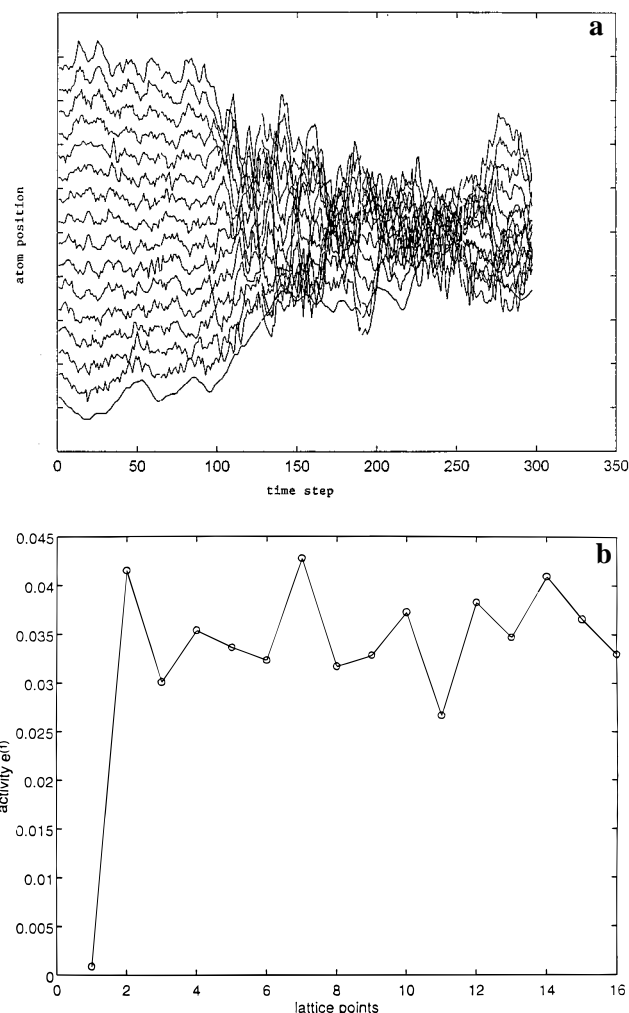
$$x_0 = 0, \quad x_j = b^0 \cos \phi_{j-1} + x_{j-1} \quad (j = 1, 2, \dots, n) \quad (40)$$

The potential  $V$  is expressed as a function of the rotation angle or equivalently the relative distances  $\zeta_j$  between the nearest neighbors:

$$\zeta_j = \frac{x_j - x_{j-1}}{b^0} \cos \phi_{j-1} \quad (j = 1, 2, \dots, n) \quad (41)$$

The local equilibrium configurations correspond to the angles  $\phi_j = 0$  and  $\pi$  or equivalently  $\zeta_j = 1$  and  $-1$ , whereas the rotational barrier is at  $\zeta_j = 0$ . This behavior is represented by a double-well potential defined by the quartic

$$V(\zeta) = \gamma(1 - \zeta^2)^2 \quad (42)$$



**Figure 10.** (a, top) Particle trajectories for a representative case where the temperature is increased from an initial value to a higher one. (b, bottom) The wavelet activities ( $e^{(1)} = \kappa_k |w_{1,k}|^2$ ) for the 16 lattice points.

with force constant  $\gamma$ . The governing dynamical equations have the form

$$\ddot{x}_1 = -\sigma + F(x_1 - x_2) + R_1(t) \quad (43)$$

$$\ddot{x}_j = -F(x_j - x_{j-1}) + F(x_{j+1} - x_j) - 2\nu\dot{x}_j + R_j(t), \quad j = 2, \dots, n-1 \quad (44)$$

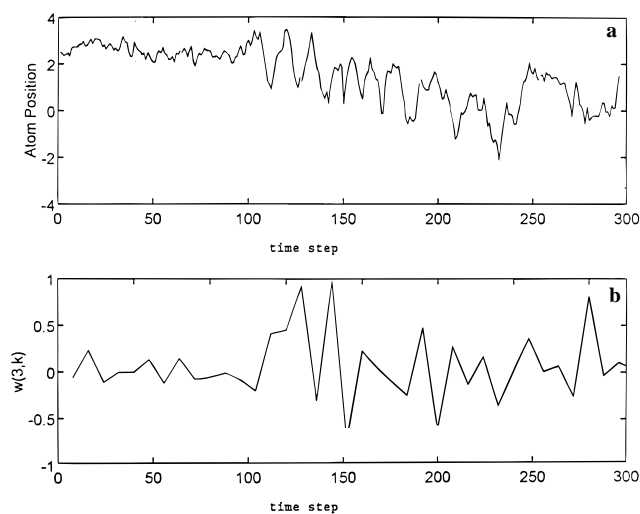
$$\ddot{x}_n = \sigma - F(x_n - x_{n-1}) - 2\nu\dot{x}_n + R_n(t) \quad (45)$$

where  $F(x) = -\partial V(x)/\partial x$  with  $V(x)$  defined in eq 42,  $\nu$  is a damping coefficient,  $R_j$  is a random force, and  $\sigma$  is a stress applied to the end points  $x_1$  and  $x_n$ . The components  $R_i$  of the random force field  $\mathbf{R} = \mathbf{R}(t)$  acting on the system satisfy the statistical relations<sup>12</sup>

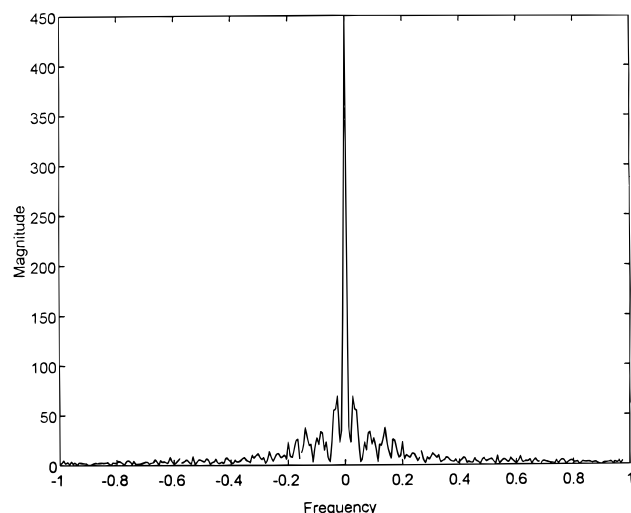
$$\langle R_i(t) \rangle = 0, \quad \langle R_i(t) R_j(t') \rangle = 4\nu T \delta(t - t') \sigma_{ij} \quad (46)$$

Here  $T$  is the temperature with Planck's constant set to unity and  $\langle \rangle$  denotes an average taken over the atoms.  $\nu$  and  $\mathbf{R}$  model temperature effects on the molecule in the sense of the Langevin equation for Brownian motion. For the numerical calculations, the polymer chain had 16 atoms, and several sets of calculations distinguished by different ambient temperature fields are performed.<sup>12</sup> Figure 10a shows the particle trajectories for a representative case where the temperature is increased from an initial value to a higher one at the prescribed 100th time step. The polymer contracts, as expected, through heating. Figure 10b plots the activity  $e^{(1)} = \sum_k |w_{1,k}|^2$  for each of the 16 lattice





**Figure 11.** (a) Time history for the 10th lattice point. The data are sampled at 300 time steps  $\Delta t$ . (b) The WT coefficients  $w_{3,k}$  at the resolution scale  $j = 3$ .



**Figure 12.** Magnitude of the FT of the time history in Figure 10a. The frequency is expressed in units of  $1/2\Delta t$ .

points. Unlike in the previous case, no significant difference exists between the lattice points. This is expected since the particles are subjected to uncorrelated white random forces. Figure 11a,b shows the time history for the 10th lattice point and the WT coefficients  $w_{3,k}$  at the resolution scale  $j = 3$ , respectively. The time instant when the temperature changes can be detected from the WT. Indeed, as seen in Figure 10b, the amplitudes of the WT coefficients  $w_{3,k}$  increase significantly after  $k = 13$ , which corresponds to  $t = 13 \times 2^3 \Delta t = 104 \Delta t$  (since  $\psi_{3,k}(t) = \psi((t/2^3) - k)$ , the time step has to be expanded by  $2^3$  at the resolution scale  $j = 3$ ). This identification is fully in accord with the actual jump at  $t = 100 \Delta t$ . Accompanying these time series is the corresponding FT magnitude plot in Figure 12. It is impossible to detect the temperature change from the FT.

#### 4. Concluding Remarks

This paper demonstrates the use of the wavelet transform (WT) for analysis of time series resulting from a numerical

simulation of molecular dynamics by analogy to its use in digital sound and image processing. The comparison of the performances of the Fourier and wavelet transforms on the same time series demonstrates the superiority of the latter for time histories with simultaneous localization in time and frequency as well as in reconstructing the original signal with a reduced data set. The conclusions are valid for both deterministic and stochastic dynamical studies. The processing of image and sound signals and the time series produced in molecular dynamics share a common issue: nonconsequential high-frequency noise or structure is present in all of these cases, and the challenge is to identify and discard the unimportant portion of the data. A major difference is, however, that the acoustic and visual signals are available and the WT post processes these. Conversely in molecular dynamics, the time series are to be generated by a numerical algorithm. The further challenge in this latter field is thus to transform the dynamical equation so that it would not generate the nonconsequential high-frequency structure from the outset. There are already efforts for using the wavelet functions as a basis for solving boundary value differential equations.<sup>18,19</sup> For dynamical equations these ideas need to be extended to initial value problems. A natural approach in this regard will follow the conversion of the dynamical equations into either discretized forms or integral equations, as suggested in the work of Beylkin.<sup>20</sup>

**Acknowledgment.** The authors acknowledge support from the Air Force Office of Scientific Research.

#### References and Notes

- (1) Morlet, J.; et al. *Geophysics* **1982**, 47, 203.
- (2) Daubechies, I. *Ten Lectures on Wavelets*; SIAM Press: Philadelphia, 1992.
- (3) Chui, C. *An Introduction to Wavelets*; Academic Press: New York, 1992.
- (4) Woods, J. W.; O'Neill, S. D. *IEEE Acoust., Speech Signal Process.* **1986**, ASSP-34, 1278.
- (5) Cetin, A. E.; Ansari, R.; Lee, S. H. *SPIE Annu. Int. Techn. Symp., Proc.*, 32nd **1988**, 974.
- (6) Mallat, S. G. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, 11, 674.
- (7) Cetin, A. E.; Gerek, Ö. N.; Ulukuş, Ş. *IEEE Trans. Circuits Syst. Video Technol.* **1993**, 3, 433.
- (8) Mallat, S. G.; Hwang, W. L. *IEEE Trans. Inf. Theory* **1991**, 38, 617.
- (9) Cetin, A. E.; Ansari, R. *IEEE Trans. Signal Process.* **1994**, ASSP-44, 194.
- (10) Permann, D.; Hamilton, I. *Phys. Rev. Lett.* **1992**, 69, 2607.
- (11) Permann, D.; Hamilton, I. *J. Chem. Phys.* **1994**, 100, 379.
- (12) Askar, A.; Owens, R. B.; Rabitz, H. A. *J. Chem. Phys.* **1993**, 99, 5316.
- (13) Askar, A.; Rabitz, H. A.; Space, B. *J. Phys. Chem.* **1995**, 99, 7330.
- (14) Space, B.; Rabitz, H. A.; Askar, A. *J. Chem. Phys.* **1993**, 99, 9070.
- (15) Askar, A. In *Mathematical and Physical Sciences*; NATO ASI Series, Series C.; Kluwer Academic Publishers: Dordrecht, 1995; Vol. 470.
- (16) See, for example, the Megawave software of CEREMADE, The University of Paris, at the ftp site mu.ceremade.daupine.fr.
- (17) Weiner, J. H.; Pear, M. R. *Macromolecules* **1977**, 10, 317.
- (18) Amaratunga, K.; Williams, J. R. *Eng. Comput.* **1993**, 10, 349.
- (19) Qian, S.; Weiss, J. J. *Comput. Phys.* **1993**, 106, 155.
- (20) Brewster, M. E.; Beylkin, G. A. Multiresolution Strategy for Numerical Homogenization, preprint.

JP961961W