# Wavelet Representations of Stochastic Processes and Multiresolution Stochastic Models

Robert W. Dijkerman and Ravi R. Mazumdar, *Member, IEEE*

*Abstract*— **Deterministic signal analysis in a multiresolution framework through the use of wavelets has been extensively studied very successfully in recent years. In the context of stochastic processes, the use of wavelet bases has not yet been fully investigated. In this paper, we use compactly supported wavelets to obtain multiresolution representations of stochastic processes with paths in $L^2$ defined in the time domain. We derive the correlation structure of the discrete wavelet coefficients of a stochastic process and give new results on how and when to obtain strong decay in correlation along time as well as across scales. We study the relation between the wavelet representation of a stochastic process and multiresolution stochastic models on trees proposed by Basseville *et al.* We propose multiresolution stochastic models on the discrete wavelet coefficients as approximations to the original time process. These models are simple due to the strong decorrelation of the wavelet transform. Experiments show that these models significantly improve the approximation in comparison with the often used assumption that the wavelet coefficients are completely uncorrelated.**

## I. Introduction

MULTISCALE signal processing is of great importance in problems arising in image processing, geophysical signal processing, acoustics, and in problems where the signal is the output of a distributed parameter system. The basic problem is one of efficient signal representation that yields a decomposition of the signal at various resolutions and at different time scales. In the deterministic context, the recent developments in the use of wavelets has led to very useful paradigms for signal representation and algorithms. This approach has origins in the work of Gabor [10] and has been developed by Meyer *et al.* [18], [12], [13] and by Daubechies [6]. The importance of wavelet bases stems from the fact that they lead to a multiresolution analysis for signals in $L^2(R)$ and offers different levels of accuracy in approximation based on a logarithmic scale. The link between wavelets and multiresolution analysis has been explored in particular by Mallat [15], [16] and Cohen [2] from the point of view of existence conditions and regularity properties. This link has been successfully exploited in image processing. On the other hand, there are few results related to the efficient multiscale representation of stochastic signals. Although the standard construction of Brownian motion from the Haar basis (a wavelet basis) is well known (see Davis [7], for example), no systematic attempt has been made to use the wavelet basis to obtain multiscale representations taking into account the statistical parameters of signals.

Recently, in work that is motivated by multiresolution analysis, Basseville *et al.* [1] have proposed a class of stochastic models defined on $n$-ary trees called "multiresolution stochastic processes." One of the aims of this paper is to explore the connection of such models to the wavelet representation of stochastic processes.

The study of the wavelet transformation of stochastic processes has been initiated by Cohen *et al.* [3], Wornell [22], Flandrin [8], [9], Ramanathan and Zeitouni [19], Tewfik and Kim [20], and Masry [17]. They were, for instance, interested in regularity issues for the spectral density and synthesizing processes with uncorrelated discrete wavelet coefficients. They showed that the correlation of the discrete wavelet coefficients of fractional Brownian motion decay fast; he also showed stationarity in the wavelet domain for stationary increments processes.

In this paper, we address the multiresolution representation of stochastic processes. Especially, we study in detail the correlation structure of the discrete wavelet coefficients of stochastic processes in general and indicate some new results on decorrelative properties of wavelets for several processes. These properties will be the motivation for the approximate multiresolution stochastic models we shall propose. The decorrelation capacity of the wavelet transform along time has already been noted and exploited for matrix inversions in the papers by Tewfik and Kim [20], [21].

We then show how the discrete wavelet coefficients in the dyadic framework can give rise to processes on trees as introduced by Basseville *et al.* [1]. These multiresolution auto-regressive models are of special interest since estimation algorithms for these processes have been developed (see for example [1], [14]). To approximate the processes in the time-scale domain, we propose and discuss several multiresolution stochastic state models. Simulation results are reported to show simple and good approximations of the time processes in a multiresolution framework. The early results of part of this paper were announced in [4].

The organization of this paper is as follows. In the next section, we discuss the discrete wavelet transform, in particular as applied to stochastic processes with sample paths in $L^2$. In Section III, we study in detail the correlation function of the discrete wavelet coefficients, and we give new results on the decorrelative capacity of the wavelet transformation. We discuss the consequences of these results and their link with

Manuscript received July 3, 1993; revised October 23, 1993. This work was supported by a grant from Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (FCAR). The associate editor coordinating the review of this paper and approving it for publication was Prof. Sergio D. Cabrera.

The authors are with Institut National de la Recherche Scientifique, INRS-Télécommunications, Université du Québec, Ile des Soeurs, Canada H3E 1H6.

IEEE Log Number 9401285.

the study of multiresolution stochastic processes conducted by Basseville *et al.* [1] in Section IV and illustrate our ideas by the multiresolution statistical modeling of several example processes in Section V. Conclusions are drawn in Section VI.

## II. The Wavelet Representation of a Process

Let us consider a compactly supported wavelet $\psi \in L^2(R)$ as introduced by Daubechies [6]. The simplest of these wavelets is the so-called Haar wavelet, which has been known for a long time. It has the form

$$\psi(x) = \begin{cases} 1 & \text{for } 0 \le x < \frac{1}{2} \\ -1 & \text{for } \frac{1}{2} \le x < 1 \\ 0 & \text{for } x < 0, x \ge 1. \end{cases} \tag{1}$$

We shall assume that the wavelet function $\psi$, is compactly supported on $[-(M-1), M] \subset [-M, M]$ with $M \in N^* = \{1, 2, 3, \ldots\}$. The wavelet $\psi$ has $M$ vanishing moments

$$\int_{-M}^{M} t^q \psi(t) dt = 0, \qquad q = 0, 1, \ldots, M-1. \tag{2}$$

The discrete wavelet coefficient $F(j,n)$ of a function $f \in L^2(R)$ at scale $j$ and time $n$ is defined by

$$F(j,n) = \int_{2^j(n-M)}^{2^j(n+M)} f(t)\psi_{j,n}(t)dt \tag{3}$$

where

$$\psi_{j,n}(t) = 2^{-\frac{j}{2}}\psi(2^{-j}t - n). \tag{4}$$

We consider the discrete wavelet transform construction of Mallat [15], using the concept of a multiresolution analysis. We briefly explain this idea since we shall use it for the multiresolution models throughout the paper. For a complete and formal treatment of the multiresolution analysis, we refer to [15] and [6].

We assume the wavelet function $\psi$ to have $L^2$ norm equal to 1. Let $O_j$ be the subspace spanned by $\{\psi_{j,n}, n \in Z\}$ for $j$ fixed. Then, we demand

$$\bigoplus_{j=-\infty}^{\infty} O_j \text{ is dense in } L^2(R) \tag{5}$$

where $\bigoplus$ denotes the direct sum. $O_j$ is called a detail space. The so-called approximation space $V_k$ is now defined by

$$V_k = \bigoplus_{j=k+1}^{\infty} O_j \tag{6}$$

and thus

$$V_k = V_{k+1} \oplus O_{k+1}. \tag{7}$$

$V_k$ can be seen as an approximation (at the resolution $k$) of $L^2(R)$. A function $f$ of $L^2(R)$ can be projected on the space $V_k$. There exists a unique scaling function $\phi \in L^2(R)$ having $L^2$ norm equal to 1 such that the translations of a dilation of $\phi$ form an orthonormal basis of $V_k$, that is, $\{\phi_{k,n}(t) = 2^{-\frac{k}{2}}\phi(2^{-k}t - n), n \in Z\}$ forms an orthonormal basis of $V_k$. The scaling function $\phi$ is compactly supported on $[0, 2M-1] \subset [0, 2M]$.

The scaling coefficient $\widehat{F}(j,n)$ of the function at scale $j$ and time $n$ is then defined by

$$\widehat{F}(j,n) = \int_{2^j n}^{2^j(n+2M)} f(t)\phi_{j,n}(t)dt. \tag{8}$$

A function $f$ of $L^2(R)$ can now be represented, for example, by the projection of $f$ on $V_k$ and its projections on $O_k$, $O_{k-1}$, $O_{k-2}, \ldots$ These projections can be obtained by calculating the inner products of $f$ with all the functions in the orthonormal bases. The set of coefficients obtained from the projections on all spaces $O_k$ ($k \in Z$) is the set of so-called discrete wavelet coefficients. This set of coefficients is sufficient to reconstruct $f$ in the $L^2$ sense since (5) holds.

The construction of the compactly supported wavelets by Daubechies [6], using the multiresolution analysis, requires that there should exist a (low-pass) filter $H$ such that

$$\begin{aligned} \phi_{0,0}(t) &= \sum_{i=0}^{2M-1} h(i)\phi_{-1,i}(t) \\ &= \sum_{i=0}^{2M-1} \langle \phi_{0,0}, \phi_{-1,i} \rangle \, \phi_{-1,i}(t). \end{aligned} \tag{9}$$

The (high-pass) filter $G$, which is uniquely determined by $H$, defines the wavelet by

$$\begin{aligned} \psi_{0,0}(t) &= \sum_{i=2(1-M)}^{1} (-1)^i h(1-i)\phi_{-1,i}(t) \\ &= \sum_{i=2(1-M)}^{1} g(i)\phi_{-1,i}(t) \\ &= \sum_{i=2(1-M)}^{1} \langle \psi_{0,0}, \phi_{-1,i} \rangle \phi_{-1,i}(t). \end{aligned} \tag{10}$$

The pair of filters $H$ and $G$ is a set of quadrature mirror filters (QMF). Since (6) holds, the dilated and translated wavelet $\psi_{j,n}(t) \in O_j$ can be projected on $V_k$ for $j - k > 0$. The use of (9) and (10) leads to

$$\begin{aligned} \psi_{j,n}(t) &= (g * h * \cdots * h)\phi_{k,i+2^{j-k}n}(t) \\ &= \sum_{i=-(2^{j-k}(M-1))}^{2^{j-k}M-(2M-1)} \langle \psi_{j,n}, \phi_{k,i+2^{j-k}n} \rangle \phi_{k,i+2^{j-k}n}(t) \\ &= \sum_{i=-2^{j-k}M}^{2^{j-k}M} \langle \psi_{j,0}, \phi_{k,i} \rangle \phi_{k,i+2^{j-k}n}(t) \\ &= \sum_{i=-2^{j-k}M}^{2^{j-k}M} c_{j-k}(i)\phi_{k,i+2^{j-k}n}(t) \end{aligned} \tag{11}$$

where $*$ denotes a convolution, and the $c_{j-k}(i)$'s are the projection coefficients of $\psi_{j,0}$ on $\phi_{k,i}$ defined by the QMF pair $H$ and $G$. Due to the orthonormality of the functions involved, the following relation holds:

$$\begin{cases} |c_{j-k}(i)| \le 1 \\ \sum_{i=-2^{j-k}M}^{2^{j-k}M} |c_{j-k}(i)|^2 = 1. \end{cases} \tag{12}$$

The sum above is only over a finite number of inner products because the wavelet and the scaling functions are compactly supported. The number of inner products depends on $j-k$ and on the length $2M$ of the support of the wavelet function. It is also easily seen that

$$\begin{aligned}|c_{j-k}(i)| &= |\langle\psi_{j-k,0},\phi_{0,i}\rangle| \\ &= |2^{-\frac{(j-k)}{2}}\int_i^{i+2M}\psi(2^{-(j-k)t})\phi(t-i)dt| \\ &\leq 2^{-\frac{(j-k)}{2}}2M||\psi||_\infty||\phi||_\infty. \end{aligned} \tag{13}$$

It can be shown that the projection coefficients of a function $f$ in $V_{k-1}$ and $O_{k-1}$ can be obtained by applying the two filters $H$ and $G$ on the set of projection coefficients of $f$ in $V_k$ and then downsampling by a factor of 2. The low-pass filter $H$ is used to obtain the coarse coefficients for $V_{k-1}$, and the high-pass filter $G$ is used to obtain the detail coefficients for $O_{k-1}$.

Let us now consider the wavelet transformation of stochastic processes. We define the discrete wavelet transform coefficient $d(j,n)$ of a second-order stochastic process $X(t), t\in[0,T]$ as the projection of the process on a dilated and translated wavelet:

$$d(j,n)=\int_{2^j(n-M)}^{2^j(n+M)}X(t)\psi_{j,n}(t)dt. \tag{14}$$

Then, $d(j,n)$ is a random variable. A sample $d_\omega(j,n)$ is the wavelet coefficient at $(j,n)$ of a sample trajectory $X_\omega(t)$, $t\in[0,T]$. We assume that $T$ is sufficiently large such that edge effects do not occur for the wavelet coefficients we consider. Reconstruction of the process from its discrete wavelet coefficients is understood in the $L^2$ sense.

The scaling coefficient $e(j,n)$ of the process at scale $j$ and time $n$ is defined by

$$e(j,n)=\int_{2^jn}^{2^j(n+2M)}X(t)\phi_{j,n}(t)dt. \tag{15}$$

Instead of the process $X(t)$, we now consider the process $d(j,n)$. We are interested to know if the wavelet coefficients can give a more efficient representation of the process than the representation in time. Without loss of generality, we can assume $X(t)$ to have zero-mean and only consider the correlation function of the process (i.e., an $L^2$ approach). An efficient representation of the process would then be obtained if many pairs of the wavelet coefficients of a process are (almost) uncorrelated with each other. This would mean that we can process a wavelet coefficient independently of many other wavelet coefficients. These ideas are further investigated in the following sections.

## III. Correlation of the Wavelet Coefficients

Let us consider the discrete wavelet coefficients of a stochastic process $X(t)$. In general, it is not true that the set of all discrete wavelet coefficients form a set of uncorrelated random variables, that is, the discrete wavelet transform of a second-order process is not equal to the Karhunen-Loéve transform of the process. To investigate the use of the wavelet transform, we consider the correlation between $d(j,n)$ and $d(k,m)$. We are interested to know if this correlation is only significant for a small number of pairs of coefficients.

We can compute, for all compactly supported wavelets and stochastic processes $X(t)$, the correlation between $d(j,n)$ and $d(k,m)$, where $j>k$. By the use of (11) and Fubini's theorem, we obtain

$$\begin{aligned}E[d(k,m)d(j,n)] &= \int_{2^k(m-M)}^{2^k(m+N)}\psi_{k,m}(s)\int_{2^j(n-M)}^{2^j(n+M)} \\ &\quad\times E[X(s)X(t)]\psi_{j,n}(t)dtds \\ &= \sum_{i=-2^{j-k}M}^{2^{j-k}M}c_{j-k}(i)\cdot\int_{2^k(m-M)}^{2^k(m+N)}\psi_{k,m}(s). \\ &\quad\times\int_{2^j(n-M)}^{2^j(n+M)}E[X(s)X(t)]\phi_{k,i+2^{j-k}n}(t)dsdt.\end{aligned} \tag{16}$$

Note that the length of effective support of the second integral is bounded by $2^k2M$ since

$$\begin{aligned}&\min(2^j(n+M),2^k(i+2^{j-k}n+2M))- \\ &\max(2^j(n-M),2^k(i+2^{j-k}n))<2^k2M\end{aligned} \tag{17}$$

and therefore, the maximum length of the second integral is only dependent on the fine scale $k$.

If we want to estimate a wavelet coefficient from wavelet coefficients at coarser scales, it would be convenient if we find that the correlation between pairs of coefficients decay quickly when the fine scale is constant and the coarse scale is increasing. This would make it feasible to predict a wavelet coefficient using only a few wavelet coefficients on the same or coarser scales.

Equation (16) shows the scale dependency for the correlation function. Note that we split up the coarse interval defined by $j$ into a set of smaller intervals. This allows us to find an upper bound for the correlation function on each small interval. We use this idea to obtain new results on the decorrelation of certain processes using the discrete wavelet transform.

A stationary structure in the correlation function of the discrete wavelet coefficients requires a notion of distance between two points (denoting two wavelet coefficients) in the dyadic tree. This distance can be either defined as the length of the shortest path from one point to another or as a function of time—and scale difference between two points. We do not define stationarity in the tree here as the general equation (16) does not indicate if and how it arises in the discrete wavelet domain. However, for special processes (such as fractional Brownian motion; see Flandrin [9]), a form of stationarity holds.

We now state some results related to particular processes. We first consider a stationary Gauss–Markov process, and then, we refer to some new results related to the decorrelation of fractional Brownian motion. Then, we give examples to show how two stationary AR processes are decorrelated using the discrete wavelet transform.

Let us consider a continuous time stationary Gauss–Markov process $X(t)$ having the correlation function $R(\tau) = e^{-\alpha_0 \tau}$, $\alpha_0 > 0$, that is, the stochastic process obeys the following differential equation:

$$dX(t) = -\frac{1}{2}\alpha_0 X(t)dt + dW(t) \tag{18}$$

with $W(t)$ standard Brownian motion. Although the class of continuous time stationary processes is much larger than the one we consider, the results of this section illustrate how processes can be decorrelated with the discrete wavelet transform.

*Theorem 3.1:* Let $X(t)$ be a stationary Gauss–Markov process with correlation function $R(\tau) = e^{-\alpha_0 \tau}$, $\alpha_0 > 0$. The correlation of the discrete wavelet coefficients $d(j,n)$ and $d(k,m)$ decay exponentially quickly across scales and along time. More precisely, for $j - k > 0$

$$|E[d(j,n)d(k,m)]| \leq \begin{cases} C(M,k)2^{-\frac{(j-k)}{2}} e^{-\alpha_0(|2^j n - 2^k m| - (2^k M + 2^j M + 2^k))} \\ \qquad \text{for } 2^j n - 2^k m - M(2^k + 2^j) \geq 0 \\ C(M,k)2^{-\frac{(j-k)}{2}} e^{-\alpha_0(|2^j n - 2^k m| - (3.2^k M + 2^j M + 2^k))} \\ \qquad \text{for } 2^j n - 2^k m + M(3.2^k + 2^j) \leq 0 \\ D(M,k)2^{-\frac{(j-k)}{2}} \\ \text{otherwise,} \end{cases} \tag{19}$$

where

$$\begin{cases} C(M,k) & = & 2M\|\psi\|_\infty \|\phi\|_\infty \frac{1}{2^k \alpha_0^2} \\ D(M,k) & = & 2M\|\psi\|_\infty \|\phi\|_\infty 2^k \\ & & \times \left[(4M)^2 + \frac{4}{\alpha_0 2^k}\left(\frac{1}{\alpha_0 2^k} + 1\right)\right] \end{cases}. \tag{20}$$

*Proof:* See Appendix A. The proof exploits (12).

Assume we have a wavelet coefficient $d(k,m)$, and we consider its dependence on wavelet coefficients $d(j,n)$'s on coarser scales ($j - k > 0$). Then, Theorem 3.1 states that for constant $k$ and $m$, the correlation decreases exponentially across scales and time. A better indication of the decorrelative properties can be obtained by working with the coefficient of variation (or normalized correlation).

In [5], we have studied the decorrelation of fractional Brownian motion (fBm), a nonstationary zero-mean Gaussian process $X(t)$, $0 < t < T$, whose correlation function is given by

$$E[X(t)X(s)] = \frac{\sigma^2}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}) \tag{21}$$

with $0 < H < 1$. We showed in [5], with a similar proof as the one for Theorem 3.1 and using the vanishing moments of the wavelet function, that the normalized correlation between the discrete wavelet coefficients of fBm decreases exponentially quickly across scales and hyperbolically quickly along time. This tells us that although the original time process has long-term dependencies, we can define simple signal processing algorithms on the set of almost uncorrelated wavelet coefficients of the process. We also showed there that in the case of standard Brownian motion (or in general for any orthogonal increments process), the correlation between discrete wavelet coefficients is equal to zero if the supports of the two analyzing wavelets are nonoverlapping.

We have studied here a class of stationary processes to illustrate the fact that both stationary and nonstationary processes can be well decorrelated by the discrete wavelet transform. Although a specific transform as the Fourier transform is very suitable to decorrelate stationary processes, the discrete wavelet transform seems to capture a wider range of processes and could be a good transform to decorrelate a signal without *a priori* knowledge of its statistical characteristics.

We consider two examples of the decorrelation of the discrete wavelet transform. For more examples, we refer to the work by Golden [11]. We consider here two discrete time stochastic processes from which we take 128 samples. Each discrete-time random variable can be seen as the scaling coefficient of a continuous-time process on a certain scale. The discrete wavelet decomposition of this set of coefficients can then be obtained using the set of QMF filters $H$ and $G$ (see [15]). The original random process (vector) is a zero-mean stationary AR(1) process:

$$X(n) = A_1 X(n-1) + W(n) \tag{22}$$

with $W(n)$ i.i.d. Gaussian noise with zero-mean and variance equal to 1 and where we have chosen $A_1 = 0.98$. Its correlation matrix is displayed in Fig. 1. Pixel $(i,j)$ in the image represents the correlation between $X(i)$ and $X(j)$. The higher the absolute value of the correlation between two samples, the whiter the pixel in the image. Fig. 2 represents the correlation matrix of a stationary AR(2) process:

$$X(n) = B_1 X(n-1) + B_2 X(n-2) + W(n) \tag{23}$$

where we have chosen $B_1 = -0.10$ and $B_2 = 0.80$. In Fig. 2, we have set the correlations that are less than zero to be equal to zero instead of plotting their absolute value in order to visualize the high frequencies of the correlation function. However, we should keep in mind that a wide band of matrix elements besides the diagonal matrix is nonzero. We clearly see the stationary structure of the AR processes.

We use a Daubechies wavelet with three vanishing moments, and we assume the process to be periodic (with period 128) for the transformation, which causes edge effects in the transformed correlation matrix. We use a discrete wavelet transform with four levels for these figures. The set of transformed coefficients is ordered with the coefficients in the lowest frequency band first, then the coefficients in the neighboring frequency band, and so on, until the band of highest frequencies is reached. We display the absolute value of the normalized correlations. In both in Figs. 3 and 4, we see the decorrelation of the wavelet transform; the correlation matrix of the transformed coefficients is "more diagonal" than the original correlation matrix, that is, the number of off-diagonal elements of the new correlation matrix, which are nonzero, is far less than is the case for the original correlation matrix. However, especially for the AR(1) process, there still exists correlation between coefficients with the same time

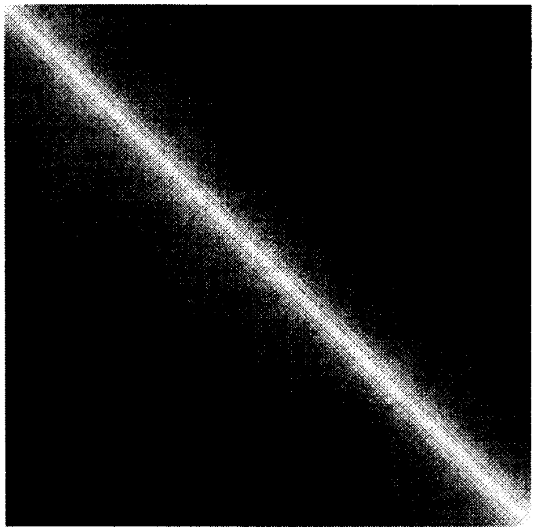


Fig. 1. Correlation matrix (128 × 128) of AR(1) process with coefficient 0.98.

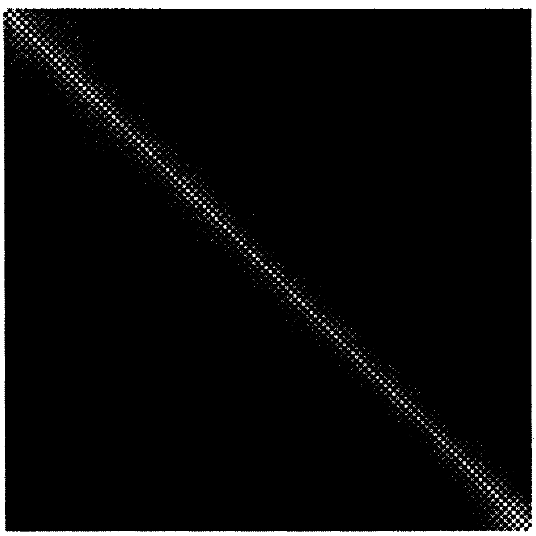


Fig. 2. Correlation matrix (128 × 128) of AR(2) process with coefficients −0.10 and 0.80.

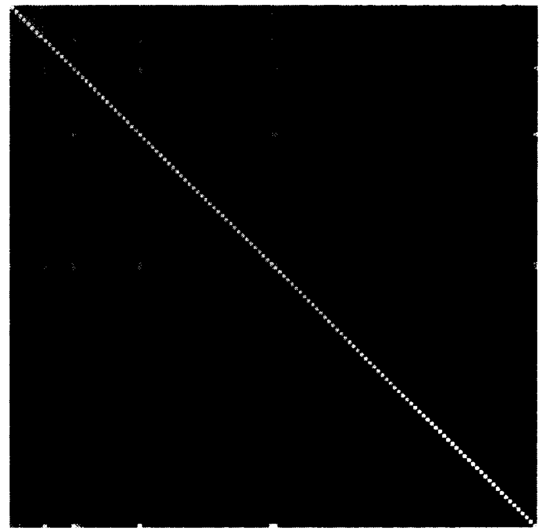


Fig. 3. Normalized transformed correlation matrix (128 × 128) of AR(1) process with four levels and three vanishing moments.

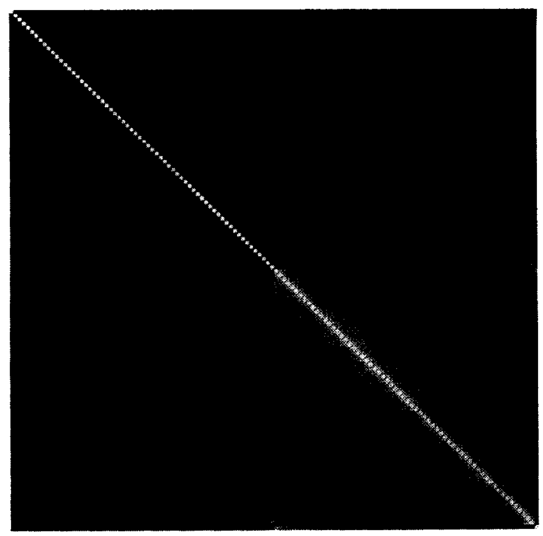


Fig. 4. Normalized transformed correlation matrix (128 × 128) of AR(2) process with four levels and three vanishing moments.

localizations but in different frequency bands. This correlation decreases with the difference between scales $(j - k)$. The correlation off the diagonal in the high-frequency band of the AR(2) process is also clearly visible.

In the next section, we consider the modeling of the processes in the transform domain and discuss whether or not we have to take the correlation between different wavelet coefficients into account.

## IV. Multiresolution Stochastic Models

In the preceding section, we have indicated that the discrete wavelet transform of two classes of stochastic processes are almost uncorrelated. We have shown that the correlation between the discrete wavelet coefficients of these processes decay exponentially quickly across scales and either hyperbolic or exponentially quickly along time. These results motivate the study of good and efficient approximate stochastic models defined in the discrete wavelet domain.

We can approximate the original time-processes by considering them to have uncorrelated discrete wavelet coefficients (as Wornell did in [22]), or we can define certain processes on the discrete wavelet coefficients. Due to the strong decay in correlation, these processes could be simple. We shall consider both approaches and discuss the quality of approximation of the original process as well as the use of these models in particular applications.

The approximate models we propose are motivated by the multiresolution stochastic models introduced by Basseville *et al.* (see for example [1]). They define stochastic processes on $n$-ary trees as models of phenomena available or appearing on different resolutions. Each state of the model is identified with a characteristic of the phenomenon at a particular resolution located around a certain time or space. We restrict ourselves here to binary trees, where each node of the tree is connected to one 'parent'-node, where its value represents the signal's feature on a coarser resolution around the same time (or space), and two 'children'-nodes, whose values represent the signal's feature on a finer resolution around the same time (space).

Basseville *et al.* [1] defined stationarity of the processes by introducing the notion of distance between nodes on the tree and developed efficient representations of these processes. Moreover, nonstationary auto-regressive models were defined, where the value at a node $(j, n)$ (at scale $j$ and time index $n$) is a linear combination of the value of the parent node $(j+1, \frac{n}{2})$ and white noise:

$$X(j,n) = A_{j,n}X\left(j+1, \left[\frac{n}{2}\right]\right) + B_{j,n}W(j,n). \tag{24}$$

The stochastic process $X(j,n), j, n \in Z$ is defined on the tree (indexed as the discrete wavelet coefficients by scale and time), and $[z]$ denotes the integer part of $z$. The stochastic variable $W(j,n)$ describes the innovation process at node $(j,n)$. Note that the variables $X(j,n), W(j,n)$ can be vectors of some desired dimension and $A_{j,n}, B_{j,n}$ appropriate matrices. Observe that our notation implies that for $n$ even, the two nodes $(j,n)$ and $(j, n+1)$ share the same parent node $(j+1, \frac{n}{2})$. In [1] and [14], estimation algorithms for these models are discussed and used. Here, we are interested in the approximation of stochastic signals by the use of similar nonstationary auto-regressive models.

A state in the multiresolution stochastic model can be identified with a scaling coefficient of the process and would give a multiresolution description of the original process. We could also consider a multiresolution stochastic model, where each state of the model is identified with a discrete wavelet coefficient of the process. We can still call this model a multiresolution stochastic state model, but now, states do not describe coarse or fine 'measurements' but difference information between two succeeding resolutions. Let us discuss these two approaches in more detail.

Assume we have a process with uncorrelated discrete wavelet coefficients $d(j,n), j,n \in Z$. The reconstruction formula for the scaling coefficients (see [15]) is itself a multiresolution stochastic model:

$$e(j,n) = \sum_{k=-\infty}^{\infty} h(n-2k)e(j+1,k) + \sum_{k=-\infty}^{\infty} g(n-2k)d(j+1,k) \tag{25}$$

where the summation is only over a finite number of terms due to the filters $H$ and $G$.

The length of the QMF filters (and therefore the length of the wavelet function) defines the complexity of the rewritten model. Golden [11] approximated the original process using the assumption of uncorrelated discrete wavelet coefficients and discussed the quality of approximation of the original process in the case of an AR(1) process and standard Brownian motion. It turned out that the AR(1) process was reasonably well approximated, but the standard Brownian motion was poorly approximated. In several examples we shall compare this approximation to the following model we propose.

Dropping the assumption of uncorrelated wavelet coefficients, we could directly define a multiresolution model on the scaling or the wavelet coefficients. We discuss this for the case of the wavelet coefficients, but in the following formulas, the $d(j,n)$'s could be replaced by $e(j,n)$'s. We propose to model the original stochastic signal by

$$\begin{aligned} d(j,n) = &\sum_{k_1=1}^{L_1} a_{k_1}(j,n)d(j, n-k_1) \\ &+ \sum_{k_2=-L_2}^{L_3} \tilde{a}_{k_2}(j,n)d\left(j+1, \left[\frac{n}{2}\right] - k_2\right) \\ &+ b(j,n)w(j,n) \end{aligned} \tag{26}$$

where $w(l,n)$ is i.i.d. Gaussian noise with variance equal to 1. The model (26) can be seen as an generalization of model (24). The assumption of uncorrelated discrete wavelet coefficients can be seen as a special case of (26).

The complexity of the rewritten model is now dependent on the choice of the values for $L_1, L_2, L_3$. The results that the discrete wavelet coefficients of certain processes are strongly decorrelated motivates the choice of small values for $L_1, L_2, L_3$. Note also that already as simplification, we only consider prediction coefficients on one coarser resolution in (26). The model for the discrete wavelet coefficients, which are defined by (26), incorporates possible correlation between different discrete wavelet coefficients. We know therefore that the original process is likely to be better approximated by (26) than by assuming that the wavelet coefficients are uncorrelated. We investigate, for some examples in the next section, if this improvement is significant. The use of (26) provides robustness in case the wavelet coefficients 'close' to each other are significantly correlated while it still defines a simple multiresolution stochastic process.

Note that such a model as (26) for the scaling coefficients instead of the wavelet coefficients does not necessarily imply an improvement compared with (24). Then, we associate with each scaling coefficient $e(j,n)$ an innovation coefficient $w(j,n)$. The set of equations for the scaling coefficients is very redundant (since scaling coefficients on a coarser scale are linear combinations of scaling coefficients on a finer scale), and therefore, the assumption of innovation at each scaling coefficient might be too strong a demand. For these reasons, we restrict ourselves here to the study of the approximation by use of (26) with small values of $L_1, L_2, L_3$.

Let us further discuss the model (26). The parameters $a_{k_1}(j,n)$, $\tilde{a}_{k_2}(j,n)$ of the models can be chosen using standard results of minimum error variance estimation theory for Gaussian processes (we do not consider higher order moments and might as well assume that all processes we study are Gaussian). We project $d(j,n)$ on the space spanned by all other $d(j, n-k_1)$'s and $d(j+1, [\frac{n}{2}] - k_2)$'s in the model equation, and we define the $a_{k_1}(j,n)$'s and $\tilde{a}_{k_2}(j,n)$'s to be equal to these projection coefficients:

$$\begin{aligned} E\big[d(j,n) \,|\, d(j,n-1),\ldots,d(j,n-L_1),\\ d\Big(j+1,\Big[\frac{n}{2}\Big]+L_2\Big),\ldots,d\Big(j+1,\Big[\frac{n}{2}\Big]-L_3\Big)\big]\\ = \sum_{k_1=1}^{L_1} a_{k_1}(j,n)d(j,n-k_1)\\ + \sum_{k_2=-L_2}^{L_3} \tilde{a}_{k_2}(j,n)d\Big(j+1,\Big[\frac{n}{2}\Big]-k_2\Big) \end{aligned} \tag{27}$$

for which we use the following result for two jointly Gaussian random vectors $Y$ and $Z$:

$$E[Y \mid Z] = E[YZ^T](E[ZZ^T])^{-1}Z. \tag{28}$$

After we have obtained the $a_{k_1}(j,n)$'s and $\tilde{a}_{k_2}(j,n)$'s, we can compute $b_{j,n}$ by imposing that the variance of $d(j,n)$ in the model should be equal to the real variance of the transform coefficient of the process.

We compare the approximation by inspecting the original correlation matrix and the correlation matrix of the approximate model of a finite set of data in the time domain (we work with a discrete time sequence now, as in the example of the previous section). The original correlation matrix is already available. The computation of the new correlation matrix requires the recursive computation of different correlation functions as we shall indicate here and discuss in more detail in Appendix B.

Assume that we have obtained the correlation matrix of the transform coefficients of a finite set of a time sequence, as in the examples at the end of the previous section. We have also obtained the model parameters of (26) using the estimation theory. Now, we have to compute the correlation of the time sequence of such a multiresolution stochastic model. Let us assume we know, for some scale $j+1$, the following correlation functions:

$$\begin{cases} E[e(j+1,n_1)e(j+1,m_1)]\\ E[d(j+1,n_1)e(j+1,m_1)]\\ E[d(j+1,n_1)d(j+1,m_1)] \end{cases} \tag{29}$$

for all $n_1$ and $m_1$ of interest (being in an appropriate subset of the integers). In practice, we start on the coarsest scale and assume the correlation functions above to be exactly equal to the correlations of the wavelet transform of the original process, that is, we do not change the upper corner of the correlation matrix of the transform coefficients. In Appendix B, we show, by use of (25), (26), and (29), that we can compute the correlation functions on the finer scale $j$:

$$\begin{cases} E[e(j,n_2)e(j,m_2)]\\ E[d(j,n_2)e(j,m_2)]\\ E[d(j,n_2)d(j,m_2)] \end{cases} \tag{30}$$

for all $n_2$ and $m_2$ of interest. Applying this computation sufficiently often, we obtain the correlation matrix of the finest scaling coefficients. This is then the new correlation matrix (in the time domain) of the approximate multiresolution stochastic model.

The approximation of the original process can be measured by the following two approaches, as used by Golden [11]. The first method is the visual inspection of the difference between the two normalized correlation matrices that are being displayed as images. A second method is the use of the so-called Bhattacharyya distance [11] between two random vectors. This metric is used to measure how close one random process is to another. In our case of zero-mean random vectors, which are represented by their correlation matrices, say, $R_1$ and $R_2$, the Bhattacharyya distance $B_d(R_1,R_2)$ reduces to

$$B_d(R_1,R_2) = \frac{1}{2}\ln\left[\left|\frac{1}{2}(R_1+R_2)\right|\right] - \frac{1}{4}\ln\left[|(R_1R_2)|\right] \tag{31}$$

where $|A|$ is the determinant of the matrix $A$. Let us assume we have a sample of one of the two random processes (vectors), where the probability is $\frac{1}{2}$ that the sample was taken from either the original random process or the approximate process. Then, the probability of error of deciding from which process the sample was is bounded above by the following quantity:

$$P(\text{error}) \le \frac{1}{2}e^{-B_d(R_1,R_2)}. \tag{32}$$

We shall investigate in the following section the quality of these multiresolution approximate stochastic models for different example processes using these measures. Note that it can be verified that $B_d(R_1,R_1) = 0$, which leads to $P(\text{error}) \le \frac{1}{2}$.

There are several applications for the use of the models we discussed above. We can think of building stochastic models for image analysis, with possibilities in the areas of predictive/transform coding of images, image restoration, or feature extraction in images, as well as fast and adequate sampling methods of processes. The models we have discussed require the availability of the discrete wavelet coefficients of a process (or at least some noisy observation of it). This means that we should have a set of (possibly noisy) scaling coefficients at a fine scale on which we can do the discrete wavelet transform. This is the case in many image processing problems, where we have an image on which we want to do multiresolution signal processing using the discrete wavelet transform. However, in the case of fusion of measurements at different resolutions from different devices, we might prefer to work directly with a model on the multiresolution measurements (which are similar to noisy scaling coefficients of a process) since it is difficult to perform a wavelet transform on an incomplete set of multiresolution measurements.

## V. Examples

In this section, we construct approximate multiresolution stochastic models for several example processes and discuss the quality of approximation of the original process.

We consider again the AR(1) (with coefficient 0.98) and AR(2) processes (with coefficients $-0.10$ and 0.80), whose correlation matrices together with their normalized transformed correlation matrices were discussed in Section III. We also consider two fractional Brownian motion processes, where the value of $\sigma^2$ is equal to 1 and the value of $H$ is 0.50 (standard Brownian motion) and 0.75. We display here the

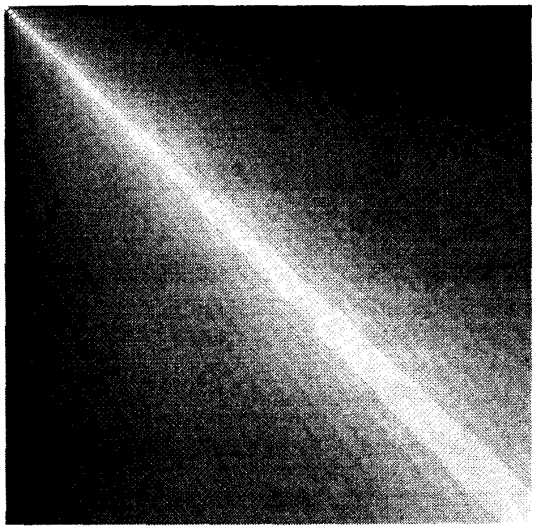


Fig. 5. Normalized correlation matrix (128 × 128) of standard Brownian motion (H = 0.50).

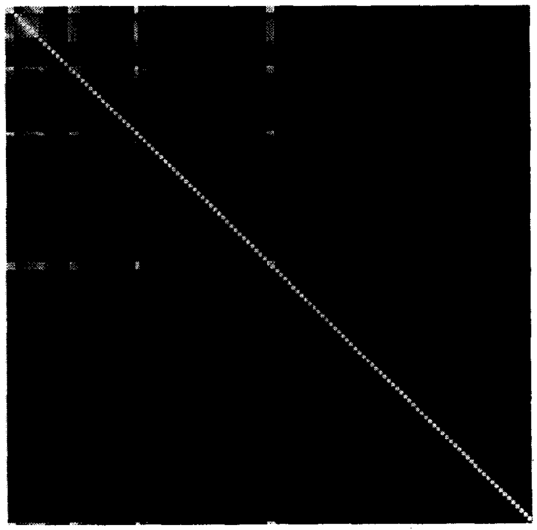


Fig. 7. Normalized transformed correlation matrix (128 × 128) of standard Brownian motion: Four levels and three vanishing moments.

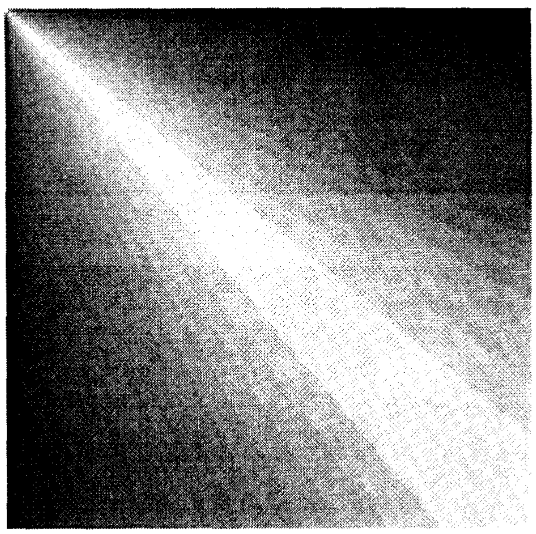


Fig. 6. Normalized correlation matrix (128 × 128) of fBm with H = 0.75.

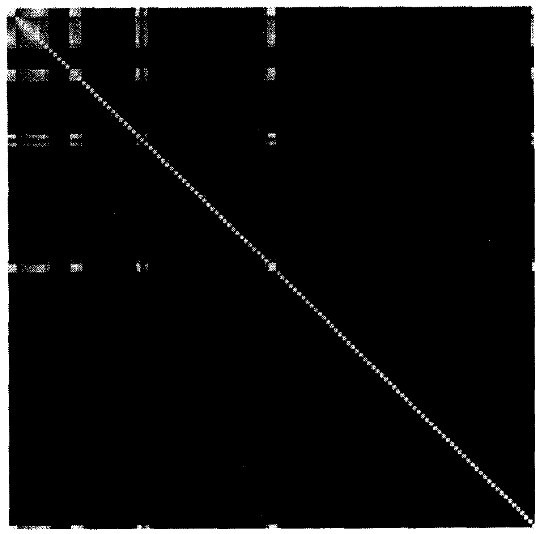


Fig. 8. Normalized transformed correlation matrix (128× 128) of fBm with H = 0.75: Four levels and three vanishing moments.

normalized correlation matrices of the last two processes and their transformed normalized correlation matrices. Again, we only plot the absolute value of the normalized correlation.

The correlation matrices in Figs. 5 and 6 show the long-term nonstationary dependencies of samples. These processes therefore differ significantly from the stationary AR processes we are studying. In both Figs. 7 and 8, we see again the decorrelation of the wavelet transform.

Now let us define multiresolution stochastic processes and quantify its approximation to an original process. We apply for each of our processes two wavelet transformations. The first one is the Haar transform, which has only one vanishing moment. For the second transform, we use a Daubechies wavelet, which has three vanishing moments.

We define four different approximate models on the wavelet coefficients for each of the processes as follows:

$$d(j,n) \perp d(k,m) \quad \text{for } (j,n) \neq (k,m), \tag{33}$$

$$d(j,n) = a(j,n)d(j,n-1) + b(j,n)w(j,n), \tag{34}$$

$$d(j,n) = \tilde{a}(j,n)d\Big(j+1, \Big[\frac{n}{2}\Big]\Big) + b(j,n)w(j,n) \tag{35}$$

and

$$d(j,n) = a(j,n)d(j,n-1) + \tilde{a}(j,n)d\Big(j+1, \Big[\frac{n}{2}\Big]\Big) + b(j,n)w(j,n). \tag{36}$$

Due to the stationary properties in the discrete wavelet domain of fractional Brownian motion and stationary processes,

TABLE I
QUALITY OF APPROXIMATE MULTIRESOLUTION MODELS FOR THE AR PROCESSES

| AR(1) | wavelet=1 | | wavelet=3 | |
|---|---|---|---|---|
| model | $B_d(R_1, R_2)$ | $\frac{1}{2}e^{-B_d(R_1,R_2)}$ | $B_d(R_1, R_2)$ | $\frac{1}{2}e^{-B_d(R_1,R_2)}$ |
| (33) | 8.8937 | $7.10^{-5}$ | 3.4653 | 0.0156 |
| (34) | 8.8017 | $8.10^{-5}$ | 1.9810 | 0.0690 |
| (35) | 5.0385 | 0.0032 | 2.4690 | 0.0423 |
| (36) | 4.3202 | 0.0066 | 1.5648 | 0.1046 |
| AR(2) | wavelet=1 | | wavelet=3 | |
| model | $B_d(R_1, R_2)$ | $\frac{1}{2}e^{-B_d(R_1,R_2)}$ | $B_d(R_1, R_2)$ | $\frac{1}{2}e^{-B_d(R_1,R_2)}$ |
| (33) | 17.0654 | $2.10^{-8}$ | 16.0557 | $5.10^{-8}$ |
| (34) | 1.8631 | 0.0776 | 1.0775 | 0.1702 |
| (35) | 16.5498 | $3.10^{-8}$ | 15.9814 | $6.10^{-8}$ |
| (36) | 1.2729 | 0.1400 | 0.9817 | 0.1873 |

TABLE II
QUALITY OF APPROXIMATE MULTIRESOLUTION MODELS FOR THE fBm's

| fBm, $H = 0.50$ | wavelet=1 | | wavelet=3 | |
|---|---|---|---|---|
| model | $B_d(R_1, R_2)$ | $\frac{1}{2}e^{-B_d(R_1,R_2)}$ | $B_d(R_1, R_2)$ | $\frac{1}{2}e^{-B_d(R_1,R_2)}$ |
| (33) | 9.9390 | $2.10^{-5}$ | 5.2017 | 0.0028 |
| (34) | 9.9390 | $2.10^{-5}$ | 2.8735 | 0.0283 |
| (35) | 5.4023 | 0.0023 | 3.1803 | 0.0208 |
| (36) | 4.7154 | 0.0045 | 2.2674 | 0.0518 |
| fBm, $H = 0.75$ | wavelet=1 | | wavelet=3 | |
| model | $B_d(R_1, R_2)$ | $\frac{1}{2}e^{-B_d(R_1,R_2)}$ | $B_d(R_1, R_2)$ | $\frac{1}{2}e^{-B_d(R_1,R_2)}$ |
| (33) | 24.4461 | $1.10^{-11}$ | 12.7440 | $1.10^{-6}$ |
| (34) | 24.6862 | $1.10^{-11}$ | 6.5792 | 0.0007 |
| (35) | 11.1640 | $7.10^{-6}$ | 6.9027 | 0.0005 |
| (36) | 10.4587 | $1.10^{-5}$ | 5.3621 | 0.0023 |

the model parameters for our examples are only dependent on $j$ and the fact if $n$ is even or odd. By use of (28), we obtain the model parameters, and by use of the procedure described in Appendix B, we can compute the correlation matrix (in the time domain) of the approximate models. In order to have a fair comparison between the approximations, we used a discrete wavelet transform with five levels for our simulations.

We show here for the AR(2) process and standard Brownian motion the normalized correlation matrices (in the time domain) of two approximate multiresolution models. We used a Daubechies wavelet having three vanishing moments. The results of the model (33) (uncorrelated discrete wavelet coefficients) and model (36) are shown in Figs. 9–12. Figs. 9 and 10 approximate Fig. 2, and Figs. 11 and 12 approximate Fig. 5. We see clearly the improvement of the approximation using model (36) over (33). In case of the Brownian motion, we see that both Figs. 11 and 12 suffer from the edge effects, which are not cancelled in the approximation. Although a wavelet with many vanishing moments is favorable for decorrelation of the process, the edge effects in the approximation are more apparent due to the longer length of the QMF filters.

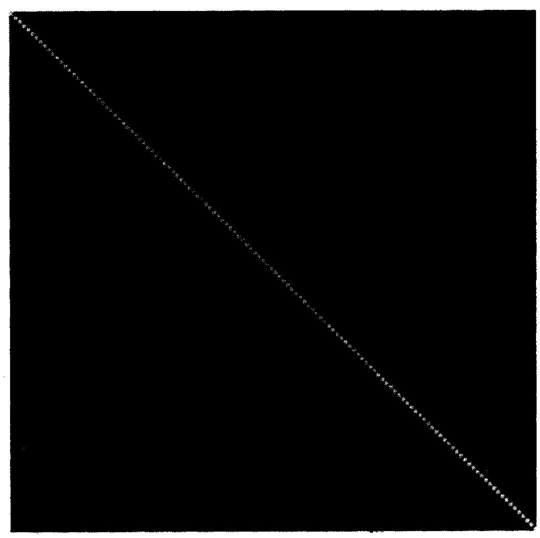


Fig. 9. Normalized correlation matrix (128 × 128) (in time domain) of approximation of AR(2) process using model (33).

We show in Tables I and II the Bhattacharyya distance between each of the approximate multiresolution models and the original processes. We denote the two different correlation matrices here by $R_1$ and $R_2$. We also show the upper bound on the probability of error of deciding from which process (the approximation or the original) a sample would be. In the tables, the following are indicated: the process, which wavelet (one or three vanishing moments) is used, and which model ((33), (34), (35), or (36)) is used.

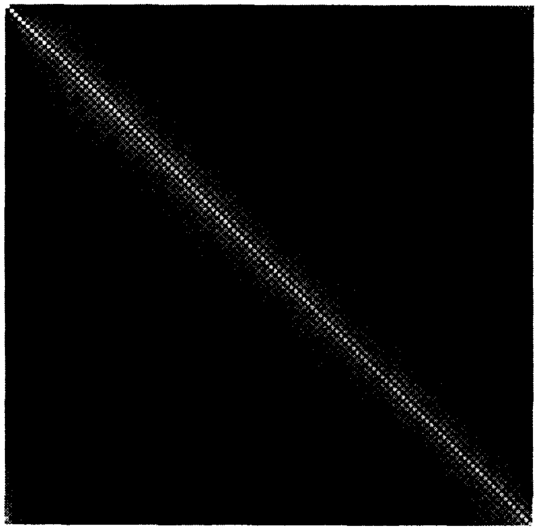


Fig. 10. Normalized correlation matrix (128 × 128) (in time domain) of approximation of AR(2) process using model (36).

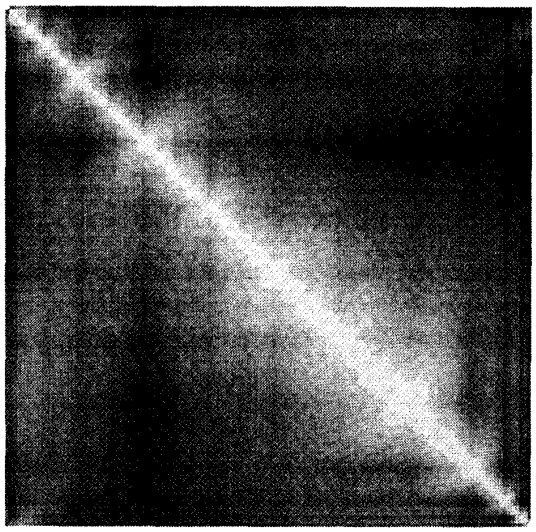


Fig. 11. Normalized correlation matrix (128 × 128) (in time domain) of approximation of standard Brownian motion using model (33).

In Tables I and II, we see that for all processes model (36) shows an improvement in approximation of the process compared with model (33). Note that models (34) and (35) alone do not improve the approximation in some cases and improve significantly in others. Note that for fBm with $H = 0.75$ with the Haar wavelet, model (34) even worsens the approximation slightly, indicating that some first-order approximations of the process might not model the overall process well. A safe procedure is, therefore, to use model (36) in all instances. The question of sufficiency of approximation using model (36) is not clear and depends on the applications, but we clearly see a marked improvement in approximation by introducing a slightly more complicated model. Note that the approximations are better for the wavelet with three vanishing moments, except

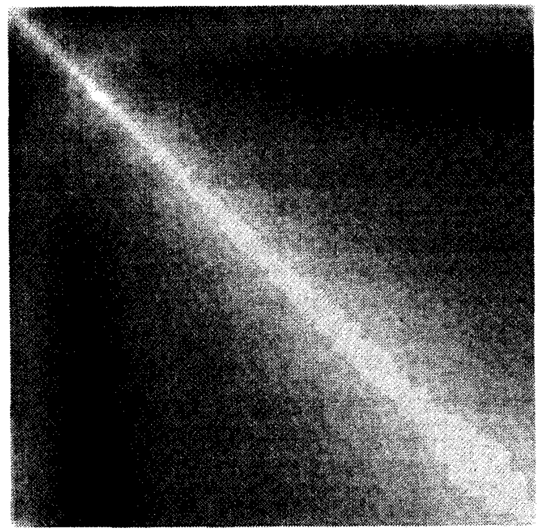


Fig. 12. Normalized correlation matrix (128 × 128) (in time domain) of approximation of standard Brownian motion using model (36).

for the AR(2) process. We can therefore conclude that the use of a wavelet with several vanishing moments, combined with a simple auto-regressive model as (36) defined on the discrete wavelet coefficients, approximates many different stochastic signals well.

## VI. Conclusion

We have studied the correlation structure of discrete wavelet coefficients of stochastic processes and have indicated how the correlation along time and across scales decreases exponentially (or hyperbolically) quickly for a class of stationary processes and fractional Brownian motion.

The study of the correlation structure of the discrete wavelet coefficients of a stochastic process enables us to construct approximate multiresolution auto-regressive models. We have discussed the approximation of stochastic processes in the time-scale domain, using multiresolution stochastic state models similar to those introduced by Basseville *et al.* [1]. We motivated models where the states of the model are identified with the discrete wavelet coefficients of the time process.

We illustrated our approach by the construction of simple approximate models for several different time processes and showed that these models significantly improve the approximation of the original process compared with the assumption of having uncorrelated discrete wavelet coefficients.

## Appendix A

*Proof of Theorem 3.1:*

Let us define $\alpha = 2^{j-k}n - m$. A simple calculation using the change of variables $s = t - \tau$ and $2^{-j}t - n = 2^{-(j-k)}\theta$ leads to

$$E[d_j(n)d_k(m)] = -\int_{-\infty}^{\infty} e^{-\alpha_0|\tau|}\int_{-\infty}^{\infty} 2^{-\frac{(j-k)}{2}}\cdot \psi(2^{-(j-k)}\theta)\psi(\theta + \alpha - 2^{-k}\tau)d\theta d\tau. \quad (37)$$

Using (11), changing the order of integration and summation, and changing variables, we obtain

$$E[d_j(n)d_k(m)] = -\int_{-\infty}^{\infty} e^{-\alpha_0|\tau|} \int_{-\infty}^{\infty} \sum_{i=-2^{j-k}M}^{2^{j-k}M} c_{j-k}(i) \times \phi(\theta - i)\psi(\theta+\alpha-2^{-k}\tau)d\theta\, d\tau = \sum_{i=-M2^{j-k}}^{M2^{j-k}} c_{j-k}(i)\int_{-\infty}^{\infty} e^{-\alpha_0 2^k|(i+\alpha)-\tau|} \cdot \times 2^k\Lambda(\tau)d\tau \tag{38}$$

with

$$\Lambda(\tau) = \int_{-\infty}^{\infty} \phi(\theta)\psi(\theta+\tau)d\theta \tag{39}$$

and

$$c_{j-k}(i) \le 2^{-\frac{(j-k)}{2}} 2M||\psi||_\infty||\phi||_\infty. \tag{40}$$

It is easy to show that the function $\Lambda(\tau)$ is supported on $[-3M, M]$ and $\forall\tau \in [-3M, M]$, $|\Lambda(\tau)| \le 1$. Furthermore, it can be verified ([20]) that this function has $M$ vanishing moments. We do not use this last property in the proof, however.

We have derived that

$$|E[d_j(n)d_k(m)]| \le 2^{-\frac{(j-k)}{2}} 2M||\psi||_\infty||\phi||_\infty 2^k \times \sum_{i=-M2^{j-k}}^{M2^{j-k}} \int_{-3M}^{M} e^{-\alpha_0 2^k|(i+\alpha)-\tau|}d\tau. \tag{41}$$

With $-M2^{j-k} \le i \le M2^{j-k}$, we consider three cases. First, if $\forall i, -M2^{j-k} \le i \le M2^{j-k}$, the relation $i+\alpha \ge M$ holds and leads to

$$\begin{aligned}&\alpha \ge M + M2^{j-k}\\ &2^jn - 2^km - M(2^k+2^j) \ge 0.\end{aligned} \tag{42}$$

Then

$$\int_{-3M}^{M} e^{-\alpha_0 2^k|(i+\alpha)-\tau|}d\tau \le \frac{1}{\alpha_0 2^k} e^{-\alpha_0 2^k(i+\alpha-M)} \tag{43}$$

and

$$\begin{aligned}|E[d_j(n)d_k(m)]| &\le 2^{-\frac{(j-k)}{2}} 2M||\psi||_\infty||\phi||_\infty 2^k \times \sum_{i=-M2^{j-k}}^{M2^{j-k}} \frac{1}{\alpha_0 2^k} e^{-\alpha_0 2^k(i+\alpha-M)}\\ &\le 2^{-\frac{(j-k)}{2}} 2M||\psi||_\infty||\phi||_\infty \frac{1}{\alpha_0} \times \int_{-2^{j-k}M-1}^{\infty} e^{-\alpha_0 2^k(i+\alpha-M)}di\\ &= 2^{-\frac{(j-k)}{2}} 2M||\psi||_\infty||\phi||_\infty \frac{1}{2^k\alpha_0^2} \times e^{-\alpha_0(|2^jn-2^km|-(2^kM+2^jM+2^k))}.\end{aligned} \tag{44}$$

In the second case, if $\forall i, -M2^{j-k} \le i \le M2^{j-k}$, the relation $i+\alpha \le -3M$ holds and leads to

$$\begin{aligned}&\alpha \le -3M - M2^{j-k}\\ &2^jn - 2^km + M(3.2^k+2^j) \le 0.\end{aligned} \tag{45}$$

Then

$$\int_{-3M}^{M} e^{-\alpha_0 2^k|(i+\alpha)-\tau|}d\tau = \int_{-3M}^{M} e^{-\alpha_0 2^k(\tau-(i+\alpha))}d\tau \le \frac{1}{\alpha_0 2^k} e^{-\alpha_0 2^k(-(i+\alpha)-3M)}, \tag{46}$$

and

$$\begin{aligned}|E[d_j(n)d_k(m)]| &\le 2^{-\frac{(j-k)}{2}} 2M||\psi||_\infty||\phi||_\infty 2^k \times \sum_{i=-M2^{j-k}}^{M2^{j-k}} \frac{1}{\alpha_0 2^k} e^{-\alpha_0 2^k(-(i+\alpha)-3M)}\\ &\le 2^{-\frac{(j-k)}{2}} 2M||\psi||_\infty||\phi||_\infty \frac{1}{\alpha_0} \cdot \times \int_{-\infty}^{2^{j-k}M+1} e^{-\alpha_0 2^k(-(i+\alpha)-3M)}di\\ &= 2^{-\frac{(j-k)}{2}} 2M||\psi||_\infty||\phi||_\infty \frac{1}{2^k\alpha_0^2} \cdot \times e^{-\alpha_0(|2^jn-2^km|-(3.2^kM+2^jM+2^k))}.\end{aligned} \tag{47}$$

In the complementary third case, there are at most $4M-1$ $i$'s for which $-3M < i+\alpha < M$. We derive

$$\begin{aligned}|E[d_j(n)d_k(m)]| &\le 2^{-\frac{(j-k)}{2}} 2M||\psi||_\infty||\phi||_\infty 2^k. \times \left[4M.4M + 4\sum_{i=0}^{\infty}\int_0^{2M} e^{-\alpha_0 2^k(i+\tau)}d\tau\right]\\ &\le 2^{-\frac{(j-k)}{2}} 2M||\psi||_\infty||\phi||_\infty 2^k \times \left[(4M)^2 + \frac{4}{\alpha_0 2^k}\sum_{i=0}^{\infty} e^{-\alpha_0 2^k i}\right]\\ &\le 2^{-\frac{(j-k)}{2}} 2M||\psi||_\infty||\phi||_\infty 2^k \times \cdot \left[(4M)^2 + \frac{4}{\alpha_0 2^k}\left(\int_{i=0}^{\infty} e^{-\alpha_0 2^k i}di + 1\right)\right]\\ &\le 2^{-\frac{(j-k)}{2}} 2M||\psi||_\infty||\phi||_\infty 2^k \times \left[(4M)^2 + \frac{4}{\alpha_0 2^k}\left(\frac{1}{\alpha_0 2^k}+1\right)\right].\end{aligned} \tag{48}$$

This completes the proof.

## Appendix B

Let us assume we know, for some scale $j+1$, the following correlation functions:

$$\begin{cases} E[e(j+1,n_1)e(j+1,m_1)]\\ E[d(j+1,n_1)e(j+1,m_1)]\\ E[d(j+1,n_1)d(j+1,m_1)],\end{cases} \tag{49}$$

$$\begin{aligned}E[d(j,m_2)d(j+1,n_1)] &= E\left[\left(\sum_{k_1=1}^{L_1} a_{k_1}(j,m_2)d(j,m_2-k_1)\right).d(j+1,n_1)\right]\\ &+ E\left[\left(\sum_{k_2=-L_2}^{L_3} \tilde{a}_{k_2}(j,m_2)d\left(j+1,\left[\frac{m_2}{2}\right]-k_2\right)\right)\cdot d(j+1,n_1)\right]\\ &+ E[b(j,m_2)w(j,m_2)d(j+1,n_1)]. \end{aligned}\tag{54}$$

for all $n_1$ and $m_1$ of interest. We show now that we can compute the correlation functions on the finer scale $j$:

$$\begin{cases} E[e(j,n_2)e(j,m_2)] \\ E[d(j,n_2)e(j,m_2)] \\ E[d(j,n_2)d(j,m_2)], \end{cases}\tag{50}$$

for all $n_2$ and $m_2$ of interest.

*Proof:* We compute three correlation functions. First of all, by use of (25), we have

$$\begin{aligned}E[e(j,n_2)e(j,m_2)] = E\Bigg[&\left(\sum_{k=-\infty}^{\infty} h(n_2-2k)e(j+1,k)\right.\\ &\left.+\sum_{k=-\infty}^{\infty} g(n_2-2k)d(j+1,k)\right)\\ &\times\left(\sum_{k=-\infty}^{\infty} h(m_2-2k)e(j+1,k)\right.\\ &\left.+\sum_{k=-\infty}^{\infty} g(m_2-2k)d(j+1,k)\right)\Bigg]\end{aligned}\tag{51}$$

which can be explicitly computed using (49).

Second, using (26), we obtain

$$\begin{aligned}&E[d(j,n_2)e(j,m_2)] =\\ &E\left[\left(\sum_{k_1=1}^{L_1} a_{k_1}(j,n_2)d(j,n_2-k_1)\right)e(j,m_2)\right]\\ &+E\left[\left(\sum_{k_2=-L_2}^{L_3}\tilde{a}_{k_2}(j,n_2)d(j+1,[\frac{n_2}{2}]-k_2)\right).\right.\\ &\left.\times\left(\sum_{k=-\infty}^{\infty} h(m_2-2k)e(j+1,k)\right)\right]\\ &+E\left[\left(\sum_{k_2=-L_2}^{L_3}\tilde{a}_{k_2}(j,n_2)d\left(j+1,\left[\frac{n_2}{2}\right]-k_2\right)\right)\cdot\right.\\ &\left.\times\left(\sum_{k=-\infty}^{\infty} g(m_2-2k)d(j+1,k)\right)\right]\\ &+E\left[b(j,n_2)w(j,n_2)\left(\sum_{k=-\infty}^{\infty} h(m_2-2k)e(j+1,k)\right.\right.\\ &\left.\left.+\sum_{k=-\infty}^{\infty} g(m_2-2k)d(j+1,k)\right)\right].\end{aligned}\tag{52}$$

This correlation function is sequentially computed. Therefore, the first term above is known, assuming that we do not use periodicity of the signal, and we disregard the terms that are not available around the boundaries. The second and third terms above can be computed using (49), and the last term is equal to zero.

Third, we compute

$$\begin{aligned}&E[d(j,n_2)d(j,m_2)]\\ &=E\left[\left(\sum_{k_1=1}^{L_1} a_{k_1}(j,n_2)d(j,n_2-k_1)\right).d(j,m_2)\right]\\ &+E\left[\sum_{k_2=-L_2}^{L_3}\tilde{a}_{k_2}(j,n_2)d\left(j+1,\left[\frac{n_2}{2}\right]-k_2\right).d(j,m_2)\right]\\ &+E[b(j,n_2)w(j,n_2)d(j,m_2)].\end{aligned}\tag{53}$$

Again, the first term is known, due to sequential computation of this correlation function. The third term is also known and equal to $(b(j,n_2))^2\delta(n_2-m_2)$. In order to compute the second term, it is sufficient to know $E[d(j,m_2)d(j+1,n_1)]$ for all possible $n_1$ and $m_2$. We have (54), which appears at the top of this page. The first term is known due to the sequential computation of this correlation function. The second term is known due to (49), and the third term is equal to zero. Therefore, we can compute $E[d(j,m_2)d(j+1,n_1)]$ and, consequently, $E[d(j,n_2)d(j,m_2)]$.

This completes the proof.

## ACKNOWLEDGMENT

The authors would like to thank V. Badrinath, École Nationale Supérieure des Télécommunications, Paris, for useful discussions. They would also like to thank a reviewer for constructive comments on an earlier version, which helped to improve the paper considerably.

## REFERENCES

[1] M. Basseville *et al.*, "Modeling and estimation of multiresolution stochastic processes," *IEEE Trans. Inform. Theory*, vol. 38, pp. 766–784, Mar. 1992.

[2] A. Cohen, "Ondelettes, analyses multiresolutions et traitement numerique du signal," Ph.D. thesis, Université Paris IX, Dauphine, 1990.

[3] A. Cohen, J. Froment, and J. Istas, "Analyse multirésolution des signaux aléatoires," *C. R. Acad. Sci. Paris*, vol. t. 312 I, pp. 567–570, 1991.

[4] R. W. Dijkerman, V. Badrinath, and R. R. Mazumdar, "Multi-scale representation of stochastic processes using compactly supported wavelets," in *Proc. IEEE-SP Int. Symp. Time–Frequency Time-Scale Anal.* (Victoria, Canada), Oct. 1992, pp. 185–188.

[5] R. W. Dijkerman and R. R. Mazumdar, "On the correlation structure of the wavelet coefficients of fractional Brownian motion," to appear in *IEEE Trans. Inform. Theory*.

[6] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, pp. 909–996, 1988.

[7] M. H. A. Davis, *Linear Estimation and Stochastic Control.* London: Chapman and Hall, 1977.

[8] P. Flandrin, "On the spectrum of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 197–199, Jan. 1989.

[9] ——, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, pp. 910–917, Mar. 1992.

[10] D. Gabor, "Theory of communication," *J. Inst. Electr. Eng.* (London), vol. 93-III, pp. 429–457, 1946.

[11] S. A. Golden, "Identifying multiscale statistical models using the wavelet transform," Master's thesis, Mass. Inst. of Technol., Lab. for Inform. Decision Syst., June 1991.

[12] P. Goupillaud, A. Grossmann, and J. Morlet, "Cycle octave and related transform in seismic signal analysis," *Geoexploration*, vol. 23, pp. 85–102, 1985.

[13] R. Kronland-Martinet, J. Morlet, and A. Grossmann, "Analysis of sound patterns through wavelet transform," *Int. J. Patt. Recogn. Artificial Intell.*, 1988.

[14] M. R. Luettgen, W. C. Karl, and A. S. Willsky, "Efficient multiscale regularization with applications to the computation of optical flow," *IEEE Trans. Image Processing*, vol. 3, no. 1, pp. 41–65, Jan. 1994.

[15] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.

[16] ——, "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 37, pp. 2091–2110, Dec. 1989.

[17] E. Masry, "The wavelet transform of stochastic processes with stationary increments and its application to fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. IT-39, pp. 260–264, Jan. 1993.

[18] Y. Meyer, *Ondelettes et opérateurs.* Paris: Hermann, 1988.

[19] J. Ramanathan and O. Zeitouni, "On the wavelet transform of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 1156–1158, July 1991.

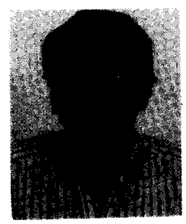
[20] A. H. Tewfik and M. Kim, "Correlation structure of the discrete wavelet coefficients of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. IT-38, pp. 904–909, Mar. 1992.

[21] ——, "Fast positive definite linear system solvers," Tech. Rep., To appear in *IEEE Trans. Signal Processing*, 1994.

[22] G. W. Wornell, "A Karhunen-Loéve-like expansion for $\frac{1}{f}$ processes via wavelets," *IEEE Trans. Inform. Theory*, vol. IT-36, pp. 859–861, July 1990.


**Robert W. Dijkerman** was born in Wageningen, The Netherlands, in 1967. He received the M.Eng. degree in applied mathematics from the University of Twente, Enschede, The Netherlands, in 1990.

He is currently a Ph.D. student at Institut National de la Recherche Scientifique, INRS-Télécommunications, Ile des Soeurs, Canada. His research interests are in the areas of statistical signal processing, stochastic modeling, and related image processing problems.



**Ravi R. Mazumdar** (S'80-M'88) was born in Bangalore, India. He received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay, India, in 1977, the M.Sc., DIC degree in control systems from Imperial College, London, England, in 1978, and the Ph.D. degree in systems science from the University of California, Los Angeles, in 1983.

From 1978 to 1979, he was employed by GEC Electrical Projects Ltd., Rugby, England. From April to October 1983, he was a Member of Technical Staff, AT&T Bell Laboratories, Holmdel, NJ. He held visiting appointments in the Department of Systems Science, UCLA, in 1983-1984 and with the Department of Applied Mathematics, University of Twente, Enschede, The Netherlands, in the Fall of 1984. From 1985-1988, he was an Assistant Professor in the Department of Electrical Engineering, Columbia University, New York, NY, where he was a member of the Center for Telecommunications Research. In July 1988, he joined INRS-Télécommunications, Montreal, Canada, where he is currently an Associate Professor. He is also an Auxiliary Professor in the Department of Electrical Engineering, McGill University, Montreal. His research interests are in the modeling and analysis of telecommunication systems, stochastic filtering and control, and applied stochastic processes.