Tomographic Reconstruction and Estimation Based on Multiscale Natural-Pixel Bases

Mickey Bhatia, William C. Karl, *Member, IEEE*, and Alan S. Willsky, *Fellow, IEEE*

***Abstract*— We use a natural pixel-type representation of an object, originally developed for incomplete data tomography problems, to construct nearly orthonormal multiscale basis functions. The nearly orthonormal behavior of the multiscale basis functions results in a system matrix, relating the input (the object coefficients) and the output (the projection data), which is extremely sparse. In addition, the coarsest scale elements of this matrix capture any ill conditioning in the system matrix arising from the geometry of the imaging system. We exploit this feature to partition the system matrix by scales and obtain a reconstruction procedure that requires inversion of only a well-conditioned and sparse matrix. This enables us to formulate a tomographic reconstruction technique from incomplete data wherein the object is reconstructed at multiple scales or resolutions. In case of noisy projection data we extend our multiscale reconstruction technique to explicitly account for noise by calculating maximum *a posteriori* probability (MAP) multiscale reconstruction estimates based on a certain self-similar prior on the multiscale object coefficients. The framework for multiscale reconstruction presented here can find application in regularization of imaging problems where the projection data are incomplete, irregular, and noisy, and in object feature recognition directly from projection data.**

I. Introduction

IN THIS PAPER we consider the solution of ill-posed tomographic reconstruction problems where the projection data are noisy and incomplete. The conventional methods for tomographic image reconstruction require high quality (i.e., noise-free) projection data to provide accurate reconstructions. Further, while those methods suited to the availability of a complete set of projection data (for example, the filtered backprojection (FBP) method) are fast, the conventional methods of coping with incomplete data lead to very computationally intensive solutions. We have developed a multiscale reconstruction technique that yields computationally tractable reconstructions from incomplete data and can be extended to yield statistically optimal reconstructions from noisy, nonstationary data with very little added computational complexity. In addition, our multiresolution framework for tomographic reconstruction is natural or desirable if the ultimate objectives are multiresolution in some way, for example if the interest is not to fully reconstruct the field but to gather information about coarse scale (i.e., aggregate) or fine scale (for example, boundary) features of the field. Using conventional techniques we would first have to reconstruct the entire field and then use postprocessing to extract such features.

To develop our multiscale reconstruction technique, we start with the natural pixel (NP) object representation [6], [7], which was originally developed for the incomplete data tomography problem. The NP representation results in a matrix based reconstruction method which has the advantage that the resulting reconstructions are devoid of many of the incomplete data artifacts present in the FBP reconstruction. The disadvantage of the NP reconstruction, or matrix-based reconstruction methods in general, is that solutions of very large, generally ill-conditioned systems of equations are required.

In this paper, we build on the NP approach by using wavelet bases to transform the NP basis functions. The standard NP system matrix, relating the input (the object coefficients) and the output (the projection data), is full. The use of wavelet bases leads to a transformation matrix, which is amenable to sparsification. We accomplish this sparsification of the NP projection operator through the novel use of one-dimensional (1-D) wavelet transforms defined in the projection domain, in contrast to general wavelet-based operator compression results. In addition to sparsification, the coarsest scale elements of this transformed matrix capture any ill conditioning in the system matrix arising from the geometry of the imaging system. We exploit this feature to partition the multiscale system matrix by scales and obtain a reconstruction procedure that only requires inversion of a well-conditioned and sparse matrix. The use of wavelet bases also enables us to formulate a multiscale tomographic reconstruction technique wherein the object is reconstructed at multiple scales or resolutions. The overall reconstruction is obtained by combining the reconstructions at different scales.

Noisy imaging problems arise in a variety of contexts (e.g., low-dose medical imaging, oceanography, and in several applications of nondestructive testing of materials) and in such cases standard matrix-based reconstruction methods (including NP) often yield unacceptable results, particularly when the noise is nonstationary. These situations generally reflect the fact that more degrees of freedom are being sought than are really supported by the data and, hence, some form of regularization is required. In contrast to the standard NP method, we are able to extend our multiscale reconstruc-

Manuscript received April 24, 1994; revised April 22, 1996. This work was supported by the Advanced Research Projects Agency through Air Force Grant F49620-93-1-0604, by the Air Force Office of Scientific Research under Grants F49620-92-J-0002 and F49620-96-10028, and by the U.S Army Research Office under Grant DAAL03-92-G-0115. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ken D. Sauer.

M. Bhatia is with J. P. Morgan & Co. Inc., New York, NY 10016 USA.

W. C. Karl is with the Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215 USA (e-mail: wckar@bu.edu).

A. S. Willsky is with Stochastic Systems Group, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

Publisher Item Identifier S 1057-7149(97)00484-3.

1057–7149/97$10.00 © 1997 IEEE

tion technique in the case of noisy projections to obtain a statistically regularized, multiscale maximum *a posteriori* probability (MAP) object estimate. We do this by realizing that for ill-posed problems the lower resolution (i.e., the coarser scale) reconstructions are often more reliable than their higher resolution counterparts and by using prior statistical models constructed *directly in the multiscale domain*, which capture such intuition. Our multiscale MAP regularized reconstructions are no more computationally intensive than our unregularized multiscale reconstructions.

This paper contrasts other multiscale tomography approaches that either concentrate on the complete data tomography problem [4], [31], [32], assume prior knowledge of the object edges to reconstruct an object from incomplete data [33], or are focused on localization of the Radon transform or radiation reduction [14], [29]. In addition, in the approaches [14], [31]–[33] the object is expanded in a two-dimensional (2-D) wavelet basis for the original spatial domain and the resulting coefficients of this expansion are then calculated from the projection data. In contrast, in our multiscale approach based on the incomplete data NP framework, a two-dimensional (2-D) multiscale representation is obtained by a *1-D wavelet expansion* of the NP basis functions.

Finally, while the work here focuses on the case of incomplete data, when complete data are available, additional efficiencies may be obtained through the use of explicit Radon transform inversion formulas, such as FBP. In such complete data cases, the multiscale methodology described herein may be applied using the FBP method as a starting point to obtain both unregularized and regularized multiscale reconstructions with the same computational complexity as the FBP reconstruction. Such application is described in detail in [4].

The paper is organized as follows. Section II contains preliminaries. In Section II-A we describe the standard tomographic reconstruction problem and in Section II-B we describe the NP reconstruction technique. We outline the theory of 1-D multiscale decomposition in Section II-C. In Section III, we develop the theory behind our wavelet-based multiscale reconstruction method starting from the NP object representation. In Section IV we build on this framework to provide a method for obtaining MAP regularized reconstructions from noisy data. The conclusions are presented in Section V.

II. Preliminaries

A. The Tomographic Reconstruction Problem

In tomography, the goal is to reconstruct an object or a field, f, from line-integral projection data [23]. For a parallel-beam imaging geometry, the projection data consists of parallel, nonoverlapping strip integrals through the object at various angles (see Fig. 1). Each angular position corresponds to a specific source–detector orientation. Suppose we have N_θ positions between 0° and 180° and N_s parallel strip integrals at each angular position. Let us label the observation corresponding to projection ℓ at angular position k by $y_k(\ell)$,

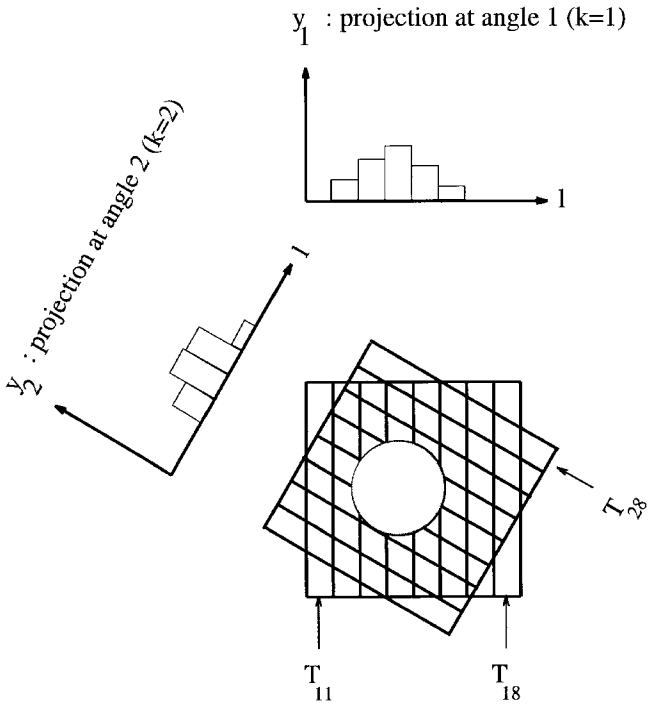

Fig. 1. Projection measurements of an object (shaded) at two different angular positions ($k = 1$ and $k = 2$). Also shown are three basis functions, T_{11}, T_{18}, and T_{28}, which are the indicator functions of the corresponding strips. Each angular projection is composed of $N_s = 8$ strips in this example.

where $k = 1, \cdots, N_\theta$ and $\ell = 1, \cdots, N_s$. Furthermore, let $T_{k\ell}(u, v)$ be the indicator function of the strip integral corresponding to this observation so that $T_{k\ell}(u, v)$ has value one within that strip and zero, otherwise. Given this notation

$$y_k(\ell) = \iint_{\Omega} f(u, v) T_{k\ell}(u, v) du dv, \quad k = 1, \cdots, N_\theta;\ \ell = 1, \cdots, N_s \quad (1)$$

where (u, v) are the usual rectangular spatial coordinates and the integration is carried over a region of interest $\mathbf{\Omega}$.

Due to practical considerations, we have to work with a discretized version of (1). By using standard discretization techniques (see, for example, [7]), the projection data at angle k is given by

$$y_k = T_k f \quad (2)$$

where T_k is an $N_s \times N_s^2$ matrix representing $\{T_{k\ell}(u, v);\ \ell = 1, \cdots, N_s\}$ and f is an N_s^2 vector representing $f(u, v)$ on an $N_s \times N_s$ square pixel lattice, and y_k is the corresponding vector of measurements $y_k(\ell)$. Thus row ℓ of T_k is the (discrete) representation of the strip function $T_{k\ell}(u, v)$ and the inner product of f with this strip yields the data contained in the corresponding entry of y_k. Finally, by combining the projection data y_k from all angles k we get the following overall observation equation:

$$y = \begin{bmatrix} y_1^T & | & y_2^T & | & \cdots & | & y_{N_\theta}^T \end{bmatrix}^T = Tf \quad (3)$$

where y is the $N_\theta N_s$ vector containing the projection data, T is the $N_\theta N_s \times N_s^2$ matrix representing the complete set of discretized basis functions $\{T_{k\ell}(u, v);\ k = 1, \cdots, N_\theta;\ \ell =$

$1, \ldots, N_s\}$ and is defined in the obvious way, and f is a N_s^2 vector representing the discretized object. The tomographic reconstruction problem then reduces to finding an estimate $\hat{f}$ of the discretized object f given the projection data contained in y.

B. Conventional Reconstruction Techniques

In this section we discuss two conventional reconstruction techniques, the widely used FBP reconstruction technique and the NP reconstruction technique used by us as a starting point for our multiscale reconstruction. In both the FBP and the NP reconstructions, the object is expanded in a nonorthogonal basis that is closely related to the data acquisition process. In particular, the estimated object is represented as a linear combination of the same functions $T_{k\ell}(u,v)$ along which the projection data are collected. Similar to (2) and (3), a discretized version of this representation may be obtained as

$$\hat{f} = \sum_{k=1}^{N_\theta} T_k^T x_k = T^T x \tag{4}$$

where $x = [x_1^T \mid x_2^T \mid \cdots \mid x_{N_\theta}^T]^T$ is a $N_\theta N_s$ vector containing the object expansion coefficients. Note that (4) can be interpreted as the backprojection operation where the coefficients in x are backprojected along the basis functions in the rows of T [23].

To complete the reconstruction the coefficients x must now be determined. The FBP and the NP methods differ in how the coefficients x are calculated from the projection data y. The standard FBP method calculates the coefficients x_k at each angle k according to the Radon inversion formula by filtering the projection data y_k at that particular angle with a ramp filter [23]. Thus, for a fixed angle k

$$x_k = R y_k \tag{5}$$

where the matrix R captures this ramp-filtering operation. Thus, (4) and (5) together represent the two operations used in the standard FBP reconstruction. Since the FBP method is based directly on the Radon inversion formula, it is valid (i.e., yields exact reconstructions) only when a continuum of noise-free line integral projections from all angles are used [23]. In practice, as indicated in (1), we only have access to sampled projection data which are collected using strips of finite width. In this case, the quality of the FBP reconstruction is a function of the quality and fineness of the corresponding projection data used. By collecting the object coefficient and data vectors at different angles we obtain the following overall equation which reflects the identical and independent processing from angle to angle performed on the projection data by the FBP method

$$x = \begin{bmatrix} R & & \bigcirc \\ & R & \\ \bigcirc & \ddots & \\ & & R \end{bmatrix} y. \tag{6}$$

An important point to note in the above equation is that the matrix R is fixed and is not a function of the imaging system (i.e., the quality and the quantity of the acquired projection data). It is this simple, regular structure in the explicit formula (6) that results in the computational efficiency of FBP.

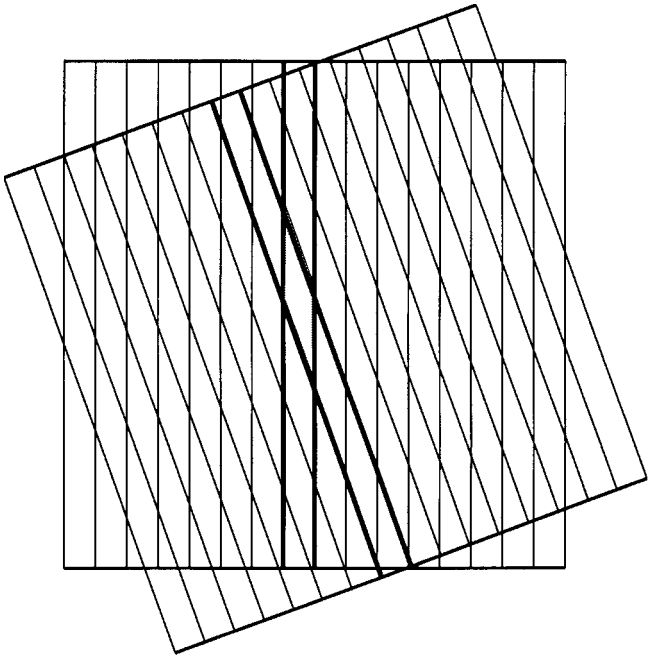

Fig. 2. Elements of the matrix C are the areas of intersection of various strips. One such area of intersection, corresponding to two strips delineated by bold lines, is shown shaded. The matrix C is full as most of these areas are nonzero.

The NP representation [6], [7], on the other hand, was originally developed for incomplete data tomography problems and belongs to the class of matrix-based reconstruction methods [24], [25]. It has found use in a variety of disciplines [1], [9], [10], [18], [28], [37]. In the NP reconstruction, an implicit equation relating the expansion coefficients, x, to the projection data, y, is derived by substituting $\hat{f}$ from (4) for f in (3) and assuming that equality holds in the resulting equation:

$$y = (TT^T)x = Cx. \tag{7}$$

The coefficients x are then calculated from the projection data y by solving the implicit equation (7). These coefficients are backprojected according to (4) as before to obtain the reconstruction $\hat{f}$. Note that in the NP reconstruction, since the matrix C is full, the processing of data is not independent from angle to angle, in contrast to FBP. The *advantage* of the NP reconstruction over the FBP is that since the matrix C is calculated for each specific acquisition geometry, the reconstruction can be customized for any imaging system, so that a complete set of angular projection data is neither assumed nor required for adequate reconstruction as in FBP.

Unfortunately, solution of the large system of (7) for the coefficients x leads to significant difficulties. The first obstacle is the sheer size ($N_\theta N_s \times N_\theta N_s$) of the matrix C. The elements of C are the areas of intersection of the strips defined by the basis functions T (see Fig. 2). Most of these areas are not zero and hence the matrix C is *full*, requiring a tremendous amount of storage. The large size of C also makes it difficult to solve for x directly from (7). In [6], [7], and [15] this problem is circumvented by two different approaches. The

first is to use iterative techniques and a suitable initial value to solve for x. This is commonly done using a Kaczmarz-like method as found in ART [19], [23]. It is well known that the computational burden of such algorithms is proportional to the number of nonzero matrix elements together with the number of iterations required for convergence, which itself depends on the conditioning of the matrix through parameter choice. The second approach is to concentrate on some specific imaging geometries which result in a matrix C that can be directly inverted in a computationally efficient manner. The problem with the second approach is that these imaging geometries may not be practical. A final major difficulty in obtaining the NP reconstruction is that there is an inherent nonuniqueness in the NP object representation arising from its tie to the data acquisition process, which results in C being rank deficient or at best being badly conditioned for most imaging geometries. None of the existing NP related work [6], [7], [15] discusses this conditioning issue brought on by the nonuniqueness of the NP representation.

The ill-conditioned nature of the matrix C can be understood at an intuitive level if one assumes an infinite field-of-view for the imaging geometry (rather than the finite rectangular field-of-view we show in Fig. 1) so that edge effects are absent. Recall that the columns of T_k^T are the basis functions of the NP representation at angle k, c.f. (4). Now the sum of all the columns of $T_{k_1}^T$ equals the sum of all the columns of $T_{k_2}^T$ since both correspond to the same indicator function of the field-of-view. This simply reflects the physical fact that both of these sums provide DC shifts of the object field. Thus, the same underlying object can be represented in a variety of ways, corresponding to different allocations of its DC component to the different angular basis sets. Hence the representation (4) is nonunique, T^T does not have full column rank, and $C = TT^T$ is not invertible. For the case of our finite field-of-view, the above discussion is exact only if k_1 and k_2 correspond to projections at 0° and 90° due to nonuniform edge effects. Even when the projections are not exactly at right angles, however, while not dropping rank, C is quite ill conditioned, as will be discussed in Section III-D.

The above discussion provides us with a preview of things to come. In Section III, we use 1-D wavelet bases to transform the NP basis functions in T into a multiscale framework. The use of wavelet bases, in addition to providing a multiscale framework, enables us to overcome the above limitations of the NP reconstruction. Use of the wavelet transform leads to a multiscale system matrix corresponding to C, which can be sparsified. Further, the coarsest multiscale basis function at any angle k turns out to be the sum of all the columns of T_k^T, responsible for the ill conditioning of the resulting multiscale matrix. We exploit this feature to partition the multiscale system matrix by scales to obtain a reconstruction procedure that requires inversion of only a well-conditioned and sparse matrix.

A final difficulty in NP arises when we consider noisy observations. In order to obtain statistically based regularized solutions to the ill-posed reconstruction problem which arises when the projection data are noisy, we need to combine (7) with a prior probabilistic model for the object coefficients x. While there is no natural way to construct a prior for these coefficients in the original projection domain that leads to a computationally efficient estimation algorithm, the transformation of these coefficients to a multiscale projection domain allows us to use simple yet powerful self-similar prior models which have been developed in this domain [4]. Specifically, in Section IV we use prior statistical models constructed directly in the multiscale projection domain that capture the intuition that for ill-posed reconstruction problems the lower resolution (i.e., the coarser scale) reconstructions are more reliable than their higher resolution counterparts. Later we will see that not only are these models conceptually and computationally simple, but they also result in good reconstructions, even in cases with nonstationary noise.

C. 1-D Wavelet Transform-Based Multiscale Decomposition

We begin with a brief review of the wavelet-based multiscale decomposition of functions. The reader is referred to [27] for details. Let $L^2(R)$ denote the vector space of measurable, square-integrable, 1-D functions $x(u)$, and let Z denote the set of integers. A multiscale approximation of $L^2(R)$ is a sequence of subspaces $\{V_j\}_{j\in Z}$ with $\cdots V_{-1} \subset V_0 \subset V_1 \cdots$ having the interpretation that the projection $A_j x(u)$ of $x(u)$ on V_j gives the approximation of x at scale j. The scales become finer with increasing j. Now, as shown in [27], there exists a unique function $\phi(u) \in L^2(R)$, called the *scaling function*, such that for each scale j, $\{\phi_{j,l}(u) = \sqrt{2^j}\phi(2^j u - l)\}_{n\in Z}$ is an orthonormal basis of V_j. Thus, the approximation of the function at scale j can be written as

$$A_j x(u) = \sum_l x^{(j)}(l)\phi_{j,l}(u) \tag{8}$$

with

$$x^{(j)}(l) = \langle x(u), \phi_{j,l}(u)\rangle \tag{9}$$

where $\langle .,. \rangle$ refers to the inner product operation. The difference in information between the approximation of the function at successive scales j and $j+1$ is captured by the *detail function* at scale j. This detail function is obtained as the projection $D_j x(u)$ of $x(u)$ on the subspace O_j which represents the orthogonal complement of V_j in V_{j+1}. Similar to the scaling function, there exists a function $\psi(u) \in L^2(R)$ called an *orthogonal wavelet*, such that for each scale j, $\{\psi_{j,l}(u) = \sqrt{2^j}\psi(2^j u - l)\}_{n\in Z}$ is an orthonormal basis of O_j. Thus the detail function at scale j, capturing the difference in information between $A_j x(u)$ and $A_{j+1}x(u)$, can be written as

$$D_j x(u) = \sum_l \xi^{(j)}(l)\psi_{j,l}(u) \tag{10}$$

where

$$\xi^{(j)}(l) = \langle x(u), \psi_{j,l}(u)\rangle. \tag{11}$$

Let h and g be functions satisfying

$$\begin{aligned} \phi(u) &= \sum_l \sqrt{2}h(l)\phi(2u - l) \\ \psi(u) &= \sum_l \sqrt{2}g(l)\phi(2u - l). \end{aligned} \tag{12}$$

The discrete approximation and detail coefficients at scale j, $x^{(j)}$, and $\xi^{(j)}$ respectively, can then be obtained from the next finer scale discrete approximation coefficients, $x^{(j+1)}$, by convolution with h and g followed by downsampling by a factor of 2:

$$\begin{aligned} x^{(j)}(l) &= (h * x^{(j+1)}(2l)) \\ \xi^{(j)}(l) &= (g * \xi^{(j+1)}(2l)). \end{aligned} \tag{13}$$

In above, $*$ refers to 1-D convolution. Conversely, the finer scale discrete approximation $x^{(j+1)}$ can be synthesized from the next coarser scale discrete approximation and detail coefficients, $x^{(j)}$ and $\xi^{(j)}$, respectively, by first up-sampling by a factor of 2 followed by convolution with h and g

$$x^{(j+1)}(l) = \sum_k h(2k-l)x^{(j)}(k) + \sum_k g(2k-l)\xi^{(j)}(k). \tag{14}$$

In practice, we do not have access to the continuous function $x(u)$ but rather to a finite number N of its samples. We assume these samples to be the approximation coefficients $x^{(J)}$ of $x(u)$ at some finest resolution J. For convenience we assume N to be a power of 2 so that $J = \log_2(N)$.

The discrete multiresolution decomposition of x is then represented in a vector form as

$$\xi = Wx^{(J)} = \begin{bmatrix} \xi^{(J-1)} \\ \xi^{(J-2)} \\ \vdots \\ \xi^{(0)} \\ x^{(0)} \end{bmatrix} \begin{matrix} \Longleftarrow \text{Finest scale detail} \\ \\ \\ \Longleftarrow \text{Coarsest scale detail} \\ \Longleftarrow \text{Coarsest scale approximation} \end{matrix} \tag{15}$$

where W is the matrix representation of the multiscale decomposition operation (13), which can be performed in an extremely efficient manner ($\mathcal{O}(N)$) [2]. The matrix W is square and invertible and, since we consider orthonormal multiresolution decomposition in this paper, $W^{-1} = W^T$. The vector $\xi^{(j)}$ represents the *detail* associated with scale j and is of length 2^j. Thus, the length of $\xi^{(j)}$ is twice that of $\xi^{(j-1)}$, which is consistent with the downsampling implied by (13). The last element $x^{(0)}$ in $Wx^{(J)}$ is the *approximation* of the signal x at the coarsest scale $j = 0$. For our purpose, this is the same as some multiple of the DC term (i.e., the sum of the elements) of $x^{(J)}$.

In our work in this paper, in addition to the Haar wavelet we use the wavelets of Daubechies [12], the separate elements of which are denoted D_n, where the length and the regularity of the wavelets increases linearly with n. Finally, since our signals are of finite length, we need to deal with the edge effects which occur at the ends of the interval in the wavelet transform. While there are a variety of ways in which to do this, such as modifying the wavelet functions at the ends of the interval in order to provide an orthogonal decomposition over the interval [13], we have chosen here to use one of the most commonly used methods, namely that of cyclically wrapping the interval [16], [27]. While this does introduce some edge effects, these are of negligible importance for the objectives and issues we wish to emphasize and explore and for the applications considered here. Further, the methods we describe can be readily adapted to other approaches for dealing with edge effects as in [13]. As a result of cyclic wrapping, the coarsest scale scaling function is the same for all wavelets D_n and is the same as the scaling function for the Haar case. In particular the coarsest scale approximation term $x^{(0)}$ is always a constant multiple of the DC component of the signal.

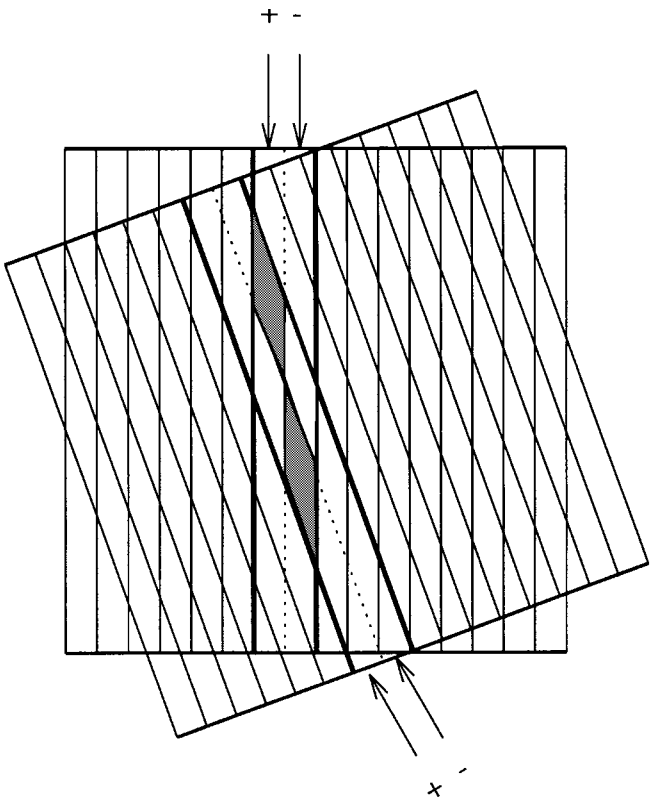

Fig. 3. Each of the two delineated strips from Fig. 2 are broken down into two substrips having a positive and a negative weight, respectively. The area of intersection of the two strips is zero in this case due to the cancellation of the positive (lightly shaded) and the negative (darkly shaded) terms.

III. A Multiscale Approach to Natural Pixel (NP) Reconstruction

A. Multiscale Transformation of the NP Basis Functions

The multiscale reconstruction is motivated by the following observations. Recall that the elements of the NP matrix C in (7) are the areas of intersection of the basis functions T. If we are able to modify these basis functions so that they are mostly orthogonal, then the corresponding areas of intersection will be nearly zero, resulting in a sparse matrix and a simplified solution for the underlying object coefficients. Specifically, suppose that we are able to modify the basis functions (i.e., strips) such that they have the form shown in Fig. 3. Each strip is a linear combination of two NP strips, one given a positive weight and the other negative. The new matrix relating the object coefficients and the projection data, according to the above choice of strips, will have as its elements the (signed) areas of intersections of the newly defined strips. It is clear from Fig. 3 that most of these elements will be zero due to the

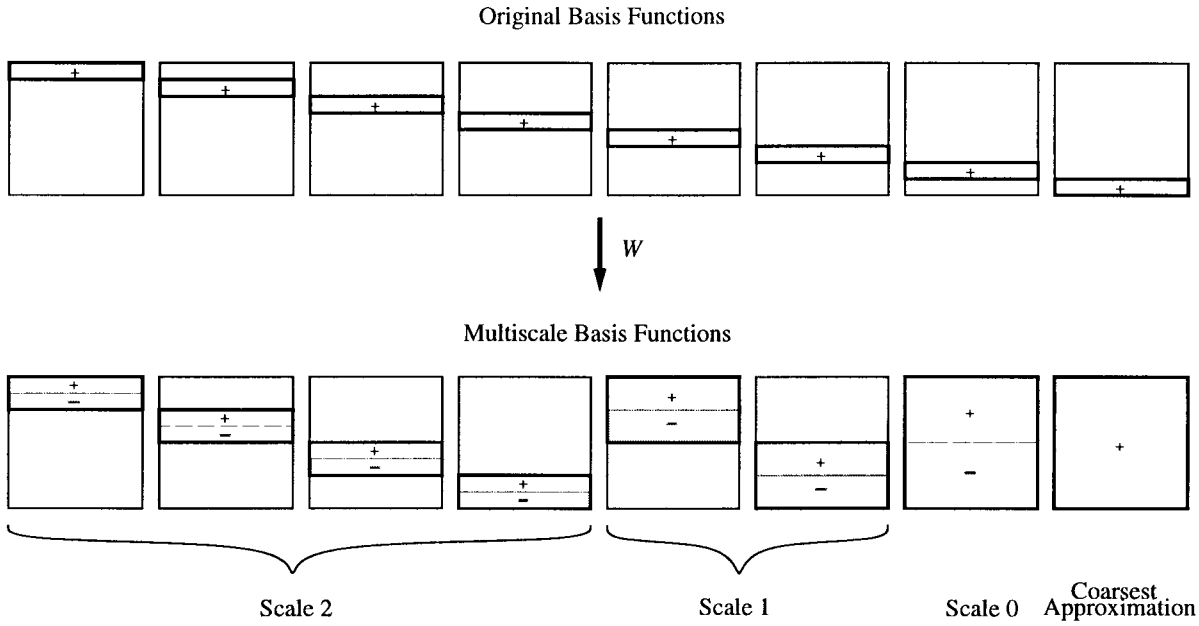

Fig. 4. Original NP basis functions contained in T_k (shown in the top half of the figure) and multiscale transformed basis functions of $\mathcal{T}_k$ (shown in the bottom half of the figure) for a fixed angle k. The multiscale basis functions are grouped into different scale components based on their spatial extent. The Haar wavelet is used for multiscale decomposition in this example. The heavy boundaries indicate the extent of support of the corresponding basis function and the shading and $+/-$ indicate the sign of the basis function over the region.

cancellation of the positive and the negative terms. Only those elements that correspond to strip intersections near the edge of the field-of-view will be nonzero. Thus we can expect this new matrix to be sparse with the degree of sparsity increasing with the size of the field-of-view (since the fraction of intersections near the edge decreases with increasing size).

The above redefinition of the basis strips with positive and negative weights is reminiscent of the finest level of a Haar transform and, in fact, we can imagine repeating the process at other levels as well. However, an important point to note is that here the Haar transform is taken only in one direction, i.e., the direction perpendicular to the long axis of the strip. This is the key to our multiscale reconstruction method: We expand the basis functions $\{T_k\}$ in a 1-D wavelet basis which then induces a corresponding 2-D multiscale object representation.[1] For a projection at a fixed angle k (and $N_s = 8$), the full such Haar transform of the original basis functions shown in Fig. 4 (including contributions from all levels) will look as shown in the bottom half of Fig. 4. A notion of scale emerges from the use of the Haar transform. The original strips have been broken down into a series of strips at multiple scales having positive and negative weights. The finest scale involves strips that have twice the width of the original strips and the coarsest scale involves strips extending over the entire field-of-view. We call the above transformed strip functions the *natural wavelet basis* because of the adaptation of the natural pixel representation. While we have used the Haar wavelet in the multiscale transformation described above and shown in Fig. 4, one can imagine using a more general wavelet for the same purpose, as discussed next.

In particular, let W be a matrix representation of the linear operator that performs a 1-D orthonormal multiscale decomposition on a discrete sequence of finite length N_s (as described in Section II-C) so that $W^{-1} = W^T$. Further, let $W_b = \text{blockdiag}(W)$ be a block-diagonal matrix with N_θ blocks along the diagonal, all equal to W (so that again $W_b^{-1} = W_b^T$). We define our general multiscale transformation of the strip basis functions as

$$\mathcal{T} \triangleq W_b T \tag{16}$$

where the matrix $\mathcal{T}$ now contains the multiscale basis functions at all the different angles

$$\mathcal{T} = \begin{bmatrix} \mathcal{T}_1^T & \mathcal{T}_2^T & \cdots & \mathcal{T}_k^T & \cdots & \mathcal{T}_{N_\theta}^T \end{bmatrix}^T \tag{17}$$

and the columns of matrices $\mathcal{T}_k^T$ represent the discretized multiscale basis functions at angle k (see Fig. 4 for the Haar case).

Now suppose we define the vectors

$$\eta \triangleq W_b y, \qquad \xi \triangleq W_b x \tag{18}$$

which contain the stacked set of wavelet coefficients of the projection data $\eta_k = W y_k$ and the object coefficients $\xi_k = W x_k$ at each angle k. Then by applying (16) to (7) we get the following relationship between the multiscale representation of the object coefficients, ξ, and the multiscale representation of the data, η:

$$\eta = \mathcal{C}\xi \tag{19}$$

where the multiscale system matrix $\mathcal{C}$ is given by

$$\mathcal{C} = W_b C W_b^T = W_b (T T^T) W_b^T = \mathcal{T}\mathcal{T}^T. \tag{20}$$

[1] Our multiscale object representation is fundamentally different from previous multiscale-related representations for tomography (for example, [31]). In these approaches a direct 2-D expansion of the object (i.e., a 2-D wavelet transform) is used, the coefficients of which are then calculated from the projection data.

Note that (20) implies that the elements of the transformed matrix $\mathcal{C}$ are the (signed) areas of intersection of the various multiscale basis functions $\mathcal{T}$. From our previous discussion we expect $\mathcal{C}$ to be *sparse*. if W reflects the use of the Haar wavelet for multiscale decomposition. In general, the use of any compactly supported Daubechies wavelet results in approximately the same effective sparsity as that achieved in the Haar case. Note that the transformed multiscale system matrix $\mathcal{C}$ in (20) is naturally obtained through a series of 1-D wavelet transforms defined in the projection domain, and is not simply a 2-D wavelet transformation of C.

Finally, by combining (18) and (4), we obtain the following representation of the object in the multiscale domain

$$\hat{f} = \sum_{k=1}^{N_\theta} \mathcal{T}_k^T \xi_k = \mathcal{T}^T \xi. \tag{21}$$

Thus, the reconstruction $\hat{f}$ is obtained by back-projecting the multiscale coefficients ξ_k at angle k along the corresponding multiscale basis functions $\mathcal{T}_k$.

Before proceeding, we note that the multiscale object representation (21) is essentially the same as that used in [4], thus permitting us to define object reconstructions at multiple scales, as we discuss in Section III-C. However, the coefficients ξ used in the representation are obtained from the NP derived relationship (19) rather than from the FBP based scheme used in [4], and it is this more complicated solution that allows us to solve incomplete data problems.

For the development to follow it will prove convenient to order the multiscale vectors η and ξ according to scales rather than projection angles, with the finest scale detail terms from all projections grouped first and the coarsest scale approximation terms grouped last. This rearrangement of (19) results in the following scale ordered and partitioned equation:

$$\begin{bmatrix} \eta^d \\ \eta^a \end{bmatrix} = \begin{bmatrix} \mathcal{C}_{dd} & \mathcal{C}_{da}^T \\ \mathcal{C}_{da} & \mathcal{C}_{aa} \end{bmatrix} \begin{bmatrix} \xi^d \\ \xi^a \end{bmatrix} = \mathcal{C}_s \begin{bmatrix} \xi^d \\ \xi^a \end{bmatrix} \tag{22}$$

where the vectors η^d and ξ^d contain all the detail terms at various scales and angles and are of length $N_\theta(N_s - 1)$, and the vectors η^a and ξ^a contain the coarsest scale approximation, i.e., the DC, terms at all angles (one for each angle) and are of length N_θ. The upper left block $\mathcal{C}_{dd}$ is a $N_\theta(N_s - 1) \times N_\theta(N_s - 1)$ symmetric matrix, the elements of which are the areas of intersection of the detail basis functions (i.e., strips) at various scales and angles. The lower right block $\mathcal{C}_{aa}$ is a $N_\theta \times N_\theta$ symmetric matrix the elements of which are the areas of intersection of the coarsest scale basis function at each angle. Finally, the off-diagonal block $\mathcal{C}_{da}$ is a $N_\theta \times N_\theta(N_s - 1)$ matrix, the elements of which are the areas of intersection of the coarsest scale approximation basis functions and the detail basis functions at various scales.

B. Multiscale Object Coefficient Determination

Let us turn our attention now to the calculation of the object coefficients ξ which, through the backprojection equation, (21), specify the reconstruction $\hat{f}$. Recall that the vector ξ consists of the object detail and approximation coefficients, ξ^d and ξ^a, respectively, which are related to the projection data through (22). For clarity, let us first consider the ideal case of an infinite field extent where edge effects are absent. In this case, as argued in Section II-B, the NP matrix C is rank deficient due to the nonuniqueness of the NP representation. Thus the multiscale matrix $\mathcal{C}_s$ (or, equivalently, $\mathcal{C}$) is also rank deficient. Since a unique solution does not exist in this case, a rational approach is to seek the minimum norm solution to (7) or, in the multiscale domain, to (22). The NP matrix equation (7) represents a large and full system of equations and so it is difficult in practice to find the minimum norm solution in this case. The multiscale relationship (22), however, has a structure that can be exploited to simplify the computations. In particular, first note that in the ideal case with no edge effects, the elements of $\mathcal{C}_{da}$, capturing the areas of intersection between the coarsest scale approximation and finer scale detail basis functions, are identically zero.[2] Further, one can show that the matrix $\mathcal{C}_{dd}$ has full rank, while the matrix $\mathcal{C}_{aa}$ is rank deficient.[3] This is hardly surprising in view of our earlier discussion in Section II-B since we have grouped all the terms contributing to the DC value of the object (and hence to the ill conditioning of $\mathcal{C}_s$) in the $\mathcal{C}_{aa}$ block. Thus, our multiscale transformation has served to "compress" and isolate the nonuniqueness that is present in the NP representation. Now, with $\mathcal{C}_{da} = 0$, the minimum norm solution to (22) is easily found since the detail and the approximation equations decouple. In particular, the minimum norm solution is given by

$$\xi^d = \mathcal{C}_{dd}^{-1} \eta^d \tag{23}$$

$$\xi^a = \mathcal{C}_{aa}^{+} \eta^a \tag{24}$$

where $\mathcal{C}_{aa}^{+}$ is the pseudoinverse of $\mathcal{C}_{aa}$ [17]. Thus, (23) captures the unique part of the solution and (24) provides a particular distribution of the DC components of the object over the coefficients in ξ^a. Before proceeding we note that since $\eta^d = \mathcal{C}_{dd}\xi^d$ represents a large but sparse system of equations we will not, in practice, find the solution to this system by explicitly calculating the inverse of $\mathcal{C}_{dd}$ (as suggested by (23)) but rather we will exploit its sparse structure and favorable conditioning to use any of the methods created especially to solve such sparse systems [17].

If we assume that the object is completely contained in the field-of-view, then we obtain the following simplification[4] of (24) for ξ^a:

$$\xi^a = \left(\frac{1}{N_\theta N_s}\right)\eta^a = \left(\frac{\mu(f)}{N_\theta N_s \sqrt{N_s}}\right) 1_{N_\theta} \tag{25}$$

where 1_p refers to a vector of length p with all elements equal to unity, and $\mu(f)$ is the total mass under the object. This

[2] This is because these elements represent the area under each wavelet (which is zero) due to the fact that the coarsest scale approximation basis functions are just indicator functions of the field-of-view.

[3] In fact, for this ideal case it is easy to see that $\mathcal{C}_{aa}$ is an $N_\theta \times N_\theta$ matrix of ones scaled by a constant.

[4] With no edge effects all elements of the $N_\theta \times N_\theta$ matrix $\mathcal{C}_{aa}$ are equal to N_s. If the object is completely contained in the field-of-view then $\eta^a = (\mu/\sqrt{N_s})1_{N_\theta}$. Further, it can be shown that if $\mathcal{C}_{aa}$ is circulant, as is the case here, then $\mathcal{C}_{aa}^{+} 1_{N_\theta} = (1/r)1_{N_\theta}$, where r is the row sum of $\mathcal{C}_{aa}$. By combining these facts (25) is obtained.

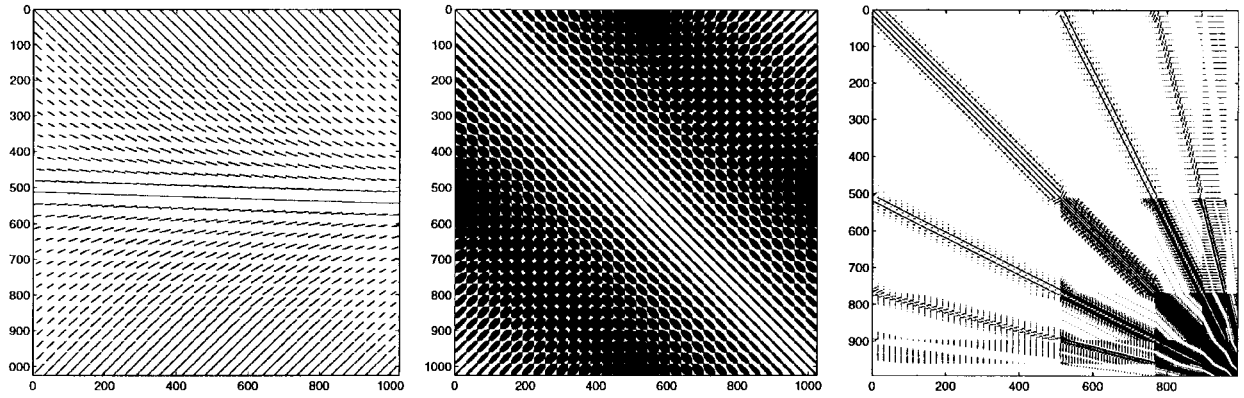

Fig. 5. Binary plot of system matrices. Left: ART system matrix, T. Middle: Natural pixel system matrix, C. Right: Multiscale system matrix, $\mathcal{C}_{dd}$, thresholded at 3.75% of the absolute maximum. An imaging geometry with $N_\theta = N_s = 32$ was assumed and the D3 wavelet is used for multiscale decomposition.

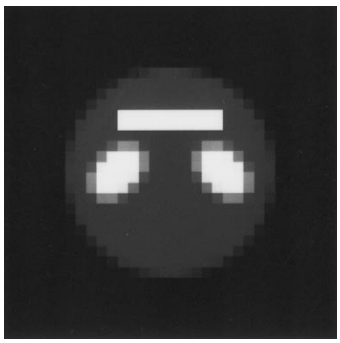

Fig. 6. Phantom used for reconstructions.

simply corresponds to spreading the DC component of the object equally among the different projections.

The development to this point has focused on the ideal case in which field-of-view edge effects are absent. In any practical situation, the field is finite and such effects arise. However, as we discuss below, they have minimal impact on the preceding development. First, due to these edge effects the off-diagonal block $\mathcal{C}_{da}$ in (22) is no longer zero. This block now will have a few nonzero elements corresponding to the areas of intersection of the basis functions near the domain boundary.[5] Even though a variety of methods exist for modifying the solutions (23) and (24) to account for this neglected coupling,[6] our experience is that practical reconstructions based on (23) and (24) (which assume $\mathcal{C}_{da} = 0$) are visually indistinguishable from ones where a correction is made for the coupling. As a result we use (23) and (24) for all the reconstructions we present in this paper.

The other impact the inclusion of edge effects has is to change the structure of $\mathcal{C}_{aa}$. In particular $\mathcal{C}_{aa}$ is no longer truly singular (unless we take views 90° apart) though it is nearly so. In any case, $\mathcal{C}_{aa}$ is still a circulant matrix with row sums nearly equal to the case when the edge effects are neglected and, hence, (25) is still valid. Finally, since the edge effects have no impact on the matrix $\mathcal{C}_{dd}$, the latter still has full rank and is well conditioned.

[5] In [5], we calculate numerical bounds on the absolute values of the elements in $\mathcal{C}_{da}$ for the Haar case.

[6] For example, in [5] we use the matrix inversion lemma [20] to refine the estimate of ξ^d in (24) by introducing the $\mathcal{C}_{da}$ coupling.

C. Object Reconstruction at Multiple Scales

Once we obtain the multiscale object coefficients ξ the object estimate is then obtained via (21) through appropriate combination of the corresponding multiscale basis functions. The multiscale nature of these basis functions naturally induce a multiscale object representation [4]. In particular we can imagine using, for example, only the coarsest scale basis functions and corresponding coefficients in (21) (effectively treating the other, finer scale coefficients as zero) to obtain a coarse reconstruction of the object. Conversely, if our interest is in fine object details, such as edges, we might only wish to use the fine scale coefficients. More generally we define the object reconstruction $\hat{f}^{(s)}$ at scale s as follows:

$$\hat{f}^{(s)} \triangleq \sum_{j=0}^{s-1} \Delta\hat{f}^{(j)} + \hat{f}^{(0)} \tag{26}$$

where $\hat{f}^{(0)}$, the reconstruction at the coarsest scale (i.e., the average or DC value), and $\Delta\hat{f}^{(j)}$, the incremental detail added in going from scale j to $j+1$, are defined as

$$\hat{f}^{(0)} \triangleq \sum_{k=1}^{N_\theta} \left(\mathcal{T}_k^{(0)}\right)^T x_k^{(0)}; \quad \Delta\hat{f}^{(j)} \triangleq \sum_{k=1}^{N_\theta} \left(\mathcal{T}_k^{(j)}\right)^T \xi_k^{(j)} \tag{27}$$

where $\xi_k^{(j)}$ is the component of ξ_k (at angle k) associated with scale j as defined in (15), $x_k^{(0)}$ is the average or the coarsest scale approximation component at angle k (see (15)), $\mathcal{T}_k^{(j)}$ is the block of $\mathcal{T}$ associated with $\xi_k^{(j)}$ in (21), and $\mathcal{T}_k^{(0)}$ is the block of $\mathcal{T}$ associated with $\xi_k^{(0)}$. Note that $\mathcal{T}_k^{(j)}$ contains the basis functions corresponding to scale j (see Fig. 4) and thus the term $\Delta\hat{f}^{(j)}$ captures the information added at scale j while $\hat{f}^{(0)}$ is just the DC (i.e., the coarsest scale) information about the object. Thus, (26) decomposes the object in a natural way into components at different scales. From (26) we can write the recursion

$$\hat{f}^{(s+1)} = \hat{f}^{(s)} + \Delta\hat{f}^{(s)} \tag{28}$$

where the reconstruction at the current scale is obtained from the corresponding reconstruction at the next coarser scale by adding the corresponding level of incremental detail. The complete reconstruction, $\hat{f}$, is equal to the finest scale reconstruction $\hat{f}^{(J)}$ where $J = \log_2(N_s)$.

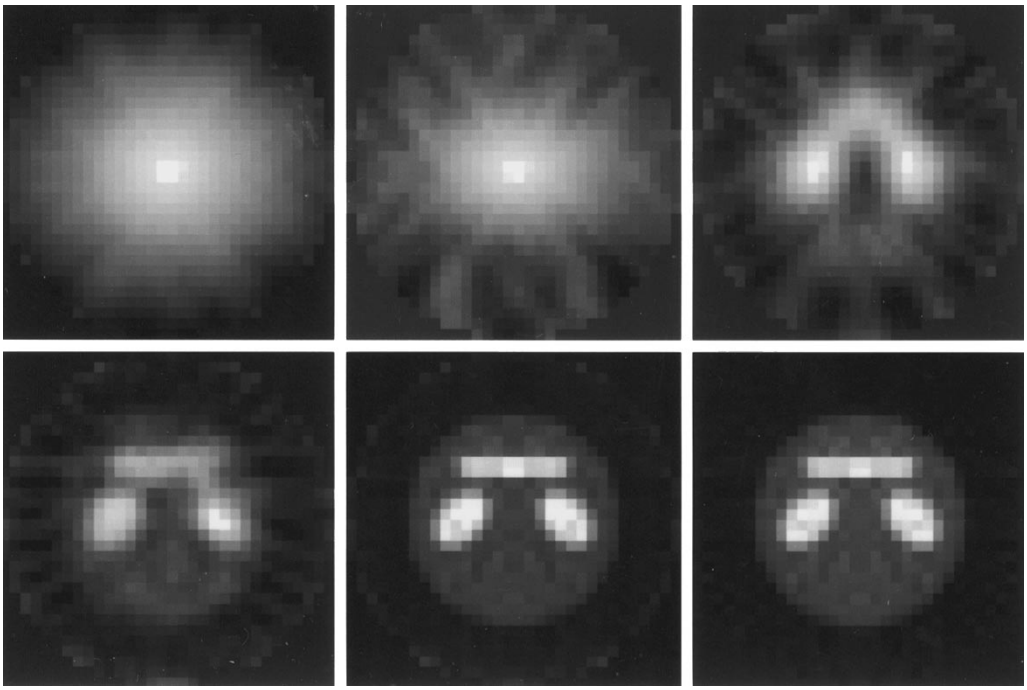

Fig. 7. Reconstructions at various scales using the D_3 wavelet and $N_\theta = N_s = 32$. Top row, left: $\hat{f}^{(1)}$. Top row, middle: $\hat{f}^{(2)}$. Top row, right: $\hat{f}^{(3)}$. Bottom row, left: $\hat{f}^{(4)}$. Bottom row, middle: $\hat{f}^{(5)}$. Bottom row, right: unwindowed ramp FBP reconstruction.

D. Computational Considerations

Our multiscale reconstruction procedure involves three steps: 1-D wavelet transformation of the data, solution of the set of (23), (24) for the multiscale coefficients, and backprojection of the results. The major computational burden lies in the second step, involving the solution of the large system of equations represented by (23) for the detail coefficients ξ^d. This will be the computational bottleneck of any matrix-based technique, including ART and NP, so we focus on this step. Typically, the solution of such large systems of equations found in matrix-based reconstruction methods is obtained by iterative methods, like the Kaczmarz method used in ART [19], [24], [25]. The computational complexity of such iterative schemes are proportional to the product of the number of computations required per iteration and the number of iterations needed for convergence. The number of computations required per iteration is itself proportional to the number of nonzero elements in the associated system matrix—which is $\mathcal{C}_{dd}$ for our approach, C of (7) for the standard NP approach, and T of (3) for ART. In addition, the number of iterations required for convergence depends on the condition number of the matrix. In particular, for poorly conditioned systems of equations, the convergence can be slow and will depend critically on the order that the elements are accessed and the value chosen for a relaxation parameter [19]. We discuss each of these points next.

1) Matrix Sparsity: While general choices of the wavelet transform operator in (16) lead to a multiscale matrix $\mathcal{C}_{dd}$ that is itself full, it appears that we may threshold a significant fraction of the elements in $\mathcal{C}_{dd}$ to zero with only slight impact on the resulting reconstructions. For example, when $N_\theta = N_s = 32$ and the D_3 wavelet is used in the definition of W_b in (16), then setting to zero all those elements of $\mathcal{C}_{dd}$ whose magnitude is below 3.75% of the absolute maximum yields reconstructions comparable to the full NP method in the noise-free, complete data case. In general, such a threshold would be chosen as large as possible consistent with the quality of the underlying data and goals of the problem. Thresholding of similar magnitude of the ART matrix T or NP matrix C produces severe artifacts in the reconstructions in addition to having much smaller effects on the sparsity of the resulting reduced matrix. This is because our transformation process has served to concentrate the "energy" in the matrix in fewer, higher amplitude elements. Exhaustive examination of these issues is a topic of current research.

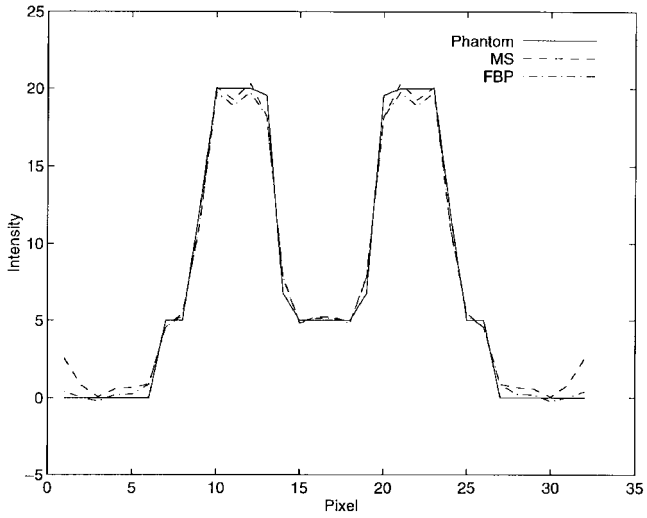

Fig. 8. A horizontal section through the phantom (solid line), the multiscale reconstruction (broken line), and the unwindowed ramp FBP reconstruction (dash-dot line).

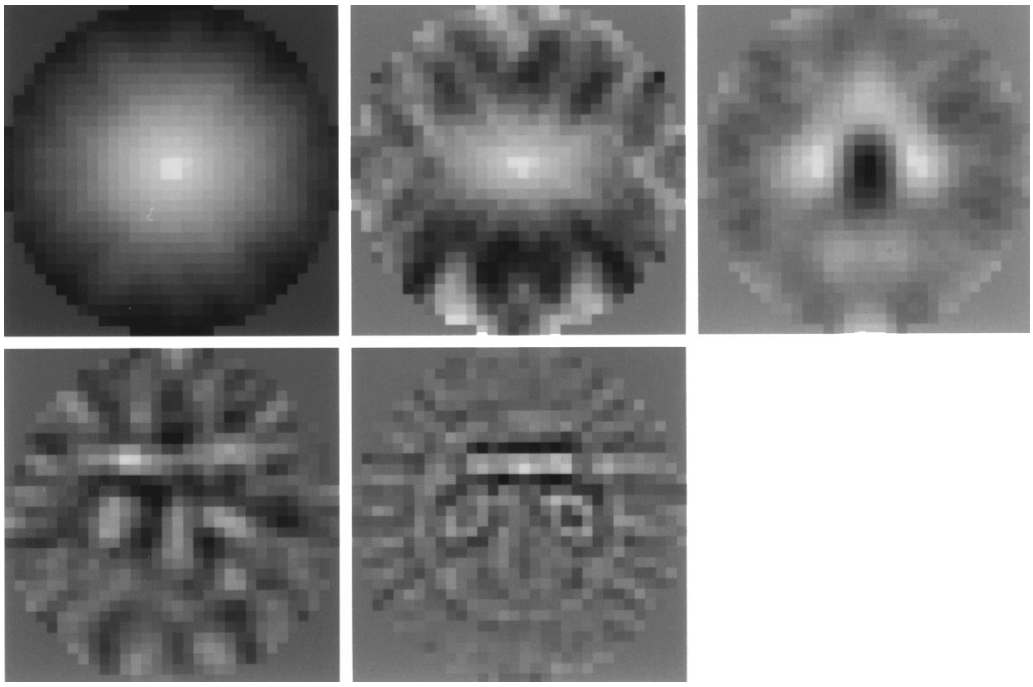

Fig. 9. Detail reconstructions at various scales, $\Delta\hat{f}^{(j)}$, using $N_\theta = N_s = 32$ and the D_3 wavelet. Top row, left: $\Delta\hat{f}^{(0)}$. Top row, middle: $\Delta\hat{f}^{(1)}$. Top row, right: $\Delta\hat{f}^{(2)}$. Bottom row, left: $\Delta\hat{f}^{(3)}$. Bottom row, middle: $\Delta\hat{f}^{(4)}$.

Fig. 5 shows the system matrices T of ART, C of the NP method, and a 3.75% thresholded version of $\mathcal{C}_{dd}$ from our multiscale NP-based method, from left to right for an imaging geometry with $N_\theta = 32$ angular projections and $N_s = 32$ strips per each angular projection. The ART matrix T is about 94% sparse (i.e., this percentage of the total number of elements are zero). In contrast, the NP matrix C is only about 39% sparse. Finally, the multiscale matrix $\mathcal{C}_{dd}$ thresholded at the 3.75% level is about 95% sparse. This implies that ART is about 25% more costly per iteration than our multiscale technique while the standard NP method will be about six times more costly than either. Also, recall that the effective degree of sparsity of $\mathcal{C}$ will naturally increase as the size of the field-of-view is increased, since relatively fewer of the pixels will be near an edge. Finally, note that most of the nonzero elements in $\mathcal{C}_{dd}$ correspond to the coarser scale terms where field-of-view edge effects are more pronounced. Thus, by focusing on reconstruction of only finer scale components, even greater gains may be obtained.

2) Matrix Conditioning: In addition to matrix sparsity (influencing the computations per iteration), matrix conditioning also has a bearing on the overall computational complexity of the reconstruction through the rate of convergence. Iterative schemes typically involve a choice of both the order in which the elements will be accessed and the value of a relaxation parameter. The choice of these parameters is typically done based on test images believed to be similar to those to be encountered, and *ad hoc*. For poorly conditioned system matrices, these choices coupled with the specific data and initial condition used will have a dramatic effect on the resulting convergence. Conversely, well conditioned systems will be largely insensitive to such effects; thus, good conditioning is highly desirable. One of the advantages of our multiscale transformation is that it results in a comparatively well-conditioned system of equations. For example, for the $N_s = N_\theta = 32$ case, the condition number of C and T are both orders of magnitude larger than that of the truncated $\mathcal{C}_{dd}$ (naturally, the full matrix $\mathcal{C}$ is a similarity transformation of C and so will have the same conditioning). Further, our experience has been that even with good choices for access order and relaxation parameter, our multiscale technique require at least half as many iterations as fast as ART or NP.

E. Examples

For all our example multiscale reconstructions the D_3 wavelet was used in the definition of W and, except where otherwise noted, the full matrix $\mathcal{C}_{dd}$ was used for simplicity. In Fig. 7 we show reconstructions $\hat{f}^{(j)}$ at various scales j, of the 32×32 phantom shown in Fig. 6, from projection data collected at $N_\theta = 32$ angles with $N_s = 32$ strips per angular projection. For this reasonably complete data case we expect the FBP reconstruction defined in (5) and the complete, finest scale, multiscale reconstruction to be similar since the NP and the FBP solutions converge in the case of complete data. This is precisely what is seen in Fig. 7 and confirmed in Fig. 8, which shows a section through the reconstructions. Finally, in line with the multiscale nature of our reconstructions, notice that the information about the phantom becomes more focused as we proceed from coarse to fine scales. In Fig. 9 we show the corresponding detail reconstructions $\Delta\hat{f}^{(j)}$ at various scales. Note that the finest scale detail reconstruction $\Delta\hat{f}^{(4)}$ (bottom row, middle in the figure) contains information about the edges and boundaries in the phantom.

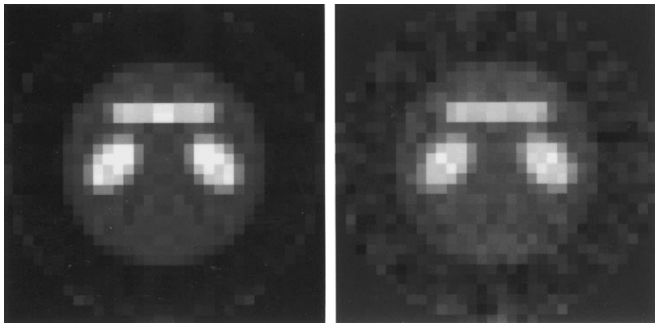

Fig. 10. Finest scale detail reconstructions using $N_\theta = N_s = 32$, and the D_3 wavelet. Left: Using the full $\mathcal{C}_{dd}$. Right: Truncating $\mathcal{C}_{dd}$ at 3.75% of its absolute maximum.

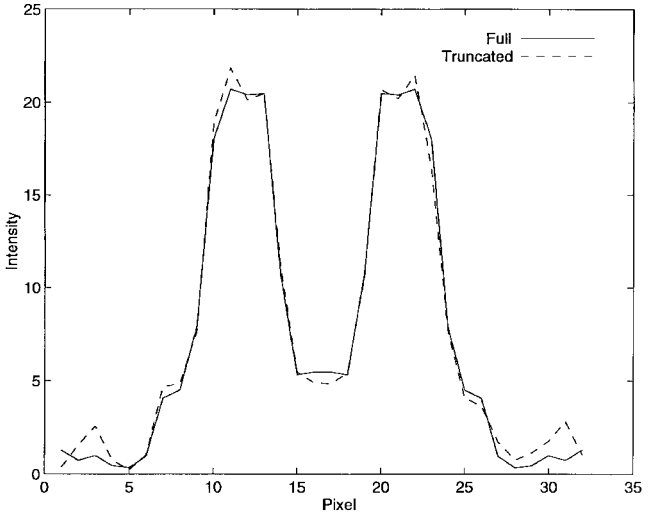

Fig. 11. Horizontal section through the full reconstruction (solid line) and the truncated reconstruction (broken line).

In Fig. 10 we show both the finest level multiscale reconstruction based on the full $\mathcal{C}_{dd}$ matrix (i.e., $\hat{f}^{(5)}$ of Fig. 7) together with the finest level reconstruction which results when $\mathcal{C}_{dd}$ is thresholded at the 3.75% level. Fig. 11 shows a section through the corresponding reconstructions.

In Fig. 12 we show an example of an incomplete data case. Here we reconstruct the phantom at different scales using $N_\theta = 5$ angular projections with $N_s = 32$ strips in each projection, and the D_3 wavelet for multiscale decomposition. We also show the corresponding (unwindowed ramp) FBP reconstruction for comparison. In this incomplete data case our NP-based multiscale reconstructions are free of many of the finest scale artifacts which arise in the FBP reconstruction. The figure also illustrates the resolution-accuracy tradeoff that is inherent in reconstructions from incomplete data, wherein the coarse scale reconstructions have fewer artifacts due to incomplete data effects at the expense of reduced resolution.

Finally, in Fig. 13, we show the finest scale detail reconstruction for the this case obtained by making the assumption that $\mathcal{C}_{dd}$ is an identity matrix. Such an assumption neglects all cross-scale and cross-angle terms in $\mathcal{C}_{dd}$ resulting in an even simpler inversion procedure. In the same figure, we also show for comparison the reconstruction based on finest scale detail from Fig. 12 that uses no approximation of $\mathcal{C}_{dd}$. We can see from the figure that if the goal is edge reconstruction, it is enough to approximate $\mathcal{C}_{dd}$ by an identity matrix. This reduces the computational complexity even further. Such an edge-oriented reconstruction, based on an identity assumption for $\mathcal{C}_{dd}$ in the multiscale framework, only requires 1-D wavelet transformation of the strip integral data and subsequent backprojection of just the fine scale coefficients.

IV. Regularized Multiscale Natural Pixel (NP) Reconstructions

In this section, we consider the estimation of an object f from *noisy* projection observations. We extend our multiscale reconstruction method presented in Section III to obtain statistically regularized estimates with the same efficiency as obtained for our unregularized estimates. This regularized solution is obtained by first solving for the MAP estimate [34] of the multiscale object coefficient vector, ξ, based on observations (19), a noise model, and a certain naturally derived multiscale prior model and then backprojecting these multiscale coefficient estimates along the corresponding multiscale basis functions as before.

In the presence of noise our original observations, (3), become

$$y = Tf + n, \qquad n \sim \mathcal{N}(0, \Lambda_n) \tag{29}$$

where the noise n is taken as an additive Gaussian vector and the notation $z \sim \mathcal{N}(m, \Lambda)$ denotes a Gaussian distribution of mean m and covariance Λ. Further, we assume that the elements of the noise vector n are uncorrelated, with equal variance at each angle, so that $\Lambda_n =$ block diag $(\lambda_1 I_{N_s}, \lambda_2 I_{N_s}, \cdots, \lambda_{N_\theta} I_{N_s})$, where λ_k is the variance of the noise at projection k and I_{N_s} refers to a $N_s \times N_s$ identity matrix. Note that this simple model allows for the possibility of nonstationary noise, arising, for example, from projections of differing fidelity.

By substituting the object representation (4) into (29), as was done in the noise-free case, we obtain the following equation (the noisy version of (7)):

$$y = Cx + n. \tag{30}$$

The multiscale decomposition of (30) using (18) and (20), followed by rearrangement in scales as described in Section III-A, results in the following partitioned equation (the noisy version of (22)):

$$\begin{bmatrix} \eta^d \\ \eta^a \end{bmatrix} = \begin{bmatrix} \mathcal{C}_{dd} & \mathcal{C}_{da}^T \\ \mathcal{C}_{da} & \mathcal{C}_{aa} \end{bmatrix} \begin{bmatrix} \xi^d \\ \xi^a \end{bmatrix} + \begin{bmatrix} \nu^d \\ \nu^a \end{bmatrix},$$
$$\begin{bmatrix} \nu^d \\ \nu^a \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \Lambda_{\nu^d} & 0 \\ 0 & \Lambda_{\nu^a} \end{bmatrix}\right) \tag{31}$$

where the vector $[(\nu^d)^T \mid (\nu^a)^T]^T$ contains the similarly scale rearranged and partitioned elements of the multiscale noise vector $\nu = W_b n$. Note that the block orthogonality of W_b means that Λ_ν will simply be a scale ordered version of Λ_n, and will still be diagonal. In particular, if $J = \log_2(N_s)$ is the scale at the finest level then

$$\Lambda_{\nu^d} = \text{block diag } (\lambda_1 I_{2^{J-1}}, \cdots, \lambda_{N_\theta} I_{2^{J-1}} \mid \lambda_1 I_{2^{J-2}}, \cdots, \lambda_{N_\theta} I_{2^{J-2}} \mid \cdots \mid \lambda_1, \ldots, \lambda_{N_\theta}) \tag{32}$$

$$\Lambda_{\nu^a} = \text{diag } (\lambda_1, \cdots, \lambda_{N_\theta}). \tag{33}$$

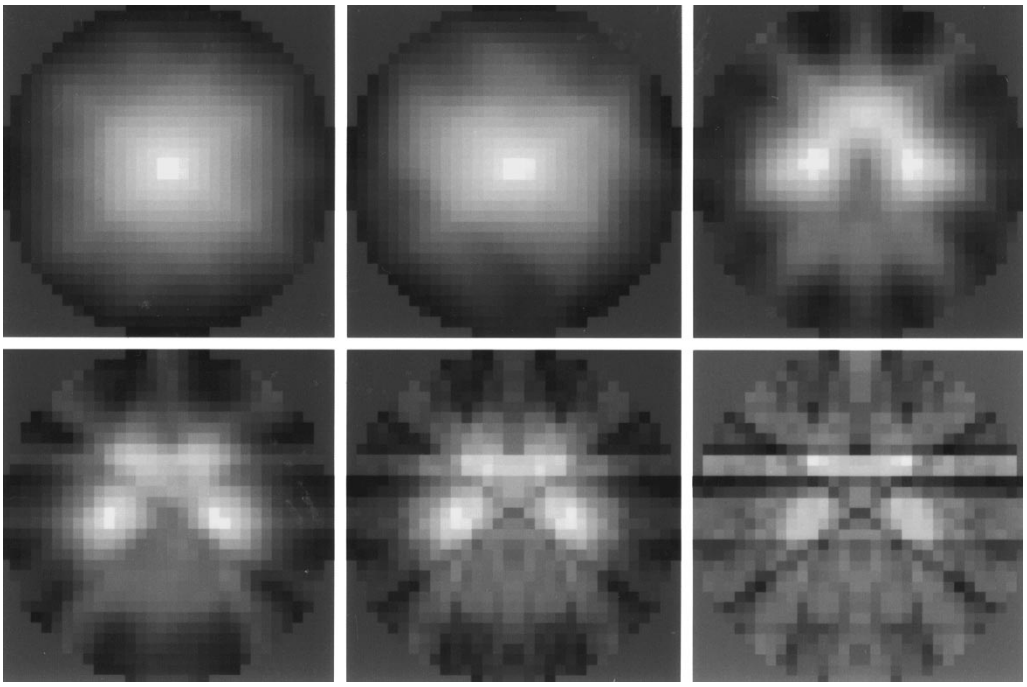

Fig. 12. Reconstruction at various scales using the D_3 wavelet and $N_\theta = 5$ and $N_s = 32$. Top row, left: $\hat{f}^{(1)}$. Top row, middle: $\hat{f}^{(2)}$. Top row, right: $\hat{f}^{(3)}$. Bottom row, left: $\hat{f}^{(4)}$. Bottom row, middle: $\hat{f}^{(5)}$. Bottom row, right: unwindowed ramp FBP reconstruction.

Recall from Section III-B that in the ideal case with no edge effects the off-diagonal blocks $\mathcal{C}_{da}$ in (31) are identically zero. More generally, as mentioned in Section III-B, due to finite field effects these blocks are not exactly zero, however reconstructions which assume these to be zero are visually indistinguishable from those that do not. As a result, in the development to follow we neglect these field-of-view edge effects and assume $\mathcal{C}_{da} = 0$ in our formulas.

A. Prior Model for the Multiscale Object Coefficients

To find the MAP estimate of ξ^d and ξ^a, we also need a prior statistical model for these quantities, i.e., we need a prior model for the multiscale object coefficients. The scale-based prior model we use for the object coefficients has been shown [4] to successfully capture the intuitively expected behavior of these coefficients while resulting in computationally efficient estimation algorithms. In particular, we assume ξ^d and ξ^a to be distributed according to

$$\begin{bmatrix} \xi^d \\ \xi^a \end{bmatrix} = \mathcal{N}\left(0, \begin{bmatrix} \Lambda_{\xi^d} & 0 \\ 0 & \Lambda_{\xi^a} \end{bmatrix}\right) \tag{34}$$

i.e., we assume ξ^d and ξ^a to be Gaussian, zero mean, independent, with variance Λ_{ξ^d} and Λ_{ξ^a}, respectively. For the variance of the approximation coefficients ξ^a, capturing the prior DC behavior of the object, we choose

$$\Lambda_{\xi^a} = \left(\frac{1}{\epsilon}\right) I_{N_\theta} \tag{35}$$

with ϵ sufficiently small (i.e., $\epsilon \to 0$) to prevent a bias in our estimate of the average (DC) behavior of the coefficients ξ, letting them be determined instead by the data, i.e., via maximum likelihood estimation.

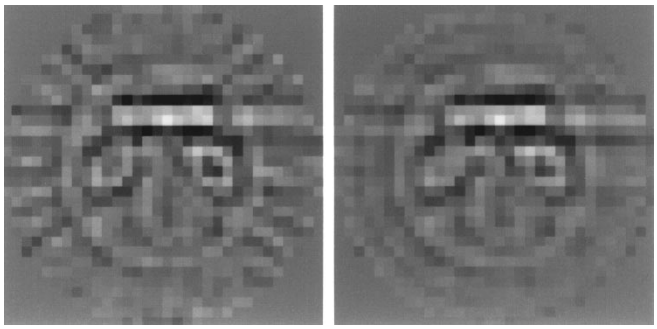

Fig. 13. Finest scale detail reconstructions using $N_\theta = N_s = 32$, and the D_3 wavelet. Left: Using the full $\mathcal{C}_{dd}$. Right: Assuming $\mathcal{C}_{dd}$ to be the identity matrix.

For the variance of the detail object coefficients ξ^d we use a self-similar prior model obtained by choosing the elements of the detail vector $\xi_k^{(m)}$ (i.e., the wavelet coefficients) at angle k and scale m as independent, $\mathcal{N}(0, \sigma^2 2^{-\rho m})$ random variables [36]. The parameter ρ determines the nature, i.e., the texture, of the resulting self-similar process while σ^2 controls the overall magnitude. This model says that the variance of the detail added in going from the approximation at scale m to the approximation at scale $m+1$ decreases geometrically with scale. If $\rho = 0$ the resulting finest level representation (the elements of x_k) corresponds to samples of white noise (i.e., the components of x_k are completely uncorrelated), while as ρ increases the components of x_k show greater long range correlation. This self-similar prior model results in a diagonal covariance matrix Λ_{ξ^d} for the detail coefficients ξ^d, the elements of which depend on the regularization parameters ρ and σ^2, i.e., the texture and the

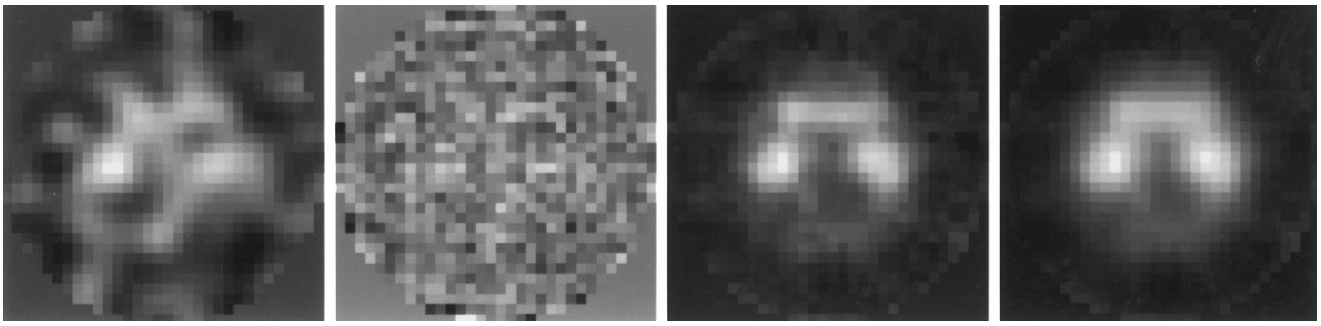

Fig. 14. Reconstructions with $N_\theta = N_s = 32$ and $\lambda_k = 360$ for k even and $\lambda_k = 36\,000$ for k odd to yield -2 dB overall SNR projection data. From left: (a) Hanning windowed FBP with $.3\pi$ cutoff filter; (b) Unregularized finest level multiscale reconstruction; (c) Regularized finest level multiscale reconstruction with $\rho = 0.2$ and $\sigma^2 = 360$; and (d) Regularized finest level multiscale reconstruction with $\rho = 0.7$ and $\sigma^2 = 360$.

overall magnitude, respectively. In particular, if $J = \log_2(N_s)$

$$\Lambda_{\xi^d} = \sigma^2 \text{block diag}\Big(2^{-\rho(J-1)} I_{2^{J-1}N_\theta} \mid 2^{-\rho(J-2)} I_{2^{J-2}N_\theta} \mid \cdots \mid I_{N_\theta}\Big). \tag{36}$$

Our self-similar, $1/f$-type prior model in the projection domain that is given above is equivalent to prior model in the original image domain which is also a $1/f$ self-similar process, but with a different, though related, fractal dimension [3], [21], [22]. Such self-similar models are commonly and effectively used in many application areas such as modeling of natural terrain and other textures, biological signals, geophysical and economic time series, etc. [8], [11], [26], [35], [36]. In addition, since the *observation noise* power is uniform across scales or frequencies, the geometrically decreasing variance of the prior model implies that the projection data most strongly influences the reconstruction of coarse scale features and the prior model most strongly influences the reconstruction of fine scale features. This reflects our belief that the fine scale behavior of the object (corresponding to high frequencies) is the most likely to be corrupted by noise.

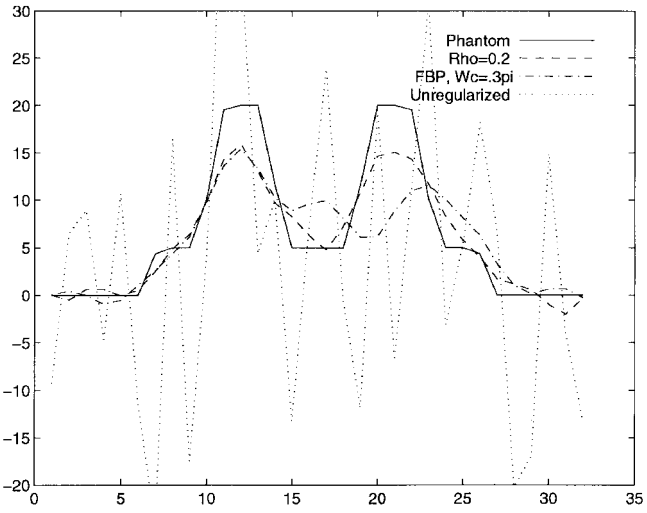

Fig. 15. Horizontal section through the phantom (solid line) and various reconstructions of the previous figure.

B. NP-Based Multiscale MAP Estimate

The MAP estimates of ξ^d and ξ^a based on the observations (31), the prior model (34)–(36) (with $\epsilon \to 0$), and assuming $\mathcal{C}_{da} = 0$ are given by

$$\hat{\xi}^d = \left(\Lambda_{\xi^d}^{-1} + \mathcal{C}_{dd}\Lambda_{\nu^d}^{-1}\mathcal{C}_{dd}\right)^{-1}\mathcal{C}_{dd}\Lambda_{\nu^d}^{-1}\eta^d \tag{37}$$

$$\hat{\xi}^a = \left(\Lambda_{\nu^a}^{-1/2}\mathcal{C}_{aa}\right)^{+}\Lambda_{\nu^a}^{-1/2}\,\eta^a \tag{38}$$

where $\Lambda_{\nu^a}^{-1/2}$ is the square root of the diagonal matrix $\Lambda_{\nu^a}^{-1}$ and we have used the fact that as $\epsilon \to 0$, $(\epsilon I_{N_\theta} + \mathcal{C}_{aa}\Lambda_{\nu^a}^{-1}\mathcal{C}_{aa})^{-1}\mathcal{C}_{aa}\Lambda_{\nu^d}^{-1} \to (\Lambda_{\nu^a}^{-1/2}\mathcal{C}_{aa})^{+}\Lambda_{\nu^a}^{-1/2}$. The structure of this estimator mirrors that of known optimal statistical estimators for stationary fields from a continuum of complete data [21], [22]. Note that in the case of angle-independent noise, where Λ_{ν^a} is a multiple of the identity, the estimate of the approximation coefficients ξ^a from (38) becomes the same as we had earlier in the unregularized case (24). Recall that (38) serves to estimate the DC component of the field. Most important, since $(\Lambda_{\xi^d}^{-1} + \mathcal{C}_{dd}\Lambda_{\nu^d}^{-1}\mathcal{C}_{dd})$ is still effectively a sparse, well-conditioned matrix due to the compression achieved in the multiscale domain, the discussion of Section III-D still holds and fast and efficient iterative algorithms [30] can be used to solve for $\hat{\xi}^d$ in (37). Note that this result is due both to the structure of the original problem, as reflected in $\mathcal{C}_{dd}$, *and* the structure of the prior, reflected in Λ_{ξ^d}. Finally, in obtaining the MAP estimates (37) and (38) we have assumed $\mathcal{C}_{da}$ to be zero. The inclusion of the effects of this neglected coupling into the MAP estimates, while negligible, is straightforward, as discussed in Section III-B.

Before proceeding, we discuss some aspects of the solution given in (37) and (38). First note that when the noise is stationary and the projections are distributed evenly, by symmetry our reconstruction is similar to applying a uniform filter to the FBP reconstruction. However, our approach directly ties the filter to prior models of both image and noise and thus can easily be adapted to other situations. One can view the change in the reconstruction induced by these new parameters as new filters on the FBP reconstruction, but our approach yields this automatically. In addition, because of the statistical basis of our approach, we obtain useful statistics (e.g., error

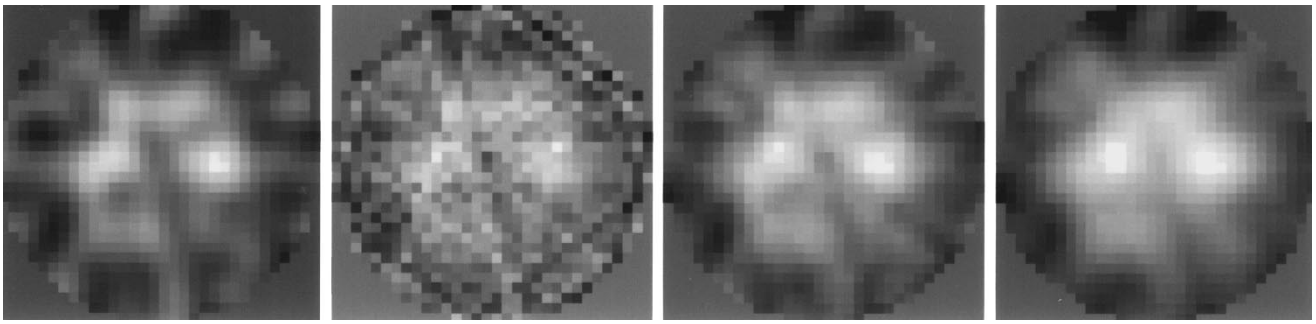

Fig. 16. Reconstructions with $N_\theta = 5$, $N_s = 32$, and $\lambda_k = 1,200$ for projection data with an SNR of about 10 dB. From left: (a) Hanning windowed FBP reconstruction with $.3\pi$ cutoff; (b) Unregularized finest level multiscale reconstruction; (c) Regularized finest level multiscale reconstruction with $\rho = 3$ and $\sigma^2 = 1.2 \times 10^6$; (d) Regularized finest level multiscale reconstruction with $\rho = 2$ and $\sigma^2 = 1.2 \times 10^5$.

variances) that can be used for tasks such as assessing the statistically proper scale of reconstruction, validating models, combining reconstructions with other information, detecting statistically significant anomalies, etc. Such applications are currently being pursued. Finally, our approach also directly applies to *nonstationary* prior models or noise without change to the algorithm. In such cases, the resulting reconstructions may again be viewed as corresponding in some sense to nonstationary, nonstandard filtering, but our approach provides it automatically.

C. Examples

Next, we show reconstructions using our regularized multiscale method in the presence of noise. The noise-free projection data are generated from the phantom of Fig. 6 and are then corrupted through the addition to projection k of independent, zero-mean Gaussian noise of variance Λ_k to yield our observations y. We also define an overall signal-to-noise ratio (SNR) for each example reconstruction as

$$\text{SNR (dB)} = 10 \log \frac{\|Tf\|^2}{N_s \sum \Lambda_k} \tag{39}$$

where Tf are the noise-free projection data. Finally, in all multiscale reconstructions we show here the Daubechies D_3 wavelet is used in the definition of the multiscale decomposition matrix W.

Fig. 14 shows a nonstationary example in which projection data was collected at $N_\theta = 32$ equally spaced angles with $N_s = 32$ strips per angular projection. Nonstationary noise was added with the standard deviation at the odd numbered angles ten times that of the even numbered angles ($\lambda_k = 360$, k odd, $\lambda_k = 36\,000$,$k =$ even) for an overall SNR of -2 dB. Finally, $\mathcal{C}_{dd}$ was thresholded so that all elements below 3.75% of the absolute maximum are set to zero. The figure shows the "best" (chosen by eye) windowed or "rolled-off" FBP reconstruction along with unregularized finest level multiscale, and various MAP regularized finest level multiscale reconstructions. In the unregularized finest level multiscale reconstruction the noise completely obscures the object. Through uniform rolling off of the ramp filter, the windowed FBP reconstruction is somewhat better, supressing some of the noise, but still losing detail. In contrast, in the regularized solution the details of the object are now easily visible. These reconstructions are obtained with essentially the same computational complexity as the unregularized solution, and naturally and automatically account for the nonstationarity in the noise. In both of the multiscale regularized reconstructions the overall magnitude of the prior model was $\sigma^2 = 360$; however, the reconstructions differ in the decay rate of the detail variance across scales, ρ. From Fig. 14 and Fig. 15, which show a section through the reconstructions, we see that as is expected, an increased regularization (i.e., smoothness) results when the value of ρ is increased from 0.2 to 0.5 (corresponding to a smoother prior texture) keeping other parameters fixed.

Fig. 16 shows a stationary example, where limited data was collected at $N_\theta = 5$ equally spaced angles with $N_s = 32$ strips per angular projection. Uniform noise was added at each angle ($\lambda_k =$ 1,200) for an overall SNR of about 10 dB. For this small example, the full $\mathcal{C}_{dd}$ matrix was used. The figure shows the best (by eye) windowed FBP reconstruction, along with the unregularized finest level multiscale, and various MAP regularized finest level multiscale reconstructions. Again, in the unregularized finest level multiscale reconstruction the noise and limited data serve to completely obscure the object. In contrast, in the regularized solutions the details of the object can now be discerned. The multiscale regularized reconstructions differ in both the the magnitude of the prior model used ($\sigma^2 = 1.2 \times 10^6$ versus $\sigma^2 = 1.2 \times 10^5$) and the decay rate of the detail variance across scales used ($\rho = 3$ versus $\rho = 2$). Note that the effect achieved for this stationary case is visually similar to the windowed FBP, but again, our method can adjust automatically to changes in noise or prior and also provides reconstructions at multiple resolutions and estimation error information, essentially for free.

V. Conclusion

We have developed a multiscale reconstruction technique based on the natural pixel (NP) approach that provides reconstructions from incomplete data yet is computationally efficient. Further, we extended this method to yield statistically optimal reconstructions from noisy data with essentially no additional computational complexity. This is in contrast to the conventional methods for image reconstruction from

incomplete data (such as NP) which only provide adequate reconstructions from high quality (i.e., noise-free) projection data, and additionally are computationally intensive. In addition, our reconstruction framework provides estimates of the field at multiple scales which is natural or desirable if the ultimate objectives are multiresolution in some way, for example if the interest is not to fully reconstruct the field but to gather information about aggregate (i.e., coarse scale) or fine scale (for example, boundaries) features of the field. Finally, the method presented here followed the standard NP development and assumed that the basis functions used to represent the underlying object were exactly and only those used to acquire the projection data (c.f. (3) and (4)). There is no fundamental reason, however, why we cannot use NP-type basis functions at more angles than we have data in the definition of (4) and construct priors that relate these sets of coefficients (which are relatively easy to capture within the context of our multiscale NP-based representation). The corresponding system matrix $\mathcal{C}$ in (20) will still be sparse and, coupled with appropriate priors, should lead to computationally feasible algorithms. Such extensions are currently under investigation.

REFERENCES

[1] J. R. Baker, T. F. Budinger, and R. H. Huesman, "Generalized approach to inverse problems in tomography: Image reconstruction for spatially variant systems using natural pixels," in *Critical Reviews in Biomedical Engineering*, vol. 20. CRC Press, pp. 47–71, 1992.

[2] G. Beylkin, R. Coifman, and V. Rokhlin, "Fast wavelet transforms and numerical algorithms I," *Commun. Pure Appl. Math.*, vol. XLIV, pp. 141–183, 1991.

[3] M. Bhatia, "Wavelet transform-based multi-resolution techniques for tomographic reconstruction and detection," Ph.D. dissertation, Dept. Elec. Eng. Comp. Sci., Mass. Inst. Technol., Sept. 1994.

[4] M. Bhatia, W. C. Karl, and A. S. Willsky, "A wavelet-based method for multiscale tomographic reconstruction," *IEEE Trans. Med. Imaging*, to be published.

[5] ——, "A wavelet-based method for multiscale tomographic reconstruction," , Lab. Inform. Decision Syst., Mass. Inst. Technol., Tech. Rep. LIDS-P-2182, 1993.

[6] M. H. Buonocore, W. R. Brody, and A. Macovski, "Fast minimum variance estimator for limited angle CT image reconstruction," *Med. Physics*, vol. 8, pp. 695–702, Sept./Oct. 1981.

[7] ——, "A natural pixel decomposition for two-dimensional image reconstruction," *IEEE Trans. Biomed. Eng.*, vol. BME-28, pp. 69–78, 1981.

[8] E. B. Cargill *et al.*,"Fractal physiology and nuclear medicine scans," in *SPIE Med. Imaging II*, vol. 914, pp. 355–361, 1988.

[9] D. J. Cha and S. S. Cha, "Improvement of a series expansion approach to interferometric tomography via natural pixel decomposition," *Opt. Eng.: J. Soc. Photo-Opt. Instr. Eng.*, vol. 32, July 1993.

[10] ——, "Natural pixel decomposition for interferometric tomographicreconstruction from ill-posed data," *Opt. Lasers Eng.*, vol. 20, pp. 325–339, 1994.

[11] C.-C. Chen, J. S. Daponte, and M. D. Fox, "Fractal texture analysis and classification in medical imaging," *IEEE Trans. Med. Imaging*, vol. 8, pp. 133–142, June 1989.

[12] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

[13] ——, "Wavelets on the interval," in *Progress in Wavelet Analysis and Applications*, Y. Meyer and S. Roques, Eds. Editions Frontieres, pp. 95–107, 1992.

[14] A. H. Delaney and Y. Bresler, "Multiresolution tomographic reconstruction using wavelets," *IEEE Trans. Image Processing*, vol. 4, pp. 799–813, June 1995.

[15] B. Dolveck-Guilpart, "Tomographic image reconstruction from a limited set of projections using a natural pixel decomposition," in *Proc. Multicentenn. Meet. Inverse Problems Inverse Methods Action*, P. C. Sabatier, Ed. NY: Springer-Verlag, Nov. 1990, pp. 54–61.

[16] N. H. Getz, "A perfectly invertible, fast, and complete wavelet transform for finite length sequences: The discrete periodic wavelet transform," in *Proc. SPIE Ann. Conf. Mathematical Imaging: Wavelet Applications in Signal and Image Processing*, July 1993.

[17] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1991.

[18] G. T. Gullberg and G. L. Zeng, "Reconstruction algorithm using singular value decomposition of a discrete representation of the exponential Radon transform using natural pixels," in *Proc. 1993 IEEE Nuclear Science Symp. and Medical Imaging Conf.*, vol. 3, pp. 1439–1443.

[19] G. T. Herman and L. B. Meyer, "Algebraic reconstruction techniques can be made computationally efficient," *IEEE Trans. Med. Imaging*, vol. 12, pp. 600–609, Sept. 1993.

[20] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[21] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[22] A. K. Jain and S. Ansari, "Radon transform theory for random fields and optimum image reconstruction from noisy data," in *Proc. ICASSP*, 1984, pp. 12A.7.1–12A.7.4.

[23] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. Piscataway, NJ: IEEE, 1988.

[24] J. Llacer, "Theory of imaging with a very limited number of projections," *IEEE Trans. Nucl. Sci.*, vol. NS-26, pp. 596–602, Feb. 1979.

[25] J. Llacer and J. D. Meng, "Matrix based image reconstruction methods for tomography," *IEEE Trans. Nucl. Sci.*, vol. NS-32, pp. 855–864, Feb. 1985.

[26] M. 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, pp. 41–64, Jan. 1994.

[27] S. G. Mallat, "A theory of multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, 1989.

[28] R. J. Michelena and J. M. Harris, "Tomographic traveltime inversion using natural pixels," *Geophys.*, vol. 56, pp. 635–644, May 1991.

[29] T. Olson and J. DeStefano, "Wavelet localization of the Radon transform," *IEEE Trans.Signal Processing*, vol. 42, pp. 2055–2067, Aug. 1994.

[30] C. C. Paige and M. A. Saunders, "LSQR: An algorithm for sparse linear equations and sparse least squares," *ACM Trans. Math. Software*, vol. 8, no. 1, pp. 43–71, Mar. 1982.

[31] F. Peyrin, M. Zaim, and R. Goutte, "Multiscale reconstruction of tomographic images," in *Proc. IEEE-SP Int. Symp. Time-Frequency Time-Scale Analysis*, Oct. 1992, pp. 219–222.

[32] B. Sahiner and A. E. Yagle, "Image reconstruction from projections under wavelet constraints," *IEEE Trans. Signal Processing*, vol. 41, pp. 3579–3584, Dec. 1993.

[33] B. Sahiner and A. E. Yagle, "Limited angle tomography using the wavelet transform," in *Proc. 1993 IEEE Nuclear Science Symp. and Medical Imaging Conf.*, Oct. 1993, pp. 219–222.

[34] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968.

[35] B. J. West and A. L. Goldberger, "Physiology in fractal dimensions," *Amer. Scientific*, vol. 75, pp. 354–365, 1987.

[36] G. W. Wornell and A. V. Oppenheim, "Estimation of fractal signals from noisy measurements using wavelets," *IEEE Trans.Signal Processing*, vol. 40, pp. 611–623, Mar. 1992.

[37] G. L. Zeng and G. T. Gullberg, "An iterative reconstruction algorithm with natural pixel basis," *J. Nucl. Med.*, vol. 34, p. 19P, 1993.

Mickey Bhatia received the B.S. in electrical engineering from the Indian Insitiute of Technology, Kharagpur, and the M.S. degree in biomedical engineering from Rutgers University, New Brunswick, NJ. He received the Ph.D. degree in radiological sciences from the Massachusetts Institute of Technology, Cambridge, in 1994.

He is currently an associate in Risk Management Research with J. P. Morgan & Co., New York, NY.

William C. Karl (M'91) received the S.M., E.E., and S.B. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, MA, and the Ph.D. degree in electrical engineering and computer science in 1991, also from MIT.

He held the position of Staff Research Scientist with the Brown–Harvard–MIT Center for Intelligent Control Systems and the MIT Laboratory for Information and Decision Systems from 1992 to 1994. He joined the faculty of Boston University, MA, in 1995, where he is currently Assistant Professor of electrical, computer, and systems engineering. Since January 1996, he has also held a joint appointment in the Department of Biomedical Engineering, Boston University. His research interests are in the areas of multidimensional and multiscale signal and image processing and estimation, geometric estimation, and medical signal and image processing.

In 1993 Dr. Karl was organizer and chair of the geometry and estimation session of the Conference on Information Sciences and Systems at Johns Hopkins University, Baltimore, MD. In 1994, he was on the technical committee for the Workshop on Wavelets in Medicine and Biology, part of the International Conference of the IEEE Engineering in Medicine and Biology Society. He is a guest editor of the 1997 special issue of the INTERNATIONAL JOURNAL ON PATTERN RECOGNITION AND ARTIFICIAL INTELLIGENCE, Processing, Analysis, and Understanding of MR Images of the Human Brain. He is also an associate editor for TRANSACTIONS ON IMAGE PROCESSING in the areas of tomography and magnetic resonance imaging.

Alan S. Willsky (S'70–M'73–SM'82–F'86) received the S.B. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1969 and 1973, respectively.

He joined the MIT faculty in 1973, and his present position is Professor of electrical engineering. From 1974 to 1981, he served as Assistant Director of the MIT Laboratory for Information and Decision Systems. He is also a founder and member of the Board of Directors of Alphatech, Inc. He has held visiting positions at Imperial College, London, L'Université de Paris-Sud, France, and the Institut de Recherche en Informatique et Systémes A1atoires, Rennes, France. His present research interests are in problems involving multidimensional and multiresolution estimation and imaging, discrete-event systems, and the asymptotic analysis of control and estimation systems. He is the author of the research monograph "Digital Signal Processing and Control and Estimation Theory," and is co-author of the undergraduate text *Signals and Systems*.

Dr. Willsky was program chairman for the 17th IEEE Conference on Decision and Control, has been an associate editor of several journals, including the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, and was guest editor of the 1992 special issue of the IEEE TRANSACTIONS ON INFORMATION THEORY on wavelet transforms and multiresolution signal analysis. He has served as a member of the Board of Governors and Vice President for Technical Affairs of the IEEE Control Systems Society and was program chairman for the 1981 Bilateral Seminar on Control Systems held in the People's Republic of China. In addition, in 1988, he was made a Distinguished Member of the IEEE Control Systems Society. He has also given several plenary lectures at major scientific meetings including the 20th IEEE Conference on Decision and Control, the 1991 1EEE International Conference on Systems Engineering, the SIAM Conference on Applied Linear Algebra in 1991, and the 1992 Inaugural Workshop for the National Centre for Robust and Adaptive Systems, Canberra, Australia. He was awarded the 1979 Alfred Noble Prize by the ASCE and the 1980 Browder J. Thompson Memorial Prize Award by the IEEE for a paper excerpted from his monograph. In 1975, he received the Donald P. Eckman Award from the American Automatic Control Council.