All-Purpose and Plug-In Power-Law Detectors for Transient Signals

Zhen Wang and Peter K. Willett, *Senior Member, IEEE*

***Abstract*—Recently, a power-law statistic operating on discrete Fourier transform (DFT) data has emerged as a basis for a remarkably robust detector of transient signals having unknown structure, location, and strength. In this paper, we offer a number of improvements to Nuttall's original power-law detector. Specifically, the power-law detector requires that its data be prenormalized and spectrally white; a constant false-alarm rate (CFAR) and self-whitening version is developed and analyzed. Further, it is noted that transient signals tend to be contiguous both in temporal and frequency senses, and consequently, new power-law detectors in the frequency and the wavelet domains are given. The resulting detectors offer exceptional performance and are extremely easy to implement. There are no parameters to tune. They may be considered "plug-in" solutions to the transient detection problem and are "all-purpose" in that they make minimal assumptions on the structure of the transient signal, save of some degree of agglomeration of energy in time and/or frequency.**

***Index Terms*—Crack detection, nonlinear detection, signal detection, sonar detection.**

I. Introduction and Context

A. Background

It is often of considerable interest to identify short-duration nonstationarities in observed signals. Applications include surveillance (e.g., [6]) in which an acoustic "transient" may indicate the presence of a threat, industrial monitoring (e.g., [16]), in which the number and severity of transients reflects machine health, and medicine (e.g., [2]). Naturally, the problem is comparatively simple if the signal to be detected is known—the only uncertainty is the time of occurrence, but knowledge of the transient is usually not available or dependable; of interest here is to detect transient signals with unknown form, location, and strength. The hypothesis test is naturally composite, with any structure open to challenge. Basically, the detector is tasked to determine whether all observations belong to a known stationary probability distribution or whether they do not.

Now, if there were nothing whatever that could be assumed about a transient signal, the detection task would be more or less hopeless. There are, fortunately, two rather qualitative properties that most transient signals possess. The first is the obvious temporal contiguity: A transient signal is often couched as a localized burst (or bursts) in time, although the duration of such a burst is unknown in most applications. The second is a tendency for most transient signals to be bandpass, that is, it is reasonable to expect most of a transient signal's energy to be contained in contiguous frequency observations, although again, there is usually little to be said about *which* frequencies.

To exploit only the former, and considering a transient event as a two-sided change (at some unknown time, the observations switches from having pdf f_0 to having pdf f_1, and at a later time, there is a return to f_0), Page's test has been explored and found to be quite useful [1], [5]. Very similar to this, Nuttall couched a transient as a contiguous burst of M bins in time, where M is known, and developed the "maximum" detector [13]. To exploit only the latter, there are detectors that begin their work on frequency domain data (usually DFT bins). Via (maximum likelihood) estimation of unknown signal parameters via the estimation–maximization (EM) algorithm, a GLRT approach is presented [21]. Of greatest interest here is Nuttall's frequency domain "power-law" detector [12], which will be introduced shortly. It is natural to use both kinds of contiguities, and for this, we have, for example, the Gaussian-mixture time-spectrogram model in [17] and the GLRT approaches arising from linear data transformations (either time-frequency or time-scale) [3], [10], [11]. These transforms are directed toward signal representation and classification, trying to distinguish signals in the transform domain.

In [23], an attempt was made to compare the performances of a number of the above transient detection approaches on a fairly wide variety of signals. Those using time contiguity alone (Page and "maximum") were perhaps the sturdiest performers overall but suffer from the need that certain parameters (signal strength or length) be prespecified and that data be prewhitened and prenormalized. Among the others, it was surprising that the most robust performance came from the simplest processor: Nuttall's frequency-domain power-law statistic. It is a very good detector indeed, and in this paper, we show a number of ways to make it better still.

B. Nuttall's Power-Law Statistic

There has been significant recent attention to Nuttall's power-law detector [12], [14] due to its simple implementation and good performance. The test is based on the following formulation. Under the signal-absent hypothesis (H_0)—that the time-domain data is complex white Gaussian noise—preprocessing by the magnitude-square DFT yields independent and identically distributed (iid) exponential random variates. Under

Manuscript received August 11, 2000; revised June 11, 2001. This work was supported by the Naval Undersea Warfare Center under Contract N66604-99-1-5021 and by the Office of Naval Research under Contract N00014-98-1-0049. The associate editor coordinating the review of this paper and approving it for publication was Dr. Vikram Krishnamurthy.

The authors are with the Department of Electrical and Systems Engineering, University of Connecticut, Storrs, CT 06269-2157 USA (e-mail: willett@engr.uconn.edu).

Publisher Item Identifier S 1053-587X(01)07771-6.

the signal-present hypothesis (H_1), the DFT observations are no longer a homogeneous population of exponentials; Nuttall's basic assumption is that there are *two* exponential populations:

$$\mathbf{H}_0\colon f(\mathbf{X}) = \prod_{k=1}^{N} \frac{1}{\mu_0} e^{-X_k/\mu_0} u(X_k)$$
$$\mathbf{H}_1\colon f(\mathbf{X}) = \prod_{k\notin S} \frac{1}{\mu_0} e^{-X_k/\mu_0} u(X_k) \cdot \prod_{k\in S} \frac{1}{\mu_1} e^{-X_k/\mu_1} u(X_k) \tag{1}$$

where

$u(\cdot)$ unit step function (unity for positive argument and zero otherwise);
N total number of FFT bins;
$\mathbf{X}$ magnitude-squared FFT bins;
S subset with size M.

It is assumed that M signal-present bins are uniformly distributed among the N FFT bins. Clearly, the precise probability law under $\mathbf{H}_1$ depends on the transient signal itself, and there is no particular reason to take (1) as fact. Nevertheless, there is considerable flexibility in (1) (mostly through the unspecified S), and the detector arising from it seems to work remarkably faithfully.

At any rate, dealing with the above model, Nuttall developed power-law statistics [12] as an approximation to the optimal detector, and these have the form

$$T(\mathbf{X}) = \sum_{k=1}^{N} X_k^{\nu} \tag{2}$$

where ν is an adjustable exponent. Notice that $\nu = 1$ is the energy detector that is optimal for $M = N$, and $\nu = \infty$, which is the maximum-magnitude FFT bin, corresponds to the GLRT for $M = 1$. Through extensive computational work, it has been found that the best compromise value for ν is 2.5 when information about M is completely unavailable. Performance of this particular power-law detector is close to the best in this class of detectors. This independence from M of the power-law is fortunate.

The above power-law statistic requires prenormalized data, meaning that μ_0 in the model must be available. As an extension of power law to unknown noise level (μ_0) cases, a constant false-alarm rate (CFAR) version was introduced [15]:

$$T_{cpl}(\mathbf{X}) = \frac{\sum_{k=1}^{N} X_k^{\nu}}{\left(\sum_{k=1}^{N} X_k\right)^{\nu}}. \tag{3}$$

Clearly, T_{cpl} is not affected by a scale factor.

The statistic (2) does a yeoman's job at detecting a wide variety of block inhomogeneities, and it might be wondered why this paper, intending to improve on it, has been written. The answer is three-fold, as follows.

1) The statistic (2) is designed with white noise of known power in mind; the fact is that the performance of (3) is disappointing in white noise, whereas for colored noise, it has very little appeal at all. We thus extend (2) in a natural way. We estimate the noise power and normalize on a bin-by-bin basis.
2) The statistic (2) is essentially optimal [12] given its frequency-domain model of (1) when there is nothing whatever known about the signal-bearing set S. However, there is some tendency for real transient signals to aggregate their energy in a band, meaning that S has some structure. The challenge is to take advantage of this tendency when it exists while avoiding any degradation in performance when it does not. We believe that we have achieved this through the simple expedient of combining contiguous DFT bins.
3) Similar to the previous point, there is a definite tendency for real transient energy to be agglomerated in the time domain, and a magnitude-square DFT essentially destroys any such information, but there is no reason why a DFT must be the preprocessing step. We investigate the (obvious) extension that a transform other than the DFT be used.

Basically, the power-law detector is as yet neither a plug-in solution nor is it as good as it can be, and we offer some remedy here.

The organization of the paper is as follows. In Section II, we first describe the detection problem for the colored noise case and derive the associated CFAR (bin-by-bin normalized) power-law statistics. It is necessary to revisit the assumptions by which the adjustable exponent ν [see (2)] is set, and we propose a measure by which it should be chosen. In Section III, we propose extensions to exploit the contiguities of the transient and thus develop new detectors both in the frequency and the wavelet domain. It is unsatisfying to report on new detector structures without advice in threshold-setting, and in Section IV, we derive both the normal and saddlepoint approximations to the signal-absent distributions, and naturally, we compare these to simulation. Numerical comparisons between the detectors are presented in Section V, and we offer concluding remarks in Section VI.

II. CFAR Power-Law Detector

A. Problem Description and the CFAR Power-Law Statistic

The focus of this section is to detect transients buried in colored noise with unknown but stationary spectrum. Clearly, the CFAR Power-Law in [15, eq. (3)] is to be applied to white noise and is not suitable here. As shown in Fig. 1, we write in a matrix a block of NL time domain observations as $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L)$, where $\mathbf{x}_i$ is a column vector of dimension N whose kth element is the time sample of index $(i-1)L + k$.[1] We immediately transform each column to its magnitude-squared frequency domain equivalent $\mathbf{X}_i$, and record $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_L)$. It is assumed that X_{ji}s

[1]To avoid possible contamination of this "reference" dataset by a transient's incipient edge, it is best to ignore a "guard" of a few blocks of data prior to that under test.

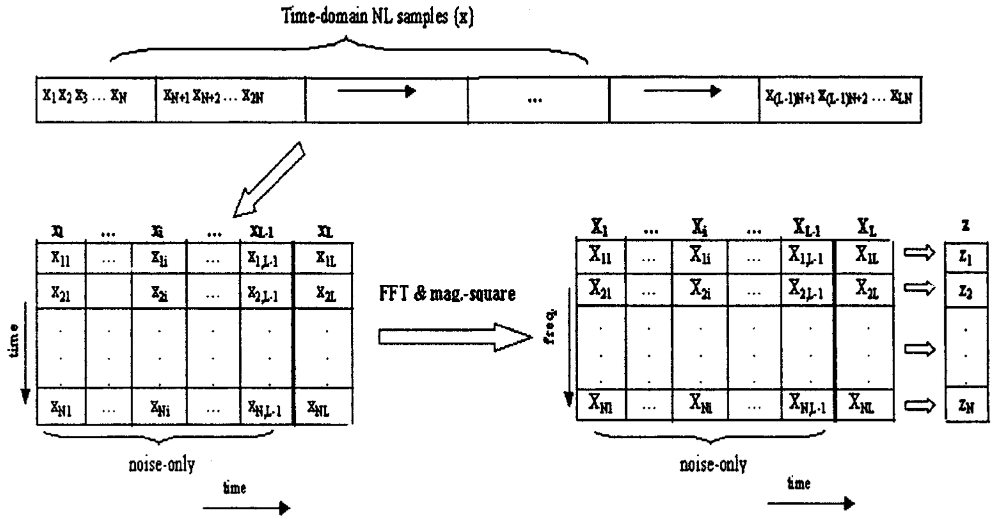

Fig. 1. Data and preliminary processing. Original time-domain sequence is reorganized into blocks, and the column-wise magnitude-square FFT is performed.

are independent and that $(\mathbf{X}_1, \ldots, \mathbf{X}_{L-1})$ are known to be noise-only samples.[2] The probability density function (pdf) of the jth element of X_i, $i = 1, 2, \ldots, L-1$ has the form

$$f(X_{ji}) = \frac{1}{\beta_j} \exp\left(\frac{-X_{ji}}{\beta_j}\right) u(X_{ji}) \tag{4}$$

where β_j are unknown but stationary.

Note that the spectral behavior of a nonwhite background is faithfully represented by the N βs. It is assumed that for the entire block of observation, the jth frequency bin maintains a mean background energy level β_j for each of the L blocks of N data. The first $L-1$ blocks provide some estimate of this level for each bin, and the goal is to test for some elevation in these levels in the *last* (Lth) block. More specifically, the pdf of the X_{jL} under hypothesis H_0 follows a distribution identical to X_{ji}, $i = 1, 2, \ldots, L-1$. On the other hand, when signal energy is present in the jth bin, the density of X_{jL} becomes

$$f(X_{jL}) = \frac{1}{\beta_j(1+s)} \exp\left(\frac{-X_{jL}}{\beta_j(1+s)}\right) u(X_{jL}) \tag{5}$$

where s is the relative signal power per bin, that is, the overall transient signal energy is $S_t = Ms$, in which M is the number of signal-energy-bearing DFT bins. Overall, we have the model

$$\mathbf{H}_0\text{: } f(\mathbf{X}) = \prod_{i=1}^{L}\prod_{j=1}^{N} \frac{1}{\beta_j} \exp\left(\frac{-X_{ji}}{\beta_j}\right) u(X_{ji})$$

$$\mathbf{H}_1\text{: } f(\mathbf{X}) = \left[\prod_{i=1}^{L-1}\prod_{j=1}^{N} \frac{1}{\beta_j} \exp\left(\frac{-X_{ji}}{\beta_j}\right) u(X_{ji})\right] \cdot \left[\prod_{j\in S} \frac{\exp\left(\frac{-X_{jL}}{\beta_j(1+s)}\right)}{\beta_j(1+s)} \prod_{j\notin S} \frac{\exp\left(\frac{-X_{jL}}{\beta_j}\right)}{\beta_j}\right] u(X_{jL}) \tag{6}$$

where S indicates the (unknown) subset with (unknown) size M out of N bins in which transient signal energy is to be found. Note that although this model may appear to have a batch flavor, Fig. 1 indicates that transient signals are to be detected on-line, although block-by-block. Each block of data that tests negative for transient signal energy joins the "window" of reference data, and hence, the least-recent block is removed to make room for it.

Following ideas similar to those frequently used in radar CFAR processing (e.g., [4]), we define the normalized magnitude-squared frequency-domain observations as

$$z_j = \frac{X_{jL}}{\frac{1}{L-1}\sum_{i=1}^{L-1} X_{ji}} \tag{7}$$

and the new power-law statistic as

$$T_{fc}(\mathbf{X}) = \sum_{j=1}^{N} z_j^{\nu} \tag{8}$$

where ν is a real exponent. Clearly, $T_{fc}(\mathbf{X})$ is non-negative and is CFAR with respect to $\{\beta_j\}$ in the model of (6). Note that normalization schemes alternative to that in (7) could be chosen. For example, one could define $z_j = X_{jL}/X_{j(R)}$ in which $X_{j(R)}$ denotes the element among $\{X_{ji}\}_{i=1}^{L-1}$ whose rank is R. It is possible that such a scheme would offer improved robustness [4], but due to its similarity to (8), we do not discuss it here.

B. SNR Analysis to Choose ν

The best value for the power ν in (8) is, in general, strongly dependent on M, which is the number of signal-present bins.

[2]This assumption of independence is in practice only approximate. In what follows, for analysis, we use the assumption; our simulations are based on time domain signals, and naturally, there is a truer representation of the dependency structure.

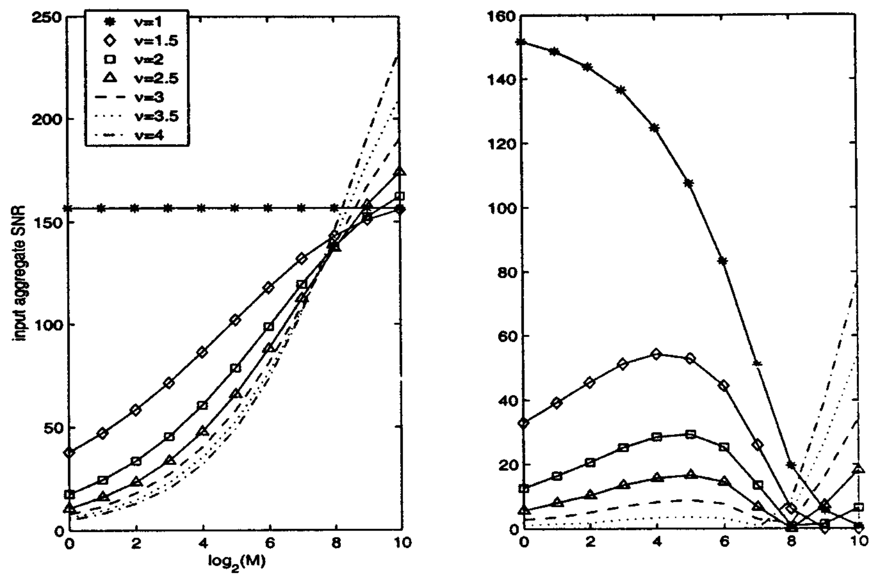

Fig. 2. SNR analysis for Nuttall's power-law statistics, with settings output $SNR = 24$, $N = 1024$. (Left) Required input aggregate SNR for power-law detectors with different ν. (Right) Input SNR-loss for different ν.

This is not at all desirable since our goal is to find a detection structure that does not depend on knowledge of such signal qualities. A clever contribution of [12] was the so-called "low-quality operating point analysis," and it was found that $2 < \nu < 3$ is a good choice over a wide range of M. This analysis depended on explicit numerical calculation of performance, and although such analysis was possible for the exponential random variables in (2), and indeed *would* be an option in (8), in some of the later statistics, it is not. Thus, here, we investigate a similar SNR analysis to suggest the best choice of ν in (8) when information about M is unavailable. Signal-to-noise ratio (SNR), which is sometimes known as deflection [19], is not a completely accurate determinant of detection performance but is a widely accepted alternative to exhaustive simulation or numerical integration. Given a statistic T, the output SNR can be expressed as

$$\text{SNR}_T = \frac{(E(T|\mathbf{H}_1) - E(T|\mathbf{H}_0))^2}{\text{Var}(T|\mathbf{H}_0)} \tag{9}$$

where $E(.|.)$ denotes the conditional expectation, and var$(.|.)$ is the conditional variance.

First, we exploit the SNR analysis to evaluate ν for the power-law statistics in [12]. The coincidence between our results and Nuttall's suggests that the SNR analysis is a reasonable method to choose the power ν for our new CFAR statistics. Based on statistic (2) and model (1), the associated SNR can be shown to be

$$\text{SNR}_{pl} = \frac{(M\Gamma(\nu+1)[(1+S_t/M)^\nu - 1])^2}{N[\Gamma(2\nu+1) - \Gamma(\nu+1)^2]} \tag{10}$$

where $\Gamma(.)$ represents the Gamma function, and S_t denotes the total signal power in M bins, which is also referred to as the input aggregate SNR. Our purpose is to evaluate the required S_t to yield fixed output SNR_{pl} for each ν. Example results from SNR analysis for the power-law detectors with different ν are shown in Fig. 2, where $N = 1024$, and the output $SNR = 24$.

The right plot in Fig. 2 shows the input SNR-loss (ISL), which, with fixed output SNR N and M is defined as

$$\text{ISL}(\nu, M) = S_t(\nu, M) - \min_{\nu}\{S_t(\nu, M)\}. \tag{11}$$

The ISL measures the input aggregate SNR that is sacrificed through use of a fixed exponent ν, as compared with the best possible exponent ν for that M or the corresponding optimal statistic. At any rate, it is immediately seen that the best value of ν, achieving minimum average signal power per bin, changes with M and sweeps through all intermediate values. When M is completely unknown, we obtain the best compromise value for ν via $\nu = \min_\nu\{\max_M\{\text{ISL}\}\}$; from the figure, we find that $\nu = 2.5$ is that choice. The tendencies and results coincide well with those obtained from the low-quality operation point analysis in [12]. Based on this, we claim, as in [12], that $2 < \nu < 3$ is a good choice for a wide range of M.

Encouraged by the above, we apply the input SNR loss analysis to the detector in (8) to select ν. It is straightforward to derive the pdf of z_j under H_0 and H_1 [8], and hence, we get, after some algebra, the result

$$\text{SNR}_{fc} = \frac{M^2(L-1)\,(\beta(L-\nu-1,\nu+1)[(1+S_t/M)^\nu - 1])^2}{N[\beta(L-2\nu-1,2\nu+1)-(L-1)\beta(L-\nu-1,\nu+1)^2]} \tag{12}$$

where $\beta(.,.)$ denotes the Beta function and S_t represents the total signal power in M bins, and $L-1$ blocks of previous DFT outputs are used for normalization. Example results for power values ν =1, 1.5, 2, 2.5, and 3, are shown in Fig. 3, where $N = 256$, output $SNR = 6$, and $L = 10$.[3] It is clear that there is no reason to explore $\nu < 1$. It is noted that the best value of ν,

[3] The larger the window size L, the better the normalization, but the more susceptible the detector to a nonstationary background. The range $6 \le L \le 32$ is often discussed [4] and provides reasonable results; we choose $L = 10$ here as representative of that range.

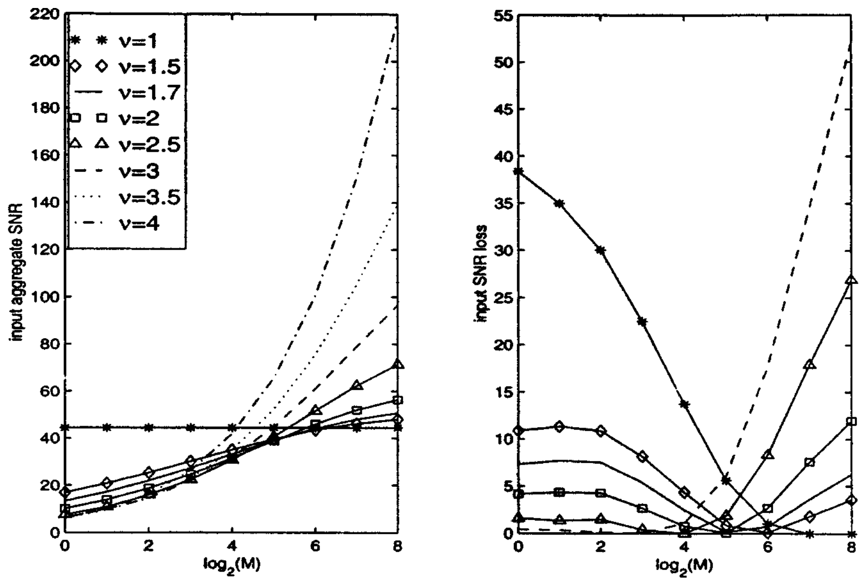

Fig. 3. SNR for CFAR power-law statistics, with settings the output $SNR = 6, N = 256$. Left figure: SNR for different ν; right figure: the input SNR loss for different ν.

providing the minimum average signal power per bin with given output SNR, changes with M. Similar results will be observed by setting different N and output SNR: $1.5 < \nu < 2$ is a good choice when information of M is completely unknown, as it yields the least ISL over M. Here, $\nu = 1.7$ appears to be the best choice.

III. Contiguity-Based Detectors

The knowledge of signal contiguity may aid in detection. Both time and frequency contiguities are exploited in both the white- (prenormalized) and colored-noise (self-normalizing) cases to improve the detection performance. For each case, the model and the corresponding statistics are described. The detector is actually a combination of a linear and a power-law processor. Since precise contiguity information is unavailable, only the cases of two and three adjacent bins are studied in this paper. Theoretical justification for these detectors is not offered; however, they make intuitive sense, and they work well. Further, although the exponent ν could be chosen differently for difference numbers of aggregated bins, we have found that this is not a major concern, and we choose ν from the single-bin analyses of the previous section.

A. Contiguity-Based Detectors in the Frequency Domain

Just as with the power-law detectors in (2), only magnitude-squared FFT outputs are of concern here. Using the contiguity tendency in frequency, we modify Nuttall's assumption that the M signal-present bins are uniformly distributed amongst the record of N to an assumption that there is a tendency that some of the M signal-occupied bins are adjacent.

1) Prenormalized Case: New random variables are obtained by combining two contiguous frequency bins. We define $U_j = X_{j-1} + X_j$, $j = 1, \ldots, N$. Assuming that the original $\{X_j\}$ are independent and exponential (that is, that this is Nuttall's model in which data are assumed already to have been normalized and whitened), U_j yields a *Gamma*$(2, \mu_0)$ random variate.

We define our new power-law detectors

$$T_{f2}(\mathbf{U}) = \sum_{j=1}^{N} U_j^{\nu} = \sum_{j=1}^{N} (X_{j-1} + X_j)^{\nu} \tag{13}$$

where $\{X_j\}$ and N have same meanings as in (1). The statistic of (13) is easily extended as

$$T_{f3}(\mathbf{U}) = \sum_{j=1}^{N} U_j^{\nu} = \sum_{j=1}^{N} (X_{j-2} + X_{j-1} + X_j)^{\nu} \tag{14}$$

to the case of three contiguous bins, and further extension is straightforward. Bins are indexed modulo $-N$.

2) Self-Normalizing Case: A similar combining process was adopted in the colored noise case by letting $U_{ji} = X_{j-1,i} + X_{j,i}$. This combining approach results in modified model and generates a new CFAR power-law detector in the frequency domain as

$$T_{fc2}(\mathbf{U}) = \sum_{j=1}^{N} \left(\frac{U_{jL}}{\frac{1}{L-1}\sum_{i=1}^{L-1} U_{ji}} \right)^{\nu}. \tag{15}$$

The similar detector T_{fc3} combines three contiguous bins.

B. Detectors in the Wavelet Domain

For time-domain observations, the DFT transforms a pure "time description" into a pure "frequency description" and, thus, clearly cannot take advantage of time contiguity. The discrete-time wavelet transform (DWT), which is an alternative to the DFT, is much more local and finds a good compromise—a time–frequency description. Hence, detectors in the wavelet domain will benefit from both temporal and frequency contiguity tendencies. The original work of Nuttall explored only the case the that preprocessing transformation was the

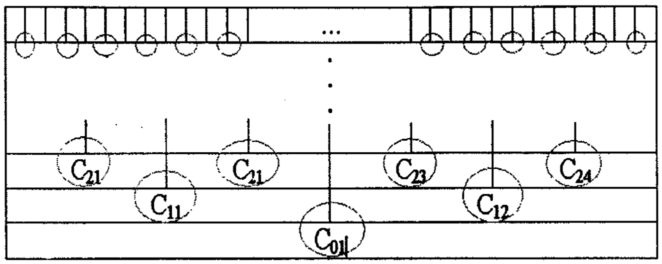

Fig. 4. Structure used in T_{w3} and T_{wc3} to combine three adjacent bins in the wavelet space. Circles illustrate the definitions $U_{kj} = C_{kj} + C_{k+1,2j-1} + C_{k+1,2j}$.

DFT. The extension to other transforms, especially the wavelet transform, is both natural and (mostly) straightforward. There are many different choices of wavelet family, and each has its proponents. However, only the simple Haar wavelet is explored due to its easy implementation, its orthogonality [20], and due to the fact that a statistic that assumes as little as possible about the transient to be detected is preferable.

1) Prenormalized Case: A derivation similar to that of Nuttall results in the power-law detector in the wavelet domain, considering that under a complex Gaussian noise assumption the magnitude-squared (orthonormal) DWT of the noise-only data obeys an iid exponential distribution. That is, we have

$$T_w(\mathbf{C}) = \sum_{k=0}^{K-1} \sum_{j=1}^{2^k} C_{kj}^{\nu} \tag{16}$$

where $K = \log_2(N)$, C_{kj} is the jth magnitude-squared DWT coefficient of k scale. The argument is that if time observations $\mathbf{x}$ follow an iid normal distribution, the corresponding DWT vector has identical pdf [22].

In the case of preprocessing by the DFT, it was argued that there is often a tendency for transient energy to crowd into contiguous frequency bins. There is similarly a tendency for transient energy to be in nearby wavelet coefficients, as each refers to a scale that roughly matches that of the transient. As shown in Fig. 4, it is natural to adopt a tree structure in the WT case, and similar to T_{f3} of (14), we define $U_{kj} = C_{kj} + C_{k+1,2j-1} + C_{k+1,2j}$, $k = 0, 1, \ldots, K-1$ and $j = 1, \ldots, 2^k$ and invoke

$$T_{w3}(\mathbf{U}) = \sum_{k=0}^{K-1} \sum_{j=1}^{2^k} U_{kj}^{\nu}. \tag{17}$$

Clearly, this detector is obtained by combining three local adjacent bins in the wavelet space. It would be possible to combine two adjacent WT samples at the same scale, but it has proven to be less effective than hoped, and we do not report it here.

2) Self-Normalizing Case: For each column time-domain vector $\mathbf{x}_i$, let the $C_{k,j,i}$s be the corresponding magnitude-squared DWT coefficients for each scale index $k = 0, 1, \ldots, K-1$, intra-block time index $j = 1, \ldots, 2^k$, and block index $i = 1, 2, \ldots, L$. Similar to the frequency domain, we have

$$T_{wc} = \sum_{k=0}^{K-1} \sum_{j=1}^{2^k} z_{kj}^{\nu} \tag{18}$$

where

$$z_{kj} = \frac{C_{k,j,L}}{\frac{1}{L-1}\sum_{i=1}^{L-1} C_{k,j,i}}.$$

We record $U_{kji} = C_{k,j,i} + C_{k+1,2j-1,i} + C_{k+1,2j,i}$, for $k = 0, 1, \ldots, K-1$, $j = 1, \ldots, 2^k$, and $i = 1, 2, \ldots, L$. As in the frequency domain, this combining approach suggests a new CFAR power-law detector

$$T_{wc3}(\mathbf{U}) = \sum_{k=0}^{K-1} \sum_{j=1}^{2^k} \left(\frac{U_{k,j,L}}{\frac{1}{L-1}\sum_{i=1}^{L-1} U_{k,j,i}} \right)^{\nu} \tag{19}$$

in the wavelet domain.

IV. Performance Analysis

Since we are interested in transients with unknown structure, location, and strength, our performance analysis will concentrate on the prediction of the threshold exceedance probability of the statistics under $\mathbf{H}_0$ and, thus, to choose a proper threshold to ensure a certain false alarm rate. We use the central limit theorem (CLT) to get the normal approximation to the distribution of the statistics. However, since the normal approximation provides poor approximation to deep tail probabilities, a procedure using saddle-point approximation is also introduced. The performance of different statistics are also studied via numerical simulation, where we set the total number of bins $N = 256$.

In the following analysis, we only consider detectors in the frequency domain; analysis in the wavelet domain is precisely equivalent, provided the transform is orthogonal.

A. Normal Approximation

- *The Statistic T_{fc}:*

  This is a summation over iid random variables and, thus, must converge to a normal distribution by the central limit theorem [18]. Recall

$$T_{fc}(\mathbf{X}) = \sum_{j=1}^{N} \left(\frac{X_{jL}}{\frac{1}{L-1}\sum_{i=1}^{L-1} X_{ji}} \right)^{\nu} = \sum_{j=1}^{N} z_j^{\nu} \tag{20}$$

  which converges in law to a normal distribution by CLT, provided $u_j = z_j^{\nu}$ has finite second moment. As shown before, z_j follows an iid $F(2, 2(L-1))$ distribution under $\mathbf{H}_0$, and thus, we can compute the mean and variance of z_j^{ν} as

$$\begin{aligned} E(z_j^{\nu}) &= (L-1)^{\nu+1}\beta(L-\nu-1, \nu+1) \\ \mathrm{Var}(z_j^{\nu}) &= (L-1)^{2\nu+1}\left[\beta(L-2\nu-1, 2\nu+1) \right. \\ &\quad \left. - (L-1)\,\beta(L-\nu-1, \nu+1)^2\right]. \end{aligned} \tag{21}$$

  Having obtained this mean and variance, we have the following result for T_{fc}: Under the assumption that z_js are

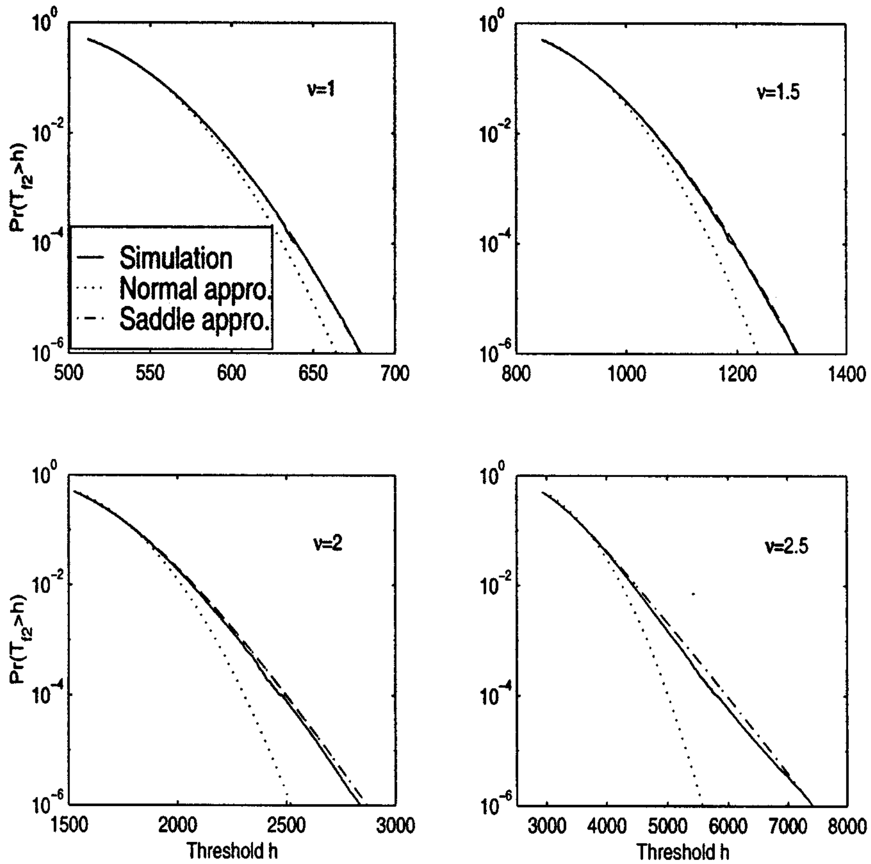

Fig. 5. Exceedance probability of T_{f2} with normal and saddlepoint approximations as a function of threshold h. The right-truncated distribution is used in the implementation of saddlepoint approximation. Here, $N = 256$, and $T = 20^{\nu}$. The approximation results are compared with the simulation results represented by the solid line.

iid $F(2, 2(L-1))$ for $j = 1, \ldots, N$, T_{fc} converges in law for large N to $\mathcal{N}(N.E(z^{\nu}), N.\text{Var}(z^{\nu}))$.

- *The Statistic* T_{f2}:

We cannot use the classical CLT directly here since the detector is a summation over dependent, although identical, random variables. However, in the following, we show that the detectors converge to normal distributions using CLT after a rewriting of the statistics. Observe that we can rewrite T_{f2} as

$$\begin{aligned} T_{f2}(\mathbf{U}) &= \sum_{j=1}^{N}(X_{j-1}+X_j)^{\nu} = \sum_{j=1}^{N} U_j \\ &= \sum_{k=1}^{N/2} U_{2k-1} + \sum_{k=1}^{N/2} U_{2k} \equiv T_o + T_e \end{aligned} \tag{22}$$

where $U_j = (X_{j-1} + X_j)^{\nu}$. Since X_j are iid $\exp(\mu_0)$ under $\mathbf{H}_0$ (we assume $\mu_0 = 1$ for convenience), we know that U_{2k-1}, $k = 1, \ldots, N/2$ are iid. Thus, T_o and T_e converge in law to $\mathcal{N}((N/2)\Gamma(\nu+2), (N/2)(\Gamma(2\nu+2) - \Gamma(\nu+2)^2))$ [recorded as $\mathcal{N}(\mu_2, \sigma_2^2)$] via the CLT.

Now, since $T_{f2} = T_o + T_e$, we know that T_{f2} follows distribution $\mathcal{N}(2\mu_2, \sigma_2^2(2+2\rho))$, where

$$\rho = \frac{E(T_oT_e) - E(T_o)E(T_e)}{\sqrt{\text{Var}(T_o)\text{Var}(T_e)}}$$

is the correlation coefficient, which is affected by the power law ν. If $\nu = 1$, we note that $T_1 = T_2$; thus, clearly, $\rho = 1$; if $\nu = 2$, our calculation reveals that $\rho = (20/21)$. For noninteger ν, a numerical method has to be used to calculate ρ. Fortunately, ρ is close to unity; thus, to simplify, we set $\rho = 1$ and thus approximate T_{f2} as convergent in law to $\mathcal{N}(2\mu_2, 4\sigma_2^2)$.

- *The Statistic* T_{f3}:

Approximating $\rho = 1$ as in the analysis of T_{f2}, we have that T_{f3} converges in law to $N(3\mu_3, 9\sigma_3^2)$, where $\mu_3 = (N/3)\Gamma(\nu+3)/2$ and

$$\sigma_3^2 = \frac{N}{3}\left(\frac{\Gamma(2\nu+3)}{2} - \left(\frac{\Gamma(\nu+3)}{2}\right)^2\right).$$

- *The Statistic* T_{fc2}:

For the detector T_{fc2} as defined in (15), we know

$$\left(\frac{U_{jL}}{\frac{1}{L-1}\sum_{i=1}^{L-1} U_{ji}}\right)$$

follows the $F(4, 4(L-1))$ distribution. Again, assuming that $\rho = 1$, we approximate T_{fc2} as convergent in law

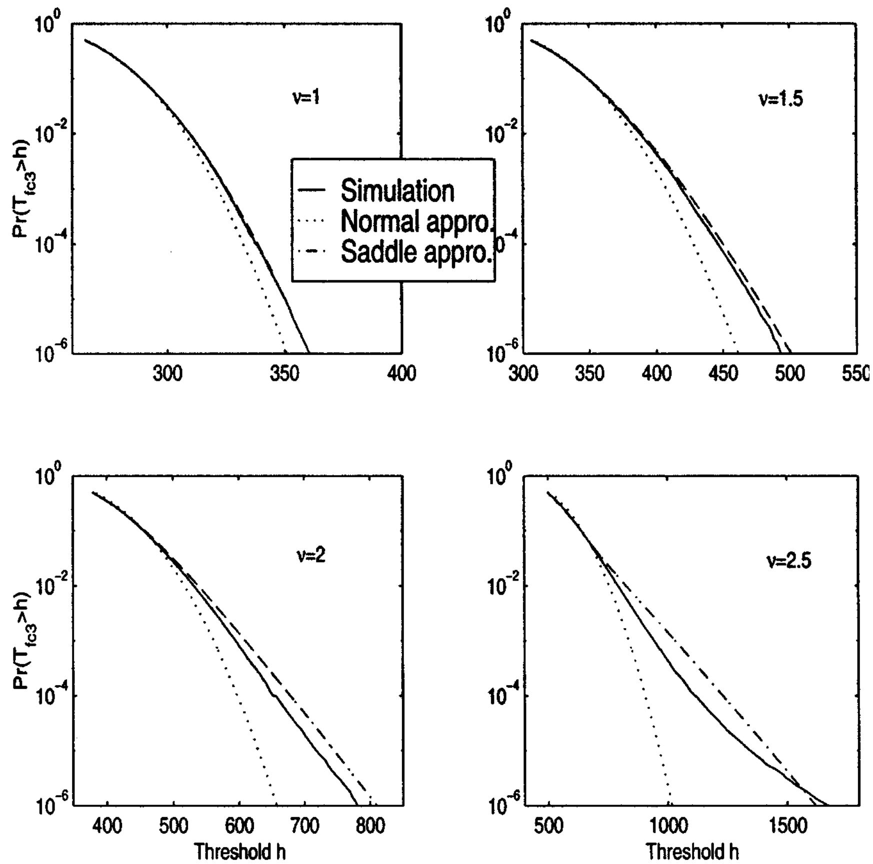

Fig. 6. Exceedance probability of T_{fc3} with normal and saddlepoint approximation as a function of threshold h. Here, $N = 256$, and $L = 10$. The right-truncated distribution is used in the implementation of saddlepoint approximation, and $T = 10^{\nu}$. The approximation results are compared with the simulation results represented by the solid line.

to $\mathcal{N}(2\mu_{c2}, 4\sigma_{c2}^2)$, where $\mu_{c2} = (N/2)E(Y)$ and $\sigma_{c2}^2 = (N/2)\text{Var}(Y)$

$$E(Y_j) = (L-1)^{\nu}\frac{\Gamma(2+\nu)\Gamma(2(L-1)-\nu)}{\Gamma(2(L-1))}$$

$$\text{Var}(Y_j) = (L-1)^{2\nu}\frac{\Gamma(2+2\nu)\Gamma(2(L-1)-2\nu)}{\Gamma(2)\Gamma(2(L-1))} - E(Y_j)^2. \tag{23}$$

- *The Statistic* T_{fc3}:
  As in the previous cases, and using

$$\mu_{c3} = \frac{N}{3}(L-1)^{\nu}\frac{\Gamma(3+\nu)\Gamma(3(L-1)-\nu)}{\Gamma(3)\Gamma(3(L-1))}$$

$$\sigma_{c3}^2 = \frac{N}{3}(L-1)^{2\nu}\left(\frac{\Gamma(3+2\nu)\Gamma(3(L-1)-2\nu)}{\Gamma(3)\Gamma(3(L-1))} - \left(\frac{\Gamma(3+\nu)\Gamma(3(L-1)-\nu)}{\Gamma(3)\Gamma(3(L-1))}\right)^2\right) \tag{24}$$

we approximate T_{fc3} as convergent in law to $\mathcal{N}(3\mu_{c3}, 9\sigma_{c3}^2)$.

B. Saddle-Point Approximation

The asymptotic normal distribution derived in the previous section is easy to work with. However, as we will see later, the normal approximation tends to estimate the tail probability relatively poorly. To obtain more accurate performance evaluation, here, we introduce the saddle-point approximation method that can be thought as a refinement of normal approximation via the indirect use of the Edgeworth expansion. Here, we state the final result, and omit the detailed development; see [7].

Let $X_1, \ldots, X_n$ be iid with pdf $f(x)$, and let $\varphi(\theta) = E(e^{\theta X})$ be the corresponding Laplace transform defined for θ. We define the sample mean $X = (1/n)\sum_{i=1}^{n} X_i$ according to [7, eq. (2.2.6)], and hence, the tail probability can be approximated via the formula

$$\Pr(\overline{X} > x) = \frac{\varphi(\theta)^n e^{-n\theta x}}{\sqrt{n}|\theta|\sigma(\theta)} \times \left\{ B_0(\lambda) + \frac{\text{sgn}(\theta)}{\sqrt{n}}\frac{\zeta_3(\theta)}{6}B_3(\lambda) + \frac{1}{n}\left[\frac{\zeta_4(\theta)}{24}B_4(\lambda) + \frac{\zeta_3(\theta)^2}{72}B_6(\lambda)\right] + O(B_0(\lambda)n^{-3/2}) \right\} \tag{25}$$

where θ is the ML estimate of θ given x, $\lambda = \sqrt{n}|\theta|\sigma(\theta)$, $B_j(\lambda)$ is the Esscher function of jth order, and $\zeta_j(\theta)$ is the jth normal-

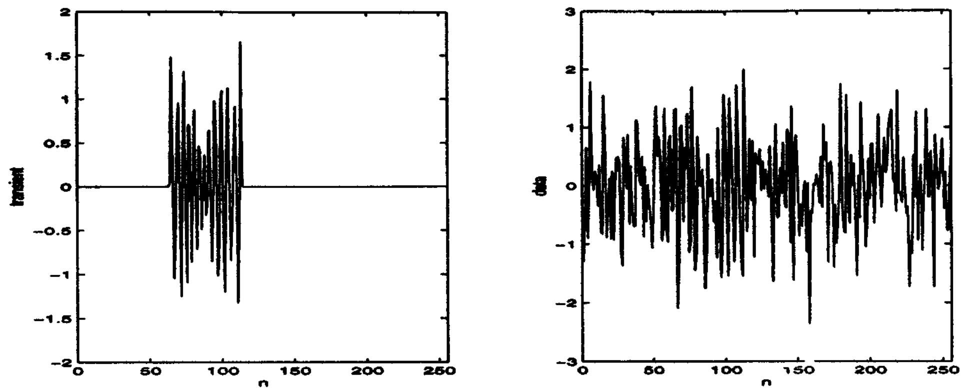

Fig. 7. Example of signal and observation process for Figs. 8 and 9. The signal (left panel) is created by passing white Gaussian noise through an FIR filter with a passband $0.4\pi < \omega < 0.6\pi$ (the number of signal-present FFT bins is approximately 25). On the right, noise is added.

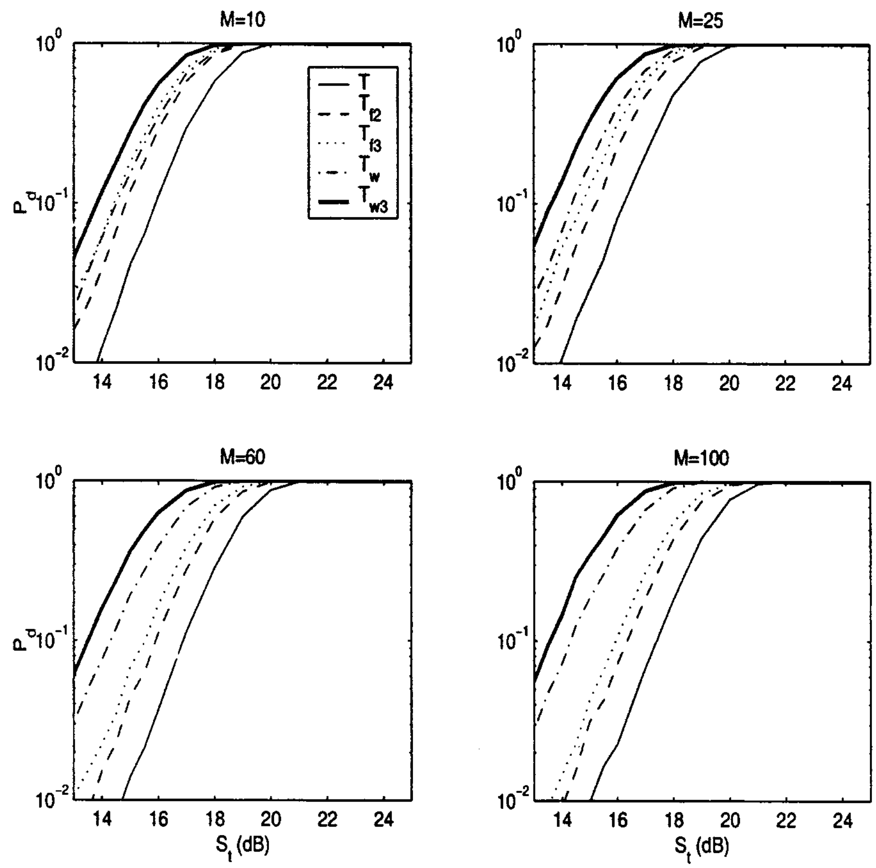

Fig. 8. Detection performances of new power-law statistics in the frequency and the wavelet domains in the prenormalized case. The exponent in each case is $\nu = 2.5$, and the transient duration is $M_t = 20$ samples; different panels refer to the number of frequency bins occupied by the signal.

ized cumulant. A detailed description of the above quantities is available in [7].[4]

C. Comparison of Approximations with Simulation

As mentioned earlier, the performance analysis in this section is focused on the exceedance probability of the statistics under the $\mathbf{H}_0$ hypothesis. The results obtained using the normal approximation and the saddlepoint approximation are shown and compared with results of numerical simulations based on 10^8 runs, for different statistics and different exponents ν, with $N = 256$.

For the white noise (prenormalized) case, we investigate the statistics T_{f2} and T_{f3} that combine two and three contiguous FFT bins correspondingly and show the performance analysis of T_{f2} in Fig. 5. We can see that the normal approximation tends to underestimate the tail probability, whereas the saddlepoint approach shows good prediction. It implies both that saddlepoint approximation is an efficient method to analyze the performances of our statistics T_{f2} and T_{f3} and that setting the correlation coefficient ρ as 1 is an acceptable approximation.

[4]In our case, for X_i obeying pdf $f(x)$, the moment generation function (MGF) does not exist. To resolve this, we simply truncate X_i to a value T.

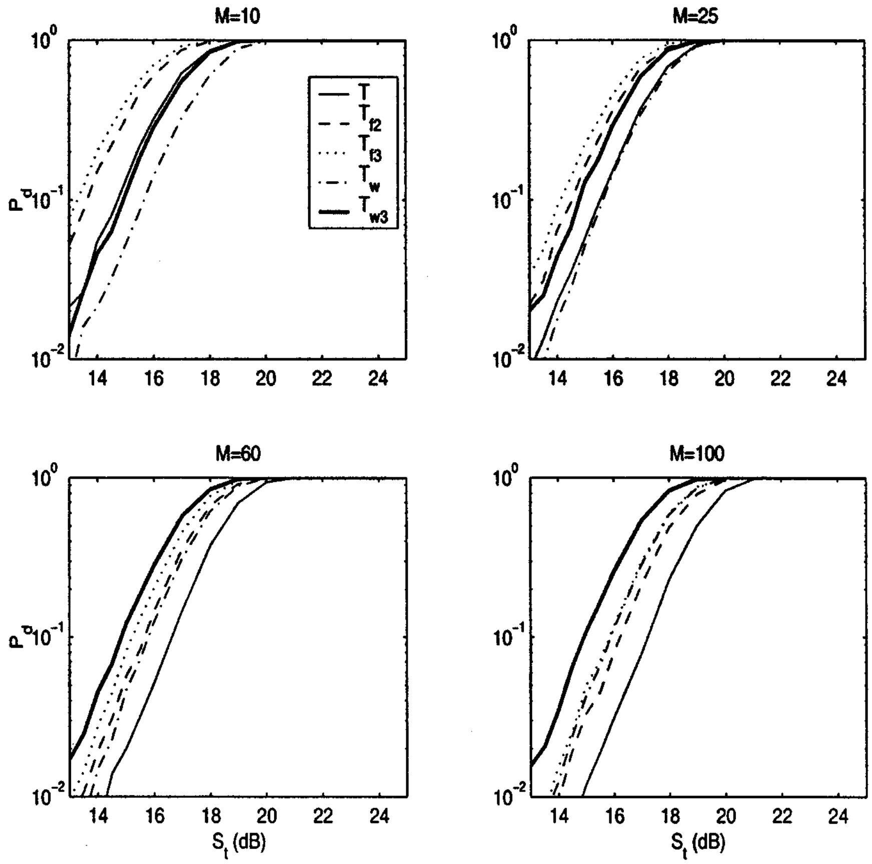

Fig. 9. Detection performances of new power-law statistics in the frequency and the wavelet domains in the prenormalized case. The exponent in each case is $\nu = 2.5$, and the transient duration is $M_t = 50$ samples; different panels refer to the number of frequency bins occupied by the signal.

For the colored noise (self-normalizing) case, we investigate the statistics T_{fc}, T_{fc2}, and T_{fc3} and show the result of T_{fc3} in Fig. 6. According to our earlier SNR analysis, we are interested in $1 < \nu < 2$. Here, we set $N = 256$ and $L = 10$ to be consistent with our later P_d simulations. We can see that the normal approximation estimates the tail probability rather poorly, whereas the saddlepoint approach shows much better prediction. It is noted, however, that even the saddlepoint approximation tends to mismatch the tail probability as ν grows large.

V. Performance Comparison

Here , we apply the detectors developed in the previous sections to numerical examples. For fixed $P_{fa} = 10^{-4}$, applying the thresholds obtained in Section IV, we compare probabilities of detection against aggregate SNR. The *aggregate* SNR is the total signal energy divided by the total noise variance, meaning that the *per-sample* (over the entire block and not just for the fuzzily-defined duration of the transient signal) SNR should be divided by this number; for example, excellent performance is available at an aggregate SNR of 20 dB, and this translates to -4 dB per sample.

Prenormalized Data: The detection performance of the improved detectors in the frequency and the wavelet domains are compared with the power law of [12] with exponent $\nu = 2.5$. Examples of the signal and noise are shown in Fig. 7 for a number of signal-containing frequency bins $M \approx 25$. From Fig. 8, in which the time-domain transient signal is of length $M_t = 20$ samples, it is clear that combining two or three contiguous FFT or wavelet bins together does improve the detection performance over different SNRs S_t. It is also noted that the detector T_{w3} based on the contiguity of wavelet bins shows advantages over all others, and the explanation is presumably that T_{w3} utilizes both temporal and frequency contiguity. However, in the case of a longer transient signal (see Fig. 9) for which the length is $M_t = 50$ samples, those transients that are more concentrated in the frequency domain ($M = 10$ and $M = 25$) are best detected by the FFT-based statistic T_{f3}. This is further explored in Fig. 10; here, contours of the probability of detection are plotted on transient-length (vertical) and bandwidth (horizontal) axes for detectors $T_w,\ _{w3},\ T_f$, and T_{f3}. It is clear that the wavelet-based detectors are more forgiving than those based on the short-time frequency transformation, but that transients of sufficiently narrow bandwidth and broad length are best served in the latter domain.

Self-Normalizing Case: The results of detectors in the case that self-normalization is required are shown in Fig. 11. From comparison with Fig. 9 (the prenormalized case), it is clear that the losses arising from the need to normalize are relatively minor. It is also gratifying that the statistic T_{wc3} is, in all these cases, the best.

VI. Summary

In [12], Nuttall derived and justified a new and easy-to-implement statistic for the detection of short-duration (transient)

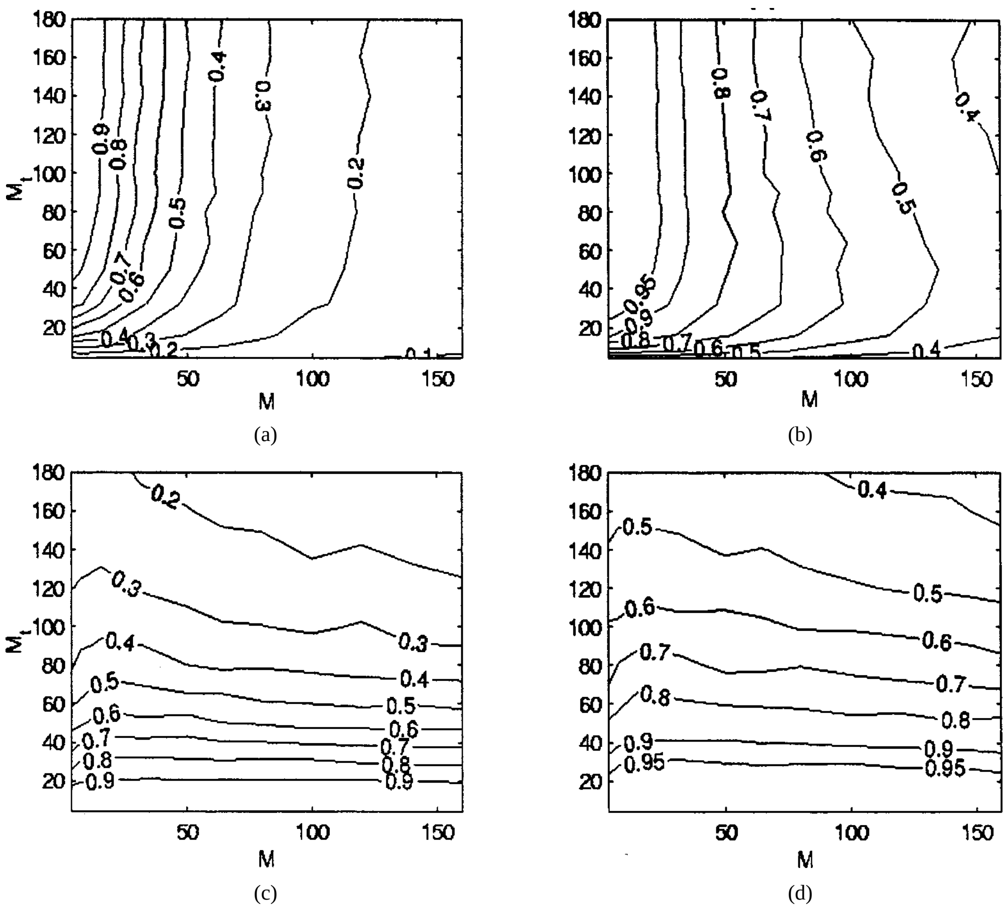

Fig. 10. Probability of detection contours for (a) T_f, (b) T_{f3}, (c) T_w, and (d) T_{w3}. These are plotted versus transient length M_t and number of signal-occupant frequency bins M.

signals: the sum of magnitude-square DFT outputs from a block of N time domain data, each raised to a power typically in the range of 2 to 3. This test has been found to be very effective indeed.

The power-law detector is almost a plug-in transient detector for all purposes but not quite: Prewhitened and prenormalized data is required. We have thus extended the power-law detector to be self-normalizing by raising to an exponent not the DFT data directly but, instead, the power in each DFT bin relative to the average power in previous DFTs. While Nuttall provided some justification both for the exponentiation and exponent in the original power-law test, their applicability in the self-normalizing case is not straightforward. Consequently, a mode of analysis ("input SNR loss") has been proposed for the choice of exponent. It is found that the optimal exponent is somewhat lower than in the original (prenormalized) power-law case.

At any rate, the self-normalizing solution works very nicely. All that is needed to have an all-purpose transient detector with simple implementation is some means to set the threshold, and this is provided via a saddle-point approximation.

Along the path to development of the improved CFAR power-law detector, it was noted that there is a tendency among real transient signals for energy to aggregate in nearby DFT bins (i.e., to be bandlimited to some degree). In their formulations, neither the original nor CFAR power-law detectors take advantage of this, and consequently, a combined-bin power-law detector is proposed. In experiments, versions of this are shown to offer significant improvement. These are made self-normalizing, and a saddlepoint approximation for threshold setting is provided. It was additionally noted that the power-law dogma of preprocessing via the DFT is open to the challenge, and indeed, a power-law processor operating on (Haar) wavelets is developed, made self-normalizing, augmented to use combined bins (since transient signals most transient signals are aggregated not just in frequency but in time/scale as well), and accorded a saddlepoint approximation for threshold setting.

Quite a few of the detectors have been developed and analyzed in this paper. We note that beyond easy choices such as window size (for CFAR) or type (wavelet/DFT or agglomerating/single-bin) and guided selection of the exponent, there are no parameters to tune. We thus consider them to be "plug-in" transient detectors, and since they make minimal assumptions on the structure of the transient signal, save that it have some degree of concentration of energy in time and/or frequency, we advertise them as "all-purpose." For reference, we give their taxonomy in Table I.

Note that the choice of prenormalized versus self-normalizing depends on the data, but in either case, our overall conclusion is that although all of these tests work well, the combined/wavelet power-law detectors (if data are prenormalized, T_{w3}, and if self-whitening and CFAR is necessary, T_{wc3}) are perhaps the finest of all. The statistics are compellingly simple to use. Take a multiresolution decomposition using the Haar

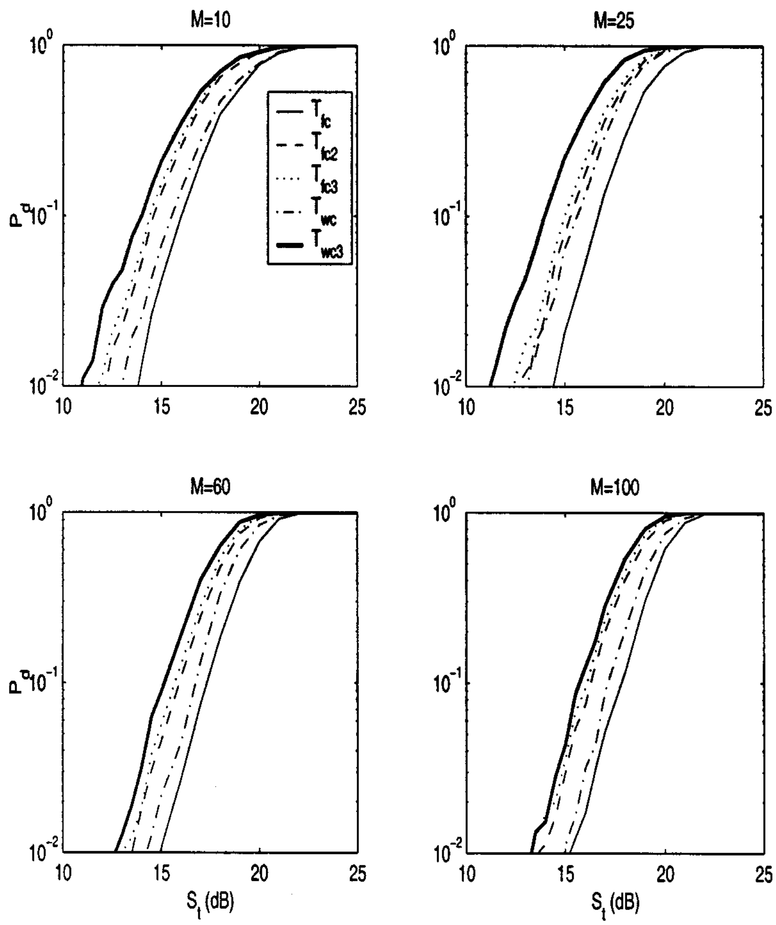

Fig. 11. Detection performance of power-law detectors in the frequency and the wavelet domains for transient detection in colored noise. Here, the time-domain transient length is $M_t = 50$, and the exponent used in all cases is $\nu = 1.5$.

TABLE I
CATEGORIZATION OF VARIOUS TRANSIENT DETECTORS DISCUSSED IN THIS PAPER. NUTTALL'S CFAR POWER-LAW IS "PARTIALLY" SELF-NORMALIZING IN THAT IT IS COMPLETELY INSENSITIVE TO SCALE BUT HAS NO MEANS TO DEAL WITH NONWHITE DATA

detector	provenance	pre-processing	bins combined	self-normalizing
T	Nuttall	DFT	1	no
T_{cpl}	Nuttall	DFT	1	partial
T_{f2}	new	DFT	2	no
T_{f3}	new	DFT	3	no
T_{fc}	new	DFT	1	yes
T_{fc2}	new	DFT	2	yes
T_{fc3}	new	DFT	3	yes
T_w	new	wavelet	1	no
T_{w3}	new	wavelet	3	no
T_{wc}	new	wavelet	1	yes
T_{wc3}	new	wavelet	3	yes

basis, normalize the magnitude-square by previous values at each scale (in the self-whitening case only), combine the result in groups of three according to Fig. 4, exponentiate (to a power 2.5 if prenormalized and to the power 1.5 if normalized), and sum. The resulting statistic can be relied on to detect quite a wide range of transient signals below (often considerably below) -3 dB on a sample-by-sample basis.

REFERENCES

[1] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[2] A. Bianchi, L. Mainardi, and E. Petrucci, "Time-variant power spectrum analysis for the detection of transient episodes in HRV signal," *IEEE Trans. Biomed. Eng.*, vol. 40, pp. 136–144, Feb. 1993.

[3] B. Friedlander and B. Porat, "Performance analysis of transient detectors based on a class of linear data transforms," *IEEE Trans. Inform. Theory*, vol. 38, pp. 665–673, Mar. 1992.

[4] P. Gandhi and S. Kassam, "Analysis of CFAR processors in nonhomogeneous background," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, pp. 427–444, July 1988.

[5] C. Han, "Transient signal detection and Page's test," Ph.D. dissertation, Univ. Conn., Storrs, 1996.

[6] T. Hemminger and Y.-H. Pao, "Detection and classification of underwater acoustic transients using neural networks," *IEEE Trans. Neural Networks*, vol. 4, pp. 712–718, Sept. 1994.

[7] J. Jensen, *Saddlepoint Approximation*. Oxford, U.K.: Clarendon, 1995.

[8] N. Johnson, S. Kotz, and N. Balakrishman, *Continuous Univariate Distribution*, 2nd ed. New York: Wiley, 1995, vol. 2.

[9] S. Kassam, *Signal Detection in Non-Gaussian Noise*. London, U.K.: Dowden & Culver, 1988, pp. 31–46.

[10] N. Lee and S. Schwartz, "Robust transient signal detection using the oversampled Gabor representation," *IEEE Trans. Signal Processing*, vol. 43, pp. 1498–1502, June 1995.

[11] S. Marco and J. Weiss, "Improved transient signal detection using a wavepacket-based detector with an extended translation-invariant wavelet transform," *IEEE Trans. Signal Processing*, vol. 45, pp. 841–850, Apr. 1997.

[12] A. Nuttall, "Detection performance of power-law processors for random signals of unknown location, structure, extent, and strength," NUWC-NPT Tech. Rep., Newport, RI, 10 751, Sept. 1994.

[13] ——, "Detection capacity of linear-and-power processor for random burst signals of unknown location," NUWC-NPT Tech. Rep., Newport, RI, 10 822, Aug. 1997.

[14] ——, "Near-optimum detection performance of power-law processors for random signals of unknown location, structure, extent, and arbitrary strengths," NUWC-NPT Tech. Rep., Newport, RI, 11 123, Apr. 1996.

[15] ——, "Performance of power-law processors with normalization for random signals of unknown structure," NUWC-NPT Tech. Rep., Newport, RI, 10 760, May 1997.

[16] Y. Ohya, Y. Takahashi, and M. Murata, "Acoustic emission from a porcelain body during cooling," *J. Amer. Ceramic Soc.*, pp. 445–448, Feb. 1999.

[17] L. Perlovsky, "A model-based neural network for transient signal processing," *Neural Networks*, vol. 7, no. 3, pp. 565–572, 1994.

[18] V. Petrov, *Limit Theorems of Probability Theory*. Oxford, U.K.: Clarendon, 1995, p. 113.

[19] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1994.

[20] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Wellesley, MA: Wellesley-Cambridge, 1996.

[21] R. Streit and P. Willett, "Detection of random transient signals via hyperparameter estimation," *IEEE Trans. Signal Processing*, vol. 47, pp. 1823–1834, July 1999.

[22] B. Vidakovic, *Statistical Modeling by Wavelets*. New York: Wiley, 1999, p. 169.

[23] Z. Wang and P. Willett, "A performance study of some transient detectors," *IEEE Trans. Signal Processing*, vol. 48, pp. 2682–2685, Sept. 2000.

Zhen Wang was born in China on January 17, 1974. She received the B.S. degree with honors from Tsinghua University, Beijing, China, in 1996 and the M.Sc. degree from the University of Connecticut, Storrs, in 2000, both in electrical engineering. She is currently pursuing the Ph.D. degree at the University of Connecticut.

Her present research interests include the detection theory and signal processing.

Peter K. Willett (SM'97) received the B.A.Sc. degree in 1982 from the the University of Toronto, Tornto, ON, Canada, and the Ph.D. degree in 1986 from Princeton University, Princeton, NJ.

He has been a Professor at the University of Connecticut, Storrs, since 1986. His interests are generally in the areas of detection theory and signal processing.

Dr. Willett is an Associate Editor for both the IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS and the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS.