FAST DFT-DOMAIN ALGORITHMS FOR NEAR-OPTIMAL TONAL DETECTION AND FREQUENCY ESTIMATION

M. D. Macleod

Signal Processing and Communications Group, Cambridge University Engineering Department, UK.

1. INTRODUCTION

The detection of tones (that is, sinusoids or, in the complex signal case, cisoids) is an important function in passive sonar. There may be many tones present, and the detector is required both to detect the tones and provide initial estimates of their frequencies, so that they can be further analysed. The background noise is typically coloured and non-Gaussian. Since detection probability rises rapidly with the Signal-to-Noise ratio (SNR), it is most important in practice for the detector to be *locally optimal*, that is, optimal at low SNR. The other important aim is to minimise computational load, since the detector may be applied to a large number of channels in parallel.

This paper considers the detection of tones in a single block of sampled data. Detection algorithms of this kind can easily be applied to sequential contiguous or overlapping blocks, for detection over longer time spans. We summarise the results for known optimal detectors and present new fast frequency-domain algorithms which acheive near-optimal detection of one or more tones in noise; the noise may be coloured, and, for the blocklengths typical in sonar, non-Gaussian. We also describe the incorporation of recently published fast frequency estimation algorithms for the detection and estimation of tonals.

2. OPTIMUM DETECTION

The input to the detector is assumed to be a block of N consecutive samples of the discrete-time signal x_n, i.e. $\{x_n;\ n = 0, 1, ..., N-1\}$. The x_n are in general complex, although the real-signal case will be considered later. The signal to be detected is modelled as the sum of p cisoids

$$s_n = \sum_{i=0}^{p} b_i \exp(j\omega_i n), \tag{1}$$

where $-\pi < \omega_i \leq \pi$ (radians per sample). $\mathbf{b} = \{b_1, ... b_p\}$ is a set of p complex amplitudes, and $\omega = \{\omega_1, ..., \omega_p\}$ a set of p frequencies. In much of the literature on tone detection the amplitude and phase notation (A_i, θ_i) is used, where $A_i = |b_i|$ and $\theta_i = \arg(b_i)$; however this makes the problem non-linear in θ_i as well as ω_i, and the use of the complex amplitudes b_i is advantageous in both analysis and implementation. (1) is an example of a *linear signal model*, which is comprehensively considered in most texts on detection theory.

We require the detector to choose one of the hypotheses

$$H_1 : (x_n = s_n + z_n) \text{ or } H_0 : (x_n = z_n) \tag{2}$$

in which $\mathbf{z} = \{z_1, ... z_N\}$ is zero mean noise. We will assume initially that the the complex amplitudes and frequencies of the tones are unknown, but that the model order p is known, and that $\mathbf{z}$ is complex, white and Gaussian, with known covariance matrix $\mathbf{R} = \sigma^2 \mathbf{I}_N$.

If it is assumed that uniform prior probability distributions are assigned to the parameters $\boldsymbol{\omega}$ and $\mathbf{b}$ (implying that any value is as likely as any other), and if it is required that the detector not only detects the tone(s) but estimates the corresponding frequencies, it can be shown [2] that the optimum detector is the *generalised likelihood ratio (LR) detector*, which uses the Maximum Likelihood (ML) joint estimates of the unknown parameters $\boldsymbol{\omega}$ and $\mathbf{b}$.

2.1 Single Tone Detection For a single complex tone ($p = 1$), the signal model becomes $s_n = b \exp(j\omega n)$. It is normal to define the "power" of a discrete-time signal as the expectation of its squared modulus, so for the single complex tone the power of the cisoid signal is $A^2 = |b|^2$ and the power of the added noise z_n is σ^2, giving a signal-to-noise ratio (SNR) of $|b|^2/\sigma^2$. The ML estimate of ω [3] is the value $\hat{\omega}$ which maximises

$$C(\omega) = \frac{1}{N}|X(\omega)|^2, \text{ where } X(\omega) = \sum_{n=0}^{N-1} x_n \exp(-j\omega n). \tag{3}$$

$X(\omega)$ in (3) is known as the Discrete Time Fourier Transform (DTFT) of the signal $\mathbf{x}$, and $(1/N)|X(\omega)|^2$ is known as the *Schuster periodogram* of $\mathbf{x}$. The generalised likelihood ratio detector therefore chooses H_1 if the maximum of the periodogram exceeds a threshold T_0^2.

2.2 Optimality of the Generalised LR Detector Whalen [2] derives the corresponding generalised LR detector for the continuous real-signal case, and also derives the locally optimum detectors under other assumptions (a locally optimum detector is one which is optimum at low SNR, which is, in practice, the important requirement for a detector). He assumes that the phase remains unknown, and shows that regardless of the prior probability distribution of the amplitude A the locally optimum detector always has the same form as the generalised LR detector. We can also show that if the prior probability distribution $p(\omega)$ of the frequency is non-uniform, the optimum detector simply maximises $p(\omega)C(\omega)$ rather than $C(\omega)$ in (3). Hence a detector of this form is optimum or near-optimum under a wide range of assumptions.

2.3 Multiple tone detection In the multiple tone case, the generalised likelihood ratio is a non-linear function of the p frequencies, with many local optima, and maximising it, to obtain the ML estimate of the frequencies, is in general difficult and computationally expensive. The problem can, however, be simplified under some important circumstances. Rife and Boorstyn [5] show that the determination of the ML estimate of any of the frequencies, say ω_i, can be carried out as if it were the only tone present, provided that the separation in frequency between it and any other tone, $|\omega_k - \omega_i|$, is $\gg 2\pi/N$; in such cases, we will describe the frequencies as "widely separated". If all the frequencies, taken in pairs, are widely separated, the non-linear optimisation in p frequencies is replaced by p non-linear optimisations in 1 frequency, for ω_1 to ω_p in turn, which is a great simplification.

In Walker [7] an asymptotically optimal algorithm is presented for the multiple tone case when the frequencies are sufficiently separated. This algorithm is misleadingly described in [1] as "finding the p largest local maxima of the periodogram"; in fact, the procedure is: first, find the maximum of the periodogram (at frequency ω_1), then subtract from the data the corresponding tone, using the ML estimates of ω_1 and b_1; find the maximum of the modified periodogram (at ω_2), subtract the corresponding tone, etc. The important difference is that when a large tone is close to a smaller one, the larger may "mask" the smaller unless the correct procedure is used.

2.4 Model order determination for widely separated frequencies In practice, it is likely that the number of tones, p, will not be known. Under the assumption of wide frequency separation, the log likelihood of hypothesis H_1 for any given value of p is a constant plus the sum of the p largest local maxima of the periodogram, calculated as explained above. Therefore, the likelihood for different candidate values of p can easily be calculated, and this could be used as a basis for selecting p. This would be an appropriate approach to use when the prior probability of a given set of tones is not the same as the product of the individual prior probabilities for each tone, for example when the signal to be detected is expected to have harmonic structure [6].

However, if we assume that the tones are produced by physically independent sources, we may reasonably assume that a hypothesis test can be applied to each possible tone independently. Starting with the largest local maximum of the periodogram, we can therefore apply a single-tone generalised likelihood ratio test to this maximum, to decide on the presence of a tone at this frequency. If we decide that there is a tone present, we subtract the ML estimate of this tone from the data and repeat the procedure, until the $(p+1)^{\text{th}}$ largest peak is lower than the threshold.

2.5 Results for real tones In the case of a single real tone (sinusoid) the signal model becomes

$$s_n = \frac{1}{2}(b\exp(j\omega n) + b^*\exp(-j\omega n)) \tag{4}$$

where the angular frequency ω is now in the range $0 < \omega \leq \pi$, and the noise z is real white Gaussian noise with variance σ^2. The amplitude of the sinusoid in (4) is $|b|$ and its power is $|b|^2/2$, giving a signal-to-noise ratio (SNR) of $|b|^2/(2\sigma^2)$.

Since for a real signal $X(\omega) = X^*(-\omega)$, detection of the complex tone with positive frequency is equivalent to detection of the tone with negative frequency, and nothing is gained by doing both. From section 2.3 we see that complex single tone detection can be applied to detect the peak near frequency ω provided that ω and $-\omega$ are "widely separated", that is $\omega - (-\omega) = 2\omega > 2\pi/N$,and also (since it is a sampled-data system) $(2\pi - \omega) - \omega > 2\pi/N$. This gives the known result [3] that the optimal real single tone detector is the same as for the complex-signal case, apart from a scale factor, provided that $\pi/N < \omega < \pi(N-1)/N$.

3. EXISTING SINGLE TONE DETECTORS

The optimum detector requires a one-dimensional search over ω to find the maximum of the periodogram (3). This search is a non-linear optimisation problem, and there are many local optima; it is normal therefore to start by searching for the maximum over a uniform discrete grid of values of ω. An obvious approach is to calculate the Discrete Fourier Spectrum (DFS) $\mathbf{X}$ of the signal $\mathbf{x}$ via the Discrete Fourier Transform (DFT)

$$X_k = \sum_{n=0}^{N-1} x_n \exp(-jnk\frac{2\pi}{N});\ k = 0, 1, ...N-1. \tag{5}$$

Comparing (5) with (3), we see that $X(k\omega_0) = X_k$, where $\omega_0 = 2\pi/N$. We therefore define the *discrete periodogram* as $(1/N)|X(k\omega_0)|^2 = (1/N)|X_k|^2$, and the search for the maximum of the periodogram can be approximated by a search for the maximum of the discrete periodogram. The DFT can, of course, be calculated efficiently over the entire frequency range using an FFT. However in both ML frequency estimation [3] and tone detection, the frequency grid spacing $\omega_0 = 2\pi/N$ obtained by the use of the discrete periodogram is too large, and results in a loss of performance. The reason for this can be seen by considering the DTFT $X(\omega)$ of a single cisoid, $x_n = A\exp(j\omega_S n)$:

$$X(\omega) = \sum_{n=0}^{N-1} A\exp(j\omega_S n)\exp(-j\omega n) = NA\exp(-j\pi\beta\frac{N-1}{N})\text{snic}(\pi\beta) \tag{6}$$

where $\beta = (\omega - \omega_S)/\omega_0$ is the frequency offset from ω_S measured in units of the DFS sample spacing, and $\text{snic}(x) \equiv \sin(x)/(N\sin(x/N))$. Putting $\omega = k\omega_0$ in (6) gives the value of DFS sample X_k. If the signal frequency ω_1 equals $M\omega_0$ with M integer, then $X_M = NA$ and $X_k = 0$ for $k \neq M$; this is the "on-bin" case. However, for other frequencies ("off-bin"), the amplitude of the largest sample of the DFS falls. In the worst case, which is when $\omega = (M+0.5)\omega_0$, it is easy to show that the maximum value of the DFS is reduced to approximately $2NA/\pi = 0.64NA$; this is the "picket-fence effect". As a result the maximum of the discrete periodogram is reduced from the true value NA^2 in the "on-bin" case to approximately $0.4NA^2$, and it is this which causes the loss in detection probability.

3.1 Zero Padding The standard way to reduce the loss in detection probability is to use *zero-padding* [3]. The N-sample input $\mathbf{x}$ is zero-padded to the αN-sample vector $\mathbf{x}^Z$ by appending $(\alpha-1)N$ zeros:

$$(x_0^Z, ...x_{N-1}^Z, x_N^Z, ..., x_{\alpha N-1}^Z) = (x_0, ..., x_{N-1}, 0, ..., 0) \tag{7}$$

If the αN-point DFT of $\mathbf{x}^Z$ is $\mathbf{X}^Z$, then $X(\omega) = X_k^Z$ for $\omega = 2\pi k/(\alpha N) = k\omega_0/\alpha$, giving a frequency resolution α times finer than the discrete periodogram. Detection can then be based on searching for the maximum of the zero-padded periodogram, $(1/N)|X_k^Z|^2$.

3.2 Windowing One solution to the picket fence and leakage effects in spectrum analysis is to use windowing. For the simple example of the Hanning window, windowing can be implemented either by multiplication in the time-domain or convolution in the frequency domain; in the frequency domain, the Hanning-windowed DFS $\mathbf{x}^W$ is given by

$$X_k^W = a(2X_k - X_{k+1} - X_{k-1}), \tag{8}$$

where a is a scale factor. In the simulations reported below, the scale factor a in (8) is set to $a = 1/\sqrt{6}$, so that the variance of X_k^W equals the variance of X_k (using the fact that the X_k are independent).

4. PERFORMANCE OF EXISTING SINGLE TONE DETECTORS

Since it can be shown [4] that the samples of the DFT of zero-mean Gaussian input noise of variance σ^2 are independent zero-mean Gaussian variables of variance $N\sigma^2$, an exact theoretical analysis of the performance of the discrete periodogram detector is possible, using the central and non-central χ^2 distributions. Although theoretical analysis of the zero-padded detector's performance for $\alpha > 1$ is intractable, its performance as a function of α can easily be determined by simulation.

The first parameter of interest is the false alarm rate as a function of detector threshold. For the standard discrete periodogram, let the probability that any given sample of the discrete periodogram exceeds a threshold T^2 be p_F. Assuming unit variance input noise, it can be shown that in theory $\ln(p_F) = -T^2$. If we were only considering single tone detection, we might wish to use the probability of a false alarm per block, given by $p_B = 1 - (1 - p_F)^L$, where L is the number of DFS samples corresponding to the chosen detector bandwidth $2\pi L/N$ rads/sample (in general, $L < N$). However, when considering multiple tone detection it is more useful to consider the probability of false alarm per bandwidth ω_0, which we denote p_0; hence for the discrete periodogram $p_0 = p_F$.

Simulations were carried out with 3000 datablocks of $N = 1024$ points each, to measure false alarm rates over a bandwidth of $2\pi 1000/1024$ in each block, i.e. with $L = 1000$. The input noise variance was 1.0, and the test thresholds were set to give false alarm probabilities per bandwidth ω_0 covering the range 3×10^{-3} to 3×10^{-5}. For the zero-padded DFS the samples become increasingly correlated as α increases, and adjacent samples of the Hanning windowed DFS are also correlated. Therefore in measuring false alarm rates any contiguous block of samples exceeding the threshold was counted as a single false alarm. It was found that to a high degree of accuracy the relationship $\ln(p_0) = -T^2 + \rho$ holds, with ρ as follows:

Detector	ZP, $\alpha = 1$	ZP, $\alpha = 2$	ZP, $\alpha = 4$	ZP, $\alpha = 8$	ZP, $\alpha = 16$	Hanning window
ρ	0	0.62	0.94	0.96	0.98	-0.05

Using thresholds derived from this table, a Receiver Operating Curve (ROC) can be determined, which is a plot of detection probability p_D against SNR for a given false alarm rate. In Figure 1 the ROCs are plotted for a false alarm rate of $p_F = 10^{-4}$ (this corresponds to a block false alarm rate of 9.5%) for zero-padded DFS detectors with $\alpha = 1, 2, 4$ and for the Hanning-windowed

DFS (with $a = 1/\sqrt{6}$). Because of the finite frequency grid spacing, the detection probability depends on signal frequency ω_S; it is maximum when $\omega_S = K2\pi/(\alpha N)$, and minimum when $\omega_S = (K+0.5)2\pi/(\alpha N)$. Figure 1 therefore shows the average detection probability, uniformly weighted over all frequencies, and Figure 2 shows the maximum and minimum detection probabilities.

It can be seen that the ROC using $\alpha = 2$ is almost as good as using $\alpha = 4$, and that either gives a significant improvement on the standard discrete periodogram. It can also be seen that although windowing is successful in reducing the variation of detection probability with frequency, it results in a significant loss of detection probability at all frequencies.

5. FAST FREQUENCY-DOMAIN APPROXIMATIONS

For a single cisoid signal the number of samples of its DFS (6) for which the spectral "energy" $|X_k|^2$ is significant is very small. This makes it possible to design efficient approximate algorithms for the calculation of the periodogram on a frequency grid denser than the standard DFS. A second advantage of working in the frequency domain is that because each sample of the DFS is the weighted sum of N input samples, the distribution of the DFS samples X_k will be very close to Gaussian, regardless of the input distribution, by the Central Limit Theorem provided that N is sufficiently large, as it is in sonar. Detector performance will therefore be insensitive to the input noise distribution.

Using standard properties of the DFT, we can show that

$$\sum_{n=0}^{N-1} x_n s_n^* = \frac{1}{N} \sum_{m=0}^{N-1} X_m S_m^* \tag{9}$$

where S is the DFT of s. Applying this to the zero-padded DFT, we have

$$X_k^Z = \sum_{n=0}^{N-1} x_n \exp(-jnk\frac{2\pi}{\alpha N}) = \frac{1}{N} \sum_{m=0}^{N-1} X_m H_m^* \tag{10}$$

where H_m is the DFT of $\exp(j2\pi nk/(\alpha N))$. (The limit of summation in (10) is $N-1$ because the zero-padded data is zero for $n > N-1$.) H_m is then given by (6), with $\omega_S = 2\pi k/(\alpha N)$, and hence $\beta = m - (k/\alpha)$. Clearly for $k = \alpha K$, we have $X_k^Z = X_K$, the standard DFS. For the other values, $k \neq \alpha K$, the magnitude of H, snic(β), decays rapidly away from the frequency of interest, ω_S. Therefore we can approximate the value of X_k^Z in (10) by summing only over those values of m for which H_m is relatively large.

The minimum useful approximation is to approximate H_m by only its two largest samples. Applying this to the case $\alpha = 2$ and $k = 2K+1$, (and approximating snic(x) by $\sin(x)/x$ and using $N \gg 1$ in (6)) the two largest values of H_m^* will be $H_K^* = (2N/\pi)\exp(-j\pi/2)$ and $H_{K+1}^* = (2N/\pi)\exp(j\pi/2)$. We then have $X_k^Z \approx (2j/\pi)(X_{K+1} - X_K)$. We therefore approximate the $\alpha = 2$ zero-padded DFS X_k^Z by using $X_k^Z = X_K$ for the "on-bin" samples ($k = 2K$)

and the approximation $X_k^Z \approx j\gamma(X_{K+1} - X_K)$ for the intermediate samples ($k = 2K + 1$). If we put $\gamma = 2/\pi$, then both the false alarm probability and detection probability will be lower for the intermediate samples than for the DFS samples. We therefore (heuristically) choose $\gamma = 1/\sqrt{2}$; because the DFS samples are independent, this makes the false alarm probability for the intermediate approximate samples the same as for the DFS samples.

For this approximate detector the log threshold offset is found to be $\rho = 0.62$ and its ROC is shown in Figures 1 and 2. It is clearly better than the DFS and much better than the windowed DFS. The computational cost (in real arithmetic operations) of this detector for $N = L = 1024$ is 58368 ops, compared with 51200 for the DFS and 132096 for the $\alpha = 2$ zero-padded DFS; the advantage increases further as L reduces.

Also shown in Figures 1 and 2 are (i) a detector approximating $\alpha = 2$ using 4 terms of the summation (10), and (ii) a detector approximating $\alpha = 3$ using 2 terms for every intermediate point. The corresponding threshold offsets are $\rho = 0.60$ and 0.86 and the computational costs are 62464 and 65536 ops. We can see that the 4-term approximation offers a slight improvement; the $\alpha = 3$, 2-term, approximation is not worthwhile, however.

6. OVERALL SYSTEM IMPLEMENTATION FOR MULTIPLE TONE DETECTION

As explained in section 2.3, a multiple tone detection system for widely separated frequencies can be implemented by repeatedly identifying the periodogram maximum and subtracting the corresponding tone from the data. Either a zero-padded DFS or the fast approximation to it described in section 4 may be used for the peak detection function. When the noise is unknown, and typically coloured as in sonar, standard approaches may be used to estimate the variance of the DFS (the "noise floor") as a function of ω.

Tonal parameter estimation for widely separated tonals may be carried out using a recently published fast, near optimal, frequency-domain algorithm [8], and computational effort may be further saved by performing the subtraction of the estimated tonal in the DFS domain. This DFS may be calculated easily using (6), and should then only be subtracted over the range of frequencies for which the magnitude of the tonal spectrum is significant in comparison to the estimated DFS noise variance. For low SNR tonals, this will be a small number of DFS samples.

The detection of tonals which are close in frequency has not been considered in detail in this paper. Preliminary investigations suggest that in passive sonar there is little advantage in using more complex algorithms for *detection*; the conditions under which detection probability is usefully improved are rare.

7. REFERENCES

1. S. M. KAY, *Modern Spectral Estimation* . Prentice Hall, New Jersey, 1987.
2. A. D. WHALEN, *Detection of Signals in Noise* . Academic Press, New York, 1971.
3. D. C. RIFE and R. R. BOORSTYN, "Single-Tone Parameter Estimation from Discrete-Time Observations", *IEEE Trans. Inform. Theory* , vol. IT-20, pp. 591-598, Sept. 1974.
4. R. J. KENEFIC, "Generalized Likelihood Ratio Detector Performance for a Tone with Unknown Parameters in Gaussian White Noise", *IEEE Trans Signal Processing* , vol. 39, pp.978-980, April 1991.
5. D. C. RIFE and R. R. BOORSTYN, "Multiple Tone Parameter Estimation from Discrete-Time Observations", *Bell Syst. Tech. Journal*, vol. 55, pp. 1389-1410, Nov. 1976.
6. T. W. EDDY, "Maximum Likelihood Detection and Estimation for harmonic sets", *J. Acoust. Soc. Am.*, vol. 68(1), pp. 149-155, July 1980.
7. A. M. WALKER, "On the Estimation of a Harmonic Component in a Time Series with Stationary Independent Residuals", *Biometrika*, Vol. 58, pp. 21-36, 1971.
8. M. D. MACLEOD, "Fast High Accuracy Estimation of Multiple Cisoids in Noise", *Signal Processing V (Proceedings of the Vth European Signal Processing Conference*, Elsevier Science Publishers, pp. 333-336, 1990.

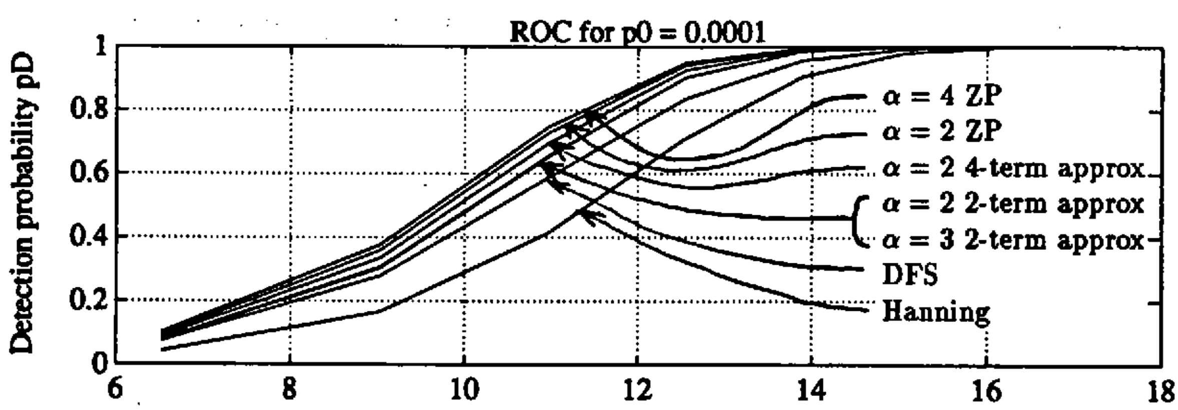

Figure 1 Spectrum SNR (dB); input SNR is 10log10(N) dB lower

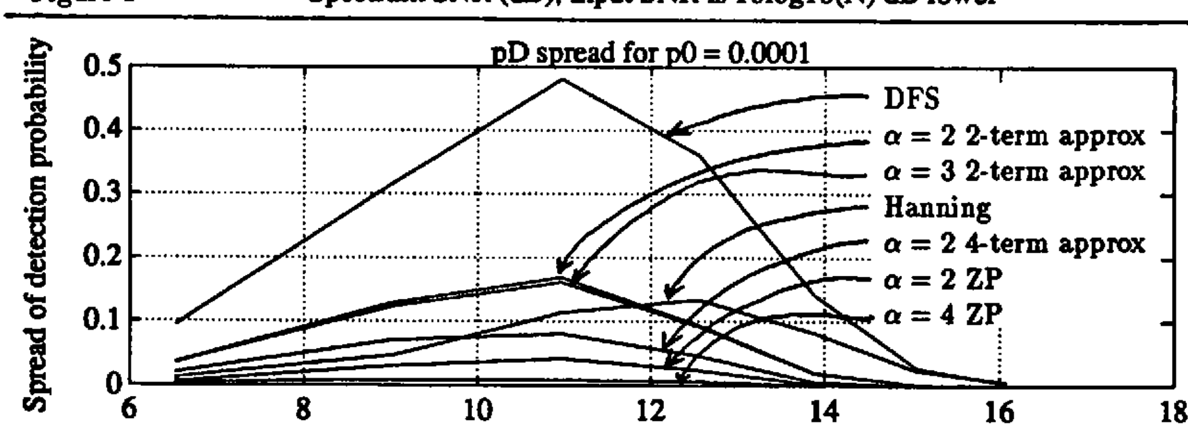

Figure 2 Spectrum SNR (dB); input SNR is 10log10(N) dB lower