RECURSIVE BAYES RISK PARAMETER ESTIMATION FROM THE CYCLIC AUTOCORRELATION MATRIX

J.Riba-Sagarra, G.Vázquez-Grau

Signal Processing Group
Dpt of Signal Theory and Communications, Polytechnic University of Catalonia
Apt 30002, 08071 Barcelona
Spain
E-mail: jriba@tsc.upc.es / gregori@tsc.upc.es

ABSTRACT†

We present a new method for recursively estimating the frequency and timing parameters of a second order cyclostationary signal with known Cyclic Autocorrelation Matrix (CAM). This problem appears in the context of RADAR as well as asynchronous communications in highly non-stationary envinonments (e.g. telemetry) due to the Doppler effect. The parameters evolution is modelled by a zero-order random walk. The estimates are obtained from the Instantaneous CAM (ICAM) of the signal. While non linear Kalman Filter Theory is a possible approach to this problem, Bayes Risk Theory is used instead. In this way the Riccati Equation and gain matrices become independent of the estimates, thus allowing a look-up table solution. The Folded Normal Density (FND)[1] is assumed as Parameters' prior. The prior is recursively updated in mean (estimates) and variance. By comparing the obtained equations with linear Kalman Filter equations we show that with few modifications a first-order random walk model can be easily incorporated to cope with highly non-stationary frequency evolutions.

1-PROBLEM STATEMENT

Consider a base band second-order cyclostationary signal x(n) of known period N. A signal y(n) is obtained from x(n):

$$y(n) = x(n - tN)\, e^{j2\pi fn} + w(n) \tag{1}$$

Parameters f (frequency) and t (timing) are unknown and w(n) is independent stationary white gaussian noise. In a RADAR context, f and t constitute the information we are seeking while in a communications context they need to be estimated to achieve signal detection. Both parameters are assumed in the interval [1/2,1/2). To tackle this problem, we define the (1xp) Instantaneous Autocorrelation Vector (IAV), (1xp) Autocorrelation Vector (AV), (cxp) Instantaneous Cyclic Autocorrelation Matrix (ICAM) and (cxp) Cyclic Autocorrelation Matrix (CAM) as[2]:

IAV: $\hat{\mathbf{r}}(n) = x(n)\, x^*(n - \mathbf{m}^T)$

AV: $\mathbf{r}(n) = E\left[\hat{\mathbf{r}}(n)\right]$

ICAM: $\hat{\mathbf{X}}(k) = \mathbf{F}^H \hat{\mathbf{r}}(kN + \mathbf{n})$

CAM: $\mathbf{X} = E\left[\hat{\mathbf{X}}(k)\right]$ (2)

where,

$$\mathbf{F} = \frac{1}{\sqrt{N}} e^{j\frac{2\pi}{N}<\mathbf{n}\mathbf{k}^T>} \qquad \mathbf{n} = [0, 1, 2, \ldots, N-1]^T \tag{3}$$

m is a (px1) vector of integers defining the lag-products of interest. **k** is a (cx1) vector of integers defining the cycle frequencies of interest. Denoting by • the Schur matrix product (component-wise), the CAM of y(n) becomes:

$$\mathbf{Y} = \mathbf{X} \bullet \mathbf{t}\mathbf{f}^H + \mathbf{W} \tag{4}$$

where: $\mathbf{t} = e^{j2\pi \mathbf{k} t}$ $\quad \mathbf{f} = e^{j2\pi \mathbf{m} f}$

$$\mathbf{W} = \Gamma \delta(\mathbf{k}) \delta(\mathbf{m}^T) \tag{5}$$

and **W** is the CAM of the additive white and stationary noise term w(n) of power Γ. The ICAMs of signal y(n) can be considered as noisy measurements of CAM:

$$\hat{\mathbf{Y}}(k) = \mathbf{Y} + \mathbf{V}(k) \tag{6}$$

where **V(k)** is a noise matrix with zero mean elements and with covariance (cpxcp) matrix **B**:

$$\mathbf{B} = E\left[\mathbf{V}(k) \otimes \mathbf{V}^H(k)\right] \tag{7}$$

The symbol (⊗) denotes the Kronecker product [2]. Instead of such a big matrix **B**, consider only the (cxp) matrix **C**:

$$\mathbf{C} = E\left[\mathbf{V}(k) \bullet \mathbf{V}^*(k)\right] \tag{8}$$

The problem is stated as follows: With **X** known, we want to estimate f and t recursively from the ICAMs (6) of y(n).

†. This work has been supported by the National Research Plan of Spain CICYT, Grant Number TIC92-0800-C05-05.

[1] FND stands for Folded Normal Density. It is obtained when we "fold" the normal density (ND) around the unit circle [1]. It is the natural counterpart of a ND for parameters in the unit circle. The variance is bounded from above by 1/12 (Uniform Density in the interval [1/2,1/2).

[2] Meaning of used notation. If **a**(b) is a column vector function of scalar b, and **b** is a row vector, then **a**(**b**) is the matrix $[\mathbf{a}(b_1),\mathbf{a}(b_2),\ldots,\mathbf{a}(b_i),\ldots]$. If **a**(b) is a row vector function of scalar b, and **b** is a column vector, then **a**(**b**) is the matrix $[\mathbf{a}^T(b_1),\mathbf{a}^T(b_2),\ldots,\mathbf{a}^T(b_i),\ldots]^T$. The sum of scalar a and vector **b** is the vector $[b_1+a, b_2+a, \ldots, b_i+a, \ldots]$.

2. RECURSIVE BAYES RISK ESTIMATOR

Denoting by $\Theta=[f,t]^T$ the unknown parameter vector, the desired estimator will be a function of ICAMs: $\hat{\Theta}(\hat{\mathbf{Y}}(k))$. First suppose we know Θ. For an estimator and a value of Θ, consider the following Euclidean Distance Risk:

$$R(\Theta,\hat{\Theta}) = E_V\left[(\Theta - \hat{\Theta}(\hat{\mathbf{Y}}))^T(\Theta - \hat{\Theta}(\hat{\mathbf{Y}}))\right] \quad (9)$$

E_V means that the expectation is taken over the noise matrix $\mathbf{V}$ distribution. Now suppose we do not know Θ, but only its prior distribution $P(\Theta)$. Bayes Risk (BR) is defined as the average of the Risk over the prior:

$$R(P,\hat{\Theta}) = E_\Theta\left[R(\Theta,\hat{\Theta})\right] = \int R(\Theta,\hat{\Theta})\,dP(\Theta) \quad (10)$$

The Bayes Risk Estimator (BRE) minimises Bayes Risk [3]:

$$\hat{\Theta}_B = \operatorname{argmin}_{\hat{\Theta}} R(P,\hat{\Theta}) \quad (11)$$

Now suppose we do not know the prior $P(\Theta)$, but only its mean $\boldsymbol{\Theta}$ and covariance Σ. Information Theory states: take the Normal Density (ND) as Maximum Likelihood (ML) estimate of the prior. Parameters f and t lie on the unit circle S^1. So, take the FND[1] as prior, and compute (10) and (11) with it. We propose a Recursive Bayes Risk Estimator (RBRE). Three steps will be followed at each iteration:

1) Compute Bayes Risk Estimator Q_B with $P(Q)$=FND (ML Prior).
2) Update prior:
   Prior mean estimate : $\mathbb{Q} = Q_B(Y(k))$.
   Prior covariance estimate Σ.
3) Predict prior.

At first, the prior covariance will be high (acquisition). After several iterations, the prior covariance will stabilise to a small value (tracking).

2.1. Compute Bayes Risk Estimator. Generally BRE becomes a non linear function of the measurements. Computing the BRE is a difficult problem. We propose the following: *first*, define the structure of the function of measurements constituting the estimator, and *second*, minimise its BR.

First; given the prior mean, and motivated by (4), we try to remove the error from the ICAM by computing the following matrix $\mathbf{Z}$:

$$\hat{\mathbf{Z}} = \hat{\mathbf{Y}} \bullet (\hat{\mathbf{t}}\,\hat{\mathbf{f}}^H)^* = \hat{\mathbf{X}} \bullet (\Delta\hat{\mathbf{t}}\,\Delta\hat{\mathbf{f}}^H)^* + \Lambda$$

where,

$$\Delta\hat{\mathbf{f}} = e^{j2\pi \mathbf{m}\,\delta f} \qquad \delta f = f - \hat{f}$$
$$\Delta\hat{\mathbf{t}} = e^{j2\pi \mathbf{k}\,\delta t} \qquad \delta t = t - \hat{t}$$
$$\Lambda = \mathbf{V} \bullet (\hat{\mathbf{t}}\,\hat{\mathbf{f}}^H)^* \quad (12)$$

---

[3] It is assumed that only the real at of the trace is taken. That is not explicitly written in (14) because mathematics becomes simpler when optimising gain matrices.

It can be easily shown that the statistic of Λ is the same as that of matrix $\mathbf{V}$:

$$\mathbf{C} = E\left[\mathbf{V} \bullet \mathbf{V}^*\right] = E\left[\Lambda \bullet \Lambda^*\right] \quad (13)$$

Now we want to estimate errors δf and δt from (a function of) $\mathbf{Z}$. Errors will be small for small prior covariance (tracking). So we chose a linear function (trace) for estimating the errors, and then update the prior [3]:

$$\hat{f}' = \hat{f} + \operatorname{tr}\left[\mathbf{T}_f^H \hat{\mathbf{Z}}\right]$$
$$\hat{t}' = \hat{t} + \operatorname{tr}\left[\mathbf{T}_t^H \hat{\mathbf{Z}}\right] \quad (14)$$

$\mathbf{T}_f$ and $\mathbf{T}_t$ are the gain matrices to be designed.

Second; choose gain matrices minimising BR. Note that we will minimise BR under the linearity constraint (14), and that in acquisition this estimator is not necessarily the best, as in general the best becomes non linear. We find only the best linear function of the matrix $\mathbf{Z}$. But in tracking, the proposed estimator is the best, as linear filtering becomes optimum for small error.[4].

Take vector $\mathbf{k}$ in (2) to be odd-symmetric: $\mathbf{k}$=-fold[$\mathbf{k}$]. Then, this symmetry appears in the phase of the column vectors of $\mathbf{Z}$. It is routine to show that under this constraint gain matrices become orthogonal: $\operatorname{tr}[\mathbf{T}_t^H\mathbf{T}_f]=0$, and so errors δf and δt are uncorrelated. Moreover, the errors are zero mean, as BRE under quadratic lost is unbiased [3]. Then, with $\hat{\Sigma} = \operatorname{diag}[\hat{\sigma}_f, \hat{\sigma}_t]$ BR can be expressed as the sum of two independent BRs:

$$R(P,\hat{\Theta}) = R_f(P,\hat{\Theta}) + R_t(P,\hat{\Theta}) \quad (15)$$

where,

$$R_f(P,\hat{\Theta}) = \sigma_f + \operatorname{tr}\left[\mathbf{A}_f\right]$$
$$\mathbf{A}_f = \mathbf{M}_t(\mathbf{T}_f \bullet \mathbf{X}^*)\mathbf{M}_f(\mathbf{T}_f \bullet \mathbf{X}^*)^H + \mathbf{T}_f^H(\mathbf{T}_f \bullet \mathbf{C}) - 2\mathbf{T}_f^H\mathbf{X}_f$$
$$R_t(P,\hat{\Theta}) = \sigma_t + \operatorname{tr}\left[\mathbf{A}_t\right]$$
$$\mathbf{A}_t = \mathbf{M}_t(\mathbf{T}_t \bullet \mathbf{X}^*)\mathbf{M}_f(\mathbf{T}_t \bullet \mathbf{X}^*)^H + \mathbf{T}_t^H(\mathbf{T}_t \bullet \mathbf{C}) - 2\mathbf{T}_t^H\mathbf{X}_t \quad (16)$$

and,

$$\mathbf{X}_f = \mathbf{X} \bullet \mathbf{P}_t\mathbf{P}_f'^H \qquad \mathbf{X}_t = \mathbf{X} \bullet \mathbf{P}_t'\mathbf{P}_f^H \quad (17)$$

Matrices $\mathbf{M}$ and vectors $\mathbf{P}$ and $\mathbf{P'}$ are defined as:

$$\mathbf{M}_f = E\left[\Delta\mathbf{f}\,\Delta\mathbf{f}^H\right] \quad \mathbf{M}_t = E\left[\Delta\mathbf{t}\,\Delta\mathbf{t}^H\right]$$
$$\mathbf{P}_f = E\left[\Delta\mathbf{f}\right] \quad \mathbf{P}_t = E\left[\Delta\mathbf{t}\right]$$
$$\mathbf{P}_f' = E\left[\delta f\,\Delta\mathbf{f}\right] \quad \mathbf{P}_t' = E\left[\delta t\,\Delta\mathbf{t}\right] \quad (18)$$

Elements in $\mathbf{M}$ and $\mathbf{P}$ can be easily expressed by means of the characteristic function [3] of the FND and elements in $\mathbf{P'}$ by the derivative of that function.

Forcing zero partial conjugate derivatives of BR (15) with respect to gain matrices $\mathbf{T}_f$ and $\mathbf{T}_t$, we obtain the optimum gains for the constrained BRE (14):

$$\mathbf{T}_f = S^{-1}\left[\mathbf{R}^{-1}\,S\left[\mathbf{X}_f\right]\right]$$

$$\mathbf{T}_t = S^{-1}\left[\mathbf{R}^{-1}\, S\left[\mathbf{X}_t\right]\right] \tag{19}$$

where S is the stacking operator [2][4]. Matrix **R** is:

$$\mathbf{R} = \left(\mathbf{M}_f^T \otimes \mathbf{M}_t\right) \bullet \left(S\left[\mathbf{X}\right] S^H\left[\mathbf{X}\right]\right) + \mathrm{diag}\left[S\left[\mathbf{C}\right]\right] \tag{20}$$

Meaning: It is instructive to consider the solution (20) for weak signals (low SNR) (high values in matrix **C**):

$$\mathbf{T}_f =< \mathbf{C} >^{-1} \bullet \mathbf{X} \bullet \mathbf{P}_t \mathbf{P}_f'^H$$
$$\mathbf{T}_t =< \mathbf{C} >^{-1} \bullet \mathbf{X} \bullet \mathbf{P}_t' \mathbf{P}_f^H \tag{21}$$

where $<\mathbf{C}>^{-1}$ means the inverse of each element in **C**. (21) is a meaningful and intuitive result. We state (without proof) its interpretation: the steps for the recursive estimation of the (f & t) parameters of a weak (low SNR) cyclostationary signal with known period and CAM are:

1) take the Wigner-Ville [5] distribution (WVD) of the signal at each period (which is the 2D Fourier Transform of the ICAM [6]),
2) take a circular and weighted correlation with the expected signal WVD (Schur product with $<\mathbf{C}>^{-1}\bullet\mathbf{X}\bullet$ in the CAM domain), (that is in accordance with previous works [5] [7]),
3) widen (convolve) the resulting function according to the prior (in order to observe a broad enough maximum) (Schur product with **P** matrices in the CAM domain)
4) take the stochastic gradient in the time-frequency plane to update the estimate (Schur product with **P'** matrices in the CAM domain).

Implicitly, our estimator is doing all that.

2.2 Update prior. Prior mean, which constitutes the recursive estimate, is updated by (14) using the gains obtained in (19). Prior covariance is updated by the corresponding minimised BRs in (16):

$$\hat{\sigma}_f'(k) = R_f(P,\hat{\Theta})$$
$$\hat{\sigma}_t'(k) = R_t(P,\hat{\Theta}) \tag{22}$$

2.3. Predict prior. Consider a zero-order random walk model for the parameters' evolution:

$$f(k+1) = f(k) + v_f(k)$$
$$t(k+1) = t(k) + v_t(k) \tag{23}$$

where $v_f(k)$ is $N[0,\beta_f]$ r.v. and $v_t(k)$ is $N[0,\beta_t]$ r.v. (In general $N[\mu,\varpi]$ means Normal Density with mean μ and variance ϖ). Predicted prior mean is its previous update. Using (23), the predicted prior covariance is obtained by:

$$\hat{\sigma}_f(k+1) = \hat{\sigma}_f'(k) + \beta_f$$
$$\hat{\sigma}_t(k+1) = \hat{\sigma}_t'(k) + \beta_t \tag{24}$$

---

[4] The (Stacking) S operator [2] reshapes matrix (cxp) elements into vector (cpx1). S^{-1} reshapes again vector elements into a matrix.

Eq. (16), (22) and (24) constitute the so-called Riccati Equation. The key point is that our Riccati equation is independent of the estimates and can therefore be computed off_line. So, the gain matrices can be stored in a look-up table. Generally, in non-linear filtering, the Riccati Equation becomes state-dependent. In our problem this does not happen because: 1) the statistic of matrix **Z** in (12) is state-independent; 2) linear filtering from **Z** has been imposed in (14).

Consider now a first-order random walk for the frequency evolution:

$$f(k+1) = f(k) + d(k) + v_f(k)$$
$$d(k+1) = d(k) + v_d(k) \tag{25}$$

where $v_d(k)$ is $N[0,\beta_d]$ r.v. and the parameter d is the frequency rate modelling the natural tendency. Along with f and t, the parameter d also has to be estimated from the ICAMs. An estimate of d will be added to the updated estimate of f (14) at iteration k to make a prediction at k+1. Apart from σ_f, two new prior covariances appear: σ_{fd} and σ_d. The final estimator and Riccati equation are found by comparison with Linear Kalman Filter equations [4] corresponding to a scalar measurement and a state vector containing the parameter and its discrete derivative (upper triangular transition matrix). The result is:

$$\alpha = \hat{\sigma}_{fd} / \hat{\sigma}_{dd}$$
$$\hat{f}(k+1) = \hat{f}(k) + \mathrm{tr}\left[\mathbf{T}_f^H \hat{\mathbf{Z}}\right] + \hat{d}(k)$$
$$\hat{t}(k+1) = \hat{t}(k) + \mathrm{tr}\left[\mathbf{T}_t^H \hat{\mathbf{Z}}\right]$$
$$\hat{d}(k+1) = \hat{d}(k) + \alpha\, \mathrm{tr}\left[\mathbf{T}_f^H \hat{\mathbf{Z}}\right]$$
$$\hat{\sigma}_f(k+1) = \hat{\sigma}_f(k) + (\alpha+1)^2 \mathrm{tr}\left[\mathbf{A}_f\right] + 2\hat{\sigma}_{fd} + \hat{\sigma}_{dd} + \beta_f$$
$$\hat{\sigma}_t(k+1) = \hat{\sigma}_t(k) + \mathrm{tr}\left[\mathbf{A}_t\right] + \beta_t$$
$$\hat{\sigma}_{fd}(k+1) = \hat{\sigma}_{fd}(k) + \hat{\sigma}_{dd}(k) + \alpha(\alpha+1)\mathrm{tr}\left[\mathbf{A}_f\right]$$
$$\hat{\sigma}_{dd}(k+1) = \hat{\sigma}_{dd}(k) + \alpha^2 \mathrm{tr}\left[\mathbf{A}_f\right] + \beta_d \tag{26}$$

2.4. Criterion for matrix C. Matrix **C** defined in (8) is based on a fourth order statistics of the incoming signal. If such information is unknown, for low SNR, matrix **C** can be approximated by the second order statistics of the ICAM of the additive Gaussian white noise, which is proven to be:

$$\mathbf{C} \approx E\left[\hat{\mathbf{W}}(k) \bullet \hat{\mathbf{W}}^*(k)\right] = \Gamma^2\left(\mathbf{1}\mathbf{1}^T + 2\delta(\mathbf{k})\,\delta(\mathbf{m}^T)\right) \tag{27}$$

and thus proportional to the squared noise power.

3. SIMULATION RESULTS

The signal x(n) used for simulations is QPSK with rectangular pulse shaping. Due to the expected frequency ambiguity, an antialiasing filter of bandwidth 20 times the symbol rate is assumed. Its output is sampled at the Nyquist rate, so the number of samples per symbol is N=20. The SNR is then SNR=EbNo-10dB. Over 500 symbols, the discrete signal frequency f changes linearly from 0.3 to 0.2.

Timing t is -0.4. Vectors **m** and **k** are:
$\mathbf{m}=[0,1,2,3,4,6,8,10,14,18]^T$ and $\mathbf{k}=[-3,-2,-1,0,1,2,3]^T$.

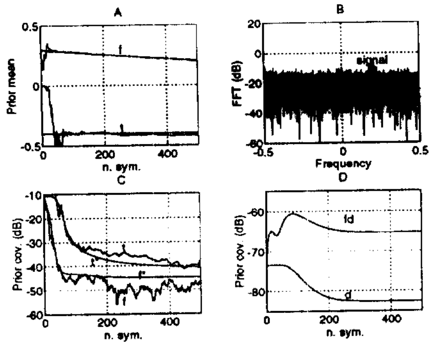

Fig.1. Prior mean and covariance evolution. EbNo=3dB.

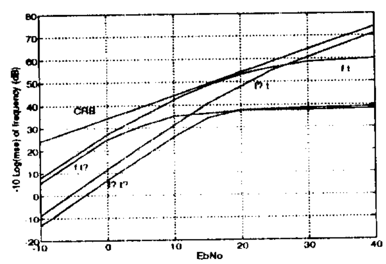

Fig.3. Small error performance vs. EbNo for different prior assumptions.

Fig.1 shows the evolution of the estimator at EbNo=3dB. Fig.1A shows a typical evolution of frequency and timing estimates. Fig.1B is an estimate of the signal power spectral density at iteration 500 (f=0.2). This plot illustrates how difficult the problem is. Fig.1C shows the evolution of frequency (f*) and timing (t*) prior covariances σ_f and σ_t given by the Riccati equation. The initial value of these covariances is 1/12, corresponding to a prior FND equal to the uniform distribution. The mean squared frequency (f) and timing (t) errors (averaged over 15 trials) of the estimates is also plotted. Real and a priori error evolutions are very close. Fig.1D shows the evolution of prior covariances σ_{fd} and σ_d. **Fig.2** shows the evolution of the estimator at EbNo=6dB and SIR=-6dB. To avoid the influence of a powerful pure sinusoid with unknown frequency, the model CAM of the desired signal plus the interfering signal has been averaged over all possible frequencies of the interfering signal. The result is a slower evolution of the Riccati equation (with respect to Fig.1). A good initial timing estimate has been assumed. **Fig.3** shows the mean squared error of the instantaneous (symbol by symbol) estimates of the frequency. Frequency gain matrix $\mathbf{T}_f$ is designed for EbNo=3dB. Cramer Rao Bound (CRB) for a data block of N=20 samples is also plotted. It is an optimistic bound as it assumes a pure sinusoid. Four different prior covariance pairs are considered.f?t? means Uniform Distribution prior for f and t. This illustrates the acquisition stage. ft: means Narrow Normal Distribution. This illustrates the tracking stage.

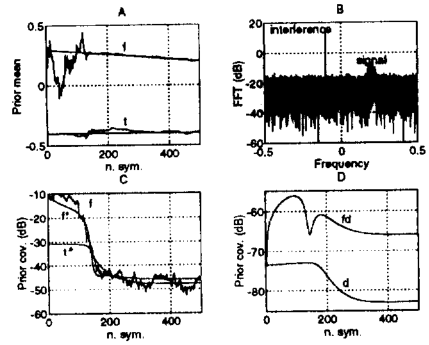

Fig.2. Prior mean and covariance evolution. EbNo=6dB. SIR=-6dB

Cases f?t and ft? illustrate the mutual information between the two parameters. For high EbNo, the curves become flat due to pattern noise. For small EbNo the curves have a slope of 2dB per dB due to the second order non-linearity of the filter.

4-CONCLUSIONS

We have presented a systematic off-line design of (cxp) gain matrices for recursively estimating the frequency and timing parameters from the instantaneous cyclic autocorrelation of a cyclostationary signal with known second order statistics. Each gain matrix minimises the Bayes Risk at each iteration considering the FND[1] as prior. The obtained estimator becomes a Non-Data-Aided candidate solution for frequency and timing recovery in highly non-stationary (high Doppler effect) environments. The quality of frequency estimates approaches the CRB for moderate EbNo during tracking, proving the importance of the timing information for the frequency estimation of a cyclostationary signal. Three further points are currently being investigated: 1) minimum information to be retained by the look-up table; we suspect that only 4(c+p) numbers (instead of cxp) per iteration will be necessary. 2) Joint estimation of signal power. 3) Criteria to select the lag-products (vector **m**) and their number.

5-REFERENCES

[1] Willsky. 'Fourier Series and Estimation on the Circle with Applications to Synchronous Communication'. IEEE Transactions on Information Theory. VOL. 20, NO. 5. September 1974.

[2] Richard Bellman. 'Introduction to Matrix Analysis'. 1970

[3] Scharf. 'Statistical Signal Processing; Detection, Estimation and Time Series Analysis'. 1990

[4] Anderson and Moore. 'Optimal Filtering'. Prentice-Hall, Englewood Cliffs, NJ, 1979.

[5] P. Flandrin. 'A Time-Frequency Formulation of Optimum Detection'. IEEE Transactions on ASSP VOL. 36, NO. 9. September 1986.

[6] W.A.Gardner. 'Statistical Spectral Analysis: A Non Probabilistic Theory'. 1988 Prentice Hall.

[7] W.A.Gardner. 'The Role of Spectral Correlation in Design and Performance Analysis of Synchronizers' . IEEE TC VOL-34 NO-8 November 1986.