# MINIMUM BER DFE EQUALIZER IN ALPHA STABLE NOISE

*Vimal Bhatia, Bernard Mulgrew and Apostolos Th. Georgiadis*

IDCOM, School of Engineering and Electronics,
The University of Edinburgh,
Edinburgh, EH9 3JL,
U.K.
Phone: +44 131 6505565, Fax: +44 131 6506554, Email: [v.bhatia, b.mulgrew, atg]@ee.ed.ac.uk

**ABSTRACT**

This paper addresses the problem of developing a least mean squares (LMS) style algorithm for minimizing bit error rate (BER) for updating linear/decision-feedback equalizers and multi-user detectors operating in impulsive noise environments characterized by an alpha stable distribution. The concepts build on earlier work in a Gaussian noise environment [1] and exploit some of the ideas developed in [2] for dealing with alpha stable noise as a mixture of a Gaussian and a Cauchy distribution. The development exploits the stable nature of the alpha distribution i.e. stability under linear transformation or filtering. An improvement in convergence and BER-performance is achieved by using minimum bit error rate (MBER) algorithm criterion over conventional LMS based design, as apparent from simulation results, thus making MBER criterion ideal for alpha stable noise environment.

## 1. INTRODUCTION

The Gaussian process has always been the dominant noise model in communications and signal processing, mainly because of the central limit theorem. In addition, the Gaussian assumption often leads to analytically tractable solutions [3]. Unfortunately, in many communication channels, the observation noise exhibits Gaussian, as well as impulsive characteristics. The sources of impulsive noise may be either natural (e.g. lightning, ice-cracking), or man-made. It may include atmospheric noise or ambient noise. It might come from relay contacts, electro-magnetic devices, electronic apparatus, or transportation systems, switching transients, and accidental hits in telephone lines [4, 5]. Most of the present day systems are optimized under the Gaussian assumption and their performance is significantly degraded by the occurrence of impulsive noise [2].

Impulsive noise is more likely to exhibit sharp spikes or occasional bursts of outlying observations than one would expect from normally distributed signals. A variety of impulsive noise models have been proposed [5, 6] However, most common model to represent impulsive phenomena is the family of $\alpha$-stable random variables [3]. Stable distributions share defining characteristics with the Gaussian distribution, such as the stability property and central limit theorems. Gaussian noise is considered as limiting case of $\alpha$-stable distribution when $\alpha = 2$.

In the following, a quick overview of stable processes is given in section-2. An overview of state-translated DFE and minimum-BER equalizer is presented in section-3 and section-4 respectively. The section-5 discusses derivations for an adaptive equalizers for $\alpha$-stable noise. Simulation technique and assumptions along with results are discussed in section-6. Finally conclusions are drawn in section-7.

## 2. THE CLASS OF STABLE RANDOM VARIABLES

The family of stable random variables (RV) is defined as direct generalization of the Gaussian law. The main characteristics of a non-Gaussian stable probability density function (pdf) is that its tails are heavier than those of the normal density. The symmetric $\alpha$-stable (S$\alpha$S) pdf is defined by means of its characteristic function $F(\omega) = \exp(\delta i\omega - \gamma|\omega|^{\alpha})$. The parameters $\alpha$, $\gamma$ and $\delta$ describe completely a S$\alpha$S distribution. The characteristics exponent $\alpha$ $(0 < \alpha \leq 2)$ controls the heaviness of the tails of the stable density; a smaller value implies heavier tails, while $\alpha = 2$ is the Gaussian case. The dispersion parameter $\gamma$ $(\gamma > 0)$ plays an analogous role to the variance and refers to the spread of the distribution. Finally, the location parameter $\delta$ is comparable with the mean of the distribution.

Theoretical justification for using the stable distribution as a basic statistical modelling tool come from the generalized central limit theorem. Unfortunately, no closed-form expressions exist for the stable density, except the Gaussian ($\alpha = 2$), Cauchy ($\alpha = 1$) and Pearson ($\alpha = \frac{1}{2}$) distributions. An important property of all stable distributions is that only the lower moments are finite. That is, if $\mathbf{x}$ is a stable RV, then $\mathbf{E_x}\{|x|^p\} < \infty$ iff $p < \alpha$. A well known consequence of this property is that all stable RV's with $\alpha < 2$ have infinite variance [2]. For further discussion on $\alpha$-stable RV and their properties refer [3].

## 3. BACKGROUND

The channel is modelled as a finite impulse response filter with an additive noise source, and the received signal at sample $k$ is

$$r(k) = \bar{r}(k) + e(k) = \sum_{i=0}^{n_a-1} a_i s(k-i) + e(k)$$

where $\bar{r}(k)$ denotes the noiseless channel output; $n_a$ is the channel length and $a_i$ are the channel tap weights; the white noise $e(k)$ has zero mean and is drawn from an alpha stable distribution with dispersion $\gamma$ and characteristic exponent $\alpha$; the symbol sequence $\{s(k)\}$ is independently identically distributed (IID) and has a 2-PAM (2 state pulse amplitude modulation) constellation.

For a conventional linear-combiner DFE the decision variable $z$ at time $k$ is a linear combination of received samples and past decisions:

$$z(k) = \mathbf{w}^T \mathbf{r}(k) - \mathbf{b}^T \hat{\mathbf{s}}_b(k)$$

where $\mathbf{r}(k) = [r(k)\, r(k-1) \ldots r(k-m+1)]^T$ is the channel observation vector, $\hat{\mathbf{s}}_b(k) = [\hat{s}(k-d-1)\, \hat{s}(k-d-2) \ldots \hat{s}(k-d-n)]^T$ is the past detected symbol vector, $\mathbf{w} = [w_0\, w_1 \ldots w_{m-1}]^T$ is the feedforward coefficient vector and $\mathbf{b} = [b_1\, b_2 \ldots b_n]^T$ is the feedback coefficient vector. The integers $d$, $m$ and $n$ will be referred to as the decision delay, the feedforward delay and feedback taps respectively. Without loss of generality, $d = n_a - 1$, $m = n_a$ and $n = n_a - 1$ will be used as this choice of DFE structure parameters which is sufficient to guarantee the linear separability of the subsets of the channel states related to the different decisions [7]. Alternatively the linear-combiner DFE can be expressed in state translated form [8]:

$$z(k) = \mathbf{w}^T (\mathbf{r}(k) - \mathbf{F}_2 \hat{\mathbf{s}}_b(k)) = \mathbf{w}^T \mathbf{r}'(k) \qquad (1)$$

where $\mathbf{F}_2$ is constructed by partitioning the channel impulse response matrix $\mathbf{F} = [\mathbf{F}_1\,\mathbf{F}_2]$, where:

$$\mathbf{F}_1 = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n_a-1} \\ 0 & a_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_1 \\ 0 & \cdots & 0 & a_0 \end{bmatrix}$$

$$\mathbf{F}_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ a_{n_a-1} & 0 & \ddots & \vdots \\ a_{n_a-2} & a_{n_a-1} & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ a_1 & \cdots & a_{n_a-2} & a_{n_a-1} \end{bmatrix}$$

Since the linear-combiner DFE is a special case of the generic DFE structure, by performing translation of eq. (1), it is reduced to the equivalent linear equalizer 'without decision feedback':

$$f'(\mathbf{r}'(k)) = \mathbf{w}^T\mathbf{r}'(k) \quad (2)$$

The decision boundary of this equivalent linear equalizer consists of M-1 hyperplanes defined by: $\mathbf{r}' : \mathbf{w}^T\mathbf{r}' = 2i - M$, $1 \le i \le M-1$. These M-1 parallel hyperplanes can always be designed properly to separate the M subsets of the translated channel states $R^{(i)}$, $1 \le i \le M$. In particular, for $M = 2$, the decision boundary, $\mathbf{r}' : \mathbf{w}^T\mathbf{r}' = 0$, is a hyperplane passing through the origin of the $\mathbf{r}'$(k)-space. It is shown, in [7], that in the state translation the channel states remain separable despite translation. The states can be made separable by applying a simple initial condition. The performance of state translated linear combiner DFE is shown to be better than conventional minimum mean square error (MMSE) DFE, however performance depends on the accuracy of the built-in channel estimator.

The Wiener or MMSE solution is often said to provide the optimal $\mathbf{w}$ and $\mathbf{b}$. It is however optimal only with respect to the mean square error criterion. Obviously, there must exist a solution $\mathbf{w}_{opt}$ which achieves the best equalization performance for the structure of eq. (2). We refer to this $\mathbf{w}_{opt}$ as the minimum bit error rate (MBER) solution of the linear-combiner DFE. The MMSE linear-combiner DFE is generally not this MBER solution. A natural question is how different the MMSE and MBER solutions can be. The difference in performance is demonstrated in [7].

## 4. MINIMUM BIT ERROR RATE EQUALIZATION

The bit error rate (BER) observed at the output of the equalizer is dependent on the distribution of the decision variable $z(k)$ which in turn is a function of the equalizer tap weights. To be more specific, the probability of error, $P_E$, is:

$$P_E = P(sgn(s(k-d))z(k) < 0)$$

The sign adjusted decision variable $z_s(k) = sgn(s(k-d))z(k)$ is drawn from a mixture process. From the definition of $z(k)$,

$$\begin{aligned} z_s(k) &= sgn(s(k-d))(\mathbf{w}^T\mathbf{F}\mathbf{s}(k) - \mathbf{b}^T\hat{\mathbf{s}}_b(k)) \\ &\quad + sgn(s(k-d))\mathbf{w}^T\mathbf{e}(k) \\ &= sgn(s(k-d))z'(k) + e'(k) \end{aligned} \quad (3)$$

$\mathbf{e}(k) = [e(k)\,e(k-1)\ldots e(k-d-n)]^T$ is the vector of noise samples; $\mathbf{s}(k) = [s(k)\,s(k-1)\ldots s(k-d-n_{n_a})]^T$ is the vector of transmitted symbols. The first term on the right hand side of eqn. (3) $sgn(s(k-d))z'(k)$, is the noise-free sign-adjusted equalizer output and is a member of a finite set with $N_z$ elements - these are the local means of the mixture. Without noise the combination of channel and DFE is a finite state machine whose state is defined by the vector $\mathbf{s}(k)$. Thus if $\mathbf{s}(k) \in \{\mathbf{s}_1 \cdots \mathbf{s}_i \cdots \mathbf{s}_{N_z}\}$, the state $\mathbf{s}_i$ uniquely defines the state of $z'(k)$, $\mathbf{r}(k)$, $s(k-d)$ and $\hat{\mathbf{s}}_b(k)$ - label these $z_i$, $\mathbf{r}_i$, $s_i$ and $\hat{\mathbf{s}}_{bi}$ respectively. Note that while $\mathbf{s}(k)$ has $N_z$ states, $s(k-d)$ has 2 possible values (2-PAM). However since $s(k-d)$ is a component of the vector $\mathbf{s}(k)$, the state of $\mathbf{s}(k)$ uniquely defines the value of $s(k-d)$. The second term $e'(k)$ is a zero mean $\alpha$-stable white noise process with dispersion $\gamma(\sum_{j=1}^{M}|w_j|^{\alpha})^{\frac{1}{\alpha}}$ and characteristic exponent $\alpha$ - defining the distribution about the local means.

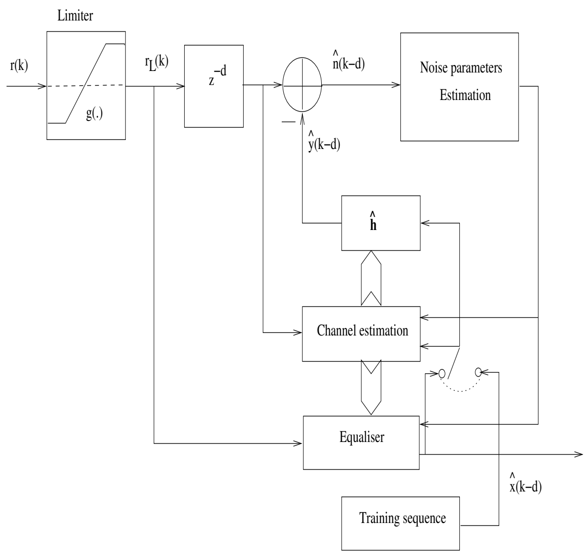


Figure 1: Receiver architecture

## 5. ADAPTIVE EQUALIZER

Consider noise density function $p(x)$ associated with the zero mean random variable $x$. The density function is symmetrical and normalized such that the variance or dispersion is unity. The associated distribution function is $P(x)$. The "generalized" error function is $Q(x) = 1 - P(x)$ and its derivative is $Q'(x) = -p(x)$.

The probability of error at the output of a linear or state translation equalizer with $N$ noise free states as a function of the weight $M$-vector $\mathbf{w}$ is:

$$P_E(\mathbf{w}) = \frac{1}{N}\sum_{i=1}^{N} Q(g_i(\mathbf{w}))$$

where $g_i(\mathbf{w})$ is the signed decision variable, normalized by the "strength" of the noise. In the Gaussian case :

$$g_i(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{r}_i s_i}{||\mathbf{w}||\sigma}$$

where $\mathbf{r}_i$ is the $i^{th}$ noise free received vector; the Euclidean norm is $||\mathbf{w}|| = (\sum_{j=1}^{M}|w_j|^2)^{\frac{1}{2}}$; $s_i$ is the transmitted symbol associated with that vector; $\sigma^2$ is the noise variance. In the $\alpha$-stable case:

$$g_i(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{r}_i s_i}{||\mathbf{w}||_{\alpha}\gamma^{\frac{1}{\alpha}}}$$

where the "$\alpha$-norm" is defined as: $||\mathbf{w}||_{\alpha} = (\sum_{j=1}^{M}|w_j|^{\alpha})^{\frac{1}{\alpha}}$

For adaptive filters, derivatives of the form $\partial P_E/\partial w_j : \forall j$ are required.

$$\begin{aligned} \frac{\partial P_E}{\partial w_j} &= \frac{1}{N}\sum_{i=1}^{N} Q'(g_i(\mathbf{w}))\frac{\partial g_i(\mathbf{w})}{\partial w_j} \\ &= -\frac{1}{N}\sum_{i=1}^{N} p(g_i(\mathbf{w}))\frac{\partial g_i(\mathbf{w})}{\partial w_j} \end{aligned}$$

No analytic expression for the density $p(x)$ exists in the general $\alpha$-stable case. However it can be approximated by a mixture of a Gaussian, $p_g(x)$ and a Cauchy $p_c(x)$ [2]:

$$\hat{p}_\alpha(x) = \beta p_g(x) \; + \; (1-\beta)p_c(x)$$

where $\beta$ is dependent on $\alpha$, $\gamma$ and the dynamic range of the receiver. Alternatively, a look-up table and interpolation can be used to evaluate the density. In the Gaussian case the derivative is given by:

$$\frac{\partial g_i(\mathbf{w})}{\partial w_j} = \frac{\partial}{\partial w_j}\left(\frac{\mathbf{w}^T}{||\mathbf{w}||}\right)\frac{\mathbf{r}_i s_i}{\sigma}$$

$$= \frac{1}{||\mathbf{w}||}\left(\mathbf{1}_j^T - \frac{\mathbf{w}^T w_j}{||\mathbf{w}||^2}\right)\frac{\mathbf{r}_i s_i}{\sigma}$$

where $\mathbf{1}_j$ is an $M$-vector with all zero elements apart from the $j$th entry which is unity. In the $\alpha$-stable case:

$$\frac{\partial g_i(\mathbf{w})}{\partial w_j} = \frac{\partial}{\partial w_j}\left(\frac{\mathbf{w}^T}{||\mathbf{w}||_\alpha}\right)\frac{\mathbf{r}_i s_i}{\gamma^{\frac{1}{\alpha}}}$$

$$= \frac{1}{||\mathbf{w}||_\alpha}\left(\mathbf{1}_j^T - \frac{\mathbf{w}^T |w_j|^{\alpha-1} sgn(w_j)}{||\mathbf{w}||_\alpha^\alpha}\right)\frac{\mathbf{r}_i s_i}{\gamma^{\frac{1}{\alpha}}}$$

Since the $\alpha$-stable case is more general we will work with it from now on. Multiply out gives:

$$\frac{\partial g_i(\mathbf{w})}{\partial w_j} = \frac{1}{||\mathbf{w}||_\alpha}\left(r_{ij} - \frac{z_i |w_j|^{\alpha-1} sgn(w_j)}{||\mathbf{w}||_\alpha^\alpha}\right)\frac{s_i}{\gamma^{\frac{1}{\alpha}}}$$

where $r_{ij}$ is the $j^{th}$ element of $\mathbf{r}_i$ and $z_i = \mathbf{w}^T\mathbf{r}_i$ i.e. the equalizer output associated with the $i^{th}$ noise free state. Collecting partial derivatives together to form a gradient vector we have:

$$\nabla P_E(\mathbf{w}) = -\frac{1}{N}\sum_{i=1}^{N} p(g_i(\mathbf{w}))\frac{1}{||\mathbf{w}||_\alpha}\left(\mathbf{r}_i - \frac{z_i\langle\mathbf{w}\rangle_\alpha}{||\mathbf{w}||_\alpha^\alpha}\right)\frac{s_i}{\gamma^{\frac{1}{\alpha}}}$$

where $\langle\mathbf{w}\rangle_\alpha$ is an $M$-vector with $j^{th}$ element is $|w_j|^{\alpha-1}sgn(w_j)$. Since the norm of the weight vector does not affect $P_E$ in the binary signalling case it can be set to unity at each iteration thus:

$$\nabla P_E(\mathbf{w}) = -\frac{1}{N\gamma^{\frac{1}{\alpha}}}\sum_{i=1}^{N} p(\frac{z_i s_i}{\gamma^{\frac{1}{\alpha}}})\,(\mathbf{r}_i - \langle\mathbf{w}\rangle_\alpha z_i)\, s_i$$

Using the kernel density ideas developed in [1] leads to an LMS-style least bit error rate (LBER) algorithm.

Filter output:

$$z(k) \; = \; \mathbf{w}^T(k)\mathbf{r}(k)$$

Update weights:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu p(\frac{z(k)s(k-d)}{\gamma^{\frac{1}{\alpha}}})$$

$$(\mathbf{r}(k) - \langle\mathbf{w}(k)\rangle_\alpha z(k))\frac{s(k-d)}{\gamma^{\frac{1}{\alpha}}}$$

The equalizer tap weights are normalized after each update. The final decision, $\hat{s}(k-d)$, is made on the filter output $\mathbf{w}^T(k)\mathbf{r}(k)$.

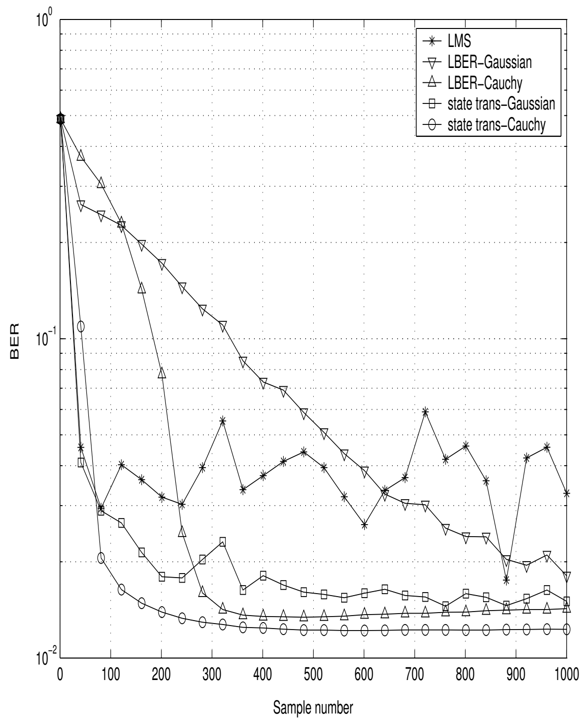


Figure 2: Convergence plot for Cauchy ($\alpha$ = 1) distributed noise

## 6. SIMULATION STUDY

The traditional performance measures are, usually, plots of the bit error ratio (BER) against the signal to noise ratio (SNR). In a Gaussian noise environment the computation of SNR involves the variance of the useful received signal, as well as the variance of the corrupting noise. In $\alpha$-stable noise environment with $\alpha < 2$, however, the variance of the noise is infinite [3] making the use of traditional BER to SNR graphs meaningless. Nevertheless, all receivers in practice have a finite input dynamic range. The limiter at the front end of the receiver is assumed to be an ideal saturation device, with transfer function

$$g(x,G) = \begin{cases} x & : \; |x| \leq G \\ sign(x)G & : \; elsewhere \end{cases}$$

G being the saturation point of the limiter. For a given saturation limit G, the SNR at the limited received signal $r_L(k)$ is always finite. Fig. 1 shows the equalizer structure used in this paper. The SNR of the limited received signal $r_L(k)$ is used for performance evaluation in environments where the noise variance is infinite. This is referred as the SNR *at the receiver*.

Simulation were performed by assuming that noise is Cauchy distributed i.e. $\alpha = 1$ and clipper, at DFE front-end, is at $\pm 4$. The variance of the noise estimate $\hat{n}(k)$ is calculated as discussed in [2]. The channel was chosen as [0.3482 0.8704 0.3482]. DFE structure was chosen to be $d$ = 2, $m$ = 3 and $n$ = 2. The legends in Fig. 2 and Fig. 3 depict: a) LMS for conventional least mean square algorithm for both feedforward and feedback taps of the equalizer, b) LBER-Gaussian for least bit error rate [1] algorithm for adapting both feedforward and feedback equalizer taps assuming that the noise is Gaussian, c) LBER-Cauchy for LBER algorithm for adapting both feedforward and feedback equalizer taps assuming Cauchy distributed noise, d) state trans-Gaussian; same structure as (b), however state translated design [7], e) state trans-Cauchy; same structure as (c) i.e. both the feedback and feedforward filter adapt assuming Cauchy distributed noise.

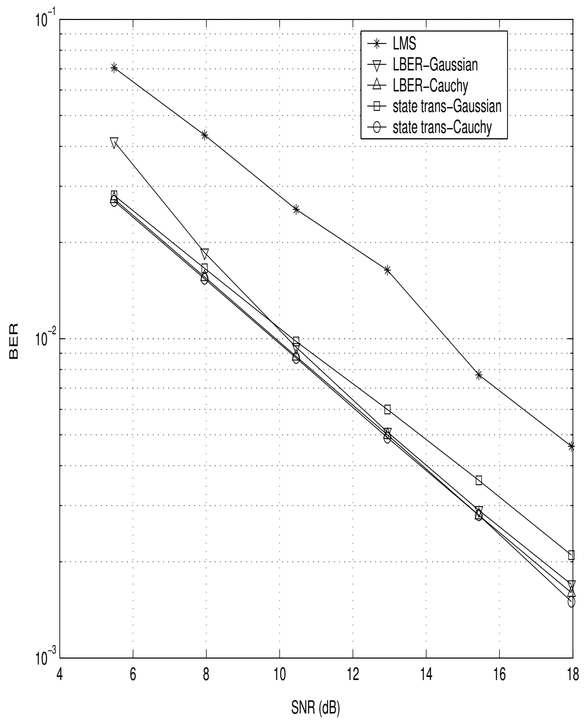


Figure 3: Performance plot for Cauchy ($\alpha = 1$) distributed noise

Fig. 2 demonstrates convergence rates for different algorithms. An ensemble of 100-runs over SNR = 7.9 dB was taken to generate this plot. Fig. 3 illustrates BER performance of the algorithms for $10^5$ input samples and over an ensemble of 1000-runs. Simulations suggest that state-translated design for Cauchy distributed noise achieves better performance than other algorithms. The convergence of the state-translated design for Cauchy distributed noise is faster and stable. The LMS algorithm, as expected, suffers from instability and poor performance for $\alpha$-stable noise, primarly because of dependence on instantaneous error. At BER of $5 * 10^{-3}$ a gain of approximately 5 dB is achieved over traditional LMS algorithm. The performance of LBER-Gaussian and state translated LBER-Gaussian algorithms suggests ability of Gaussian mixtures to model $\alpha$-stable distribution [9].

## 7. CONCLUSIONS

An minimum bit error rate adaptive algorithm for impulsive noise modelled as $\alpha$-stable noise has been proposed in this paper. It is shown that for minimum bit error design, the adaptation is a function of the noise density function. The comparison between various adaptive algorithms working in identical channel, noise and DFE structure has been drawn. The LBER-Cauchy and the state trans-Cauchy has faster convergence than the other adaptive algorithms in Cauchy noise environments, which is a special form of $\alpha$-stable noise. Extensive simulations strongly suggest that the state-translated design for the $\alpha$-stable noise has better convergence and BER performance than the other algorithms. It is also interesting to observe that the adaptive algorithms based on Gaussian noise assumption despite slow convergence in impulsive noise environment perform closer to those designed with Cauchy noise assumption. This can be attributed to the fact that $\alpha$-stable noise can be modelled as Gaussian mixture [9, 10] and the effect of the limiter at front end. Lastly as expected LMS performs poorer that other algorithms in $\alpha$-stable noise environments. From this discussion we can safely conclude that the adaptation cost function based on noise distribution results in improved performance.

## 8. ACKNOWLEDGMENT

Vimal acknowledges the support of EPSRC for this work.

## REFERENCES

[1] B. Mulgrew and S. Chen, "Adaptive Minimum-BER Linear Decision Feedback Equalisers for Binary Signalling," *Signal Processing*, vol. 81, July 2001, pp. 1479-1489.

[2] A.T. Georgiadis and B. Mulgrew, "Adaptive Bayesian Decision Feedback Equaliser for Alpha-Stable Noise Environments," *Signal Processing*, vol. 81, Aug. 2001, pp. 1603-1623.

[3] M. Shao and C. L. Nikias, "Signal Processing with fractional lower order moments: Stable processes and their applications," *Proceedings of the IEEE*, vol. 81, July 1993, pp. 986-1009

[4] G. A. Tsihrintzis and C. L. Nikias, "Performance of optimum and suboptimum receivers in the presnece of impulsive noise modeled as an alpha-stable process," *Proceedings of MILCOM'93*, vol. 2, July 1993, pp. 658-662.

[5] L. F. Lind and N. A. Mufti, "Efficient method for modelling implusive noise in a communication system," *IEE Electronics Letters*, vol. 32, Aug. 1996, pp. 1440-1441.

[6] I. Mann, S. McLaughlin, W.Henkel, R. Kirby, and T. Kessler, "Impulse generation with appropriate amplitude, length, interarrival, and spectral characteristics," *IEEE Journal on selected areas in com.*, vol. 20, July 2002, pp. 901-912.

[7] S. Chen, B. Mulgrew, E. S. Chng and G. J. Gibson, "Space translation properties and the minimum-BER linear-combiner DFE," *IEE Proc. Communications*, vol. 145, 1998, pp. 316-322.

[8] A. P. Clark, L. H. Lee and R. S. Marshall, "Developments of the conventional non-linear equaliser," *IEE Proceedings Part F*, vol. 129, 1982, pp. 85-94.

[9] E.E. Kuruoglu, C. Molina and W.J. Fitzgerald, "Approximation of alpha-stable probability densities using finite Gaussian mixtures," *Proceedings of EUSIPCO 98, Signal Processing IX: Theories and Applications*, vol. 2, Sept. 1998, pp. 989-992.

[10] R.J. Kozick, R.S. Blum, and B.M. Sadler, "Signal processing in non-Gaussian noise using mixture distributions and the EM algorithm," *Thirty-First Asilomar Conference on Signals, Systems and Computers*, vol. 1, Nov. 1997, pp. 438–442.