# MLSE for CPM Signals in a Fading Multipath Channel *


Linda M. Zeger
Lucent Technologies

Hisashi Kobayashi
Princeton University



## Abstract

We perform sequence estimation for CPM signals transmitted in a time varying multipath channel. The EM (Expectation–Maximization) algorithm, an iterative procedure for producing maximum likelihood estimates, is applied to handle the unknown channel. In order to enable implementation of the EM algorithm in this system, a simplification of this algorithm is derived. Channel estimates derived from training data or estimates of previously transmitted information symbols are used initially. In subsequent iterations tentative estimates of current information symbols are used in addition to improve the channel estimates, which in turn improve sequence estimation. Simulation results of the simplified EM algorithm applied to GSM are presented.



**keywords:** EM algorithm, GSM.


## I Introduction

The EM algorithm [1] can be applied to simultaneously perform sequence estimation and handle an unknown channel by averaging over possible realizations of unknown channel parameters at each point in time [2]. The convergence properties of the EM algorithm [1] imply that the final sequence estimate is the MLSE (maximum likelihood sequence estimate), where the likelihood of the received samples given the transmitted data sequence is not dependent on an estimate of the channel. Application of the EM algorithm can thus potentially enable performance improvement over equalization methods that are based on a single possibly inaccurate estimate of the channel at each point in time. Recent application of the EM algorithm to sequence estimation [2] considers a received signal that is linear in the transmitted symbols and a channel that is either a random phase channel or a random amplitude fading channel. The average over unknown channel parameters for these two models was performed analytically.

In this paper, we apply the EM algorithm to estimate data sequences transmitted by continuous phase modulation (CPM) in a time varying multipath channel. Our channel model has random time varying parameters for the amplitude, phase shift, and time delay for each multipath component. The averaging over these multipath parameters can not be performed analytically, and is computationally intractable. Therefore, we derive a simplified version of the EM algorithm to enable implementation. This simplified EM algorithm is equivalent to the original EM algorithm when an adequate amount of training data is used.

In Section II the modulation and channel models are specified. The EM algorithm is applied to sequence estimation in Section III, and an implementable version of the EM algorithm is derived for these models in Section IV. The construction of the prior density used in the EM algorithm is discussed in Section V. Simulation results based on the GSM system are presented in Section VI, and conclusions are drawn in Section VII.

## II The Model

The transmitted data sequence is denoted by $C_n$ for $n = 1, 2, 3, \ldots$ where $C_n \in \{-1, 1\}$, and the sequence is denoted collectively by $\mathbf{C}$. The transmitted signal uses some form of CPM, and is thus given by

$$X(t) = \cos(\omega_c t + \sum_n C_n q(t - nT)), \qquad (1)$$

where $T$ denotes the bit period, and $\omega_c$ is the carrier frequency. The continuous function $q(t)$ can be represented as the integral of a baseband pulse.

We consider a general multipath model with a total of $M$ paths. The received signal is then

$$V(t) = \sum_{i=1}^{M} \alpha_i(t) X(t - \tau_i(t), \theta_i(t)) \qquad (2)$$

where the phase shift, amplitude, and delay of the $i^{th}$ path are denoted respectively by $\theta_i(t)$, $\alpha_i(t)$, and $\tau_i(t)$. White Gaussian noise is added to the multipath fading model (2), and the resulting $V(t)$ can then be represented as

$$V(t) = \Re\{Y(t) \exp(j\omega_c t)\}. \qquad (3)$$

The complex envelope $Y(t)$ of the received signal is then

$$Y(t) = \sum_{i=1}^{M} \alpha_i(t) \exp\left[j\left(\Phi_i(t, \mathbf{C})\right)\right] + n_I(t) + jn_Q(t), \qquad (4)$$

where the phase is

$$\Phi_i(t, \mathbf{C}) = \theta_i(t) - \omega_c \tau_i(t) + \sum_n C_n q(t - \tau_i(t) - nT), \qquad (5)$$

and the inphase and quadrature noise components $n_I$ and $n_Q$ are independent WGN processes with noise density $N_o$.

We denote the channel parameters collectively by $\beta$:

$$\beta = \beta(t) = \{\alpha_1(t), \theta_1(t), \tau_1(t), \ldots \alpha_M(t), \theta_M(t), \tau_M(t)\}. \qquad (6)$$


*This work has been supported, in part, by the National Science Foundation, the NJ Commission on Science and Technology, and the Ogasawara Foundation for the Promotion of Science and Engineering. The second author has also been supported by Asahi Kasei Co.



0-7803-5582-2/99/$10.00 ©1999 IEEE

The receiver should produce the MLSE of the transmitted sequence $\mathbf{C}$, given the samples of $Y(t)$, denoted collectively by $\mathbf{y}$: the likelihood $f(\mathbf{y} \mid \mathbf{C})$ must be maximized with respect to $\mathbf{C}$. However, the received signal samples $\mathbf{y}$ depend on unknown channel parameters $\beta$. Conventional methods compute an estimate $\bar{\beta}$ of $\beta$, and then, given $\bar{\beta}$, produce the sequence estimate of $\mathbf{C}$ that maximizes $f(\mathbf{y} \mid \mathbf{C}, \bar{\beta})$. Bit errors in the estimate of $\mathbf{C}$ arise that are due to errors in the channel estimate $\bar{\beta}$.

# III The EM Algorithm for Sequence Estimation

In order to remove bit errors due to uncertainty in the channel, we apply the EM algorithm to take an average over the unknown channel parameters. The EM algorithm is used here to produce estimates $\mathbf{C}$ that maximize $f(\mathbf{y} \mid \mathbf{C})$ by averaging the logarithm of another likelihood function $f(\mathbf{y}, \beta \mid \mathbf{C})$ over $\beta$. We now summarize application of the EM algorithm to sequence estimation in the presence of a general unknown channel with parameters denoted by $\beta$. Equations (7) to (9) below parallel the formulation in [2], where application of the EM algorithm to sequence estimation was first described. We then focus on application to the modulation and channel models specified in Section II.

The EM algorithm consists of repeating two steps until convergence, at which point the estimate of $\mathbf{C}$ will be a local maximum of $f(\mathbf{y} \mid \mathbf{C})$. The E (Expectation) step at the $(p+1)^{st}$ iteration computes the expected log likelihood

$$Q(\mathbf{C}^{p+1} \mid \mathbf{C}^{p}) = E\left[\ln f(\mathbf{y}, \beta \mid \mathbf{C}^{p+1}) \mid \mathbf{y}, \mathbf{C}^{p}\right], \quad (7)$$

as a function of $\mathbf{C}^{p+1}$, given $\mathbf{y}$ and the estimate $\mathbf{C}^{p}$ of $\mathbf{C}$ from the previous ($p^{th}$) iteration. The conditional density $f(\beta \mid \mathbf{y}, \mathbf{C}^{p})$ is used to take the expectation over the unknown parameters $\beta$. The M (Maximization) step of the $(p+1)^{st}$ iteration determines the transmitted sequence $\mathbf{C}^{p+1}$ that maximizes $Q(\mathbf{C}^{p+1} \mid \mathbf{C}^{p})$ given $\mathbf{C}^{p}$. This estimate $\mathbf{C}^{p+1}$ will then be used in the E step in the $(p+2)^{nd}$ iteration in $f(\beta \mid \mathbf{y}, \mathbf{C}^{(p+1)})$.

As shown in [2], the E step (7) is reduced to

$$Q(\mathbf{C}^{p+1} \mid \mathbf{C}^{p}) = \int \ln f(\mathbf{y} \mid \beta, \mathbf{C}^{p+1}) f(\beta \mid \mathbf{y}, \mathbf{C}^{p}) d\beta, \quad (8)$$

where the density $f(\beta \mid \mathbf{y}, \mathbf{C}^{p})$ can be expressed as

$$f(\beta \mid \mathbf{y}, \mathbf{C}^{p}) = f(\mathbf{y} \mid \beta, \mathbf{C}^{p}) \rho(\beta). \quad (9)$$

The EM algorithm is now applied to CPM and the multipath channel model described in Section II. The density $f(\mathbf{y} \mid \beta, \mathbf{C})$ can be obtained from (4), and omitting the normalization constant is :

$$f(\mathbf{y} \mid \beta, \mathbf{C}) = \exp - \sum_{l=1}^{K} \Bigg\{ \quad (10)$$

$$\left[ y_I(t_l) - \sum_{i=1}^{M} \alpha_i(t_l) \cos\left(\Phi_i(t_l, \mathbf{C})\right) \right]^2$$

$$+ \left[ y_Q(t_l) - \sum_{i=1}^{M} \alpha_i(t_l) \sin\left(\Phi_i(t_l, \mathbf{C})\right) \right]^2 \Bigg\} / 2N_o,$$

where $K$ is the number of samples of the received signal, and the inphase and quadrature components of the samples taken at time $t_l$ are denoted by $y_I(t_l)$ and $y_Q(t_l)$ respectively. We denote the negative logarithm of (10) by:

$$\lambda(\beta, \mathbf{y}, \mathbf{C}) = -\ln f(\mathbf{y} \mid \beta, \mathbf{C}). \quad (11)$$

Substitution of (11), evaluated at $\mathbf{C}^{p+1}$, into (8) yields

$$Q(\mathbf{C}^{p+1} \mid \mathbf{C}^{p}) = -\int \lambda(\beta, \mathbf{y}, \mathbf{C}^{p+1}) f(\beta \mid \mathbf{y}, \mathbf{C}^{p}) d\beta. \quad (12)$$

Equation (9) is used in (12), where construction of the prior density $\rho(\beta)$ is discussed in Section V. The E and M steps are repeated until $\mathbf{C}^{p+1} = \mathbf{C}^{p}$, when convergence is achieved, at which point the estimated sequence $\mathbf{C}^{p+1}$ is a maximum of $f(\mathbf{y} \mid \mathbf{C})$.

The EM algorithm could be directly implemented if the multiple integral in (12) could be performed analytically, as it can for the modulation and channel models considered in [2], which produce a likelihood $f(\mathbf{y} \mid \beta, \mathbf{C})$ simpler than that of (10). However, for modulations such as GMSK with the channel model of Section II, calculation of $Q(\mathbf{C}^{p+1} \mid \mathbf{C}^{p})$ would require numerical integration for every realization of the sequence $\mathbf{C}^{p+1}$, which would be computationally intractable.

# IV Simplified EM Algorithm

We now simplify the EM algorithm to enable evaluation of (12). This simplification is accomplished by deriving a Gaussian approximation to (9) as a starting point. We first note that the sequence $\mathbf{C}$ is divided into subsequences of symbols such that the channel varies little over the length of a subsequence. Assuming a constant but unknown channel during each subsequence, we use the EM algorithm separately within each such subsequence of symbols to compute the MLSE of that subsequence. We let a subscript $s$ denote the number labeling each subsequence.

We begin by deriving a Gaussian approximation to $f(\mathbf{y}_s \mid \beta, \mathbf{C_s}^p)$. We consider the estimate $\tilde{\beta}_s^p$ at iteration $p$ of the true $\beta$ during transmission of subsequence $s$ that maximizes the likelihood $f(\mathbf{y}_s \mid \beta, \mathbf{C_s}^p)$, where $\mathbf{C}_s^p$ denotes the current estimate of the transmitted symbols in that subsequence, and $\mathbf{y}_s$ denotes the corresponding samples of the received signal. Thus $\tilde{\beta}_s^p$ is defined by

$$\tilde{\beta}_s^p = \tilde{\beta}_s^p(\mathbf{y_s}, \mathbf{C}_s^p) = \arg\min_{\beta} \lambda(\beta, \mathbf{y_s}, \mathbf{C_s}^p), \quad (13)$$

where $\lambda(\beta, \mathbf{y_s}, \mathbf{C_s}^p)$ is obtained from (10) and (11). In order to enable this minimization to be performed numerically, we require CPM so that the phase $\Phi_i(t_l, \mathbf{C_s}^p)$ as given by (5) is continuous. When $\beta$ is close to $\tilde{\beta}_s^p$, a Taylor series expansion of the exponent in (10) about $\tilde{\beta}_s^p$ yields

$$f(\mathbf{y}_s \mid \beta, \mathbf{C}_s^p) \approx \exp - \left[ (\beta - \tilde{\beta}_s^p)^t \mathbf{A}_s^p (\beta - \tilde{\beta}_s^p)/2 \right], \quad (14)$$

where an overall constant factor independent of $\beta$ has been omitted because it is irrelevant in the M step, and the superscript $t$ denotes transpose. The elements of the positive definite matrix $\mathbf{A}_s^p$ are given by

$$\mathbf{A}_{sij}^p = \frac{1}{2N_o}\frac{\partial^2\lambda(\beta,\mathbf{y_s},\mathbf{C_s}^p)}{\partial\beta_i\partial\beta_j}\,|_{\tilde{\beta}_s^p}\,. \tag{15}$$

We consider a Gaussian prior density $\rho_s(\beta)$:

$$\rho_s(\beta) \approx \exp -\left[(\beta-\bar{\beta}_{sp})^t\mathbf{B}_s(\beta-\bar{\beta}_{sp})/2\right]. \tag{16}$$

The mean $\bar{\beta}_{sp}$ and inverse variance matrix $\mathbf{B}_s$ depend on the speed at which the channel changes, and are derived explicitly in Section V. The posterior density (9) for the channel parameters is therefore the product of the Gaussian densities (14) and (16), and thus can be expressed as the Gaussian

$$f(\beta\mid\mathbf{y_s},\mathbf{C}_s^p) \approx \exp -\left[(\beta-\hat{\beta}_s^p)^t\hat{\mathbf{A}}_s^p(\beta-\hat{\beta}_s^p)/2\right], \tag{17}$$

where the estimated posterior mean $\hat{\beta}_s^p$ and the inverse posterior variance matrix $\hat{\mathbf{A}}_s^p$ in subsequence $s$ at iteration $p$ are

$$\hat{\beta}_s^p = (\hat{\mathbf{A}}_s^p)^{-1}\left(\mathbf{A}_s^p\tilde{\beta}_s^p+\mathbf{B}_s\bar{\beta}_{sp}\right) \tag{18}$$

$$\hat{\mathbf{A}}_s^p = \mathbf{A}_s^p+\mathbf{B}_s. \tag{19}$$

Equation (18) is thus a weighted sum of the estimate $\bar{\beta}_{sp}$ obtained from the prior density, which for example is derived from training data, and the estimate $\tilde{\beta}_s^p$, which is derived from the estimate of the information sequence $\mathbf{C}_s^p$ through (13). At the $p^{th}$ iteration the estimates $\tilde{\beta}_s^p$ and $\mathbf{A}_s^p$ derived from the likelihood function are updated, while the estimates $\bar{\beta}_{sp}$ and $\mathbf{B}_s$ derived from the prior density remain fixed throughout all iterations for subsequence $s$.

Using the approximation (17) in (12), it is seen that most of the contribution to (12) comes from values of $\beta$ near the peak $\hat{\beta}_s^p$ of $f(\beta\mid\mathbf{y_s},\mathbf{C}_s^p)$. We consider a region $R(\hat{\mathbf{A}}_s^p)$ around $\hat{\beta}_s^p$ with size determined by $\hat{\mathbf{A}}_s^p$ such that most of the contribution to $Q(\mathbf{C}_s^{p+1}\mid\mathbf{C}_s^p)$ comes from $\beta$ in $R(\hat{\mathbf{A}}_s^p)$. The region $R(\hat{\mathbf{A}}_s^p)$ grows with $N_o$, since matrix elements of both $\mathbf{B}_s$ and $\mathbf{A}_s^p$ are inversely proportional to $N_o$. Equation (12) can then be approximated by

$$Q(\mathbf{C}_s^{p+1}\mid\mathbf{C}_s^p) \approx -\int_{R(\hat{\mathbf{A}}_s^p)}\lambda(\beta,\mathbf{y_s},\mathbf{C}_s^{p+1})f(\beta\mid\mathbf{y_s},\mathbf{C}_s^p)d\beta. \tag{20}$$

In the maximization of $Q(\mathbf{C}_s^{p+1}\mid\mathbf{C}_s^p)$ in the M step, we consider the result of minimizing $\lambda(\beta,\mathbf{y_s},\mathbf{C_s}^{p+1})$ with respect to $\mathbf{C}_s^{p+1}$ for various values of $\beta$, and note that a range of $\beta$ will produce the same optimal subsequence $\mathbf{C}_s^{p+1}$. We define $S$ to be the set of all $\beta$ such that the same subsequence $\mathbf{C}_s^{p+1}$ minimizes $\lambda(\beta,\mathbf{y_s},\mathbf{C}_s^{p+1})$ as minimizes $\lambda(\hat{\beta}_s^p,\mathbf{y_s},\mathbf{C_s}^{p+1})$:

$$S=\{\beta : \arg\min_{\mathbf{C}_s^{p+1}}\lambda(\beta,\mathbf{y_s},\mathbf{C}_s^{p+1}) = \arg\min_{\mathbf{C}_s^{p+1}}\lambda(\hat{\beta}_s^p,\mathbf{y_s},\mathbf{C}_s^{p+1})\}. \tag{21}$$

If

$$R(\hat{\mathbf{A}}_s^p)\subset S, \tag{22}$$

then every $\beta$ in the integral in (20) produces the same subsequence $\mathbf{C}_s^{p+1}$ upon minimizing $\lambda(\beta,\mathbf{y_s},\mathbf{C}_s^{p+1})$. Therefore, for the purpose of finding the optimal subsequence $\mathbf{C}_s^{p+1}$, only $\lambda(\hat{\beta}_s^p,\mathbf{y_s},\mathbf{C_s}^{p+1})$ needs to be minimized and the integral in (20) need not be evaluated. Hence, the M step is reduced to finding the optimal sequence $\mathbf{C}_s^{p+1}$:

$$\mathbf{C}_s^{p+1} = \arg\min_{\mathbf{C}^{p+1}}\lambda(\hat{\beta}_s^p,\mathbf{y_s},\mathbf{C}^{p+1}), \tag{23}$$

and the Viterbi Algorithm can be used to perform this minimization.

The E step is then reduced to computing $\hat{\beta}_s^p = \hat{\beta}_s^p(\mathbf{y_s},\mathbf{C_s}^p)$ from (18), since the integration need not be performed. Therefore, when (22) is satisfied, the EM algorithm, reduces to iteratively estimating $\beta$ and $\mathbf{C}$. This reduced EM algorithm is thus similar to joint sequence and parameter estimation discussed in [4]. We have shown here that our iterative algorithm produces the sequence estimate that maximizes the likelihood $f(\mathbf{y}\mid\mathbf{C})$, since it is equivalent to the EM algorithm when (22) holds.

# V Initialization

Since the EM algorithm converges to a local maximum of $f(\mathbf{y}\mid\mathbf{C})$, a good initial estimate $\mathbf{C}_s^1$ must be provided to ensure convergence to the global maximum of the transmitted data sequence. We start the algorithm in each subsequence with the E step by constructing the initial channel parameter density denoted by $f(\beta\mid\mathbf{y_s},\mathbf{C_s}^0)$. There is no previous estimate $\mathbf{C_s}^0$, and thus the density $f(\beta\mid\mathbf{y_s},\mathbf{C_s}^0)$ is actually independent of any estimate of the transmitted information subsequence. Therefore, we equate the posterior density to a prior density $\rho_s(\beta)$

$$f(\beta\mid\mathbf{y_s},\mathbf{C_s}^0) = \rho_s(\beta) \tag{24}$$

in the initial E step of each subsequence. In addition, the prior density $\rho_s(\beta)$ will also be used in $f(\beta\mid\mathbf{y_s},\mathbf{C_s}^p)$ in (12) through (9) in subsequent iterations of the E step.

We now describe methods for constructing $\rho_s(\beta)$. We consider a Gaussian density (16). We first discuss very slowly and very rapidly varying channels from which $\rho_s(\beta)$ is derived from previous subsequences and training data respectively. Intermediate situations are considered at the end of this section.

When the channel fades slowly, information from one subsequence can be used to form an initial estimate of the channel in the next subsequence. If a mobile moves very slowly, so that the channel changes very little between that user's successive subsequences, then the assignment

$$\rho_s(\beta) = f(\beta\mid\mathbf{y_{s-1}},\mathbf{C_{s-1}}^P), \tag{25}$$

can be used, where $f(\beta\mid\mathbf{y_{s-1}},\mathbf{C_{s-1}}^P)$ denotes the posterior channel density obtained from the final ($P^{th}$) EM iteration of subsequence number $s-1$. The prior density $\rho_s(\beta)$ is then given by the Gaussian density specified by equations (17) to (19) evaluated at the final iteration of subsequence $s-1$, so that $\rho_s(\beta)$ has the mean and inverse variance

$$\bar{\beta}_{s\rho} = \hat{\beta}^P_{s-1} \quad (26)$$
$$\mathbf{B}_s = \hat{\mathbf{A}}^P_{s-1}. \quad (27)$$

When the channel changes rapidly enough that a given user's consecutive subsequences are not highly correlated, equation (25) is no longer accurate. In this case, training data can be used to construct $\rho_s(\beta)$. Given a training data sequence $\mathbf{C}_{sT}$ of symbols in each subsequence $s$ and the corresponding samples of the received signal denoted collectively by $\mathbf{y}_{sT}$, we let the expected value of $\beta$, denoted by $\bar{\beta}_{sT}$, equal the maximum likelihood estimate of $\beta$ in subsequence $s$ derived from this training data:

$$\bar{\beta}_{sT} = \bar{\beta}_{sT}(\mathbf{y}_{sT}, \mathbf{C}_{sT}) = \arg\min_{\beta} \lambda(\beta, \mathbf{y}_{sT}, \mathbf{C}_{sT}), \quad (28)$$

where $\lambda(\beta, \mathbf{y}_{sT}, \mathbf{C}_{sT})$ is given by (11). In order to determine the variance of $\beta$, we consider the Cramer–Rao inequality [3], which states that the covariance matrix of an unbiased estimator is at least as large as the inverse Fisher information matrix. Hence, we choose the variance of $\rho_s(\beta)$ to equal the inverse Fisher information matrix. The matrix elements $J_{ij}(\bar{\beta}_{sT})$ of the Fisher information matrix $\mathbf{J}_s$ are

$$J_{ij}(\beta) = E\left[\frac{\partial \ln f(\mathbf{y}_{sT} \mid \beta, \mathbf{C}_{sT})}{\partial \beta_i} \frac{\partial \ln f(\mathbf{y}_{sT} \mid \beta, \mathbf{C}_{sT})}{\partial \beta_j}\right], \quad (29)$$

where $E$ denotes the expectation over the possible values of $\mathbf{y}_{sT}$. Therefore, the Gaussian prior density based on training data has mean and inverse variance equal to

$$\bar{\beta}_{s\rho} = \bar{\beta}_{sT} \quad (30)$$
$$\mathbf{B}_s = \mathbf{J}_s.$$

In the first ($p = 0$) iteration of each subsequence (24) indicates that the prior density alone is used for $f(\beta \mid \mathbf{y}_s, \mathbf{C}_s{}^0)$. Equations (18) and (19) then reduce to

$$\hat{\beta}^0_s = \bar{\beta}_{s\rho}. \quad (31)$$
$$\hat{\mathbf{A}}^0_s = \mathbf{B}_s. \quad (32)$$

When training data is used as in (30), the first iteration of our simplified EM algorithm is thus analogous to the current equalization method used in GSM: the channel is estimated from training data, and then the transmitted sequence is estimated from this possibly imperfect estimate of the channel. While in GSM this estimate of the transmitted sequence is the final estimate, our simplified EM algorithm uses this first sequence estimate to improve the estimate of the channel parameters, which in turn is used to improve the sequence estimate.

The condition (22) for simplification of the EM algorithm for $p = 0$ reduces to

$$R(\mathbf{B}_s) \subset S. \quad (33)$$

We note that $\hat{\mathbf{A}}^p_s \geq \mathbf{B}_s$, with equality holding at $p = 0$. This matrix inequality is defined in the sense that $\hat{\mathbf{A}}^p_s - \mathbf{B}_s$ is nonnegative definite, and thus $R(\hat{\mathbf{A}}^p_s) \subseteq R(\mathbf{B}_s)$. Hence, if we ensure that (33) holds, then (22) will also be satisfied. When training data is transmitted and $\mathbf{B}_s = \mathbf{J}_s$, computation of the matrix elements of $\mathbf{J}_s$ using (10) in (29) shows their dependence on the number of symbols in the training sequence and the SNR. Thus the length of the training sequence can be selected so that (33) and hence (22) will be satisfied for the lowest desired SNR.

Above we discussed situations in which $\rho_s(\beta)$ is constructed from training data inserted in each subsequence for a rapidly varying channel and from channel parameter estimates from the previous subsequence for a very slowly fading channel. A third scenario combines the two situations for mobiles of intermediate speeds: training data could be sent only when the receiver detects an unacceptable error rate, or it could be sent periodically. A prior Gaussian density $\rho_{s'}(\beta)$ with mean and variance given by (30) would be used for subsequences $s'$ that contain training data. Subsequences $s$ without training data derive the prior density from the previous subsequence since the channel does not change too much over the time of a subsequence; thus $\rho_s(\beta)$ is approximated by a Gaussian with mean given by (26). However, at intermediate speeds the variance would be larger than that given in (27), because the greater variations in the channel over the time of a subsequence considered here makes it more likely that in subsequence $s$ beta will be farther from $\bar{\beta}_{s\rho} = \hat{\beta}^P_{s-1}$ than described by the inverse variance $\hat{\mathbf{A}}^P_{s-1}$. Thus in order to pick a larger variance matrix for $\rho_s(\beta)$ here, the inverse variance matrix $\mathbf{B}_s$ could be chosen to be some multiple of $\hat{\mathbf{A}}^P_{s'-1}$ or $\mathbf{J}_{s'}$ from the most recent subsequence, labeled by $s'$, in which training data was sent.

The information bit rate could potentially be increased by sending training data only in some subsequences, and by choosing the length of the training data sequence so that it is no longer than necessary to satisfy (33) for the lowest desired SNRs, as well as to produce a close enough initial estimate $\mathbf{C}^1_s$ of $\mathbf{C}_s$ that will yield the correct local maximum.

# VI Simulations

Simulations of our simplified EM algorithm were performed in a rapidly varying channel with the GSM system. Each subsequence over which the EM algorithm is run corresponds to a GSM time slot. The true channel parameters are chosen independently in a user's successive time slots. A two path multipath model is used for the true channel, where each path undergoes independent Rayleigh fading. The time delay $\tau$ was randomly picked from a uniform distribution with range from .9 $\mu$sec to 7.2 $\mu$sec. Equation (30) is used to form $\rho_s(\beta)$ from training data, and the Viterbi algorithm is used to find the optimal sequence $\mathbf{C}^{p+1}$ through (23) in the M step.

The probability of bit error is plotted in Figure 1 as a function of SNR. It is the differences in the BERs of the three curves, rather than the absolute BERs that is of interest

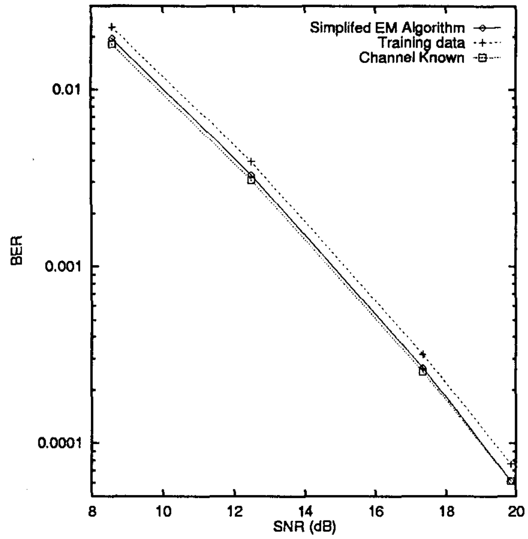


Figure 1: BER vs. SNR for a rapidly varying multipath channel in GSM with use of a discrete multipath model. The three curves correspond to use of the simplified EM algorithm, use of training data alone to estimate the channel, and the ideal case of a receiver to which the channel is exactly given.

here. No coding or interleaving was used, and the absolute BERs depend on the channel models. At each SNR the BERs were calculated from either 40,000 or 80,000 independent time slots; repeated runs demonstrated high accuracy in the differences in the BERs for the three curves.

The upper curve in Figure 1 displays the BER when training data alone is used to estimate the channel, and then the estimated sequence is produced from this channel estimate, as is done in the initial iteration of the simplified EM algorithm. This initial EM iteration is thus analogous to the method of sequence estimation currently used in GSM. In fact, the BER from the current GSM system, if it did not use coding or interleaving, would be expected to be higher than that of this curve: All curves in Figure 1 were derived from the discrete multipath model (2), as opposed to the finite impulse response used currently in GSM to model the channel. Since estimation of a finite impulse response from a training data sequence requires a linear approximation to the nonlinearly modulated GMSK signal used in GSM, this estimation incurs error from this linear approximation. It was shown [5] that use of a multiray model to parameterize the channel similar to our discrete multipath model (6) eliminates this error and yields an improvement of 2 to 4 dB relative to use of the finite impulse to characterize the channel.

The solid middle curve in Figure 1 displays the BER when the simplified EM algorithm is used to estimate the sequence. The lower curve displays a lower bound on the lowest possible BER that can be achieved for this channel: it is the BER when the channel is known exactly by the receiver. The simplified EM algorithm is seen to significantly decrease the BER relative to that when training data alone is used to estimate the channel. The EM algorithm essentially removes the bit errors due to uncertainty in the channel, as seen by the fact that the EM algorithm's BER is almost as low as the BER when the channel is exactly known. The EM algorithm decreased the BER by 14% to 20%, with the largest decreases at highest SNR, relative to that when the sequence estimate is based on a single estimate of the channel derived from training data and the discrete multipath model.

Most of the decrease in the BER with the EM algorithm takes place in the first EM iteration following the initial step based on the training data. While there is some additional decrease in bit errors following the second EM iteration, little if any improvement is seen in subsequent EM iterations. Therefore, a practical receiver could be built by performing only one or two EM iterations following the initial sequence estimate based on training data alone.

# VII Conclusions

We have derived a simplified EM algorithm to find the MLSE of CPM signals in a fading multipath channel. When enough training data is provided for the lowest SNR considered, the EM algorithm reduces to iteratively estimating the channel parameters and the transmitted sequence. Initial channel parameter estimates are obtained from training data for rapidly fading channels, or from previous subsequences for slowly fading channels. Subsequent iterations allow for improvement in symbol estimation over that derived from the initial channel estimates. Simulations for GSM show that the simplified EM algorithm reduces the BER so that it is almost as low as that which would be obtained from an exactly known channel.

# References


[1] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM Algorithm", *J. Royal Statistical Soc., Ser. B*, vol. 39, no. 1, pp.1-38, 1977.

[2] C. N. Georghiades and J. C. Han, "Sequence Estimation in the Presence of Random Parameters Via the EM Algorithm", *IEEE Trans. Commun.*, vol. 45, pp. 300-308, March 1997.

[3] T.M. Cover and J.A. Thomas, "Elements of Information Theory", New York, 1991.

[4] H. Kobayashi, "Simultaneous Adaptive Estimation and Decision Algorithm For Carrier Modulated Data Transmission Systems", *IEEE Trans. Comm. Tech.*, pp. 268-280, June 1971.

[5] J. Chen, A. Paulraj, and U. Reddy, "Multichannel Maximum–Likelihood Sequence Estimation Equalizer for GSM Using a Parametric Channel Model" *IEEE Trans. Commun.*, vol. 47, pp. 53-63, Jan. 1999.