Hidden Markov Modeling of Flat Fading Channels

William Turin, *Senior Member, IEEE,* and Robert van Nobelen

***Abstract*— Hidden Markov models (HMM's) are a powerful tool for modeling stochastic random processes. They are general enough to model with high accuracy a large variety of processes and are relatively simple allowing us to compute analytically many important parameters of the process which are very difficult to calculate for other models (such as complex Gaussian processes). Another advantage of using HMM's is the existence of powerful algorithms for fitting them to experimental data and approximating other processes. In this paper, we demonstrate that communication channel fading can be accurately modeled by HMM's, and we find closed-form solutions for the probability distribution of fade duration and the number of level crossings.**

***Index Terms*— Fading channels, hidden Markov models, parameter estimation.**

I. Introduction

MANY papers and books are devoted to modeling fading communication channels. A common feature of the models is that they all have memory. The most widely used model describes fading as a Gaussian process [3]. However, this model is difficult to use in applications. For example, there are no closed form expressions for characteristics associated with the model such as the probability density function (PDF) of fade durations and the probability distribution of the number of fades inside a fixed time interval. Several approximations for these probability distributions are available in the literature [9], [12], [16].

Hidden Markov models (HMM's), on the other hand, allow us to calculate many important system parameters (such as probabilities of various error sequences and other performance characteristics) in closed form [27]. Also, there are many efficient statistical algorithms for fitting HMM's to experimental data and for approximating various stochastic processes with HMM's [1], [15], [22], [27].

Many papers use HMM's to model channels with memory. Gilbert initiated the study of the HMM for real communication channels error statistics [8]. His model is popular because of its simplicity. However, measurements of various channels showed a necessity of using more complex models [7], [10], [21].

Many experimental results demonstrate that HMM's can be used to model error sequences in digital communications over fading channels [6], [18]–[20], [23]. Several papers are devoted to modeling the fading process itself with HMM's [17], [25], [24]. The important problem arises when approximating one process with the other is how to select a measure of the approximation quality: a distance between the processes. The most general approach is to fit a high-order multidimensional probability distribution of one process to the other which can be accomplished by minimizing a distance measure between the distributions (such as mean-squared error, Kullback–Leibler divergence [11], or χ^2). This approach can be computationally expensive [22]. Alternatively, we can simulate one process and treat the simulation results as experimental data for estimating parameters of the other process.

Since the method of fitting multidimensional probability distributions requires calculation of complex multidimensional integrals, the method of moments is often used. This method consists of calculating the probability distribution moments for both processes (such as means, autocorrelation functions, etc.) and finding parameters of the approximating process from the corresponding system of equations. This approach is simple, but often leads to a poor approximation.

The other question that should be asked when approximating with HMM's is how to choose the model structure, the transition probability matrix structure, and a type of observation probability distribution. In this paper we compare different models in terms of the complexity of fitting the models to the fading process and difficulty of their use in applications.

Our paper is organized as follows. In Section II we analyze the most widely used model for the Rayleigh fading [3]. In Section III we consider various Markov models for the fading envelope. Modeling with HMM's is addressed in Section IV. Section V describes various algorithms for HMM parameter estimation. In Sections VI and VII we demonstrate that there are closed-form solutions for the fade duration distribution and level crossing number distributions for the fading HMM which are not available for the Gaussian model. In Section VII we introduce an HMM for modeling the combination of fading and additive noise.

II. Fading Channel Model

Let $x(t)$ be a low-pass equivalent of the transmitted signal with the in-phase component $x_I(t) = \mathrm{Re}\{x(t)\}$ and quadrature component $x_Q(t) = \mathrm{Im}\{x(t)\}$. Consider a frequency-nonselective fading channel with additive noise $n(t)$. This channel can be modeled as [13, p. 716]

$$y(t) = c(t)x(t) + n(t) \tag{1}$$

where $y(t)$ is the received signal and fading is modeled by the complex random process $c(t)$. Different models are based on different assumptions about $c(t)$ and $n(t)$.

Usually it is assumed that $n(t)$ is zero-mean complex additive white Gaussian noise (AWGN), and $c(t)$ is a complex stationary zero-mean Gaussian process with independent and identically distributed (i.i.d.) real and imaginary parts. The

Manuscript received September 1997; revised February 1998.

The authors are with AT&T Labs-Research, Florham Park, NJ 07932-0971 USA.

Publisher Item Identifier S 0733-8716(98)08658-2.

PDF of a sequence $\boldsymbol{c}_k = (c(t_1), c(t_2), \ldots, c(t_k))$ has the form

$$f(\boldsymbol{c}_k) = (2\pi)^{-k}|\boldsymbol{D}|\exp\Big(-0.5\mathrm{Re}\{\boldsymbol{c}_k\boldsymbol{D}\boldsymbol{c}_k^H\}\Big) \quad (2)$$

where $|\boldsymbol{D}|$ denotes the determinant of $\boldsymbol{D}$, $\boldsymbol{c}_k^H$ denotes the conjugate transpose of $\boldsymbol{c}_k$, and $\boldsymbol{D}^{-1}$ is the process covariance matrix

$$\boldsymbol{D}^{-1} = [R(t_j - t_i)]_{k,k} \quad (3)$$

$R(\tau)$ is the process autocorrelation function. The noise multidimensional distribution has a similar form.

This model is called a Rayleigh fading because its envelope $\alpha(t) = |c(t)|$ is Rayleigh distributed

$$\Pr\{\alpha(t) < a\} = 1 - \exp\Big(-0.5a^2/\mu\Big).$$

Different models of fading channels are based on different assumptions about the power spectral density $S(f)$ or, equivalently, the autocorrelation function $R(\tau)$ of $c(t)$. In this study we consider the model in which

$$S(f) = \mu/\pi\ \sqrt{f_D^2 - f^2}, \quad R(\tau) = \mu J_0(2\pi f_D|\tau|) \quad (4)$$

where $J_0(\cdot)$ is the Bessel function of the first kind, f_D is the maximal Doppler frequency, and μ is the power of the fading process $c(t)$. We will refer to this model as Clarke's model [3]. The other models include rational functions [12], [16], [19], simple irrational functions [12], [16], and Gaussian PDF's [16]. It is well known that stationary Gaussian stochastic processes can be modeled by filtering white Gaussian noise $w(t)$ [13].

Alternatively, it can be simulated by the following equation [9]:

$$c(t) = 2\sum_{i=1}^{N} e^{j(i\pi/N+1)}\cos\ \left(2\pi f_D t\ \cos\ \frac{2\pi i}{4N+2}\right) + \sqrt{2}\cos 2\pi f_D t. \quad (5)$$

In practical systems, the transmitted signal $x(t)$ has the form

$$x(t) = \sum_k\ x_k p(t - k\Delta) \quad (6)$$

where $1/\Delta$ is the symbol rate, x_k is the transmitted symbol value (a complex number corresponding to a point of the signal constellation), and $p(t)$ represents the shaping pulse. If we assume that $c(t)$ is slowly varying so that it is nearly constant over a symbol duration Δ, the sampled output of the coherent demodulator followed by the receiver matched filter can be approximated by

$$y_k = c_k x_k + n_k \quad (7)$$

where n_k is the sample of the filtered Gaussian noise and $c_k = c(k\Delta)$ is a sample of the fading process.

A. Fading Envelope Model

For slow fading, the communication system performance depends mainly on the value of the envelope $\alpha(t) = |c(t)|$. The envelope can be described by its multidimensional PDF which can be obtained from (2) by using the equation [24]

$$f\ (\boldsymbol{\alpha}_k) = (2\pi)^{-k}\Big|\boldsymbol{D}\Big|\alpha_1\alpha_2\cdots\alpha_k\ \int_0^{2\pi}\ \int_0^{2\pi}\ \cdots \cdot \int_0^{2\pi}\exp\Big(-0.5\boldsymbol{\alpha}_k\boldsymbol{G}\boldsymbol{\alpha}_k^T\Big)\ d\theta_1\ d\theta_2\cdots d\theta_k \quad (8)$$

where $\alpha_i = \alpha(t_i), \theta_i = \theta(t_i), \boldsymbol{G}$ is the matrix whose ijth element is $g_{ij} = d_{ij}\cos(\theta_i - \theta_j)$, where d_{ij} is the element in the ith row and jth column of D, and $\boldsymbol{\alpha}_k = (\alpha_1, \alpha_2, \ldots, \alpha_k)$.

If $k = 1$ we have the Rayleigh PDF

$$f(\alpha_1) = \alpha_1\mu^{-1}\exp\Big(-0.5\alpha_1^2/\mu\Big), \qquad \mu = R(0) = \sigma^2. \quad (9)$$

If $k = 2$, we have [9, p. 50]

$$f(\alpha_1, \alpha_2) = \frac{\alpha_1\alpha_2}{\mu^2(1-\lambda^2)}\exp\left[-\frac{\alpha_1^2 + \alpha_2^2}{2\mu(1-\lambda^2)}\right] \cdot I_0\left(\frac{\alpha_1\alpha_2}{\mu}\ \frac{\lambda}{1-\lambda^2}\right) \quad (10)$$

where $\lambda = R(t_2 - t_1)/R(0), I_0(\cdot)$ is the modified Bessel function of the first kind.

Obviously, this model is difficult to use in applications in which distribution of high-dimension k is needed. In these applications, the Monte Carlo method is usually applied. It is possible, however, to approximate the model with an HMM which is simpler to apply.

III. Markov Models

Markov processes are popular in modeling fading because they can model processes with memory and their theory is well developed. We consider different model structures commonly used in describing fading channels.

A. Multiple Markov Chains

Multiple Markov processes $\boldsymbol{z}_1^\infty$ are processes with finite memory. If the process has memory m, the conditional PDF of z_t, given all past observations $\boldsymbol{z}_1^{t-1}$, depends only on the m previous observations $\boldsymbol{z}_{t-m}^{t-1}$ (where symbol $\boldsymbol{z}_i^j$ denotes $z_i, z_{i+1}, \ldots, z_j$), i.e.,

$$f\Big(z_t\Big|\boldsymbol{z}_1^{t-1}\Big) = f\Big(z_t\Big|\boldsymbol{z}_{t-m}^{t-1}\Big).$$

Since the correlation function $R(\tau)$ tends to zero and c_k is Gaussian it can be approximated with a Markov process if m is large enough so that $R(m\Delta) \approx 0$ (because c_k and c_{k+m} become uncorrelated and hence independent). The envelope transitional PDF has the form

$$a\Big(\alpha_k\Big|\boldsymbol{\alpha}_{k-m}^{k-1}\Big) = f\Big(\boldsymbol{\alpha}_{k-m}^{k}\Big)/f\Big(\boldsymbol{\alpha}_{k-m}^{k-1}\Big) \quad (11)$$

where $\boldsymbol{\alpha}_{k-m}^{k-1} = (\alpha_{k-m}, \alpha_{k-m+1}, \ldots, \alpha_{k-1})$. The size m of the process memory can be determined using an approximation accuracy measure (an example of such a measure is given in Section V-B). For the simple Markov process $(m = 1)$ the transition PDF is the Rician PDF.

If we perform the envelope quantization into N levels, the quantized process $\{\rho_k\}$ can be approximated by the Markov

chain with N states. In particular, for the simple Markov chain $(m = 1)$ we have the transition probabilities

$$a_{ij} = \int_{q_{L,i}}^{q_{H,i}} \int_{q_{L,j}}^{q_{H,j}} f(\alpha_1, \alpha_2)\, d\alpha_1\, d\alpha_2 \Big/ \int_{q_{L,i}}^{q_{H,i}} f(\alpha)\, d\alpha \tag{12}$$

where $f(\alpha_1, \alpha_2)$ is given by (10). If the quantization intervals are small, we can write $a_{ij} = a(\rho_j|\rho_i)(q_{H,j} - q_{L,j})$, for $j < N$ and $a_{iN} = 1 - \Sigma_{j=1}^{N-1} a_{ij}$.

Reference [24] suggested that the first-order model approximation is satisfactory for Rayleigh fading whose power spectral density is given by (4). This approximation, however, is satisfactory for relatively short intervals only.

To approximate the Bessel function arising in (4) we need Markov chains with larger memory (we address the question of fitting the autocorrelation function in Section IV-A).

Since the number of states grows exponentially with the process memory, this approach is not practical.

B. Quantized Autoregressive and Moving Average (ARMA)

Using standard methods of the infinite impulse response (IIR) filter design [14] we can approximate the fading power spectral density with a rational function of the form

$$S(f) = \frac{|\sum_{i=0}^{q} d_i z^i|^2}{|1 + \sum_{i=1}^{p} h_i z^i|^2}, \qquad d_0 \neq 0 \tag{13}$$

where $z = e^{-2\pi j f}$. In this case $c(t)$ can be modeled by the complex ARMA process

$$c_k = \sum_{i=1}^{p} h_i c_{k-i} + v_k, \quad v_k = \sum_{i=0}^{q} d_i n_{k-i} \tag{14}$$

where n_i are i.i.d. Gaussian variables. If $q = 0$ we have an AR process $(v_k = d_0 n_k)$. The ARMA approximation of the fading is a special case of the Markov process whose state is defined by the vectors $\boldsymbol{c}_{k-p}^{k-1} = (c_{k-p}, c_{k-p+1}, \ldots, c_{k-1}), \boldsymbol{n}_{k-q}^{k-1} = (n_{k-q}, n_{k-q+1}, \ldots, n_{k-1})$. To obtain a Markov chain with the finite number of states, we need to quantize these vectors. This approach is more directly related to approximating the autocorrelation function than the one considered in the previous section. Moreover, it allows us to use the standard methods of filter design for building the model. In the case of a Butterworth filter [16], [19] we have a simple Markov process. The fading envelope $\alpha_k = |c_k|$ is a function of the Markov chain and, therefore, is a special case of an HMM. The ARMA model complexity grows linearly with process memory. This model, however, is difficult to use for calculating model statistics. For example, there is no closed-form expression for the fade duration and level crossing distributions for the ARMA model.

As in the case of the multiple Markov chains, the size of this model transition probability matrix grows exponentially with the process memory. The matrix is large, but sparse.

C. Birth-and-Death Processes

Birth-and-death processes are a special case of a Markov model [4], [17], [25]. These models assume that the quantized fading amplitude from the current ith level can jump only to the adjacent levels $a_{ij} = 0$ if $|j - i| > 1$. If a_{ii} is large, this structure allows us to model slowly varying processes.

It is possible to improve this model by splitting each state i into two i_s and i_e, depending on the transition slope [17]. i_s corresponds to the start of fading below the level q_i and i_e corresponds to its end.

The accuracy of this model depends on the selection of quantization levels. The quantization levels must satisfy two conditions: a_{ij} given by (12) should be close to zero for $|j - i| > 1$ and the original model state duration distributions should be close to exponential distributions (since the state durations of a Markov chain have exponential probability distributions). These two conditions are difficult to satisfy. In order to fit the exponential state duration distribution the number of quantization levels must be large, but in this case the probabilities a_{ij} of transitions to the nonadjacent levels $(|j-i| > 1)$ are not negligibly small. The model approximation can be improved by allowing nonadjacent level jumps.

D. Monte Carlo Method

Since it is difficult to evaluate the integrals which are needed for the Markov process approximation, the model parameters can be estimated using the results of computer simulation. After sampling and quantizing the envelope $\alpha(t) = |c(t)|$ we obtain a sequence $\boldsymbol{\rho}_1^T$. Now we can apply the well-known methods of fitting Markov chains to experimental data [2]. The state transition probabilities are estimated by

$$\hat{a}_{ij} = n_{ij}/n_i, \quad n_i = \sum_{j=1}^{n} n_{ij} \tag{15}$$

where n_{ij} is the number of transitions from state i to state j.

IV. Modeling Fading with HMM

An HMM is a probabilistic function of a Markov chain and can be defined as $\{S, X, \boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B}(x)\}$ where $S = \{s_1, s_2, \ldots, s_n\}$ is the set of the Markov chain states, X denotes the HMM output (observation) set, $\boldsymbol{\pi}$ is a vector of state initial probabilities, $\boldsymbol{A} = [a_{ij}]_{n,n}$ is a matrix of state transition probabilities $(a_{ij} = \Pr\{s_j|s_i\})$, and $\boldsymbol{B}(x) = \mathrm{diag}\{b_j(x)\}$ is a diagonal matrix of the output $x \in X$ conditional probability densities in state s_j. If X is discrete, $\boldsymbol{B}(x)$ is a matrix of probabilities $(b_j(x) = \Pr\{x|s_j\})$. Without loss of generality, we denote states by their indexes $(s_i = i)$.

Alternatively, HMM can be represented by the so-called state–space equations

$$u_{k+1} = G(u_k, \xi_k) \tag{16}$$

$$x_k = H(u_k, \eta_k) \tag{17}$$

where ξ_k and η_k are i.i.d. variables. Indeed, it follows from (16) that $\{u_k\}$ is a Markov process, generally with an infinite number of states. If $G(u, v)$ has a finite discrete range (number of possible values which it can take), however, then we have

a Markov chain with the finite number of states and state-transition probabilities

$$a_{ij} = \Pr\{u_{k+1} = j | u_k = i\} = \Pr\{G(i,\xi) = j\}. \quad (18)$$

According to (17), observations x_k are conditionally independent variables, given the state sequence, and have the following PDF:

$$F_i(x) = \Pr(x_k < x | u_k = i) = \Pr\{H(i,\eta) < x\}. \quad (19)$$

Conversely, for any HMM we can find $G(i,\xi)$ and $H(i,\eta)$ by inverting (18) and (19), respectively.

In order to understand the limitations of the HMM approximation, let us study the class of autocorrelation functions of HMM's.

A. HMM Autocorrelation Function

To calculate the autocorrelation function

$$R(\tau) = E(x_k x_{k+\tau})$$

we evaluate the probability densities $p(\boldsymbol{x}_1^k)$ of an HMM output sequences $\boldsymbol{x}_1^k = (x_1, x_2, \ldots, x_k)$. These probability densities have the form [21]

$$p(\boldsymbol{x}_1^k) = \boldsymbol{\pi}\boldsymbol{P}(x_1)\boldsymbol{P}(x_2)\cdots\boldsymbol{P}(x_k)\mathbf{1} = \boldsymbol{\pi}\prod_{i=1}^{k}\boldsymbol{P}(x_k)\mathbf{1} \quad (20)$$

where $\boldsymbol{P}(x) = \boldsymbol{A}\boldsymbol{B}(x)$ is the matrix PDF of x and $\mathbf{1}$ is the column matrix of ones.

Using these equations, we can write

$$R(0) = \boldsymbol{\pi}\boldsymbol{E}(x^2)\mathbf{1}, \quad R(\tau) = \boldsymbol{\pi}\boldsymbol{E}(x)\boldsymbol{A}^{\tau-1}\boldsymbol{E}(x)\mathbf{1},$$
$$R(-\tau) = R(\tau), \qquad \text{for } \tau > 0$$

where

$$\boldsymbol{E}(x) = \int_{-\infty}^{\infty} x\boldsymbol{P}(x)\,dx, \quad \boldsymbol{E}(x^2) = \int_{-\infty}^{\infty} x^2\boldsymbol{P}(x)\,dx$$

are the matrix expectations of x and x^2, respectively.

It follows from these equations that the z-transform of $R(\tau)$ for $\tau > 0$ is a rational function

$$\boldsymbol{R}(z) = \sum_{\tau=1}^{\infty} R(\tau)z^{-\tau} = \boldsymbol{\pi}\boldsymbol{E}(x)(\boldsymbol{I}z - \boldsymbol{A})^{-1}\boldsymbol{E}(x)\mathbf{1}. \quad (21)$$

Expanding it into partial fractions we obtain

$$\boldsymbol{R}(z) = \sum_{j=1}^{r}\sum_{i=1}^{m_j} D_{ij}(z - \lambda_j)^{-i} \quad (22)$$

where λ_j are the matrix $\boldsymbol{A}$ eigenvalues. Thus

$$R(\tau) = \sum_{j=1}^{r}\sum_{i=1}^{m_j} D_{ij}\binom{\tau-1}{i-1}\lambda_j^{\tau-i}, \qquad \tau > 0. \quad (23)$$

In particular, if all the eigenvalues are different, $R(\tau)$ is a mixture of exponential functions

$$R(\tau) = \sum_{j=1}^{r} D_{1j}\lambda_j^{\tau-1}, \qquad \tau > 0. \quad (24)$$

It follows from these equations that an HMM power spectral density is a rational function of $z = e^{-2\pi jf}$

$$S(f) = \boldsymbol{\pi}\boldsymbol{E}(x^2) - [\boldsymbol{\pi}\boldsymbol{E}(x)\mathbf{1}]^2 + \boldsymbol{\pi}\boldsymbol{E}(x)[(\boldsymbol{I} - \boldsymbol{Q}z)^{-1} + (\boldsymbol{I} - \boldsymbol{Q}z^{-1})^{-1}]\boldsymbol{E}(x)\mathbf{1}$$

where

$$\boldsymbol{Q}^{\tau-1} = \boldsymbol{A}^{\tau-1} - \mathbf{1}\boldsymbol{\pi}.$$

According to (23), the autocorrelation function must have the form

$$R(\tau) = \sum_{j=1}^{r}[P_j(\tau)\cos\nu_j\tau + Q_j(\tau)\sin\nu_j\tau]q_j^{\tau}, \qquad \tau > 0$$

where $P_j(\tau)$ and $Q_j(\tau)$ are polynomials $q_j = |\lambda_j|$ and $\nu_j = \arg(\lambda_j)$. If all eigenvalues are different $R(\tau)$ takes the form

$$R(\tau) = \sum_{j=1}^{r} P_j q_j^{|\tau|}\cos\nu_j\tau.$$

Obviously this class of functions is rich enough to approximate any autocorrelation function. Let us consider now different methods of approximating the fading process with HMM's.

V. HMM Parameter Estimation

Once we selected a class of models, we need to fit a model from this class to the fading process. There are several methods for fitting the model.

A. Method of Moments

The method of moments consists of finding the model's parameters by equating moments of two models and solving the corresponding equations. Usually one of the models is represented by its experimental data. Examples of such an approximation of the Rayleigh fading by an HMM are given in [17].

There are several problems with the method of moments. One of the problems is that the system of equations for the moments is often ill posed. The moments are the same for quite different models. The model structure is usually selected using our intuition and the model accuracy must be evaluated separately. The other problem is that the moment selection is quite arbitrary. For example, we can find an HMM whose autocorrelation function is close to that of the fading process. However, this does not mean that the multidimensional probabilities associated with these processes are close.

Therefore, the method of moments could be used for selecting initial values of the model parameters which are then improved by more sophisticated statistical algorithms.

B. Approximating Multidimensional Probability Densities

One of the most powerful methods of approximating a stochastic process with an HMM consists of fitting multidimensional probability distributions of an HMM to that of the

original process. The HMM parameters can be obtained as those which minimize the Kullback–Leibler divergence

$$\theta = \arg\min K(f||p_\theta) \tag{25}$$

where

$$K(f||p_\theta) = \int_{X^k} f(\boldsymbol{x}_1^k) \log \frac{f(\boldsymbol{x}_1^k)}{p_\theta(\boldsymbol{x}_1^k)}\, d\boldsymbol{x}_1^k \tag{26}$$

$f(\cdot)$ and $p_\theta(\cdot)$ are given by (8) and (20), respectively and θ is the HMM parameter vector. The minimum in (25) can be obtained iteratively by the expectation-maximization (EM) algorithm [27]: its version for fitting HMM's is called the Baum–Welch algorithm [1].

We would like to point out that the computational efficiency of the algorithm depends on the nature of statistical data. Obviously, for slow fading, an HMM approximation should be close to the birth-and-death process. Since the model matrix is sparse, direct application of the previous equations for the slow fading data is very inefficient. We can improve the algorithm efficiency by using matrix fast exponentiation [27]. Another improvement can be achieved by taking advantage of the following property of the Baum–Welch algorithm: if $a_{ij,p} = 0$ at some iteration step p, then $a_{ij,p+1} = 0$. Therefore, at each iteration we can replace small elements with zeroes and apply the sparse matrix multiplication algorithms.

Alternatively, we can start with the birth-and-death process approximation. If the state transitions satisfy the Markovian property, but the state duration distributions are not exponential, we have a semi-Markov process approximation which can be transformed into an HMM as follows [21, p. 48].

Let

$$\overline{\boldsymbol{A}} = \begin{bmatrix} 0 & \overline{a}_{12} & \cdots & \overline{a}_{1m} \\ \overline{a}_{21} & 0 & \cdots & \overline{a}_{2m} \\ \cdot & \cdot & \cdots & \cdot \\ \overline{a}_{m1} & \overline{a}_{m2} & \cdots & 0 \end{bmatrix} \tag{27}$$

be the transition process transition probability matrix. Suppose that we were able to approximate the state duration distributions with the phase-type matrix-geometric distribution

$$p_i(x) = \boldsymbol{\mu}_i \boldsymbol{A}_i^{x-1} \boldsymbol{b}_i \tag{28}$$

where $\boldsymbol{A}_i$ is a square matrix, $\boldsymbol{\mu}_i$ is a row vector, and $\boldsymbol{b}_i$ is a column vector such that

$$\boldsymbol{E}_i = \begin{bmatrix} \boldsymbol{A}_i & \boldsymbol{b}_i \\ \boldsymbol{\mu}_i & 0 \end{bmatrix}$$

is a stochastic matrix (that is all its elements are nonnegative and each row sums to one $\boldsymbol{E}_i\mathbf{1} = 1$). Then the semi-Markov process is equivalent to an HMM whose state transition probability matrix is given by

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & \overline{a}_{12}\boldsymbol{b}_1\boldsymbol{\mu}_2 & \cdots & \overline{a}_{1m}\boldsymbol{b}_1\boldsymbol{\mu}_m \\ \overline{a}_{21}\boldsymbol{b}_2\boldsymbol{\mu}_1 & \boldsymbol{A}_2 & \cdots & \overline{a}_{2m}\boldsymbol{b}_2\boldsymbol{\mu}_m \\ \overline{a}_{m1}\boldsymbol{b}_m\boldsymbol{\mu}_1 & \overline{a}_{m2}\boldsymbol{b}_m\boldsymbol{\mu}_2 & \cdots & \boldsymbol{A}_m \end{bmatrix}. \tag{29}$$

In this presentation, the diagonal matrices $\boldsymbol{A}_i$ represent the transitions between states of the HMM that correspond to the ith quantization level of the fading process (that is the probability of observing the level in those states is equal to one).

The matrix-geometric distribution parameters can be estimated using the Baum–Welch algorithm if we notice that the state holding process is a binary HMM whose state transition probability matrix is $\boldsymbol{E}_i$.

To illustrate this method, we performed simulation using Clarke's model of the Rayleigh fading channel with the parameters $f_D = 100$ Hz, the normalized fade rate $f_D T = 0.01$. We simulated 1 million steps of the fading amplitude to estimate the state transition probability matrix $\overline{\boldsymbol{A}}$, where the states are represented by the quantization intervals (0, 0.0149), (0.0149, 0.0329), (0.0329, 0.0725), (0.0725, 0.1601), (0.1601, 0.3535), (0.3535, 0.7805), (0.7805, 1.7234), (1.7234, ∞).

The state duration distributions are typically not exponential (as also shown in [17]). As an example, the state duration distribution of the sixth quantization interval is shown in Fig. 1, together with the matrix-geometric distribution approximation using a five-state model. The matrix $\boldsymbol{E}_6$ is

$$\boldsymbol{E}_6 = \begin{bmatrix} 0.9414 & 0.0249 & 0.0165 & 0.0173 & 0.0000 \\ 0.0070 & 0.6763 & 0.0015 & 0.3152 & 0.0000 \\ 0.0062 & 0.2393 & 0.7532 & 0.0013 & 0.0000 \\ 0.0390 & 0.0026 & 0.0071 & 0.7487 & 0.2026 \\ 0.0829 & 0.0000 & 0.9171 & 0.0000 & 0.0000 \end{bmatrix}.$$

The quality of the fit is illustrated in Fig. 1. The quantized fading level process is described by the HMM whose transition probability matrix is given by (29).

VI. Fade Duration Distribution

As we pointed out before, it is possible to find a closed-form solution for the fading-duration distribution of an HMM fading model. To calculate fade-duration probability distribution we can use the method of matrix probabilities [21]. Let

$$\boldsymbol{P}_a = \int_R^\infty \boldsymbol{P}(x)\, dx = \boldsymbol{A}\, \mathrm{diag}\left\{ \int_R^\infty b_j(x)\, dx \right\} \tag{30}$$

be the matrix probability of the fading to be above the level R and

$$\boldsymbol{P}_b = \int_{-\infty}^{R-0} \boldsymbol{P}(x)\, dx \tag{31}$$

be the matrix probability of the fading to be below level R. Then the probability that the fade duration is equal to τ can be written as

$$p_R(\tau) = \boldsymbol{\pi}\boldsymbol{P}_a\boldsymbol{P}_b^\tau\boldsymbol{P}_a\mathbf{1}/\boldsymbol{\pi}\boldsymbol{P}_a\mathbf{1}. \tag{32}$$

The probability distribution of intervals between consecutive fades can be found similarly

$$q_R(\nu) = \boldsymbol{\pi}\boldsymbol{P}_b\boldsymbol{P}_a^\nu\boldsymbol{P}_b\mathbf{1}/\boldsymbol{\pi}\boldsymbol{P}_b\mathbf{1}.$$

As we see, these probability distributions are matrix geometric. The comparison between the cumulative matrix-geometric distribution of fade durations

$$p_R(\tau < t) = \boldsymbol{\pi}\boldsymbol{P}_a(\boldsymbol{I} - \boldsymbol{P}_b^t)\mathbf{1}/\boldsymbol{\pi}\boldsymbol{P}_a\mathbf{1}$$

and simulated by both Clarke's model and HMM is presented in Fig. 2.

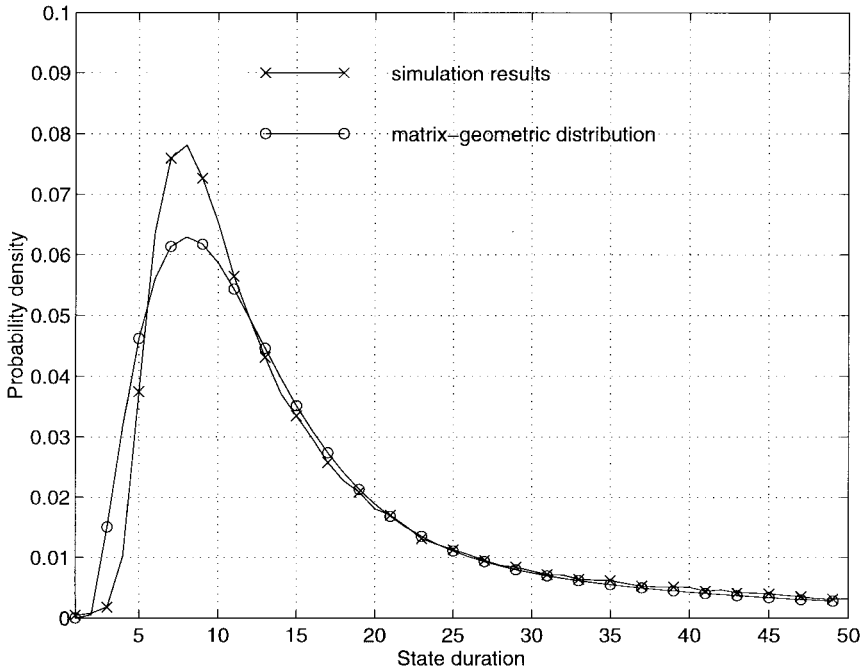


Fig. 1. A typical state duration distribution and its approximation.

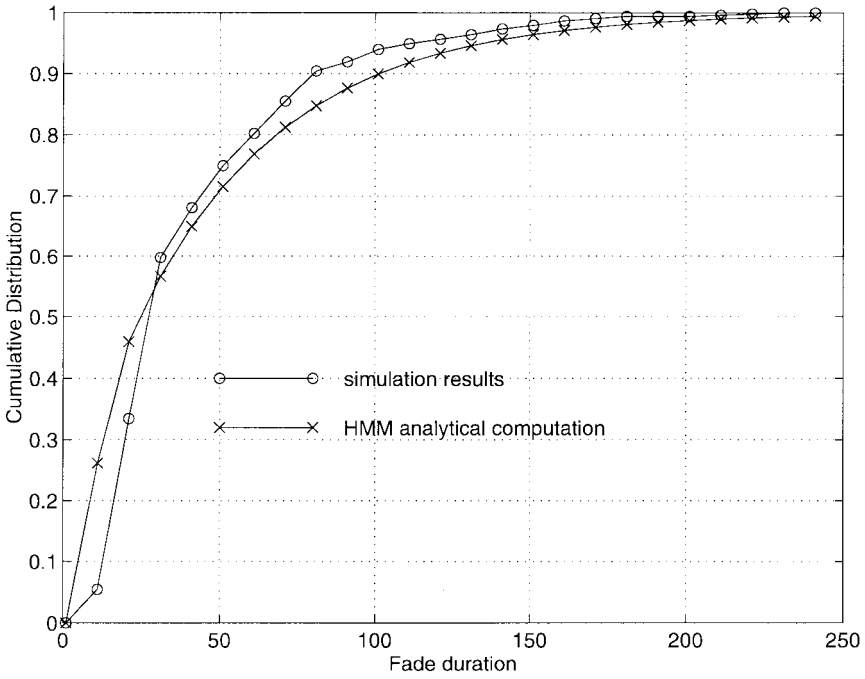


Fig. 2. Fade-duration distribution.

The mean duration of fades can be easily found using (31)

$$\overline{\tau}_{\mathrm{HMM}} = \sum_{\tau=1}^{\infty} \tau p_R(\tau) = \boldsymbol{\pi P}_a \boldsymbol{P}_b (\boldsymbol{I} - \boldsymbol{P}_b)^{-1} \mathbf{1} / \boldsymbol{\pi P}_a \mathbf{1}$$

and the mean duration between the fades is equal to

$$\overline{\nu}_{\mathrm{HMM}} = \boldsymbol{\pi P}_b \boldsymbol{P}_a (\boldsymbol{I} - \boldsymbol{P}_a)^{-1} \mathbf{1} / \boldsymbol{\pi P}_b \mathbf{1}.$$

If $\boldsymbol{\pi}$ is a stationary probability vector, these equations can be simplified

$$\overline{\tau}_{\mathrm{HMM}} = \boldsymbol{\pi P}_b \mathbf{1} / \boldsymbol{\pi P}_a \mathbf{1},\ \overline{\nu}_{\mathrm{HMM}} = \boldsymbol{\pi P}_a \mathbf{1} / \boldsymbol{\pi P}_b \mathbf{1}.$$

There is a closed-form solution for the mean duration of fades [12, p. 189]

$$\overline{\tau}_J = \frac{\eta}{R}\left[\exp\left(\frac{R^2}{2\sigma^2}\right) - 1\right]$$

where $\eta = \sigma_1 \sigma^{-2}$ and σ_1 is the standard deviation of the process first derivative. Fig. 3 compares $\overline{\tau}_{\mathrm{HMM}}$ with $\overline{\tau}_J$ for different levels R.

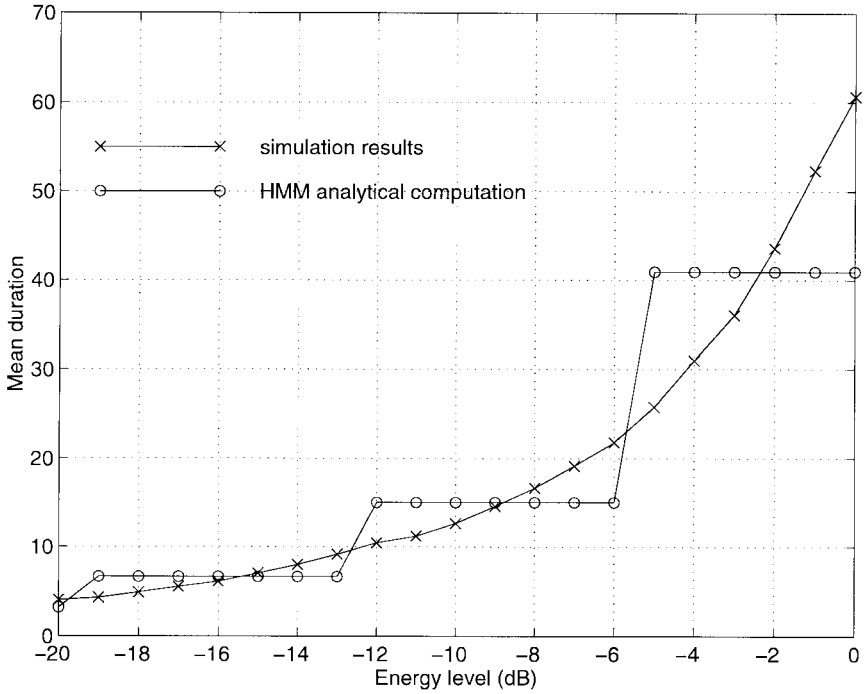


Fig. 3. Mean duration of fades.

VII. Level Crossing Number Distribution

The other important parameter which characterizes the fading process is the number of fades below a certain level during a fixed time interval. This parameter is closely related to the number of the level crossings by the fading process (if we neglect the interval boundary condition the number of fades is one half of the number of crossings). It is not difficult to derive this distribution for the HMM.

Indeed, let ξ_t be a characteristic function of the fading amplitude below a level R

$$\xi_t = \begin{cases} 1, & \text{if } \alpha(t) \geq R \\ -1, & \text{otherwise.} \end{cases}$$

The number of the level crossings is equal to the number of transitions $-1 \to 1$ and $1 \to -1$ in the binary HMM ξ_t which is described by the matrix probabilities $\boldsymbol{P}_a$ and $\boldsymbol{P}_b$.

To derive the number of level crossing probability distribution it is convenient to consider the level transition process $\eta_t = \xi_{t-1}\xi_t$. This process is also an HMM with the matrix probabilities

$$\boldsymbol{P}(0) = \begin{bmatrix} \boldsymbol{P}_a & 0 & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{P}_b \\ \boldsymbol{P}_a & 0 & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{P}_b \end{bmatrix},$$

$$\boldsymbol{P}(1) = \begin{bmatrix} 0 & \boldsymbol{P}_b & 0 & 0 \\ 0 & 0 & \boldsymbol{P}_a & 0 \\ 0 & \boldsymbol{P}_b & 0 & 0 \\ 0 & 0 & \boldsymbol{P}_a & 0 \end{bmatrix}. \tag{33}$$

Since η_t is an HMM the number of crossings probability distribution is a matrix binomial distribution $p_n(m) = \boldsymbol{p}_n(m)\mathbf{1}$ which can be evaluated by the following forward algorithm [21, p. 76]

$$\begin{aligned} \boldsymbol{p}_1(0) &= \boldsymbol{\pi}_\eta \boldsymbol{P}(0), \quad \boldsymbol{p}_1(1) = \boldsymbol{\pi}_\eta \boldsymbol{P}(1) \\ \boldsymbol{p}_n(m) &= 0, \quad \text{for } m < 0 \quad \text{or } m > n \\ \boldsymbol{p}_n(m) &= \boldsymbol{p}_{n-1}(m)\boldsymbol{P}(0) + \boldsymbol{p}_{n-1}(m-1)\boldsymbol{P}(1). \end{aligned} \tag{34}$$

It is not difficult to verify that the stationary distribution of the transition HMM is

$$\boldsymbol{\pi}_\eta = \left(\boldsymbol{\pi}\boldsymbol{P}_a^2, \boldsymbol{\pi}\boldsymbol{P}_a\boldsymbol{P}_b, \boldsymbol{\pi}\boldsymbol{P}_b\boldsymbol{P}_a, \boldsymbol{\pi}\boldsymbol{P}_b^2\right). \tag{35}$$

These equations can be simplified since matrices $\boldsymbol{P}(0)$ and $\boldsymbol{P}(1)$ contain many zeroes

$$\begin{aligned} \boldsymbol{q}_1^{(1)}(0) &= \boldsymbol{\pi}\boldsymbol{P}_a^2, \quad \boldsymbol{q}_1^{(2)}(0) = \boldsymbol{\pi}\boldsymbol{P}_b^2, \\ \boldsymbol{q}_1^{(1)}(1) &= \boldsymbol{\pi}\boldsymbol{P}_b\boldsymbol{P}_a, \quad \boldsymbol{q}_1^{(2)}(1) = \boldsymbol{\pi}\boldsymbol{P}_a\boldsymbol{P}_b \\ \boldsymbol{q}_n^{(1)}(m) &= [\boldsymbol{q}_{n-1}^{(1)}(m) + \boldsymbol{q}_{n-1}^{(2)}(m-1)]\boldsymbol{P}_a \\ \boldsymbol{q}_n^{(2)}(m) &= [\boldsymbol{q}_{n-1}^{(1)}(m-1) + \boldsymbol{q}_{n-1}^{(2)}(m)]\boldsymbol{P}_b \end{aligned} \tag{36}$$

and

$$p_n(m) = \boldsymbol{q}_n^{(1)}(m)\mathbf{1} + \boldsymbol{q}_n^{(2)}(m)\mathbf{1}. \tag{37}$$

The corresponding cumulative probability distribution is compared with the results of computer simulation in Fig. 4.

It is not difficult to show that the mean number of level crossings equals to [21, p. 70]

$$m_T = T\boldsymbol{\pi}(\boldsymbol{P}_b\boldsymbol{P}_a + \boldsymbol{P}_a\boldsymbol{P}_b)\mathbf{1} = 2T\boldsymbol{\pi}\boldsymbol{P}_a\boldsymbol{P}_b\mathbf{1}.$$

Thus, the level crossing rate is

$$z_{\text{HMM}} = 2\boldsymbol{\pi}\boldsymbol{P}_a\boldsymbol{P}_b\mathbf{1} \tag{38}$$

while for the Clarke's model it is [12, p. 182]

$$z_J = \eta^{-1} R \exp\left(-\frac{R^2}{2\sigma^2}\right)$$

we compare z_{HMM} and z_J in Fig. 5.

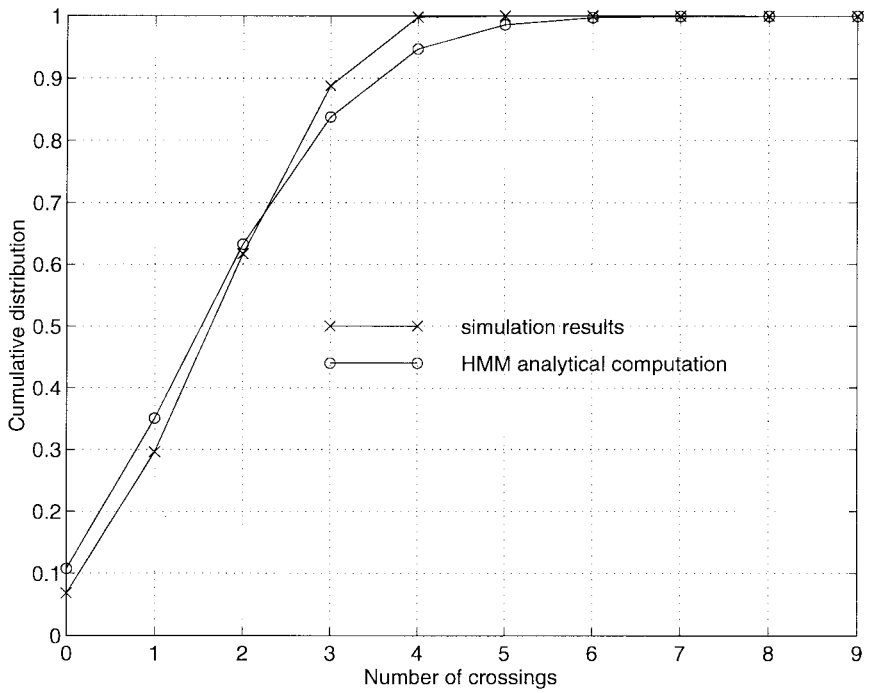


Fig. 4. Level crossing number distribution.

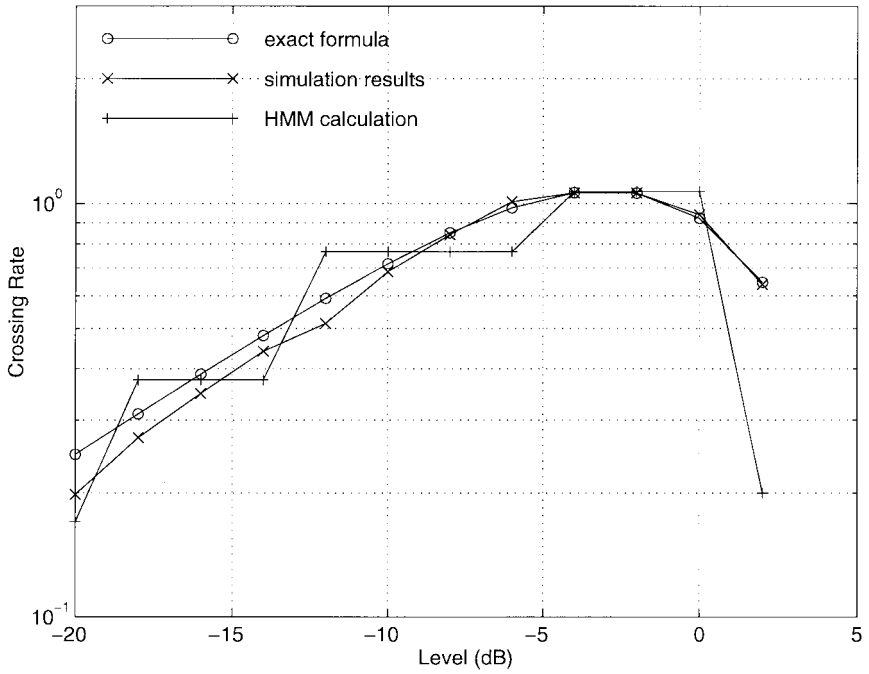


Fig. 5. Level crossing rates.

VIII. Channels with Fading and AWGN

Let us consider now modeling the slow-fading channel output according to (7). If we assume that the source sequence $\{x_k\}$ is generated by an HMM, then the channel output $\{y_k\}$ is an HMM whose state space is the Cartesian product of the states of the source and fading HMM's and its transition probability matrix is a Kronecker product of the source and fading transition probability matrices

$$\boldsymbol{A}_y = \boldsymbol{A}_x \otimes \boldsymbol{A}_c.$$

The output conditional probability is given by

$$b_j(y_k) = p(c_k x_k + n_k | j), \qquad j = (s_k^x, s_k^c)$$

where s_k^x and s_k^c are the states of the source and fading, respectively. This PDF has the form

$$b_j(y_k) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} b_{s_k^x}(u)\, b_{s_k^c}(v) f_n(y_k - uv)\, du\, dv$$

where $b_{s_k^x}(\cdot)$ and $b_{s_k^c}(\cdot)$ are source and fading conditional probability densities and $f_n(\cdot)$ is the noise PDF.

If the fading is Markovian, the output conditional PDF is just a convolution

$$b_j(y_k) = \int_{-\infty}^{\infty} b_{s_k^x}(u) f_n(y_k - uc_k)\, du.$$

If the source is discrete

$$b_{s_k^x}(u) = \sum_{i=1}^{m} b_i \delta(u - X_i)$$

then

$$b_j(y_k) = \sum_{i=1}^{m} b_i f_n(y_k - X_i c_k).$$

IX. Conclusion

There are many reasons for modeling fading with HMM's. It is convenient to have a common model for different types of fading. The class of HMM's is broad enough to approximate accurately various types of fading. We have demonstrated that different approaches to modeling Rayleigh fading represent special cases of an HMM.

If fading is modeled by an HMM, then bit errors and block errors in data communications over fading channels can be modeled by an HMM.

On the other hand, HMM's are comparatively simple allowing us to find closed-form expressions for various fading characteristics which are not available in the closed form for other models. We illustrate this by considering probability distributions of fade durations and the number of level crossings.

We have also considered relationships between various models and HMM's and shown how to approximate these models with HMM's.

Acknowledgment

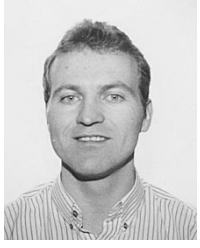
The authors would like to thank the reviewers for their helpful suggestions.

References

[1] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Statist.*, vol. 41, pp. 164–171, 1970.
[2] P. Billingsley, "Statistical methods in Markov chains," *Ann. Math. Statist.*, vol. 32, pp. 12–40, 1961.
[3] R. H. Clarke, "A statistical theory of mobile radio reception," *Bell Syst. Tech. J.*, vol. 47, pp. 957–1000, 1968.
[4] R. Coco, "A queueing theoretic approach to multipath fading—I," Bell Labs., Tech. Memo., Apr. 1983.
[5] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Statist. Soc.*, vol. 76, pp. 341–353, 1977.
[6] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol. 42, pp. 1977–1997, Sept. 1963.
[7] B. D. Fritchman, "A binary channel characterization using partitioned Markov chains," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 221–227, Apr. 1967.
[8] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. J.*, vol. 39, pp. 1253–1265, Sept. 1960.
[9] W. C. Jakes, *Microwave Mobile Communications*. New York: Wiley, 1974.
[10] L. N. Kanal and A. R. K. Sastry, "Models for channels with memory and their applications to error control," *Proc. IEEE*, vol. 66, no. 7, pp. 724–744, July 1978.
[11] S. Kullback, *Information Theory and Statistics*. New York: Wiley, 1959.
[12] W. C. Y. Lee, *Mobile Communications Engineering*. New York: McGraw-Hill, 1982.
[13] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1989.
[14] L. Rabiner and B. Gold, *Theory and Applications of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
[15] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice Hall, 1993.
[16] S. O. Rice, "Distribution of the duration of fades in radio transmission: Gaussian noise model," *Bell Syst. Tech J.*, vol. 37, pp. 581–635, May 1958.
[17] M. Sajadieh, F. R. Kschischang, and A. Leon-Garcia, "A block memory model for correlated Rayleigh fading channels," in *Proc. IEEE Int. Conf. Commun.*, Dallas, TX, June 1996, pp. 282–286.
[18] S. Sivaprakasam and K. S. Shanmugan, "An equivalent Markov model for burst errors in digital channels," *IEEE Trans. Commun.*, vol. 43, pp. 1347–1355, Apr. 1995.
[19] F. Swarts and H. C. Ferreira, "Markov characterization of channels with soft decision outputs," *IEEE Trans. Commun.*, vol. 41, pp. 678–682, May 1993.
[20] S. Tsai, "Markov characterization of the HF channel," *IEEE Trans. Commun. Technol.*, vol. COM-17, pp. 24–32, Feb. 1969.
[21] W. Turin, *Performance Analysis of Digital Transmission Systems*, 2nd ed. New York: Computer Science, 1998.
[22] ———, "Fitting probabilistic automata via the EM algorithm," *Commun. Statist.-Stochastic Models*, vol. 12, no. 3, pp. 405–424, 1996.
[23] W. Turin, and M. M. Sondhi, "Modeling error sources in digital channels," *IEEE J. Select. Areas Commun.*, vol. 11, pp. 340–347, Mar. 1993.
[24] H. S. Wang and P.-C. Chang, "On verifying the first-order Markovian assumption for a Rayleigh fading channel model," *IEEE Trans. Veh. Technol.*, vol. 45, pp. 353–357, May 1996.
[25] H. S. Wang and N. Moayeri, "Finite-state Markov channel—A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, pp. 163–171, Feb. 1995.
[26] M. Zorzi and R. R. Rao, "On the statistics of block errors in bursty channels," *IEEE Trans. Commun.*, vol. 45, pp. 660–667, June 1997.
[27] W. Turin, *Digital Transmission Systems: Performance Analysis and Modeling*. New York: McGraw-Hill, 1998

William Turin (SM'83) received the M.S. degree in mathematics from Odessa State University, Odessa, U.S.S.R., in 1958, and the Ph.D. degree in mathematics from the Institute for Problems of Mechanics of the Academy of Sciences of the U.S.S.R., Moscow, in 1966.

From 1960 to 1979 he was associated with the Moscow Institute of Electrical Engineering and Telecommunications, first as an Assistant Professor and later an Associate Professor. From 1980 to 1981 he was a Senior Research Scientist in the Department of Psychology, New York University, New York, NY. Since 1981 he has been a member of Technical Staff at AT&T Bell Laboratories, Holmdel and Murray Hill, NJ. Currently, he is a Technology Consultant at AT&T Labs-Research, Florham Park, NJ. His research interests include digital signal processing, handwriting and speech recognition and compression, computer simulation, error correcting codes, microwave radio, and Markov processes. He is the author of four books and numerous papers.



Robert van Nobelen was born in Haarlem, the Netherlands, in 1968. He received the B.E. (hons), M.E., and Ph.D. degrees in electrical and electronic engineering from the University of Canterbury, Christchurch, New Zealand, in 1991, 1993, and 1996, respectively.

He is currently a Senior Member of Technical Staff in the Communications Research Department of AT&T Labs-Research, Florham Park, NJ. His research interests are convolutional, trellis, and turbo coding for fading channels, channel modeling techniques, and radio-link and MAC protocols.