Generalized Markov Modeling for Flat Fading

Fulvio Babich, *Member, IEEE*, Owen E. Kelly, *Member, IEEE*, and Giancarlo Lombardi, *Member, IEEE*

***Abstract*—This paper investigates the properties of a method for obtaining Markov models of unspecified order to be applied to narrow-band fading channels with additive white Gaussian noise. The models are obtained by applying the context tree pruning algorithm to experimental or simulated sequences. Fading environments are identified in which the extension from first-order to higher order models is justified. The paper presents, as examples, the evaluation of the covariance function and the packet error distribution.**

***Index Terms*—Context trees, error analysis, fading channels, Markov processes.**

I. Introduction

EFFICIENT channel utilization in designing wireless services is helped by the availability of accurate tractable channel models. Markov models are often employed, due to their consolidated theory.

Starting from early works of Gilbert [1] and Elliot [2], an effective representation of digital communication channels with memory is given in finite state Markov channels (FSMC's). An FSMC is a discrete valued time-varying channel, whose variation is determined by a finite-state Markov chain: each state corresponds to a known memoryless channel [3]. For narrow band systems in flat fading with additive white Gaussian noise (AWGN), the signal-to-noise ratio (SNR) at the receiver is proportional to channel attenuation (fading). Fading can be discretized by sampling and quantization wherein one *fading state* is associated to each interval of attenuation values. If a binary modulation format is used on the channel, a binary symmetric channel (BSC) is associated to each fading state [4]. An FSMC model, specific for narrow-band fading channels, was introduced by Wang [4], describing fading state evolution as a first-order Markov chain, and it was applied to the evaluation of system-related channel properties [5]–[7].

In [8] and [9], it is pointed out that a first-order model is not always satisfactory. In [8], the context tree pruning (CTP) algorithm is proposed as a method to build higher order Markov models to improve FSMC description. Such a method can be applied to fit models to a wide variety of discrete phenomena.

In this letter, state definitions that the CTP selects for Rayleigh fading are shown; the Ricean case has also been fully investigated. A range of fading rates and threshold values are considered in building models, and the results are summarized according to state space size and structure. Being non-Markovian, the quantized fading channel theoretically requires an infinite-order CT model for exact description, however tractable models with an order higher than one can be obtained for several relevant fading regimes. They exhibit a degree of accuracy and complexity that is sufficient for evaluating system-related parameters. The autocorrelation coefficient (ACC) of quantized fading and the packet error distribution (PED) are evaluated to demonstrate model accuracy.

Paper approved by S. Roy, the Editor for Communication Theory/Systems of the IEEE Communications Society. Manuscript received July 7, 1998; revised April 18, 1999 and September 12, 1999. The work of F. Babich and G. Lombardi was supported by MURST (Italy) "ex quota 40%."

F. Babich and G. Lombardi are with the Dipartimento di Elettrotecnica, Elettronica e Informatica, Università di Trieste, I-34127 Trieste, Italy (e-mail: babich@stelvio.univ.trieste.it; lombardi@stelvio.univ.trieste.it).

O. E. Kelly was with WINLAB, Rutgers, The State University of New Jersey, Piscataway, NJ 08854-8060 USA. He is now with Nortel Networks, Ottawa, ON K1Y 4H7, Canada (e-mail: oekelly@ieee.org).

Publisher Item Identifier S 0090-6778(00)03638-2.

II. CTP-Based Markov Models

Let $z(t)$ be the envelope of a narrow-band Ricean fading process, modeled according to Clarke's model [10] and having maximum Doppler spread f_D and $\mathrm{E}\{z^2(t)\} = 2$. Let the instantaneous "fading power" $x(t) = z(t)^2/2$ be sampled with period T and quantized with respect to a set of thresholds $\{A_j\}_{j=1}^{L-1}$ to obtain an L-ary discrete time fading process φ

$$\varphi_n = \sum_{j=1}^{L-1} \mathbf{1}\left[\frac{z(nT)^2}{2} > A_j\right], \qquad n = 0, 1, \cdots \quad (1)$$

with $\varphi_n \in \mathcal{S} = \{0, 1, \cdots, L-1\}$, $A_0 \triangleq 0$, $A_L \triangleq +\infty$. If binary quantization is considered, the channel is either *good* or *bad* depending on whether fading power is greater than or less than $A = 10^{-F/10}$, where F (decibels) is the *fade margin*. For the process φ, the *fading rate* is defined as $f_D T$ and the symbol φ_n is the fading state at time nT.

In the terms of [11], a *context* is a short sequence of fading states and contexts can have different lengths. A CT model is defined by a set of contexts $\mathcal{C}$ together with transition probabilities $p(\varphi|c)$, $\forall\, \varphi \in \mathcal{S}$, $c \in \mathcal{C}$. The context set has the property that for any given sequence $\varphi^n = \{\varphi_1, \cdots, \varphi_n\}$ and any value of n, the suffix of φ^n matches exactly one context $c \in \mathcal{C}$ (with length $|c|$) as follows: $c = \{\varphi_{n-k+1}\}_{k=1}^{|c|}$. The CT depth (memory) d is the length of the longest context in $\mathcal{C}$. Transition probabilities are such that $p(\varphi_{n+1}|\varphi^n) = p(\varphi_{n+1}|c(\varphi^n))$, where $c(\varphi^n)$ is the unique context of φ^n. Every CT model may be converted simply to a Markov chain without estimation of additional parameters. Its state set $\mathcal{M}$ is also made of short sequences (contexts) either coincident with the ones in $\mathcal{C}$ or made by their extension. The number of free parameters is $|\mathcal{C}|(L-1)$. If the data generating source is Markovian and the given sequence is sufficiently long, the CTP algorithm [11] is capable of selecting the correct CT model. For this work, CTP-based models were chosen over other CT models because the model family has a recursive structure that facilitates efficient selection of the context set, $\mathcal{C}$.

The FSMC model is completed by specifying the transmission mode. If transmission power is constant, then at the receiver

$\mathrm{SNR}(t) = \varrho x(t) = \varrho z(t)^2/2$, where ϱ is the average value of $\mathrm{SNR}(t)$. Assume binary modulation. Define the *crossover probability* $e(s)$ as the average error probability on the BSC associated with the occurrence of fading state s, $s \in S$. Crossover probabilities are evaluated as in [4]

$$e(s) = P(\text{error}|s) = \frac{1}{\pi_s} \int_{A_s}^{A_{s+1}} p_e(\varrho\xi) f(\xi)\, d\xi \qquad (2)$$

where $\pi_s = \int_{A_s}^{A_{s+1}} f(\xi)\, d\xi$ is the probability of fading state s, $f(\xi)$ is probability density of fading power (noncentral χ^2 with two degrees of freedom for Ricean fading, exponential for Rayleigh fading) and $p_e(\varrho\xi)$ is the error probability associated to the given modulation format, when SNR is equal to $\varrho\xi$ and AWGN is present on the channel.

Observe that each Markov state c (context) *does not hide* the fading state, being $s = s(c) = c(1)$. All the evaluations that can be performed for a first-order model can be repeated with higher order CT models (Section IV), using the *same* techniques. It is worth repeating that CT models determine *fading state evolution* and not quantities directly related to the fading state itself.

III. Model Structure

Fading has been simulated according to the filtering method proposed in [12] and the resulting process has been quantized on several levels. CTP has been applied to the discretized fading, choosing $M = 7$ as maximum context length (depth). However, recall that discretized fading process can be *exactly* represented only by an *infinite* depth CT. The increase of training sequence length N and of depth M can lead to models with increasing complexity and accuracy [8]. Training length $N = 5 \cdot 10^5$ samples and depth constraint $M = 7$ were chosen as a common basis for model comparison. Using that combination, it happens that the maximum context length d was strictly less than 7 in all models constructed over the range of fading environments investigated. Section IV shows that models within this scope can predict significant channel properties.

Fig. 1 shows some properties of context sets for models fitted to Rayleigh fading, assuming binary quantization, using threshold $-F$, and letting F (fade margin) vary in the range $0 \le F \le 20$ (decibels). Results for Rice fading or multiple-level quantization continue the trends of Fig. 1.

Fig. 1(a) concerns the depth (memory) of the models obtained in different fading environments. The plotted contours separate regions where a memoryless model (depth 0), a first-order Markov model (depth 1), or more complex models result from CTP. For the Rayleigh case, when fading is slow ($f_D T \lesssim 0.01$), a first-order model is supplied, coincident with the Wang model, while for fast fading ($f_D T \gtrsim 0.4$), a memoryless model results for practical values of F ($F \gtrsim 6$). Such conclusions hold for any number of levels and are found in the two- and three-level cases with the aid of statistical tests in [13]. On the other side, the intermediate range $0.01 \lesssim f_D T \lesssim 0.4$ is characterized with larger models. Note that information provided by the training sequence is insufficient to discover higher order models when $f_D T \le 0.01$, where it appeared that too long sequences are needed.

Fig. 1(b) shows the number of contexts $|\mathcal{C}|$ in each model. For CTP, the value of depth M is chosen *a priori* and it bounds the depth $d \le M$. With alphabet size L, a depth d CT model generates a process that admits an exact representation as the Markov chain with L^d states. By CTP, such a process can be generally represented using many fewer than L^d states: a Markov model with order (depth) d needs the estimation of $k_m = L^d(L-1)$ parameters, whereas the equivalent CT model needs only $k_c = |\mathcal{C}|(L-1)$.

Fig. 1(c) shows the state reductions allowed using CTP, as quantified by the ratio $k_m/k_c = L^{d(\mathcal{C})}/|\mathcal{C}|$. Where $|\mathcal{C}| \ll L^{d(\mathcal{C})}$, it indicates that only a few (judiciously selected) long memory states are used. In some environments, a CT model has an order of magnitude fewer states. The meaning of this gain can be explained as follows. Central limit theorem arguments [14] state that a k-dimensional model estimated from n training data has redundancy

$$\mathrm{E}\left\{-\log \frac{\hat{p}(\varphi_{n+1}|c(\varphi^n))}{p(\varphi_{n+1}|c(\varphi^n))}\right\} = r(k, n) + \mathcal{O}(1/n)$$

where $r(k, n) = (k/2n)\log_2(n/2\pi)$ and $\mathrm{E}\{\cdot\}$ is the expectation operator. If the unknown process were of similar or less complexity than an Mth-order Markov chain, and models were estimated using both CT parameterization and Mth-order parameterization, then $r(k_m, n) = (k_m/k_c) r(k_c, n)$. That is, the redundancy performance using the traditional parameterization is worse by a factor of k_m/k_c. Performance could also be compared by specifying equal model fidelity $r(k_c, n_c) = r(k_m, n_m)$ and asking how much extra data is required by the nonparsimonious fitting method, i.e., solve for n_m. For large n, the exact solution (available using Lambert's W function [15]) is bounded by $n_m > (k_m/k_c) n_c$. From Fig. 1(c), one sees that a traditional Markov model can require even an order of magnitude more training than CTP.

Fig. 1(d) depicts a classification of models in terms of state space structure. Four kinds of model structure are put on evidence by data analysis: memoryless, first-order Markovian, runlength, and general. The term *runlength* denotes models for which every sequence in the context set has the form $a^n b$ with $a, b \in \mathcal{S}$, $|b - a| \le 1$ (Fritchman-like) [3]. The term *general* indicates models for which none of the preceding highly structured state spaces is adequate; general contexts contain arbitrary patterns (e.g., 01 101). As the number of quantization levels increases, the region of Fritchman-like and general models expands to higher fade margins, but covers approximately the same range of fading rates. Among these complex models, the region of Fritchman-like models becomes narrower, that is, transitions between nonadjacent fading states become more relevant as quantization is refined.

IV. Applications

CTP-based models can be applied to the evaluation of channel or system properties such as the PED; the ACC of the quantized fading process [9], the distribution of time in a given channel state (in $\mathcal{S}$), examined in [16], the level crossing statistics [17], digital channel properties [6], [18], and protocol performance evaluation and optimization [19].

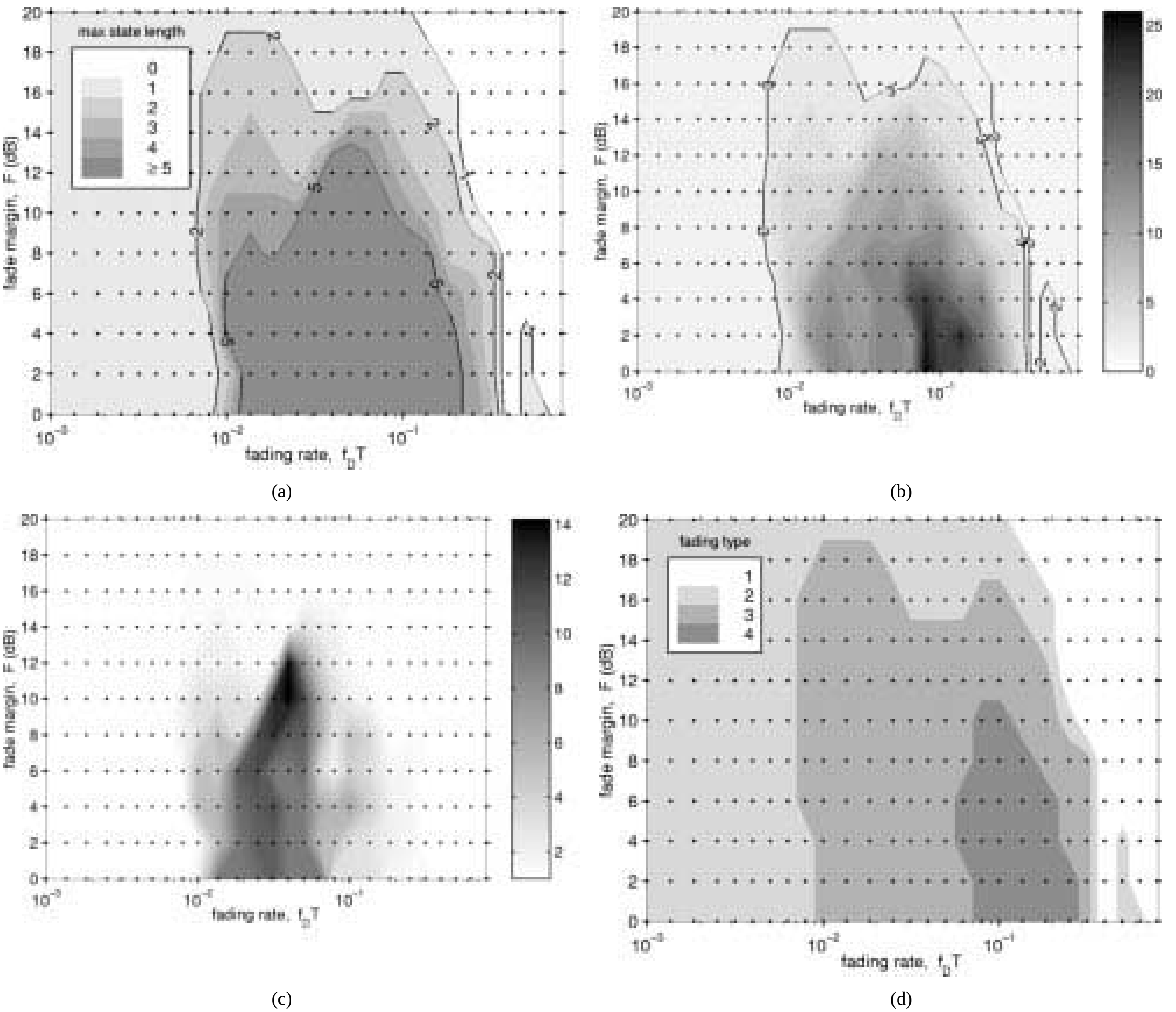

Fig. 1. Survey CT model properties for two-level quantized Rayleigh fading ($K = 0, L = 2, M = 7$) as a function of the fading rate $f_D T$ and the fade margin F. Each data point represents a CT model that has been fitted to $5 \cdot 10^5$ simulated samples of the indicated fading environment: (a) maximum context length; (b) number of contexts; (c) ratio between the number of parameters k_m needed to specify a traditional Markov model and the number of parameters k_c needed to specify the equivalent CT model; and (d) model type—1) memoryless, 2) first-order Markov, 3) runlength or Fritchman-like, and 4) general.

A. PED

PED $P(m, n)$, that is the probability of exactly m errors within an

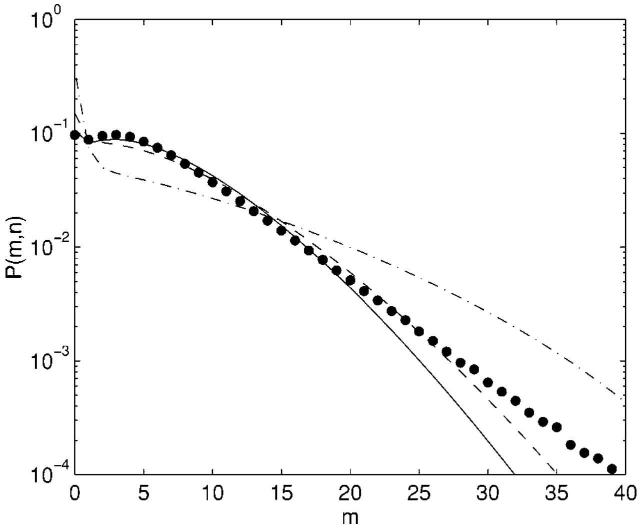

Fig. 2. Comparison between PED behaviors, obtained by direct simulation ("•"), by using first-order CT (approximately Wang) model (– · – · – ·), and by higher order CT models (——— : $\mathcal{M}_1$, – – – : $\mathcal{M}_2$), for the case treated in the text.

is 2-FSK with noncoherent detection [20, Sect. 4.3.1]. From (2), assuming $p_e(\varrho\xi) = (1/2)\, e^{-(1/2)\varrho\xi}$ [20, eq. (4.3.19)] and $f(\xi) = e^{-\xi}$ for Rayleigh fading, the average crossover probability in state k [i.e., when the context is c_k and the fading state is $s(c_k) \in \mathcal{S}$] is

$$\begin{aligned} e_k &= \frac{1}{\pi_{s(c_k)}} \int_{A_{s(c_k)}}^{A_{s(c_k)+1}} \frac{1}{2}\, e^{-(1/2)\varrho\xi}\, e^{-\xi}\, d\xi \\ &= \frac{1}{\pi_{s(c_k)}(2+\varrho)} \left[\exp\left(-A_{s(c_k)}\left(1+\frac{\varrho}{2}\right)\right)\right. \\ &\qquad \left. - \exp\left(-A_{s(c_k)+1}\left(1+\frac{\varrho}{2}\right)\right)\right] \end{aligned}$$

with $s(c_k) \triangleq c_k(1)$ (Section II) and $k = 0, \cdots, |\mathcal{M}| - 1$. The parameter values for simulation are ϱ

If $\varphi'_{\rm CT}$ is a fitted CT process, the evaluation is performed from (7), as in [9]

$$\begin{aligned} \mathrm{E}\{\varphi'_{\rm CT}[n]\varphi'_{\rm CT}[n+m]\} &= \sum_{i,\, j=0}^{|\mathcal{M}|-1} \overline{A}_{s(c_i)}\overline{A}_{s(c_j)} P(c[n]=c_i,\, c[n+m]=c_j) \\ &= \overline{\mathrm{A}}\mathrm{T}^m(\mathrm{p} * \overline{\mathrm{A}})^\top \\ \mathrm{E}\{\varphi'_{\rm CT}\} &= \overline{\mathrm{A}}\mathrm{p}^\top \end{aligned} \tag{8}$$

where $c[n]$ is the context at time nT, $(\cdot)^\top$ is the transpose operator, $\overline{\mathrm{A}} = \{\overline{A}_{s(c_k)}\}_{k=0}^{|\mathcal{M}|-1}$ and $\mathrm{p} * \overline{\mathrm{A}} = \{p_k \overline{A}_{s(c_k)}\}_{k=0}^{|\mathcal{M}|-1}$ are row vectors. Fig. 4 refers to the binary quantized Rayleigh process φ. Note that less complex models are required in this case.

V. Conclusions

The CTP algorithm can be applied to quantized "fading power" sequences, fitting nonhidden Markov models with generic state space structure. CTP is optimal in the sense that it among Markov descriptions with observable states, it provides nearly the smallest parameterization. For several fading regimes, tractable models with higher order than the well-known first-order model can be discovered, by using training sequences with nonexcessive length (residence time distribution in any state will be matrix-geometric [22], not geometric). Higher order models appear to be useful in specific applications, related to error behavior or to the analog channel properties, while low-order models with few states suffice for protocol evaluation.

CTP-based models give good results for applications that are conditioned by limited channel memory, such as evaluation of PED and related parameters in channels with fading that is not either fast or slow. In this case, accurate CTP models can be extracted having a weak dependence on the training sequence length. When the effect of very long memory must be considered, for example, fade length statistics in slow fading channels, CTP models may require excessive training data and grow to inconvenient size. Hidden Markov Models [17] accommodate that situation with more compact parameterization, though at the expense of state visibility and by the use of more elaborate model fitting tools.

References

[1] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. J.*, vol. 39, pp. 1253–1266, Sept. 1960.

[2] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol. 42, pp. 1977–1997, Sept. 1963.

[3] L. N. Kanal and A. R. K. Sastry, "Models for channels with memory and their applications to error control," *Proc. IEEE*, vol. 66, pp. 724–744, July 1978.

[4] H. S. Wang and N. Moayeri, "Finite-state Markov channel: A useful model for radio communication channel," *IEEE Trans. Veh. Technol.*, vol. 44, pp. 163–171, Feb. 1995.

[5] M. Zorzi, R. R. Rao, and L. B. Milstein, "ARQ error control on fading mobile radio channels," *IEEE Trans. Veh. Technol.*, vol. 46, pp. 445–455, May 1997.

[6] 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.

[7] H. Bischl and E. Lutz, "Packet error rate in the noninterleaved Rayleigh channel," *IEEE Trans. Commun.*, vol. 43, pp. 1375–1382, Feb./Mar./Apr. 1995.

[8] F. Babich, O. E. Kelly, and G. Lombardi, "A context-tree based model for quantized fading," *IEEE Commun. Lett.*, vol. 3, pp. 46–48, Feb. 1999.

[9] C. C. Tan and N. C. Beaulieu, "First-order Markov modeling for the Rayleigh fading channel," in *Proc. IEEE GLOBECOM'98*, Sydney, Australia, Nov. 8–12, 1998.

[10] J. D. Parson, *The Mobile Radio Propagation Channel*. New York: Halsted Press, 1992.

[11] M. J. Weinberger, J. J. Rissanen, and M. J. Feder, "A universal finite memory source," *IEEE Trans. Inform. Theory*, vol. 41, pp. 643–652, May 1995.

[12] J. I. Smith, "A computer generated multipath fading simulation for mobile radio," *IEEE Trans. Veh. Technol.*, vol. VT-24, pp. 39–40, Aug. 1975.

[13] G. Lombardi, "Measurement, simulation and analysis of indoor mobile channel," Ph.D. dissertation (in Italian), Dip. di Elettrotenica, Elettronica e Informatica, Univ. of Trieste, Trieste, Italy, 1998.

[14] J. Rissanen, "Fisher information and stochastic complexity," *IEEE Trans. Inform. Theory*, vol. 42, pp. 40–47, Jan. 1996.

[15] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Adv. Comput. Math.*, vol. 5, no. 4, pp. 329–359, 1996.

[16] F. Babich, O. E. Kelly, and G. Lombardi, "A variable-order discrete model for the mobile propagation channel," in *Broadband Wireless Communications*, M. Luise and S. Pupolin, Eds. London, U.K.: Springer-Verlag, 1998, pp. 259–266.

[17] W. Turin and R. Van Nobelen, "Hidden Markov models for fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1809–1817, Dec. 1998.

[18] M. Zorzi, "Outage and error events in bursty channels," *IEEE Trans. Commun.*, vol. 46, pp. 349–356, Mar. 1998.

[19] O. E. Kelly and F. Babich, "Protocol optimization using discrete channel statistics," in *Proc. IEEE 1998 Int. Conf. Universal Personal Communications (ICUPC'98)*, Florence, Italy, Oct. 5–9, 1998, pp. 469–473.

[20] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1985.

[21] K. S. Miller, *Multidimensional Gaussian Distributions*. New York: Wiley, 1964.

[22] A. Csenki, "Dependability for systems with a partitioned state space," in *Lecture Notes in Statistics, vol. 90*. New York: Springer-Verlag, 1994.