

Fading Channel Modeling via Variable-Length Markov Chain Technique

Wuttipong Kumwilaisak, *Member, IEEE*, C.-C. Jay Kuo, *Fellow, IEEE*, and Dapeng Wu, *Senior Member, IEEE*

Abstract—Channel characterization and modeling are essential to the wireless communication system design. A model that optimally represents a fading channel with a variable-length Markov chain (VLMC) is proposed in this paper. A VLMC offers a general class of Markov chains whose structure has a variable order and a parsimonious number of transition probabilities. The proposed model consists of two main components: 1) the optimal fading partition under the constraint of a transmission policy and 2) the derivation of the best VLMC representation with respect to the Kullback–Leibler (K–L) distance of fading samples. The fading partition is used to discretize a continuous fading channel gain. The optimal discretization criterion is developed based on the cost function of fading channel statistics and the transmission policy used in the system. Once a continuous fading channel gain is discretized, a VLMC is then used to model the channel. To obtain the optimal VLMC representation, we use the K–L distance of the discretized fading samples as the optimization criterion. The K–L distance of the discretized fading samples is used to determine the appropriate transition probabilities characterizing the optimal VLMC. Last, we show simulation results that demonstrate the accuracy and the effectiveness of the proposed fading channel representation in modeling the Rayleigh fading as well as the lognormal fading.

Index Terms—Channel modeling, fading partition, Kullback–Leibler (K–L) distance, Lagrangian optimization, variable-length Markov chain (VLMC).

I. INTRODUCTION

THE KNOWLEDGE of wireless fading channels plays an important role in the wireless communication system design. This is particularly true in allocating resources and selecting a suitable transmission policy to match the instantaneous fading channel condition [16], [23], [51], [52]. To represent a fading channel, one may use a mathematical model [1], [2], [28], [49] in terms of a probability density function (pdf) or a discrete Markov model. To analyze the error performance of a digital modulation scheme over a fading channel, the pdf of a *continuous* channel gain is typically used. However, it

is difficult to use the continuous channel model to analyze the performance of upper layer protocols (e.g., medium access control, transport layer protocols) due to the complex integration involved [21], [24], [26], [41], [48]. To simplify the fading channel modeling and reduce the analysis complexity, a *discrete* Markov model is often adopted as an alternative [12], [15], [24], [34], [40], [43], [45], [51], [52].

There are a couple of challenging issues to be addressed in making the discrete Markov model an accurate and useful tool for the wireless fading channel representation. They are elaborated below. First, the exact fading envelope or the fading channel SNR distribution is unknown in practice. Thus, the fading channel model with a discrete Markov model under the assumption of a given fading SNR distribution (e.g., the Rayleigh, Rician, or Nakagami distribution [39]) may not match well with actual measurements. Second, the first-order Markov model, which is the most popular discrete Markov model in representing fading channels [12], [15], [24], [34], [40], [43], [45], [51], [52], may not accurately model fading channels in all environments [40]. For example, a different Doppler rate may cause a different coherence time. Then, the search of the optimal order of a discrete Markov chain in fading channel modeling has to be studied. Third, the number of parameters characterizing a discrete Markov model should be small to simplify its structure and the performance analysis of a wireless communication system design. It is possible to have multiple Markov models representing the same fading channel with a given number of transition probabilities at the same time. The challenge is to derive a discrete Markov model that best represents the statistical property of a fading process from several Markov models. Last, the transmission policy that is used by a wireless communication system, e.g., adaptive modulation [16], should be considered in discrete Markov modeling.

To address the above issues, in this paper, we present a model that optimally represents a fading channel with a variable-length Markov chain (VLMC) [5]. A VLMC provides a general class of Markov chains that has a variable order and parsimonious number of transition probabilities. Our main idea can be simply stated as follows. The proposed modeling technique consists of two main parts: 1) optimal fading partitioning under the constraint on a transmission policy and 2) derivation of the best VLMC representation with respect to the Kullback–Leibler (K–L) distance criterion. In part 1), there are three steps of the fading partition, which is used to discretize a continuous fading channel gain. First, a kernel density estimation (KDE) scheme is used to estimate the pdf of the fading channel gain. Second, the quantization of the continuous channel gain is formulated as a constrained optimization problem. Third, a

Manuscript received November 16, 2005; revised October 23, 2006, May 16, 2007, July 14, 2007, and August 5, 2007. This work was supported by King Mongkut's University of Technology Thonburi Research Fund. The review of this paper was coordinated by Prof. E. Bonek.

W. Kumwilaisak is with the Department of Electronics and Telecommunication Engineering, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand (e-mail: wuttipong.kum@kmutt.ac.th).

C.-C. J. Kuo is with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: cckuo@sipi.usc.edu).

D. Wu is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: wu@ece.ufl.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TVT.2007.907305

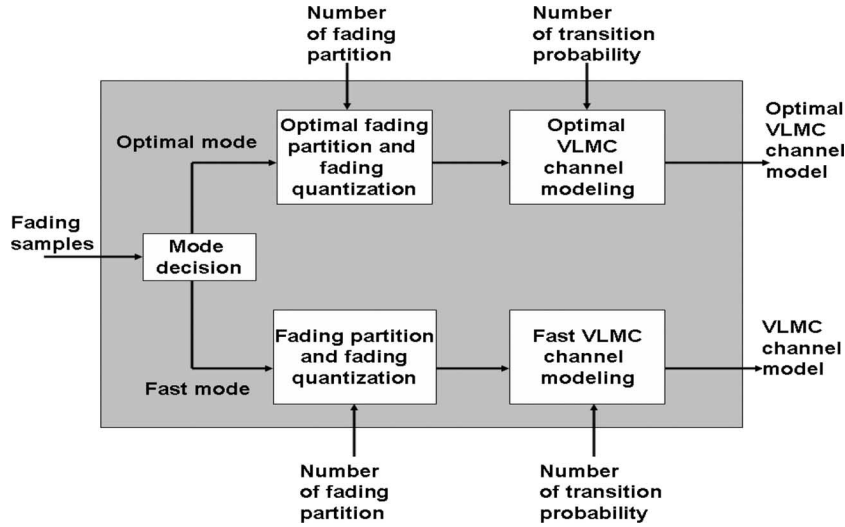


Fig. 1. Diagram of the proposed VLMC channel model.

relaxation method based on the Lagrange multiplier is used to solve the constrained optimization problem to provide the optimal solution to the fading partition. Once a continuous fading channel gain is quantized, a VLMC is used to model the channel. To obtain the optimal VLMC representation in part 2), we use the K–L distance as the optimization criterion and formulate the VLMC modeling as a constrained optimization problem. The Lagrangian optimization technique is then used to select the optimal VLMC representation that best approximates the statistics of the fading channel under a constraint on the total number of transition probabilities. When the fast channel modeling is desired, some steps of the optimal channel modeling method can be omitted and replaced with a fast VLMC channel modeling algorithm. Fig. 1 shows the flow of our proposed algorithm. There are three main inputs of the proposed channel modeling: 1) the number of fading partitions; 2) the number of transition probabilities characterizing VLMC; and 3) a set of fading SNR samples. The number of fading partitions and the number of transition probabilities are adjustable parameters. They control the complexity of the VLMC fading channel model.

The rest of this paper is organized as follows. Previous work in fading channel modeling with a discrete Markov process is reviewed in Section II. The fading partition under a transmission policy using the relaxation method is presented in Section III. The optimal fading modeling with a VLMC based on the K–L distance and the Lagrangian optimization technique is presented in Section IV. The fast VLMC channel modeling is also described in Section IV. Some properties of the VLMC are discussed in Section V. Applications of the proposed fading model are described in Section VI. Simulation results for the proposed channel modeling are shown in Section VII. Last, concluding remarks are given in Section VIII.

II. REVIEW OF PREVIOUS WORK AND RESEARCH MOTIVATION

Wireless fading channel modeling with discrete Markov models has been thoroughly examined in the past. We perform

a brief review on previous work and then describe our research motivation.

A. First-Order Markov Model

The efforts in [12], [15], [24], [34], [40], [43], [45], and [50]–[52] have focused on the use of the first-order Markov model to represent the fading channel and use the derived result in the design of a wireless communication system. Elliot [12] and Gilbert [15] first proposed the use of a finite-state Markov chain (FSMC) model to approximate a lossy channel with memory. The mapping from a physical fading channel to the first-order discrete Markov model was proposed by Wang and Moayeri [44] with the Rayleigh fading statistics. Sajadieh *et al.* [34] used the first-order continuous-time Markov double chain to model the Rayleigh fading channel and compute the Markov state dwell time. Similarly to [44], Zhang and Kassam [50] examined the first-order Markov model to represent the Rayleigh fading channel. The error trace in the network layer was studied using the first-order Markov model by Konrad *et al.* [24]. The application of the first-order Markov model to protocol design and performance analysis was considered in [40], [43], [51], and [52]. The effort included the design of forward error correction codes and the throughput analysis of the automatic repeat request (ARQ) scheme.

B. Other Markov Models

There have been some attempts to use higher order Markov models [3], [4] and the hidden Markov model (HMM) to represent fading channels. Babich *et al.* [4] used the context tree approach to model fading channels, where the derived Markov model has a variable order. Babich and Lombardi [3] used a contingency table to determine the order of the Markov chain and used the narrow-band Rayleigh and Rician statistics to compute state transition probabilities. The HMM was adopted to represent fading channels in [37] and [42]. Sivaprakasam and Shanmugan [37] used the Baum–Welch algorithm to estimate the HMM parameters. The fading channel modeling with the

HMM was also discussed by Turin and van Nobelen [42], where HMM parameters were computed from computer-generated data. They used the derived HMM for fading statistical analysis (e.g., the fade duration), which is difficult to compute based on the continuous fading channel model.

C. Research Motivation

The first-order Markov assumption for fading channel modeling was verified by Wang and Chang [45], which was, however, questioned by Tan and Beaulieu [40]. Even with the possible limitation of the first-order Markov model in representing the fading channel, Tan and Beaulieu [40] did not attempt to get a better fading channel model using a higher order Markov model under the constraint on the number of transition probabilities. Although higher order Markov models and HMMs were used to enhance the fading modeling accuracy, the optimal Markov model representation (e.g., the optimal order and the optimal number of states) has not been addressed. Moreover, a fading partition that takes into account the accuracy of channel parameters and the objective of a transmission policy was not considered before. These are the main research motivations of this paper.

III. OPTIMAL FADING PARTITION UNDER A TRANSMISSION POLICY

Before presenting the channel model, we describe a general methodology for a modem-layer channel modeling. A modem-layer channel consists of a digital modulator, a fading channel, and a demodulator [48], and it can be modeled by an FSMC whose states are characterized by different bit error rates (BERs). For example, a modem-layer channel with the Rayleigh fading and a certain Doppler spectrum can be converted to an FSMC with different BER states. The basic idea of this modeling process can be stated as follows.

- 1) Discretize the continuous Rayleigh random variable into a discrete random variable based on some optimal criterion (e.g., the minimum mean square error).
- 2) Map the resulting discrete random variable or SNR to a set of discrete BER values for a given modulation scheme, for example, BPSK.
- 3) Estimate the state transition probabilities that reflect the Doppler spectrum.

This procedure gives the states (in terms of different BER values) and the transition probability matrix of the Markov chain.

Being different from the above modem-layer channel model that assumes a fixed modulation scheme such as BPSK, our channel model is more general in the sense that it can incorporate advanced transmission policies, e.g., adaptive modulation, so that the mapping from an SNR to a BER is dynamic depending on the SNR value.

We first consider a set of fading sampled SNRs obtained from a fading process as $\Gamma = \{\gamma_1, \dots, \gamma_L\}$, where γ_i is the SNR sample i of the fading process, and L is the total number of SNR samples in Γ . The pdf of fading SNRs can be derived from SNR samples in Γ , which are not necessary to be in the

Estimated Fading SNR Distribution

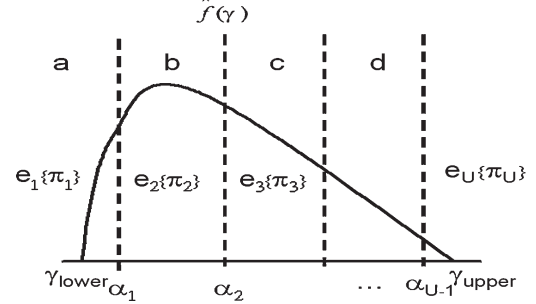


Fig. 2. Partition of a fading channel SNR range.

log-scale domain (i.e., decibel domain). In other words, we can directly use the fading gain in the linear-scale domain to derive the pdf of fading SNRs. This is highly dependent on the availability of the measured fading data. Here, we use the KDE [46] to estimate the pdf of a fading SNR, which can be expressed as

$$\hat{f}(\gamma) = \frac{1}{L \times h} \sum_{i=1}^L K\left(\frac{\gamma - \gamma_i}{h}\right)$$

where $K(\cdot)$ is an arbitrary pdf (e.g., the normalized Gaussian distribution with zero mean and unit variance) and has a bandwidth equal to h . To compute the optimal h to give the best match between the estimated and actual pdf's of the fading process, an iterative approach called the solved equation method [46] is used. Note that the pdf estimation can be done by other methods such as linear regularization [31].

Consider a fading partition scheme that partitions the whole fading SNR range of the estimated pdf to be U intervals as

$$\Delta = \{\alpha_1, \dots, \alpha_{U-1}\}, \quad \gamma_{\text{lower}} \leq \alpha_1 \leq \dots \leq \alpha_{U-1} \leq \gamma_{\text{upper}}$$

where $[\alpha_{i-1}, \alpha_i]$ is the interval i bounded by α_{i-1} and α_i , and γ_{lower} and γ_{upper} are the endpoints on the left and on the right of the estimated pdf, respectively, as shown in Fig. 2. Each fading partition is viewed as one channel state, and it has its own fading channel parameters for characterizing the channel condition. For example, if we partition the whole range of a fading SNR to two intervals, one of them represents the good channel state (i.e., with the low bit error probability), whereas the other represents the bad channel state (i.e., with the high bit error probability). In other words, the physical meaning of the fading partition is to discretely categorize the fading channel based on channel conditions. These partitions representing different fading channel conditions are characterized by fading channel parameters, such as the average probability of bit errors.

To compute the optimal fading partition, we jointly consider two criteria: 1) the accuracy of fading channel parameters characterizing the fading partition and 2) the objective of the wireless communication system design. For the first criterion, the fading channel parameters of the model should be as close as possible to those of the actual fading. For the second criterion, we need to meet the objective of the wireless communication system design when the specific transmission policy is adopted.

We use $C_a(\Delta)$ and $C_b(\Delta)$ to denote the cost functions for the accuracy of fading channel parameters and the objective of the wireless communication system, respectively. We consider $C_a(\Delta)$ first. The bit error probability of the fading channel, which is a function of the fading channel condition and the transmission policy, is used as a parameter in $C_a(\Delta)$. We try to minimize the difference of the bit error probability between the actual fading and that obtained from the fading partition. Then, we have

$$C_a(\Delta) = \sum_{i=1}^U \int_{\alpha_{i-1}}^{\alpha_i} (e_i\{\pi_i\} - P_b\{\gamma, \pi_i\})^2 \times \hat{f}(\gamma) d\gamma \quad (1)$$

where $P_b\{\gamma, \pi_i\}$ is the bit error probability of the actual fading channel with transmission policy π_i when $\text{SNR} = \gamma$ and $\alpha_0 = \gamma_{\text{lower}}$, $\alpha_U = \gamma_{\text{upper}}$, $\hat{f}(\gamma)$ is the estimated fading SNR pdf, and $e_i\{\pi_i\}$ is the average bit error probability characterizing the fading partition $[\alpha_{i-1}, \alpha_i]$ with transmission policy π_i . Transmission policy π_i specifies the transmission scheme (e.g., a modulation type) for the fading partition $[\alpha_{i-1}, \alpha_i]$. In different fading partitions, it is not necessary to have the same transmission policy. Here, we adopt the adaptive modulation scheme [16] as the transmission policy, which uses possibly different modulation schemes under different fading channel conditions.

For $C_b(\Delta)$, our objective of a transmission policy is to limit the average bit error per symbol, which is the product between the number of bits per symbol of the modulation and the bit error probability of the actual fading channel. The cost function can be written as

$$C_b(\Delta) = \sum_{i=1}^U \int_{\alpha_{i-1}}^{\alpha_i} B_i \times P_b\{\gamma, \pi_i\} \hat{f}(\gamma) d\gamma \quad (2)$$

where B_i is the number of bits per symbol of the modulation that is used in the interval i (e.g., if transmission policies π_i and π_{i+1} of fading interval i and $i+1$ are BPSK and QPSK and B_i and B_{i+1} are one and two bits, respectively).

Based on the cost functions defined above, our problem can be formulated as follows.

1) *Problem 1—Optimal Fading Partition Under a Transmission Policy:* Find the optimal fading partition Δ^* such that

$$\min_{\forall \Delta} C_a(\Delta) \quad (3)$$

subject to

$$C_b(\Delta) \leq C_{b,\text{target}} \quad (4)$$

where $C_{b,\text{target}}$ is the targeted limitation of the bit error per symbol.

The solution to Problem 1 can be obtained by minimizing a Lagrangian cost function, which is defined as

$$J_p(\lambda_p, \Delta) = C_a(\Delta) + \lambda_p C_b(\Delta) \quad (5)$$

where $J_p(\lambda_p, \Delta)$ is the combined cost function of $C_a(\Delta)$ and $C_b(\Delta)$, and λ_p is the Lagrange multiplier. There are two sets of

variables in the cost function given in (5), i.e., the fading partition $\Delta = \{\alpha_1, \dots, \alpha_{U-1}\}$ and a set of bit error probabilities for the fading partition, i.e., $E = \{e_1(\pi_1), \dots, e_{U-1}(\pi_{U-1})\}$.

To minimize $J_p(\lambda_p, \Delta)$ in (5), we can first take its derivative with respect to $e_i\{\pi_i\}$ and set it to zero, which leads to

$$e_i\{\pi_i\} = \frac{\int_{\alpha_{i-1}}^{\alpha_i} P_b\{\gamma, \pi_i\} \hat{f}(\gamma) d\gamma}{\int_{\alpha_{i-1}}^{\alpha_i} \hat{f}(\gamma) d\gamma}, \quad i = 1, 2, \dots, U. \quad (6)$$

Similarly, we can take the derivative of $J_p(\lambda_p, \Delta)$ in (5) with respect to α_i and set it to zero, which yields

$$\begin{aligned} & (e_i\{\pi_i\} - P_b\{\alpha_i, \pi_i\})^2 - (e_{i+1}\{\pi_{i+1}\} - P_b\{\alpha_i, \pi_{i+1}\})^2 \\ & = \lambda_p [B_{i+1} (1 - P_b\{\alpha_i, \pi_{i+1}\}) - B_i (1 - P_b\{\alpha_i, \pi_i\})] \\ & \quad i = 1, 2, \dots, U. \end{aligned} \quad (7)$$

Thus, we obtain a system of nonlinear equations (6) and (7), which can be solved by the relaxation method as described below.

Algorithm 1: Relaxation method to obtain the optimal fading partition

- Step 1) Select a value for λ_p , and initialize with a fading partition (for example, the uniform partition).
- Step 2a) Compute $e_i\{\pi_i\}$, $i = 1, \dots, U-1$, from the given fading partition using (6).
- Step 2b) Use the value of $e_i\{\pi_i\}$, $i = 1, \dots, U-1$ from Step 2a) to compute the fading partition α_i , $i = 1, \dots, U-1$, using (7).
- Step 2c) We iterate between Steps 2a) and 2b) until a converged fading partition is reached. This converged fading partition is the optimal fading partition for the given λ_p value.
- Step 3) Go to Step 1) to adjust the value of λ_p so that the solution meets the constraint of the averaged bit error per symbol.

From the above algorithm, an iterative approach, where several λ_p 's are tried until an appropriate solution is found, is required. Note that we use the bit error probability and the average bit error per symbol as parameters to set up the above optimization framework. However, it is possible to modify the optimization framework by accommodating various channel parameters and transmission policies of interest. Note that from our experiment, the number of iterations, which is commonly required to obtain a fading partition, is about ten.

IV. VLMC MODELING OF FADING CHANNELS

In this section, we explain the method to fit the discretized fading samples to the VLMC. The objective in VLMC modeling is to obtain the model that behaves as close as possible to the actual fading process. We review a concept of the VLMC in Section IV-A. The method in fitting discretized fading samples with the VLMC is described in Sections IV-B and C. The fast VLMC channel modeling is given in Section IV-D.

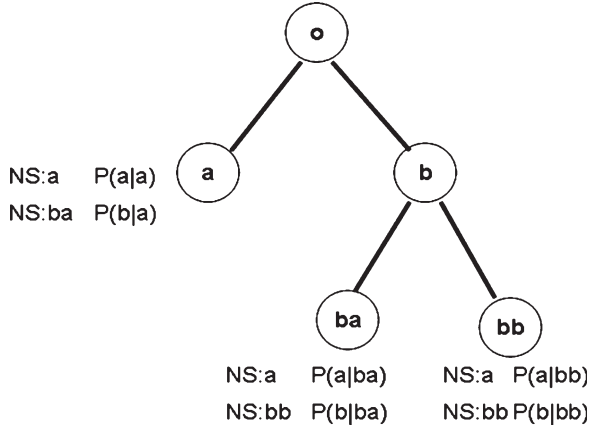


Fig. 3. Context tree structure of a VLMC. NS stands for the next state corresponding to a current state and a transition probability.

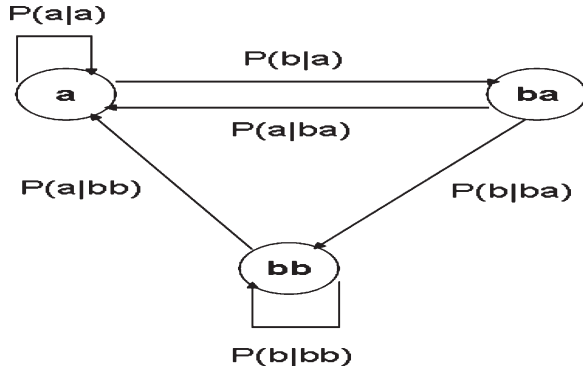


Fig. 4. Finite state machine representation of the VLMC.

A. VLMC

The VLMC [6], [47] is a Markovian process whose structure has a sparse memory with some states that are lumped together. The structure of the VLMC can be characterized by a variable order and a parsimonious number of transition probabilities. The VLMC can be represented by either a context tree or a finite state machine. We can derive a finite state machine representation directly from a context tree representation of the VLMC. An example of a context tree and a finite state machine representation of the VLMC is illustrated in Figs. 3 and 4, respectively. The VLMC in the example has a variable order, namely, one state is characterized by the first-order Markov chain (i.e., state a), whereas the others are characterized by the second-order Markov chain (i.e., states ba and bb).

To describe the VLMC in detail, we first define the finite categorical space χ as an arbitrary set of letters that represent all possible data values that are generated from a Markov process. For example, the finite categorical space in Fig. 3 is $\chi = \{a, b\}$. The context function and the variable-length memory can be introduced to serve as tools in VLMC construction [6].

Definition 1: Let $(\mathbf{x}_t)_{t \in \mathbb{Z}}$ be a time series, where $\mathbf{x}_{-l+1+t}^t = x_{-l+1+t}, \dots, x_t$, with value $x_t \in \chi$. We define a function, which is denoted by Ψ , that maps an infinite sequence (an infinite past) to a possibly shorter string as

$$\Psi : \mathbf{x}_{-\infty}^t \rightarrow \mathbf{x}_{-l+1+t}^t$$

where

$$l = l(\mathbf{x}_{-\infty}^t) = \min \{k; P\{x_{t+1}|\mathbf{x}_{-\infty}^t\} = P\{x_{t+1}|\mathbf{x}_{-k+1+t}^t\}, \forall x_{t+1} \in \chi\}.$$

Then, $\Psi(\cdot)$ is called a *context function*, and $\Psi(\mathbf{x}_{-\infty}^t) = \mathbf{x}_{-k+1+t}^t$ is called a *context* of the process time t with length k . Let $0 \leq \zeta \leq \infty$ be the smallest integer such that

$$l(\mathbf{x}_{-\infty}^t) \leq \zeta, \quad \forall \mathbf{x}_{-\infty}^t \in \chi^\infty.$$

Then, ζ is called the *order* of the context function $\Psi(\cdot)$, and, if $\zeta \leq \infty$, $(\mathbf{x}_t)_{t \in \mathbb{Z}}$ is called a *VLMC* of order ζ .

It is obvious from the above definition that the VLMC of order ζ is a Markov chain of order ζ with an additional structure having variable order l inside its structure. The context represents how previous data values of the Markov process contribute to the knowledge of the incoming data values, and the length of the context corresponds to the number of past data. For example, the context function for the VLMC given in Figs. 3 and 4 is

$$\Psi(\mathbf{x}_{-\infty}^{t-1}) = \begin{cases} a, & x_{t-1} = a, \mathbf{x}_{-\infty}^{t-2} \text{ arbitrary} \\ ba, bb, & x_{t-1} = b, x_{t-2} \in \{a, b\} \\ \mathbf{x}_{-\infty}^{t-3} \text{ arbitrary.} \end{cases}$$

Thus, the corresponding transition probabilities can be expressed as

$$P\{x_t | x_{t-1}, \dots\} = \begin{cases} P\{x_t | x_{t-1} = a\} \\ P\{x_t | x_{t-1} = b, x_{t-2} \in \{a, b\}\}. \end{cases}$$

Given the contexts of a VLMC, we can represent it as a context tree that is defined as follows.

Definition 2: Let $\Psi(\cdot)$ be a context function of the VLMC. Its context tree τ is defined as

$$\tau = \{\omega_l : \omega_l = \Psi(\mathbf{x}_{-\infty}^t), \omega_l u \notin \tau, \forall u \in \chi, l \leq \zeta\}$$

where ω_l is a context of length l corresponding to the nodes of the tree, and $\omega_l u$ is the concatenation of context ω_l and alphabet u , where $u \in \chi$.

The context tree structure of the VLMC does not have to be complete. It depends only on those contexts that have an influence on the occurrence of incoming data values, thus reducing the number of parameters of a complete Markov model. The terminal nodes of the tree (i.e., the node that does not possess children nodes or the node that does not possess children nodes equaling $|\chi|^{l+1}$, where l is the length of a context corresponding to the considering node) will represent the state of the VLMC, whereas the depth of the tree corresponds to the order of the VLMC. Fig. 3 gives a tree structure representing the second-order three-state VLMC with three terminal nodes and one internal node (node o). Only terminal nodes are used in the VLMC channel model, where there are a total of three states of the VLMC corresponding to terminal nodes. The depth of the tree is equal to two. This means that the maximum order of the VLMC is equal to two. In contrast, the complete second-order Markov model has four terminal nodes, which are nodes corresponding to aa , ab , ba , and bb . As we can see, the VLMC

lumps nodes aa and ab of the complete second-order Markov model together. It implies that the structure of the VLMC is less complex than a higher order Markov model.

In this paper, we use the VLMC to model fading channels by exploring the time-varying nature of the Doppler rate or the coherence time. For example, we can use a high-order or a low-order Markov chain to characterize a channel with long or short coherence time while maintaining low description complexity with the VLMC. In other words, when the fading is changing with time, the structure and the parameters of the VLMC fading channel model can optimally capture these variations. In the next two subsections, we propose a method to obtain the optimal VLMC representation of fading channels constrained by the number of transition probabilities.

B. Fading Channel Modeling: Problem Formulation

A VLMC model can be built based on a set of measured fading SNR values and a finite categorical space χ , where each letter represents an interval in a fading partition, as described in Section III. To be more specific, we use $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_L\}$ to denote a set of measured fading SNR values and $\mathcal{A} = \{x_1, x_2, \dots, x_L\}$ to represent a letter sequence that is obtained by assigning a letter in χ to every measured fading sample in Γ . To give an example, if γ_1 is between $[\alpha_0, \alpha_1]$ in Fig. 2, we assign the letter “ a ” to x_1 . We call the letters in \mathcal{A} the discretized fading samples.

To characterize a context tree or a VLMC, we need to compute the probability and the transition probability of a context. The probability of context ω_l can be written as

$$P\{x_k^{k+1-1} = \omega_l\} = \frac{N(\omega_l)}{L - l + 1} \quad (8)$$

where ω_l is a context with length l in \mathcal{A} , and $N(\omega_l)$ is the number of contexts ω_l obtained from \mathcal{A} . The transition probability can be expressed as

$$P\{x_{k+l} = u_0 | x_k^{k+1-1} = \omega_l\} = \frac{N(\omega_l u_0)}{N(\omega_l)} \quad (9)$$

where $N(\omega_l u_0)$ is the number of contexts $\omega_l u_0$, which is the concatenated context of context ω_l and letter u_0 , and $P\{x_{k+l} = u_0 | x_k^{k+1-1} = \omega_l\}$ is a transition probability given that the past sequel of letters is equal to context ω_l with the incoming data equal to $u_0 \in \chi$. To obtain $N(\omega_l)$ and $N(\omega_l u_0)$, the sliding window with length l is used to count the number of contexts that are existing in \mathcal{A} .

The closeness of two transition probabilities with contexts ω_l and $\omega_l u$ can be measured using the K–L distance [11], [27], i.e.,

$$\begin{aligned} & \rho(P\{x_{k+l} = u_0 | x_k^{k+1-1} = \omega_l u\} \| \\ & P\{x_{k+l} = u_0 | x_k^{k+1-1} = \omega_l\}) \\ &= P\{x_{k+l} = u_0 | x_k^{k+1-1} = \omega_l u\} \\ & \times \ln \frac{P\{x_{k+l} = u_0 | x_k^{k+1-1} = \omega_l u\}}{P\{x_{k+l} = u_0 | x_k^{k+1-1} = \omega_l\}}. \end{aligned} \quad (10)$$

The smaller the K–L distance, the closer these two transition probabilities. When the K–L distance is close to zero, there is

little contribution in the knowledge improvement of incoming letter u_0 from the concatenated context $\omega_l u$, where $u \in \chi$, as compared with context ω_l alone. To give an example, suppose that the K–L distance in (10) between $\omega_l u = ab$ and $\omega_l = a$ is close to zero. This shows that contexts ab and a have almost the same contribution to the knowledge improvement of incoming letter $u_0 \in \chi$. Therefore, there is no need to concatenate letter b to context a , and the Markov state corresponding to context ab is not included in the VLMC structure.

There could be more than one VLMC to represent the same fading channel. To select the optimal one, the tradeoff between the model complexity and its accuracy in reflecting the stationary process of the fading channel has to be balanced. The complexity of a VLMC model can be measured in terms of the number of states and the number of transition probabilities. The summation of the K–L distances of terminal nodes of a context tree for the VLMC channel model Υ can be written as

$$\begin{aligned} D_{\Upsilon} & (P\{x_{k+l} = u_0 | x_{k-1}^{k+1-1} = \omega_l u\} \| \\ & P\{x_{k+l} = u_0 | x_{k-2}^{k+1-1} = \omega_l\}) \\ &= \sum_{\forall \omega_l u \in \tau_t} \rho(P\{x_{k+l} = u_0 | x_{k-1}^{k+1-1} = \omega_l u\} \| \\ & P\{x_{k+l} = u_0 | x_{k-2}^{k+1-1} = \omega_l\}) \end{aligned} \quad (11)$$

where τ_t is a set of terminal nodes of the context tree of VLMC Υ , and $l \leq \zeta$ is the length of a context that is associated with a terminal node. The smaller the summed K–L distance of all terminal nodes is, the better the VLMC represents the statistical property of a fading process that is represented by the letters in χ . We have the following theorem about the summed K–L distance of terminal nodes.

Theorem 1—Convexity of the Summed K–L Distance of Terminal Nodes: $D_{\Upsilon}(P\{x_{k+l} = u_0 | x_{k-1}^{k+1-1} = \omega_l u\} \| P\{x_{k+l} = u_0 | x_{k-2}^{k+1-1} = \omega_l\})$ is a convex function in a set of transition probabilities of VLMC terminal nodes.

The proof of Theorem 1 is given in the Appendix.

The problem of determining the optimal VLMC channel representation can be formulated as follows.

Problem 2—Optimal VLMC Channel Representation: Find the optimal VLMC channel representation Υ^* such that

$$\begin{aligned} \min_{\Upsilon} & D_{\Upsilon} \left(P\{x_{k+l} = u_0 | x_{k-1}^{k+1-1} = \omega_l u\} \| \right. \\ & \left. P\{x_{k+l} = u_0 | x_{k-2}^{k+1-1} = \omega_l\} \right) \end{aligned} \quad (12)$$

subject to

$$N_{\Upsilon} \leq N_{\Upsilon, \text{target}}$$

where $D_{\Upsilon}(P\{x_{k+l} = u_0 | x_{k-1}^{k+1-1} = \omega_l u\} \| P\{x_{k+l} = u_0 | x_{k-2}^{k+1-1} = \omega_l\})$ is the summed K–L distance between all terminal nodes and their parent nodes corresponding to a context tree of VLMC (Υ), and $N_{\Upsilon, \text{target}}$ is the number of transition probabilities that are allowed to characterize the optimal VLMC channel representation.

C. Fading Channel Modeling: Problem Solution

Due to the convexity of the summed K–L distance of the VLMC terminal nodes, we can use the Lagrangian optimization technique to solve Problem 2 in search for the optimal VLMC model. For the minimization problem in Problem 2, we can define the following cost function:

$$J_c(\lambda_c) = D_T \left(P \{ x_{k+l} = u_0 | \mathbf{x}_{k-1}^{k+1-1} = \omega_l u \} \parallel P \{ x_{k+l} = u_0 | \mathbf{x}_{k-2}^{k+1-1} = \omega_l \} \right) + \lambda_c N_T \quad (13)$$

where λ_c is the Lagrange multiplier. The solution procedure consists of two steps: 1) constructing the summed K–L distance characteristic curve and 2) searching the optimal VLMC representation. They are detailed below.

1) *Constructing the Summed K–L Distance Characteristic Curve:* The summed K–L distance characteristic curve characterizes the relationship between the summed K–L distance of terminal nodes from a context tree of a VLMC model and its number of transition probabilities. It is built by considering several possible context trees representing the same discretized fading process \mathcal{A} . Here, we limit the maximum depth of a context tree (or the maximum order of the VLMC) to ξ_{\max} to control the size of a context tree. The algorithm for constructing the summed K–L characteristic curve is described below.

Algorithm 2: Construction of the summed K–L distance characteristic curve

Initialization: Initialize the root node of a tree.

Step 1) Let $l = 1$ be the size of a sliding window and ω_l be a context with length l . From \mathcal{A} , compute $P\{\mathbf{x}_{k-1}^{k+1-1} = \omega_l\}$, $P\{x_{k+l+1} = u_0 | \mathbf{x}_{k-1}^{k+1-1} = \omega_l\}$, and $P\{x_{k+l+1} = u_0 | \mathbf{x}_k^{k+1-1} = \omega_l u\}$ through a sliding window $\forall u_0, u \in \chi$, where $\omega_l u$ is a context with length $l+1$ obtained from concatenating context ω_l with u . Then, add nodes corresponding to ω_l in \mathcal{A} to the tree.

Step 2) Repeat Step 1) until $l = \xi_{\max}$, which will give τ_0 . Then, go to Step 3).

Step 3) For every context from \mathcal{A} , compute $\rho(P\{x_{k+l} = u_0 | \mathbf{x}_{k-1}^{k+1-1} = \omega_l u\} \parallel P\{x_{k+l} = u_0 | \mathbf{x}_{k-2}^{k+1-1} = \omega_l\})$, which is the K–L distance between $P\{x_{k+l} = u_0 | \mathbf{x}_{k-1}^{k+1-1} = \omega_l u\}$ and $P\{x_{k+l} = u_0 | \mathbf{x}_{k-2}^{k+1-1} = \omega_l\}$ for specific context $\omega_l u$ and ω_l from parameters that are obtained in Step 1).

Step 4) Select threshold T_{th} and prune nodes corresponding to context $\omega_l u$ if

$$\left| \rho \left(P \{ x_{k+l} = u_0 | \mathbf{x}_{k-1}^{k+1-1} = \omega_l u \} \parallel P \{ x_{k+l} = u_0 | \mathbf{x}_{k-2}^{k+1-1} = \omega_l \} \right) \right| < T_{th}. \quad (14)$$

Repeat the pruning process for every node in τ_0 until the condition in (14) is not satisfied for all remaining nodes.

Step 5) From the pruned tree, compute the summed K–L distance of the terminal nodes of the tree. Repeat Steps 4) and 5) for various threshold values, which begins with τ_0 .

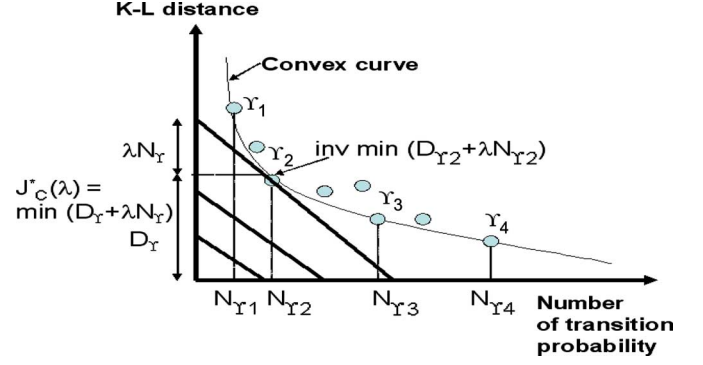


Fig. 5. Summed K–L distance characteristic curve and the process in using the Lagrangian optimization to search for the optimal solution.

For different T_{th} , we obtain various pairs of the summed K–L distance and the corresponding number of transition probabilities, which gives the summed K–L distance characteristic curve. An example is illustrated in Fig. 5.

2) *Optimal VLMC Model:* We use the bisection algorithm [31] to search for λ_c^* from the summed K–L distance characteristic curve, which provides the optimal VLMC representation based on the cost function $J_c(\lambda_c)$. Let $D_T(\lambda_c)$ and $N_T(\lambda_c)$ be the summed K–L distance defined in (11) and the number of transition probabilities of the VLMC channel model, respectively. When the Lagrange multiplier is equal to λ_c^* , it gives the optimal VLMC model. The bisection search algorithm can be stated below.

Algorithm 3: The bisection algorithm to search for the optimal VLMC

Initialization: Initialize with two values λ_u and λ_l such that $N_T(\lambda_u) \leq N_{T,target} \leq N_T(\lambda_l)$.

Step 1) Set $\lambda_{next} = ((D_T(\lambda_l) - D_T(\lambda_u)) / (N_T(\lambda_l) - N_T(\lambda_u))) + \epsilon$, where ϵ is an arbitrarily small positive number, which is added to ensure that the smallest number of channel parameters is selected if λ_{next} is a singular slope value. Perform the optimization process in (13) for $\lambda_c = \lambda_{next}$.

Step 2) If $N_T(\lambda_{next}) = N_{T,target}$, then stop. Otherwise, set $\lambda_l = \lambda_{next}$ if $N_T(\lambda_{next}) \geq N_{T,target}$, or set $\lambda_u = \lambda_{next}$ if $N_T(\lambda_{next}) \leq N_{T,target}$. Go to Step 1).

D. Fast VLMC Channel Modeling Algorithm

The methods described in Sections III and IV-C1 and 2 find the optimal representation of a fading channel with a fixed number of transition probabilities. The optimal VLMC provides a bound on the representation of the fading channel in terms of statistical properties, which are measured by the K–L distance. In this section, we describe a fast algorithm to compute the VLMC channel model to significantly reduce the complexity of channel modeling. The input parameters to this process are the number of fading partitions and the number of transition probabilities. The algorithm can be stated as follows.

Algorithm 4: Fast VLMC channel modeling

Initialization: We are given the number of fading partitions U , the number of transition probabilities, and a set of fading samples, where all fading samples are discretized based on their SNR intervals.

- Step 1) Perform Steps 1) to 3) as given in Algorithm 3. If the total number of transition probabilities corresponding to the current context tree does not exceed the given number of transition probabilities, we can terminate the algorithm and use the VLMC channel model corresponding to the context tree as a channel model. If this condition fails to hold, go to Step 2).
- Step 2) Prune the node of the tree corresponding to context $\omega_l u$ that has the lowest $|\rho(P\{x_{k+l}=u_0 | \mathbf{x}_{k-1}^{k+1-1} = \omega_l u\} \| P\{x_{k+l}=u_0 | \mathbf{x}_{k-2}^{k+1-1} = \omega_l\})|$. Then, compute the number of transition probabilities of the pruned tree. If it does not exceed the given number of transition probabilities, terminate the algorithm and use the VLMC channel model corresponding to the context tree as a channel model. Otherwise, repeat Step 2) until the desired VLMC channel model is obtained.

V. PROPERTIES OF THE VLMC CHANNEL MODEL

Several properties of the VLMC channel model are examined in this section. With a given fading partition, we may have new contexts added to the VLMC channel model. The mutual information of discretized fading samples is studied in the following theorem, which sheds light on the correlation between discretized fading samples.

Theorem 2—Mutual Information Property of the VLMC: Consider a fading process, which is stationary and ergodic, and two VLMC models Υ and Υ_e , where Υ_e is derived from Υ by adding node/nodes to the context tree of Υ , which results in an order increase. Let M_e and M be orders of Υ_e and Υ , respectively. The letter x_i , which is an element of \mathcal{X} , represents a discretized fading sample at time i . Then, we have

$$\mathcal{I}_e(x_{M+1}; \mathbf{x}_1^M) \geq \mathcal{I}(x_{M+1}; \mathbf{x}_1^M) \quad (15)$$

where $\mathcal{I}_e(x_{M+1}; \mathbf{x}_1^M)$ and $\mathcal{I}(x_{M+1}; \mathbf{x}_1^M)$ denote the mutual information [11] of discretized fading samples at time $M+1$, given the knowledge of the past discretized fading samples from time 1 to M obtained from Υ_e and Υ , respectively.

The proof of Theorem 2 is given in the Appendix. This theorem implies that the discretized fading samples before adding new contexts decorrelate faster than those after adding new contexts. In other words, the correlation among the discretized fading samples increases when they are represented by a higher order VLMC. This result is also proven by experiments as shown in Figs. 12 and 13 in Section VII for the lognormal fading environment.

The next two theorems are developed to study the effect from changing the number of fading partitions to the property of discretized fading samples. First, we study the mutual infor-

mation of the VLMC when the number of fading partitions is increased by splitting the partition. Then, we derive the bound on the mutual information reduction when the fading partitions are merged.

Theorem 3—Asymptotic Mutual Information of the VLMC After the Split of One Interval: Consider a stationary and ergodic fading process. Two subintervals $(\alpha_{i-1}, \delta]$ and $(\delta, \alpha_i]$ are obtained by splitting the interval $(\alpha_{i-1}, \alpha_i]$ into two. Define

$$p_{i,1} = \int_{\alpha_{i-1}}^{\delta} \hat{f}(\gamma) d\gamma, \quad p_{i,2} = \int_{\delta}^{\alpha_i} \hat{f}(\gamma) d\gamma,$$

$$p_i = p_{i,1} + p_{i,2} = \int_{\alpha_{i-1}}^{\alpha_i} \hat{f}(\gamma) d\gamma,$$

$$\mathcal{H}_K = -\sum_{j=1}^K p_j \ln p_j$$

$$\mathcal{H}_{K+1} = -\sum_{j=1, j \neq i}^K p_j \ln p_j - p_{i,1} \ln p_{i,1} - p_{i,2} \ln p_{i,2}$$

where p_i is the average probability of the partitioned interval i ; $p_{i,1}$ and $p_{i,2}$ are the average probabilities of the two subintervals that are obtained from splitting the interval i ; and \mathcal{H}_K and \mathcal{H}_{K+1} are entropies of discretized fading samples before and after splitting the partitioned interval, respectively. Furthermore, let M be the maximum order of the VLMC and L be the number of discretized fading samples from the fading process used for constructing the VLMC with the size sufficiently large. Then, the mutual information of the discretized fading samples x_{M+1} and x_M, \dots, x_1 after splitting $\mathcal{I}_{K+1}(x_{M+1}; \mathbf{x}_1^M)$ increases with probability 1 if

$$\begin{aligned} p_{i,1} \ln \left(\frac{p_{i,1} + p_{i,2}}{p_{i,1}} \right) + p_{i,2} \ln \left(\frac{p_{i,1} + p_{i,2}}{p_{i,2}} \right) \\ \geq (K+1)^L e^{-L\mathcal{H}_{K+1}} \mathcal{H}_{K+1}. \end{aligned} \quad (16)$$

Theorem 4—Asymptotic Mutual Information of the VLMC After Merging Two Intervals: Consider a stationary and ergodic fading process. Let $(\alpha_{i-1}, \alpha_{i+1}]$ be the interval obtained by merging two neighboring intervals $(\alpha_{i-1}, \alpha_i]$ and $(\alpha_i, \alpha_{i+1}]$. Define

$$p_i = \int_{\alpha_{i-1}}^{\alpha_i} \hat{f}(\gamma) d\gamma, \quad p_{i+1} = \int_{\alpha_i}^{\alpha_{i+1}} \hat{f}(\gamma) d\gamma$$

$$p_{i,i+1} = p_i + p_{i+1} = \int_{\alpha_{i-1}}^{\alpha_{i+1}} \hat{f}(\gamma) d\gamma.$$

In other words, p_i and p_{i+1} are the average probabilities of intervals i and $i+1$, respectively, and $p_{i,i+1}$ is the average probability of the interval after merging i and $i+1$. Furthermore,

we can express entropies of discretized fading samples before and after merging fading partitions as

$$\mathcal{H}_{K+1} = -\sum_{j=1}^{K+1} p_j \ln p_j$$

$$\mathcal{H}_K = -\sum_{j=1, j \neq i, i+1}^{K+1} p_j \ln p_j - p_{i,i+1} \ln p_{i,i+1}$$

respectively. Let M be the maximum order of the VLMC and L be the number of discretized fading samples from the fading process used for constructing the VLMC when the size is sufficiently large. Then, the mutual information of the discretized fading samples y_L and $\mathcal{I}_K(y_{M+1}; \mathbf{y}_1^M)$ will be decreased at most by the following amount:

$$K^L H_K e^{-LH_K} + p_i \ln \frac{p_{i,i+1}}{p_i} + p_{i+1} \ln \frac{p_{i,i+1}}{p_{i+1}}. \quad (17)$$

The proofs of Theorems 3 and 4 are given in the Appendix.

When we split an interval in the fading partition, the correlation between discretized fading samples may increase or decrease. However, if the conditions in Theorem 3 are met, we can guarantee with probability 1 that the correlation between discretized fading samples increases. Theorem 3 also implies that, to achieve a good approximation of the statistical property of a fading process, the process of fading partition and VLMC channel modeling should be jointly considered. Theorem 4 gives a lower bound on the mutual information reduction when the number of fading partitions is reduced. In other words, the correlation of discretized fading samples after merging intervals in a fading partition will reduce at most by that corresponding to the derived lower bound of the mutual information reduction.

The next theorem specifies the relationship of the K-L distances of parent and child nodes in a context tree in terms of mutual information and entropy.

Theorem 5—Average K-L Distance Between Parent and Child Nodes of a Context Tree: Let M be the maximum order of the VLMC, u the letter of χ representing the partition, and ω_{M-1} a context with length $M-1$. Define $D_{\text{avg}}(P\{x_{M+1} = u_0 | \mathbf{x}_1^M \omega_{M-1} u\} || P\{x_{M+1} = u_0 | \mathbf{x}_2^M \omega_{M-1}\})$, which is the average K-L distance between $P\{x_{M+1} = u_0 | \mathbf{x}_1^M \omega_{M-1} u\}$ and $P\{x_{M+1} = u_0 | \mathbf{x}_2^M \omega_{M-1}\}$ for all contexts ω_{M-1} with length $M-1$. $\mathcal{I}(x_{M+1} = u_0; \mathbf{x}_2^M = \omega_{M-1})$ and $\mathcal{I}(x_{M+1} = u_0; \mathbf{x}_1^M = \omega_{M-1} u)$ are the mutual information of u_0 and ω_{M-1} , as well as u_0 and $\omega_{M-1} u$, respectively. $\mathcal{H}_p(P\{x_{M+1} = u_0 | \mathbf{x}_1^M = \omega_{M-1} u\} || P\{x_{M+1} = u_0 | \mathbf{x}_2^M \omega_{M-1}\})$ is the cross entropy between $P\{x_{M+1} = u_0 | \mathbf{x}_2^M = \omega_{M-1}\}$ and $P\{x_{M+1} = u_0 | \mathbf{x}_1^M = \omega_{M-1} u\}$, and $\mathcal{H}(x_{M+1} = u_0 | \mathbf{x}_2^M = \omega_{M-1})$ is the conditional entropy of u_0 given the knowledge of context ω_{M-1} . Then, we have

$$D_{\text{avg}}(P\{x_{M+1} = u_0 | \mathbf{x}_1^M = \omega_{M-1} u\} || P\{x_{M+1} = u_0 | \mathbf{x}_2^M = \omega_{M-1}\}) = \Delta \mathcal{I} + \Delta \mathcal{H} \quad (18)$$

where

$$\Delta \mathcal{I} = \mathcal{I}(x_{M+1} = u_0; \mathbf{x}_2^M \omega_{M-1}) - \mathcal{I}(x_{M+1} = u_0; \mathbf{x}_1^M = \omega_{M-1} u) \quad (19)$$

$$\Delta \mathcal{H} = \mathcal{H}_p(P\{x_{M+1} = u_0 | \mathbf{x}_1^M \omega_{M-1} u\} || P\{x_{M+1} = u_0 | \mathbf{x}_2^M = \omega_{M-1}\}) - \mathcal{H}(x_{M+1} = u_0 | \mathbf{x}_2^M = \omega_{M-1}). \quad (20)$$

The proof of Theorem 5 is given in the Appendix. By Theorem 5, we see that the K-L distance between parent and child nodes of a context tree is the difference between the information gained and the uncertainty reduction by increasing the order of the VLMC. When they are equal, there is no need to increase the order of the VLMC any longer since there is no more uncertainty reduction of the discretized fading process (i.e., the increase in mutual information).

VI. APPLICATIONS OF THE VLMC CHANNEL MODELING

In this section, we show the usage of the VLMC channel model in two example applications, namely, the computation of the average fade duration and the signal outage probability.

A. Average Fade Duration

The average fade duration of the small-scale fading (i.e., the Rayleigh fading channel) plays an important role in the communication system design. It is related to the packet length selection, the signal outage probability, and the size of the interleaver that is used in error correction codes, as explained below. In the digital broadcasting context, the packet length is often chosen to be longer than the average fade duration, and an error-correction code is used to correct the burst errors. In the full duplex mobile data communication system or the random access system, the packet length can be selected to be shorter than the average fade duration, and the ARQ scheme is used to retransmit lost packets. For the signal outage probability, we can use it to adaptively adjust the power of the transmitted signal along with the fading channel conditions and the distance between the transmitter and the receiver. The power adjustment can ensure the probability that the strength of the received signal is small in the fade period. Last, the size of the interleaver should be larger than the fade duration to overcome burst errors so that corrupted bits can be recovered by error correction codes.

It is difficult to obtain an analytical solution of the average fade duration using the mathematical characterization of the small-scale fading channel since it demands the knowledge of the joint pdf between an SNR and its derivative [16], which is difficult to obtain. Even if the joint pdf is available, a complex integration has to be performed to analytically or numerically obtain the average fade duration. In contrast, the computation of the average fade duration of the small-scale fading using the proposed VLMC channel model requires only addition, multiplication, and division operations. We show how to compute the average fade duration using the VLMC channel model below.

Suppose that the VLMC consists of J_{tot} states corresponding to U fading intervals, where fading interval j corresponds to the

SNR range between γ_{j-1} and γ_j . Note that $J_{tot} \geq U$ because there may be more than one Markov state corresponding to one fading interval. For example, suppose that $U = 2$. The fading partitions are represented by letters a and b . If there are three states of the VLMC corresponding to two fading partitions, which are a , ba , and bb ; therefore, $J_{tot} = 3$. Each state is characterized by its context and transition probabilities. Based on [25], the average time that the fade stays in state i can be computed as

$$\eta_i = \frac{1}{P_{\text{sum},i}}$$

where η_i is the average time that the SNR level stays in state i , and $P_{\text{sum},i}$ is the summation of the transition probabilities from state i to other states. Furthermore, to compute the average fade duration that the SNR is lower than γ_j , we can sum up the average time for all states corresponding to the SNR intervals below γ_j . In other words, suppose that Ω_{γ_j} is a set of the VLMC states corresponding to SNR intervals below γ_j . Then, the average fade duration below γ_j is equal to

$$\Lambda_{\gamma_j} = \sum_{\forall i \in \Omega_{\gamma_j}} \eta_i$$

where Λ_{γ_j} is the average fade duration below γ_j . As shown above, the average fade duration computation requires only addition, multiplication, and division operations without involving a complex integration process.

B. Signal Outage Probability

The signal outage probability [39] can be computed by the multiplication of the average fade duration and the average number of fades per second. It gives the probability of the received envelope level being less than γ_j . The average number of fades per second across γ_j can be expressed as

$$\Phi_{\gamma_j} = \sum_{\forall i \in \Theta_{\gamma_j}} P_{\text{cross},i} \times p_i$$

where Θ_{γ_j} is a set of VLMC states corresponding to SNR intervals above γ_j , $P_{\text{cross},i}$ is the summation of transition probabilities from state i to the states whose SNR is below γ_j , and p_i is the state probability of VLMC state i . Thus, the signal outage probability with an SNR lower than γ_j is equal to $\Lambda_{\gamma_j} \Phi_{\gamma_j}$.

VII. SIMULATION RESULTS

In this section, we present a sequence of computer simulation results to demonstrate the main concepts discussed in previous sections.

A. Channel Discretization Based on the Fading SNR Distribution

The proposed fading partition scheme will be justified in this subsection. First, we study the fading SNR distribution estimation using the proposed method as described in Section III. It will help us to determine a suitable number of fading samples used for channel modeling. The closeness

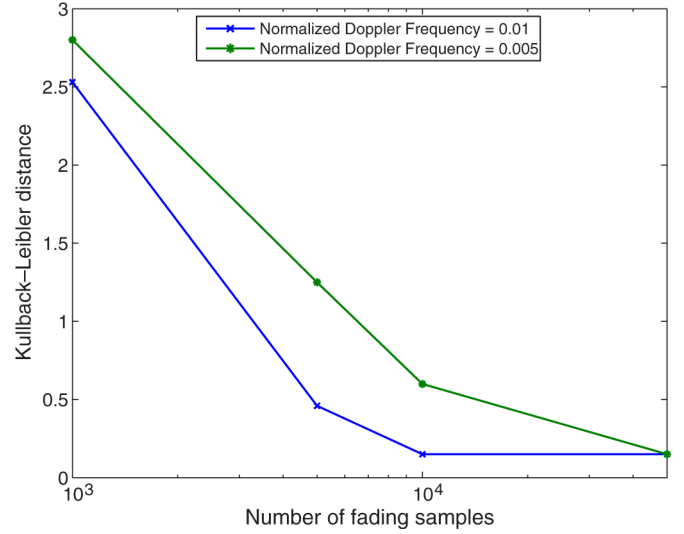


Fig. 6. K–L distance between the Rayleigh fading channel and the estimate fading SNR distribution with different number of fading samples.

between the estimated and the actual fading SNR distributions can be objectively measured using the following K–L distance:

$$D(\hat{f}(\gamma) \| f(\gamma)) = \int_{-\infty}^{\infty} \hat{f}(\gamma) \ln \frac{\hat{f}(\gamma)}{f(\gamma)} d\gamma$$

where $\hat{f}(\gamma)$ and $f(\gamma)$ denote the estimated and the actual fading SNR distributions, respectively.

Fig. 6 shows the effect of the number of fading SNR samples used on the K–L distance in an environment with a Rayleigh fading channel and a normalized Doppler frequency equal to 0.01 and 0.005. The more the number of fading samples used, the smaller the K–L distance. Moreover, in the fading environment with a larger normalized Doppler frequency, fewer samples are required to obtain the same K–L distance. In other words, to obtain an accurate estimation of the fading SNR distribution, a fading environment with a larger normalized Doppler frequency requires fewer fading samples than a fading environment with a lower normalized Doppler frequency.

Next, we explore the performance of our proposed fading partition algorithm as in Section III. Here, we use the probability of bit errors, as described in Section III, as our criterion. First, we study the fading partition scheme when λ_p in (5) is set to zero. In other words, the cost function that is given in (2) is ignored. The number of intervals in a fading partition is set at four, 10, and 16. The transmission policy in each interval is fixed, which is chosen to be the BPSK modulation. Fifty thousand samples of the Rayleigh fading gain with a normalized Doppler frequency equal to 0.01 are generated in our simulation. The Newton–Raphson method [33] is used to solve (7). To show the channel modeling capability via the fading partition, we compare the probability of bit errors of the Rayleigh fading channel and that obtained from the fading partition model in Fig. 7(a), where the effect on the interval number is shown. When the interval number increases, the performance curves of the fading partition are closer to those of the Rayleigh fading. Fig. 7(b) shows the performance comparison of the proposed

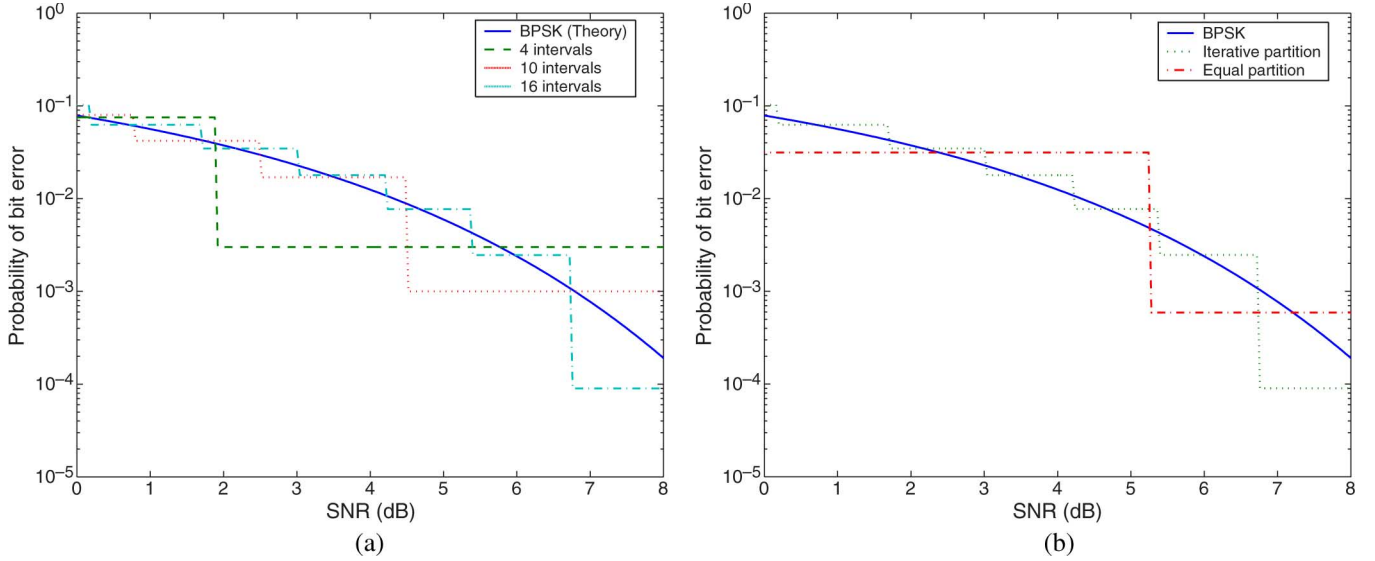


Fig. 7. Comparison of probabilities of bit errors. (a) Rayleigh fading channel and the proposed iterative fading partition mechanism with four, 10, and 16 intervals. (b) Rayleigh fading channel, the proposed iterative fading partition mechanism with 16 intervals, and the equal partition scheme.

TABLE I
RESULTS OF THE OPTIMAL FADING PARTITION, WHERE $e_1\{\pi_1\}$ AND $e_2\{\pi_2\}$ ARE THE PROBABILITIES OF BIT ERRORS IN INTERVALS $[\alpha_0, \alpha_1)$ AND $[\alpha_1, \alpha_2)$, RESPECTIVELY

| α_1^* (dB) | $C_a(\Delta^*)$ | $C_b(\Delta^*)$ | $e_1\{\pi_1\}$ | $e_2\{\pi_2\}$ |
|-----------------------------|-----------------|-----------------|----------------|----------------|
| -3.0456 ($\lambda_p = 0$) | 0.0024 | 0.1869 | 0.1909 | 0.0922 |
| -1.4367 | 0.0026 | 0.1636 | 0.1697 | 0.0789 |
| -1.0253 | 0.0028 | 0.1574 | 0.1640 | 0.0751 |
| -0.5354 | 0.0030 | 0.1498 | 0.1571 | 0.0702 |
| 0.0852 | 0.0033 | 0.1402 | 0.1484 | 0.0635 |

fading partition scheme and the equal-partition scheme. The equal-partition scheme equally divides the whole SNR range except the first and last intervals, in which $\alpha_0 = -66$ dB and $\alpha_U = 32$ dB (i.e., $\alpha_{i+1} - \alpha_i = \alpha_{i+2} - \alpha_{i+1}, \forall i, i \neq 0$ and $i + 2 \neq U$). As we can see, the proposed iterative technique provides a better approximation in terms of the probability of bit errors, particularly in the low SNR ranges, which have a high probability of occurrence in the Rayleigh fading environment.

Last, we consider the complete cost function as given in (5). We consider a fading partition that is characterized by $\alpha_0 < \alpha_1 < \alpha_2$, where $\alpha_0 = -66$ dB, and $\alpha_2 = 32$ dB. Thus, there is only one free parameter, i.e., α_1 . The transmission policy is adaptive modulation with two choices (BPSK and QPSK), and BPSK and QPSK modulation schemes are used in $[\alpha_0, \alpha_1)$ and $[\alpha_1, \alpha_2)$, respectively. Table I shows the results of the optimal fading partition under the Rayleigh fading environment with a normalized Doppler frequency equal to 0.01 and a different α_1 value. As shown in the table, when $\lambda_p = 0$, we get the lowest cost function, which is $C_a(\Delta^*)$. In other words, we do not have to consider the tradeoff between $C_a(\Delta^*)$ and $C_b(\Delta^*)$ but concentrate on $C_a(\Delta^*)$ alone in this simple case. When $\lambda_p = 0$, the fading partition optimization only minimizes the closeness of the average probabilities of bit errors between the actual fading environment and the approximate fading channel representation by the VLMC. When $\lambda_p > 0$, $C_a(\Delta^*)$ and $C_b(\Delta^*)$ are considered at the same time. The optimal solution corresponding to λ_p is the tradeoff between the accuracy in approximating the channel characteristics in terms of the average probabilities of bit

errors and the designed transmission policy. Different λ_p values correspond to different α_1^* values (i.e., with different fading partitions), as shown in the first column of Table I. As α_1^* increases, $e_1\{\pi_1\}$ is decreasing since the channel condition corresponds to the SNR interval from $-\infty$ to α_1^* , which includes channel conditions of a higher SNR (i.e., a good channel condition). As α_1^* increases, $e_2\{\pi_2\}$ is also decreasing since the channel condition corresponds to the SNR interval from α_1^* to ∞ , which excludes channel conditions of a lower SNR (i.e., a bad channel condition). Thus, we see the decrease in the average bit error probabilities from the top to the bottom of Table I.

B. Channel Approximation With Discretized Fading SNR Samples

In the last subsection, we studied how to approximate a fading SNR distribution and the evaluation of the fading partition. Here, we study the channel approximation of a fading channel with discretized fading SNR samples. The environment we consider is in the presence of Rayleigh fading or lognormal fading. The Rayleigh fading and the lognormal fading are often used to model the short-term and long-term fading processes, respectively [39]. Fifty thousand samples of each fading channel are generated in the simulation.

For the lognormal fading, the envelope level that is experienced at location i can be expressed as [39]

$$\gamma_i = \rho \times \gamma_{i-1} + (1 - \rho) \times n_i \quad (21)$$

where γ_i and γ_{i-1} are the envelope levels in decibels at positions i and $(i - 1)$, respectively ρ is the correlation between the envelope levels, and n_i is a Gaussian random variable with zero mean and standard deviation σ . For the typical suburban propagation at 900 MHz, it has been experimentally verified by Gudmundson [18] that $\sigma = 7.5$ dB and $\rho = 0.82$ at a distance of 100 m (i.e., two envelope levels, which separate with a distance equal to 100 m, have a correlation equal to 0.82). For the typical urban propagation at 1700 MHz, Gudmundson [18] has

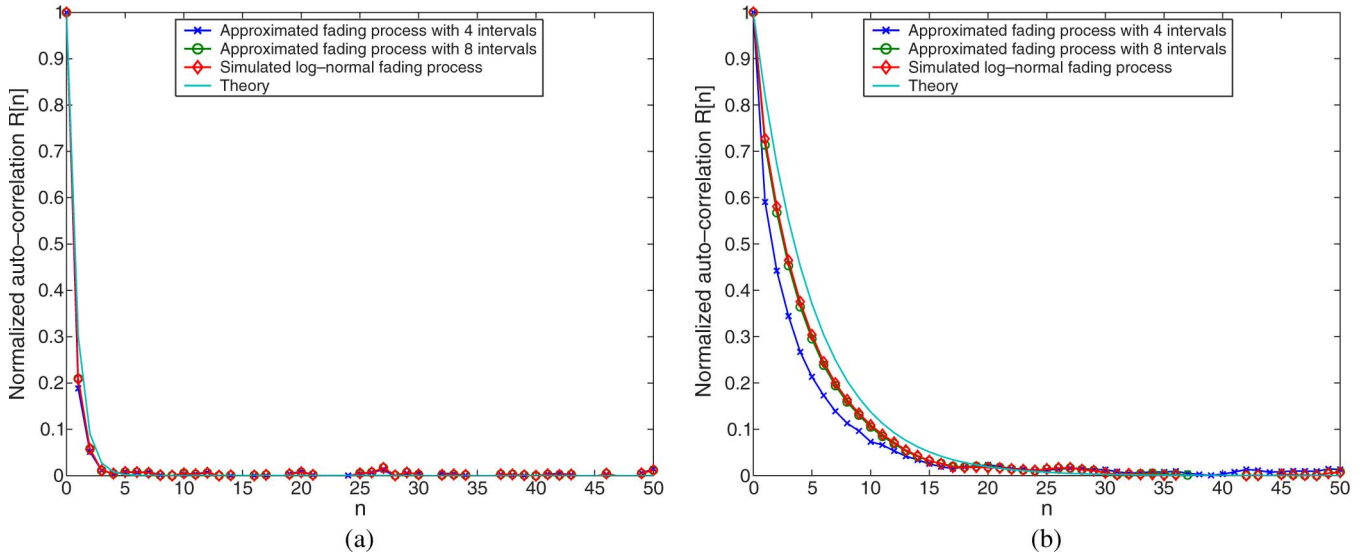


Fig. 8. Normalized autocorrelation comparison between the lognormal fading process and the discretized fading process with a different number of intervals. (a) $\rho = 0.3$ and $\sigma = 4.3$ dB. (b) $\rho = 0.82$ and $\sigma = 7.5$ dB.

also experimentally verified that $\sigma = 4.3$ dB and $\rho = 0.3$ at a distance of 10 m (i.e., two mean average levels, which separate with a distance equal to 10 m, have a correlation equal to 0.3). The normalized autocorrelation of the envelope level from the lognormal fading SNR can be expressed as [18], [39]

$$R_s(i) = \rho^i. \quad (22)$$

The Rayleigh fading process is generated with different normalized Doppler frequencies. We assume that the angle of arrival of the multipath signal is uniform. The fading conditions, which correspond to normalized Doppler frequencies 0.01 and 0.001, respectively, are considered. The autocovariance coefficient of the fading power under the Rayleigh fading environment with the uniform angle of arrival can be expressed as [39]

$$\mu(i) = J_0^2(2\pi f_D i) \quad (23)$$

where $\mu[i]$ is the autocovariance coefficient of the fading power under the Rayleigh fading channel, and $J_0(2\pi f_D i)$ is the zero-order Bessel function of the first kind with normalized Doppler frequency f_D .

We partition the possible range of a fading SNR into two, four, and eight intervals with the method described in Section III. The condition of a discretized fading channel is represented by the mean value of the SNR of each fading interval. The mean value of the SNR in the i th interval can be expressed as

$$\gamma_{i,\text{avg}} = \frac{\int_{\alpha_{i-1}}^{\alpha_i} \gamma \hat{f}(\gamma) d\gamma}{\int_{\alpha_{i-1}}^{\alpha_i} \hat{f}(\gamma) d\gamma}, \quad i = 1, 2, \dots, U. \quad (24)$$

For example, if the fading SNR is located in the interval $[\alpha_{i-1}, \alpha_i]$, we will represent it with $\gamma_{i,\text{avg}}$.

Fig. 8 compares the normalized autocorrelation of the lognormal fading process that is obtained from the discretized fading process, the simulated lognormal process, and the correlation function that is computed from (22). Again, when the fading

partition has more intervals, the normalized autocorrelation coefficient is closer to the theoretical value. From the simulation results, it is sufficient to have eight intervals in the fading channel discretization to approximate the statistical property of the true fading channel.

Fig. 9 compares the autocovariance coefficient [32] of the fading power that is obtained from the discretized fading, the simulated Rayleigh fading channel, and the correlation that is computed from (23). When the number of intervals increases, the autocovariance coefficient of the fading channel approaches the simulated Rayleigh fading channel in fading with $f_D = 0.001$ in Fig. 9(a) and fading with $f_D = 0.01$ in Fig. 9(b). The partition with eight intervals produces a satisfactory result.

As shown in Figs. 8 and 9, we see that the statistical properties of the quantized fading process with eight partition intervals are close to the theoretical results in the Rayleigh and lognormal fading channels. Therefore, VLMC channel modeling with eight partition intervals can provide a good approximation of the actual fading channel.

C. VLMC Channel Modeling

In this subsection, we evaluate the fading channel modeling with the optimal VLMC. We also examine the performance of VLMC channel modeling based on the fast algorithm described in Section IV-D. The characteristics of a fading channel, including the normalized autocorrelation, the autocovariance coefficient, and the derived parameters such as the average fade duration and the signal outage probability, are used in the evaluation. Furthermore, we investigate the complexity of VLMC channel modeling by comparing Wang's model [44], the higher order Markov chain, and the proposed VLMC model. Recall that a higher order Markov chain is a Markov chain whose states have the same order. For example, for a second-order Markov chain, all its states have an order of two. In contrast, different states in the VLMC may have a different number of orders.

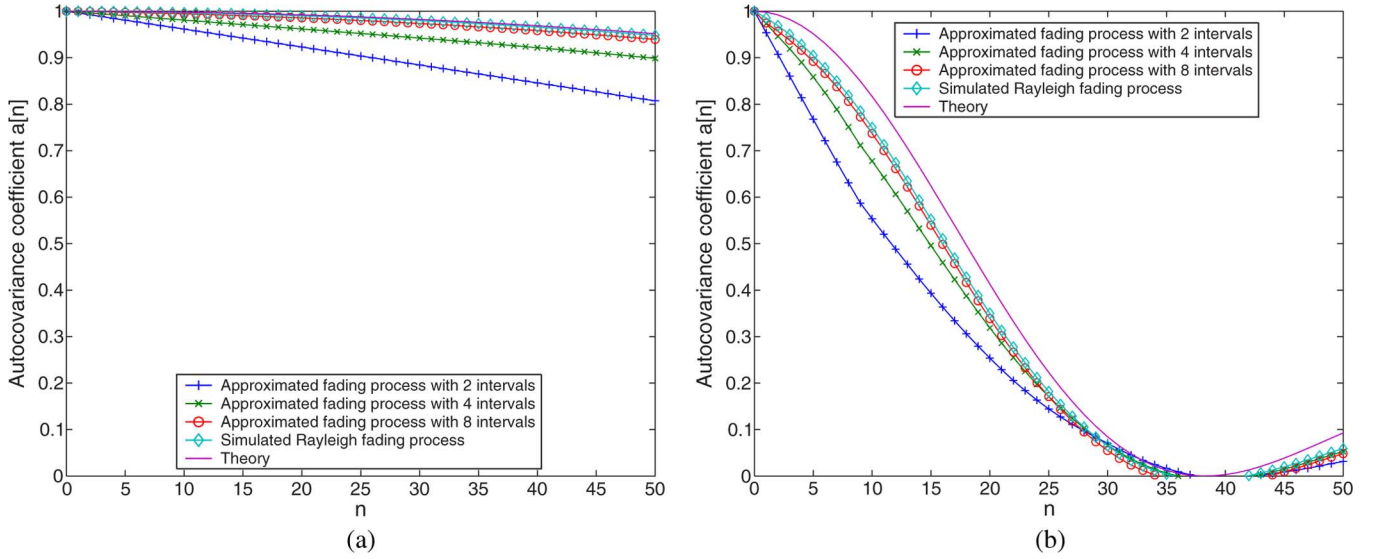


Fig. 9. Autocovariance coefficient comparison between the Rayleigh fading channel and the discretized fading process with two, four, and eight intervals for (a) fading with the normalized Doppler frequency 0.001 and (b) fading with the normalized Doppler frequency 0.01.

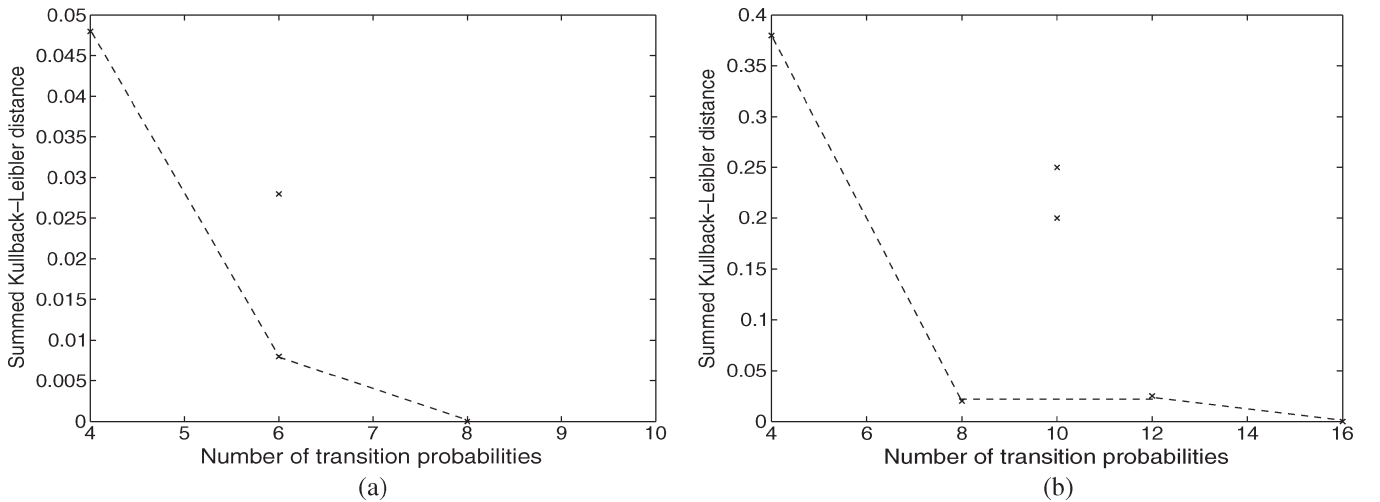


Fig. 10. Summed K-L distance characteristic curves of the lognormal fading process with two intervals in (a) urban environment with $\rho = 0.3$ and $\sigma = 4.3$ dB, and (b) suburban environment with $\rho = 0.82$ and $\sigma = 7.5$ dB.

TABLE II
RESULTS OF THE OPTIMAL VLMC FADING MODELING OF THE LOGNORMAL FADING CHANNEL IN A SUBURBAN ENVIRONMENT WITH TWO FADING PARTITIONS

| Maximum VLMC order | Parameters of complete higher order Markov model | Parameters of VLMC | Average VLMC order |
|-----------------------|--|--------------------|--------------------|
| 1 (Wang's Model [44]) | 4 | 4 | 1 |
| 2 | 12 | 8 | 1.5 |
| 3 | 28 | 10 | 1.8 |
| 4 | 60 | 16 | 2.5 |
| 5 | 124 | 37 | 3.2 |

TABLE III
RESULTS OF THE OPTIMAL VLMC FADING MODELING OF THE LOGNORMAL FADING CHANNEL IN A SUBURBAN ENVIRONMENT WITH EIGHT FADING PARTITIONS

| Maximum VLMC order | Parameters of complete higher order Markov model | Parameters of VLMC | Average VLMC order |
|-----------------------|--|--------------------|--------------------|
| 1 (Wang's Model [44]) | 44 | 44 | 1 |
| 2 | 224 | 169 | 1.81 |
| 3 | 867 | 270 | 2.65 |
| 4 | 2742 | 960 | 3.72 |
| 5 | 7207 | 3769 | 4.64 |

First, the summed K-L distance characteristic curve of the lognormal fading channel is shown in Fig. 10 with the number of intervals set to two. Fig. 10(a) and (b) corresponds to the lognormal fading channel in an urban and a suburban environment, respectively. We see from this figure that the

lognormal fading in an urban environment requires fewer model parameters than that in a suburban environment. Note that the out-of-curve points are obtained by suboptimal VLMC channel models that have the same number of transition probabilities as the optimal VLMC that is located on the summed K-L distance characteristic curve.

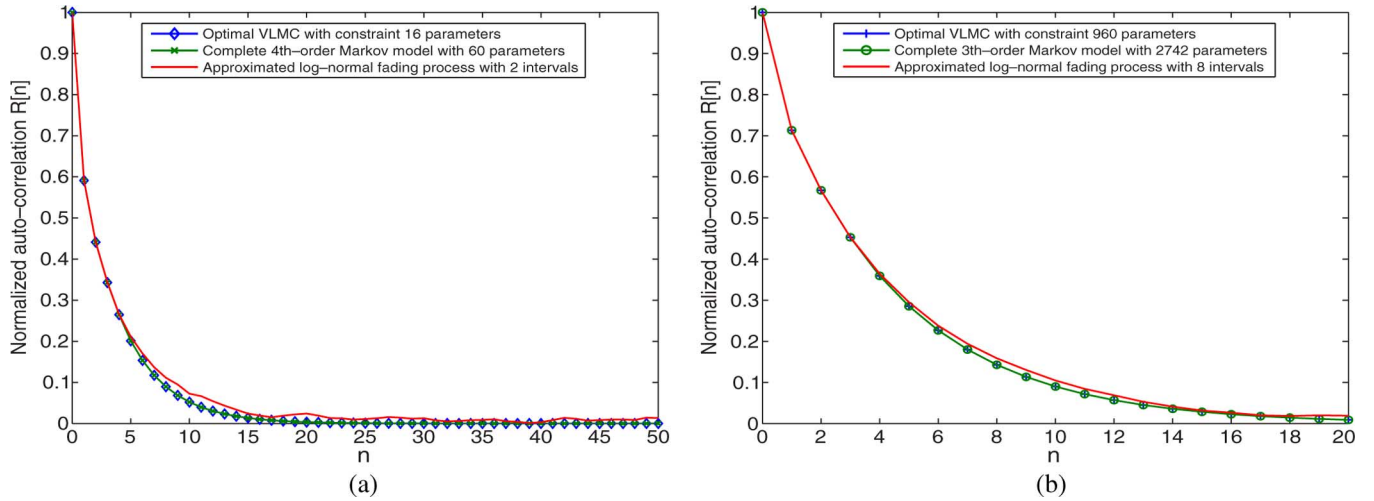


Fig. 11. Statistical property comparison between the optimal VLMC and the complete higher order Markov model of the lognormal fading channel in a suburban environment. (a) Two fading partitions. (b) Eight fading partitions.

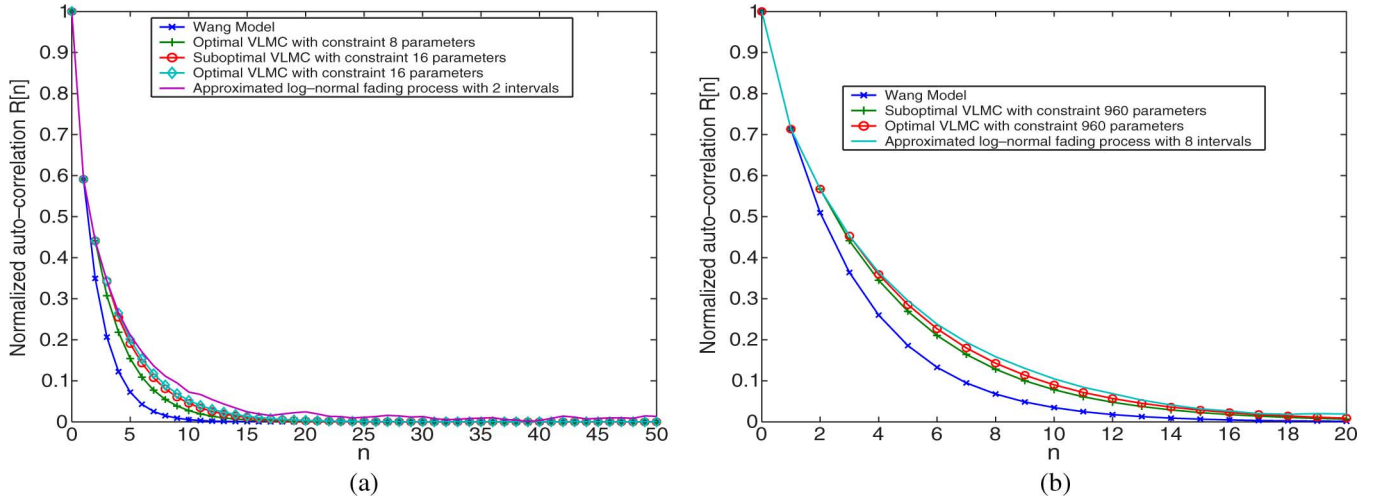


Fig. 12. Comparison of the normalized autocorrelation of several channel models for the lognormal fading process in a suburban environment with $\rho = 0.82$ and $\sigma = 7.5$ dB. (a) Two intervals. (b) Eight intervals.

Tables II and III show the results of the optimal VLMC in the modeling of the lognormal fading channel in a suburban environment with the number of fading partitions equal to two and eight, respectively. These two tables justify the complexity of the VLMC structure when compared with the higher order Markov chain. We show the maximum order and the average order of the optimal VLMC by varying the number of transition probabilities. The average order is computed by averaging the order of all states of the VLMC. We see that the proposed VLMC model greatly reduces the number of transition probabilities compared to the higher order Markov chain. It implies that the optimal VLMC structure is less complex than the full Markov chain. When the maximum order is one, the proposed VLMC model is comparable with Wang's model. However, since the order of Wang's model is only equal to one, its channel modeling capability is severely limited.

It is worthwhile to point out that these experiments are conducted by approximately fixing the statistical properties of the optimal VLMC fading model and the complete higher order Markov model. Fig. 11 compares the normalized autocorrelation

between the optimal VLMC model and the complete higher order Markov model of the same maximum order. As shown in the figure, the normalized autocorrelations of the optimal VLMC model and the complete higher order Markov model are very close to each other.

Figs. 12 and 13 show the performance of the optimal VLMC model in approximating the normalized autocorrelation of discretized fading processes with two and eight intervals in a suburban and an urban environment, respectively. The performance metric is the normalized autocorrelation. We compare the VLMC results with the popular Wang model [44], [45] and the suboptimal VLMC under the same number of transition probabilities (or the same model parameters), where the suboptimal solution is obtained from the fast-fading channel modeling algorithm given in Section IV-D. We see from these figures that the optimal VLMC model with a larger number of transition probabilities approximates the discretized fading process better than that with a smaller number of transition probabilities. This can be seen from the closeness of the normalized autocorrelation. Moreover, the performance of the

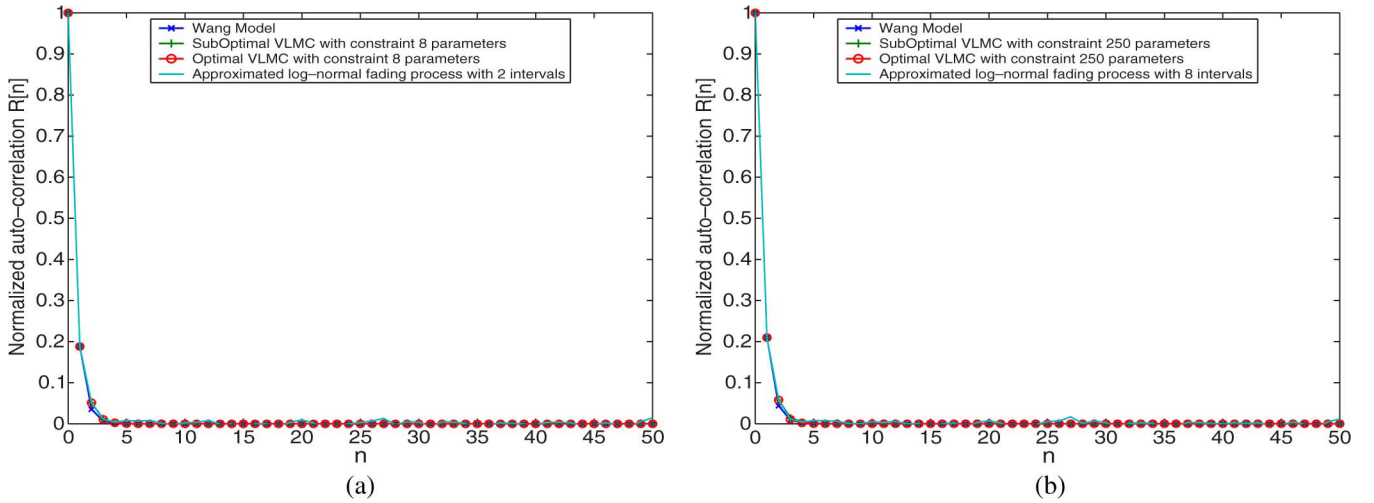


Fig. 13. Comparison of the normalized autocorrelation of several channel models for the lognormal fading process in an urban environment with $\rho = 0.3$ and $\sigma = 4.3$ dB. (a) Two intervals. (b) Eight intervals.

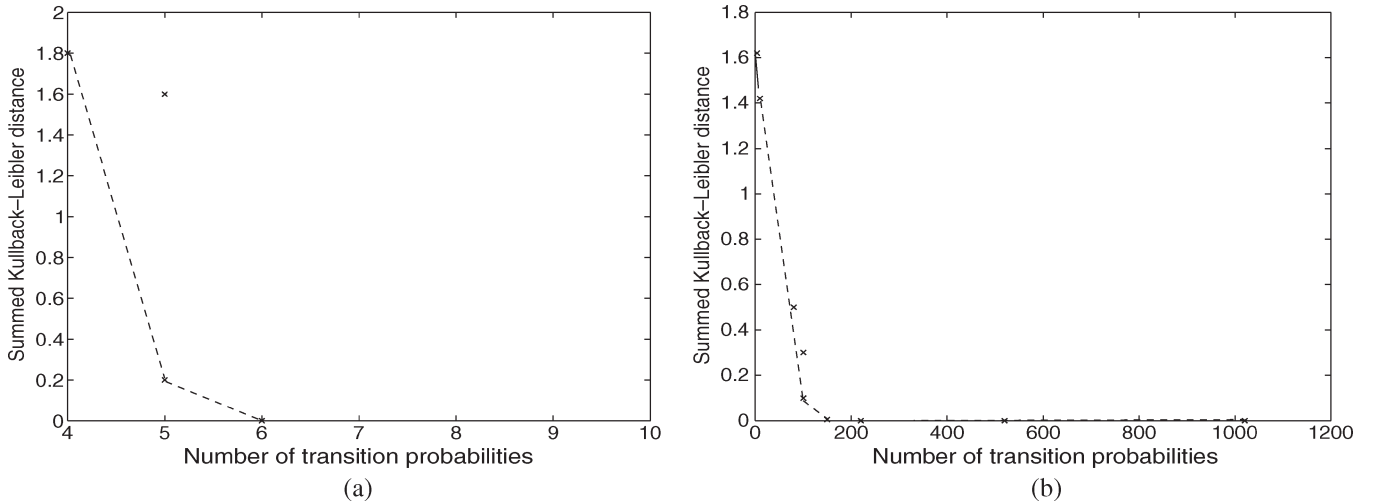


Fig. 14. Summed K-L distance characteristic curve of the Rayleigh fading process with two partitions. (a) Fading environment with the normalized Doppler frequency equal to 0.001. (b) Fading environment with the normalized Doppler frequency equal to 0.01.

TABLE IV
RESULTS OF THE OPTIMAL VLMC FADING MODELING OF THE RAYLEIGH FADING CHANNEL WITH NORMALIZED DOPPLER FREQUENCY EQUAL TO 0.01 AND TWO FADING PARTITIONS

| Maximum VLMC order | Parameters of complete higher order Markov model | Parameters of VLMC | Average VLMC order |
|-----------------------|--|--------------------|--------------------|
| 1 (Wang's Model [44]) | 4 | 4 | 1 |
| 2 | 10 | 8 | 1.6 |
| 3 | 19 | 13 | 2.125 |
| 4 | 31 | 16 | 2.636 |
| 5 | 47 | 22 | 3.14 |

TABLE V
RESULTS OF THE OPTIMAL VLMC FADING MODELING OF THE RAYLEIGH FADING CHANNEL WITH NORMALIZED DOPPLER FREQUENCY EQUAL TO 0.01 AND EIGHT FADING PARTITIONS

| Maximum VLMC order | Parameters of complete higher order Markov model | Parameters of VLMC | Average VLMC order |
|-----------------------|--|--------------------|--------------------|
| 1 (Wang's Model [44]) | 22 | 22 | 1 |
| 2 | 61 | 48 | 1.68 |
| 3 | 126 | 81 | 2.23 |
| 4 | 231 | 115 | 2.636 |
| 5 | 395 | 164 | 3.2 |

optimal VLMC model is significantly better than Wang's model in the suburban environment, as shown in Fig. 12. The performance of the fast-fading channel modeling is close to that of the optimal VLMC. However, the VLMC and Wang's model provide almost the same performance in the urban environment, as shown in Fig. 13. In other words, increasing the VLMC order for the lognormal fading channel modeling in an urban environ-

ment does not help in improving the channel modeling accuracy. Thus, the choice of the modeling VLMC structure depends on the fading environment and the desired modeling accuracy.

Similarly, we conduct the simulation for the Rayleigh fading environment. The autocovariance coefficient is used as the performance metric. The summed K-L distance characteristic curves of the Rayleigh fading channel under

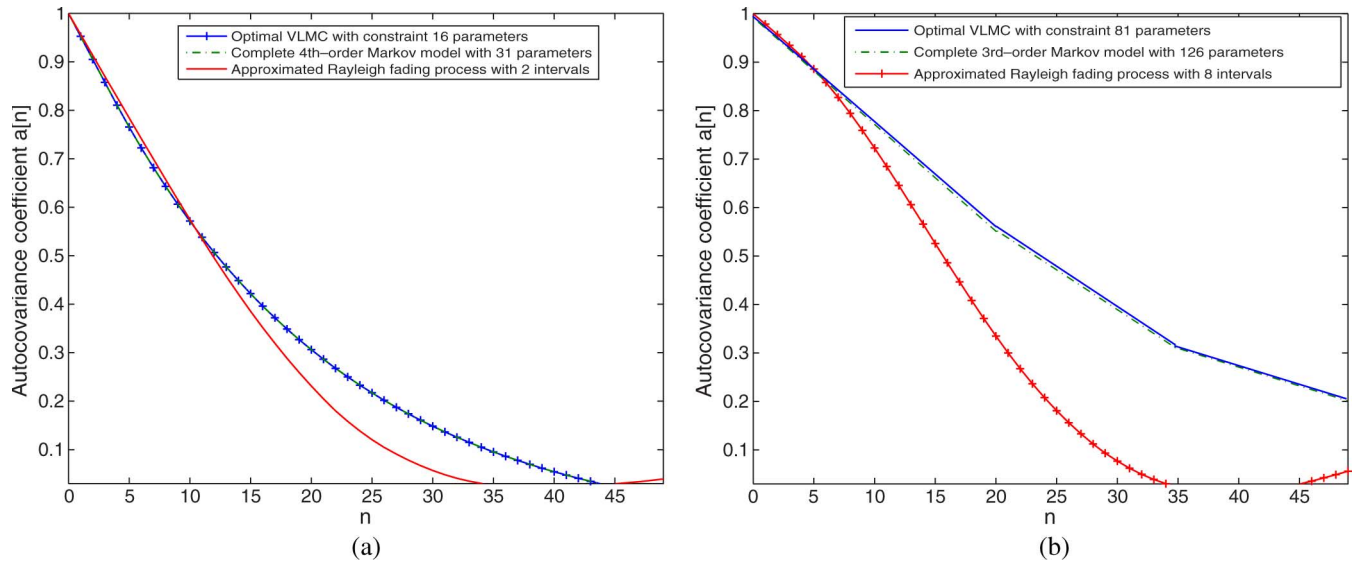


Fig. 15. Statistical property comparison between the optimal VLMC and the complete higher order Markov model of the Rayleigh fading channel with normalized Doppler frequency equal to 0.01. (a) Two fading partitions. (b) Eight fading partitions.

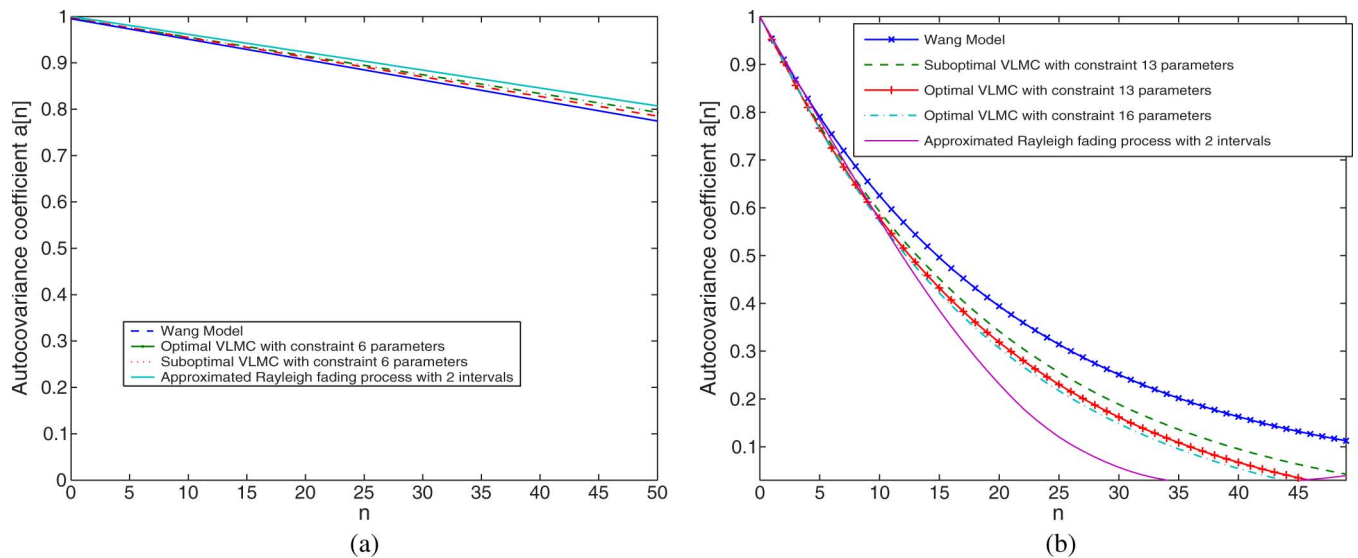


Fig. 16. Comparison of the autocovariance performance of several VLMC channel models for a discretized Rayleigh fading process with two intervals and normalized Doppler frequency equal to (a) 0.001 and (b) 0.01.

different fading environments are shown in Fig. 14(a) and (b), where the normalized Doppler frequencies are 0.001 and 0.01, respectively. We see that the Rayleigh fading channel under fading channel conditions with low dynamic change (i.e., a lower normalized Doppler frequency) requires fewer transition probabilities than that in fading channel conditions with high dynamic change. In other words, a small number of transition probabilities are sufficient to model the Rayleigh fading with low dynamic change. Tables IV and V show the results of the optimal VLMC in modeling the Rayleigh fading channel with a normalized Doppler frequency equal to 0.01 and number of fading partitions equal to two and eight, respectively. The complexity of the VLMC is justified by comparing the number of transition probabilities with that of the complete higher order Markov model. Results show that our VLMC greatly reduces the number of transition probabilities compared to the higher

order Markov model. When compared with Wang's model, the VLMC requires more parameters; however, it provides much better approximation performance, as discussed below.

These experiments are conducted by approximately fixing the statistical properties of the optimal VLMC fading model and the complete higher order Markov model. Fig. 15 compares the autocovariance coefficients between the optimal VLMC model and the complete higher order Markov model of the same maximum order. As shown in the figure, the autocovariance coefficients of the optimal VLMC and the complete higher order Markov model are very close to each other.

We consider a discretized Rayleigh fading channel with two and eight intervals. Figs. 16 and 17 show how well the optimal VLMC can model a discretized fading channel. The optimal VLMC with a larger number of transition probabilities models the discretized fading channel better than that with

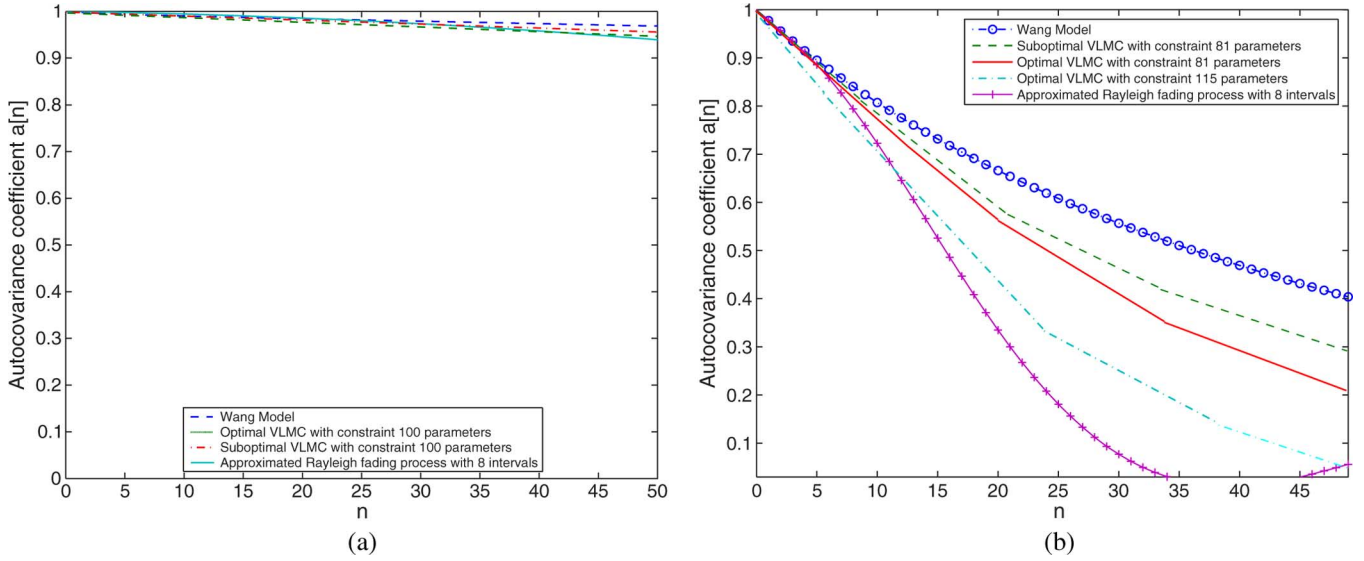


Fig. 17. Comparison of the autocovariance performance of several VLMC channel models for a discretized Rayleigh fading process with eight intervals and normalized Doppler frequency equal to (a) 0.001 and (b) 0.01.

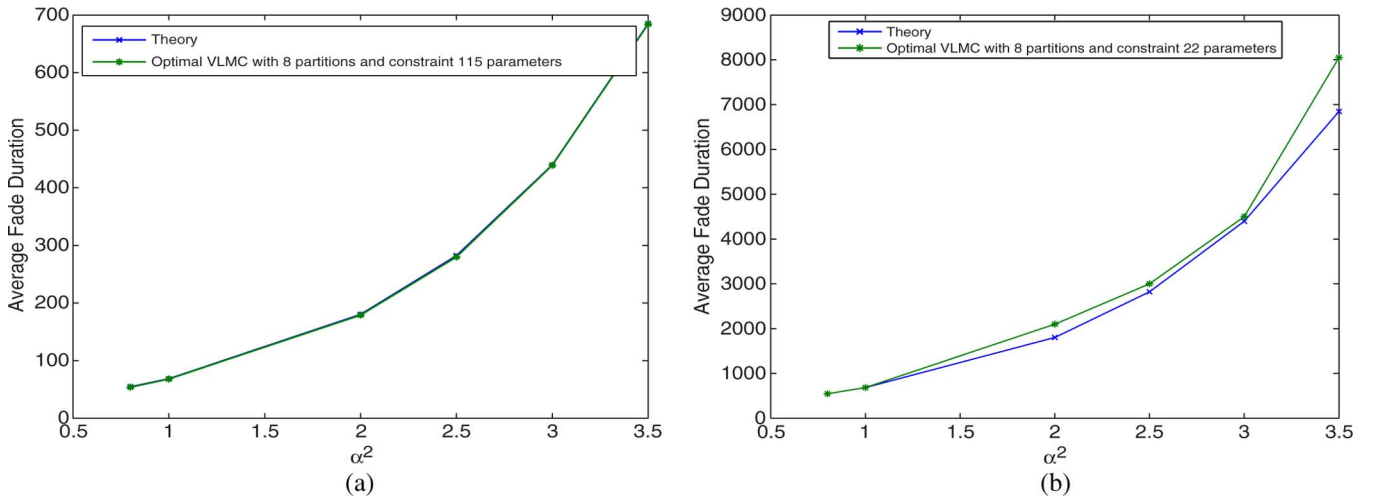


Fig. 18. Comparison of the average fade duration for the Rayleigh fading using the optimal VLMC channel model and theoretical analysis with normalized Doppler frequency equal to (a) 0.01 and (b) 0.001.

a smaller number of transition probabilities. This can be seen by the closeness of the autocovariance coefficient. As compared with Wang's model, the optimal VLMC offers a better fading channel model. This is particularly true in a fading environment with high dynamic change. However, in a fading environment with low dynamic change, the performances of different methods in modeling the discretized fading channel are almost the same. Note that the suboptimal VLMC models in Figs. 16 and 17 are based on the fast-fading modeling algorithm. Although the accuracy of the statistical properties of the suboptimal VLMC is not as good as that of the optimal VLMC, their performances are somehow close. If the channel modeling complexity is a major concern, the tradeoff between the complexity and the performance can be considered.

D. Application of the VLMC Model

In this section, we consider the application of the VLMC model to the computation of the average fade duration and the

signal outage probability, which are important parameters in the wireless communication system design. We use the Rayleigh fading channel as an example since there exist closed-form solutions for the average fade duration and the signal outage probability so that the correctness of VLMC channel modeling can be verified. In Figs. 18 and 19, we compare the average fade duration and the signal outage probability that are obtained from the optimal VLMC and the theoretical analysis. The average fade duration is measured in terms of the number of symbols. The average fade duration under α can be computed as [39]

$$\bar{t} = \frac{e^{\alpha^2} - 1}{\alpha \times f_D \times \sqrt{2\pi}}$$

where f_D is the normalized Doppler frequency. The parameter γ is the ratio between the fading envelope level and the root

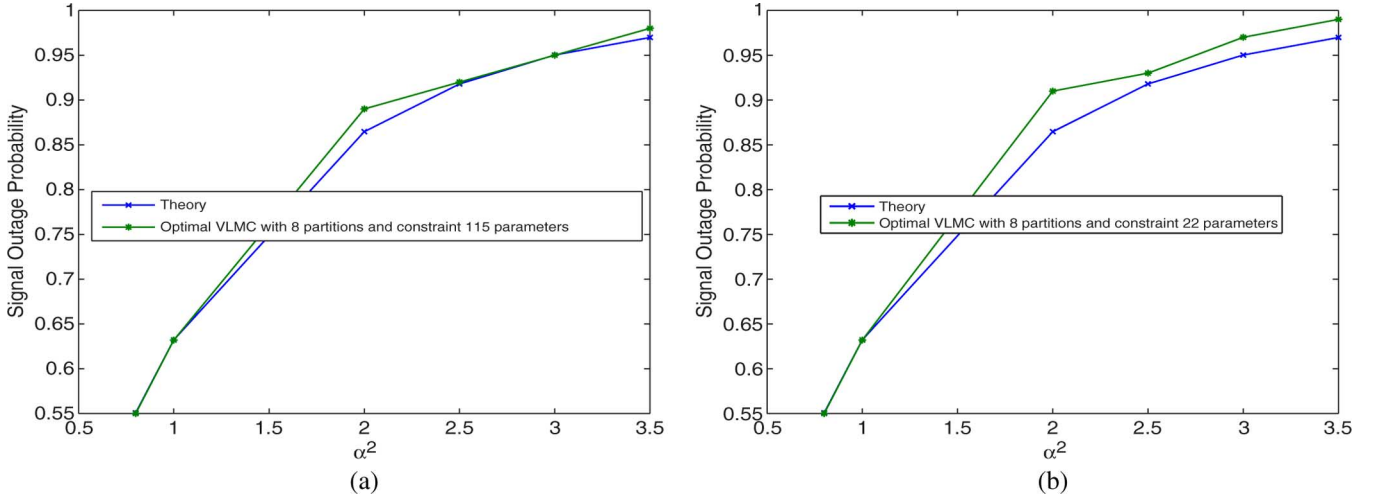


Fig. 19. Comparison of the signal outage probability for the Rayleigh fading using the optimal VLMC channel model and theoretical analysis with normalized Doppler frequency equal to (a) 0.01 and (b) 0.001.

mean square envelope level, which can be expressed as

$$\alpha = \frac{\gamma}{\gamma_{\text{rms}}} \quad (25)$$

where γ is the fading envelope level, $\gamma_{\text{rms}} = \sqrt{\gamma_{\text{avg}}^2}$ is the root mean square envelope signal level, and γ_{avg}^2 is the average power of a fading channel, which can be defined as

$$\gamma_{\text{avg}}^2 = \int_{-\infty}^{\infty} \vartheta^2 \times f(\vartheta) d\vartheta \quad (26)$$

where ϑ is the random variable of the fading envelope signal, and $f(\vartheta)$ is the pdf of the fading envelope signal.

The signal outage probability is the probability of the received envelope level being less than γ , which can be expressed as [39]

$$P\{z \leq \gamma\} = 1 - e^{-\alpha^2} \quad (27)$$

where z is the received envelope level, and α is defined in (25).

We compute the fade duration for every envelope level corresponding to a fading partition. The optimal VLMC channel model is derived based on eight intervals. The average power of the transmitted signal is set as $\gamma_{\text{avg}} = 10$ dB. The numbers of transition probabilities in Figs. 18 and 19(a) and (b) are set to 115 and 22, respectively. They show the average fade duration and the signal outage probability when the Rayleigh fading has high dynamic change (with a normalized Doppler frequency equal to 0.01) and low dynamic change (with a normalized Doppler frequency equal to 0.001). We see that the average fade duration and the signal outage probability of the optimal VLMC are close to the theoretical values in both fading environments with high and low dynamic changes. Thus, the proposed VLMC model can be used to compute the fading parameters.

VIII. CONCLUSION

Optimal fading channel modeling with VLMC was studied in this paper, where the K–L distance was chosen as the optimization criterion. The proposed scheme consists of two main components. One is the fading partition, which takes the transmission policy into account. The other is the derivation of the optimal VLMC that represents a fading channel under the constraint of the number of transition probabilities. The optimal VLMC model was obtained by constructing the summed K–L distance characteristic curve and selecting the optimal VLMC using the Lagrangian optimization. A fast VLMC channel modeling algorithm was proposed to speed up the process of channel modeling. Several properties of the optimal VLMC fading channel were stated and proved. It was shown by simulation that the optimal VLMC provides an excellent fading channel approximation with reasonable complexity. As compared with the higher order Markov model, the complexity of the VLMC model is significantly lower in achieving the same approximation performance. As compared with the popular Wang's model, the VLMC model can provide a broad range of performance–complexity tradeoff. Last, we showed the application of the VLMC channel modeling to the determination of fading parameters with good accuracy, such as the average fade duration and the signal outage probability.

APPENDIX

Proof of Theorem 1: From [11, Th. 2.7.2], we know that $\rho(P\{x_{k+l} = u_0 | \mathbf{x}_{k-1}^{k+l-1} = \omega_l u\} \| P\{x_{k+l} = u_0 | \mathbf{x}_{k-2}^{k+l-1} = \omega_l\})$ is convex in a pair of transition probabilities. Since the summation of convex functions is also a convex function [7], the summed K–L distance of the VLMC terminal nodes is convex in a set of transition probabilities of VLMC terminal nodes. ■

Proof of Theorem 2: We can express mutual information $\mathcal{I}_e(x_{M+1}; \mathbf{x}_1^M)$ and $\mathcal{I}(x_{M+1}; \mathbf{x}_1^M)$ as [11]

$$\mathcal{I}_e(x_{M+1}; \mathbf{x}_1^M) = \mathcal{H}_e(x_{M+1}) - \mathcal{H}_e(x_{M+1} | \mathbf{x}_1^M) \quad (28)$$

$$\mathcal{I}(x_{M+1}; \mathbf{x}_1^M) = \mathcal{H}(x_{M+1}) - \mathcal{H}(x_{M+1} | \mathbf{x}_1^M) \quad (29)$$

where $\mathcal{H}_e(x_{M+1})$ and $\mathcal{H}(x_{M+1})$ are the entropies of x_{M+1} obtained from Υ_e and Υ , respectively. We use $\mathcal{H}_e(x_{M+1}|\mathbf{x}_1^M)$ and $\mathcal{H}(x_{M+1}|\mathbf{x}_1^M)$ to denote conditional entropies of x_{M+1} , given the past M quantized fading samples obtained from Υ_e and Υ , respectively. It is clear that $\mathcal{H}_e(x_{M+1}) = \mathcal{H}(x_{M+1})$ since the number of intervals in a fading partition is still the same. Moreover, we have

$$\mathcal{H}(x_{M+1}|\mathbf{x}_1^M) = \mathcal{H}_e(x_{M+1}|\mathbf{x}_{M-M_e+1}^M) \geq \mathcal{H}_e(x_{M+1}|\mathbf{x}_1^M)$$

by using the fact that the order of Υ is less than Υ_e and the entropy property from [11]. Therefore, we conclude that $\mathcal{I}_e(x_{M+1}; \mathbf{x}_1^M) \geq \mathcal{I}(x_{M+1}; \mathbf{x}_1^M)$, which proves the theorem. ■

Proof of Theorem 3: It is assumed that the VLMC process of consideration is stationary and ergodic. Let χ_K and χ_{K+1} be finite categorical spaces representing all K and $K+1$ partitions, respectively. Let y_i and x_i be quantized fading samples i obtained before and after interval splitting. The mutual information after interval splitting will increase if

$$\begin{aligned} I_{K+1}(x_L; \mathbf{x}_{L-M-1}^{L-1}) - I_K(y_L; \mathbf{y}_{L-M-1}^{L-1}) \\ = \mathcal{H}_{K+1} - \mathcal{H}_K - \mathcal{H}_{K+1}(x_L|\mathbf{x}_{L-M-1}^{L-1}) \\ + \mathcal{H}_K(y_L|\mathbf{y}_{L-M-1}^{L-1}) > 0 \end{aligned} \quad (30)$$

where $I_{K+1}(x_L; \mathbf{x}_{L-M-1}^{L-1})$ and $I_K(y_L; \mathbf{y}_{L-M-1}^{L-1})$ are the mutual information measured after and before interval splitting, respectively. Since this process is stationary, we have

$$\begin{aligned} I_{K+1}(x_L; \mathbf{x}_{L-M-1}^{L-1}) &= I_{K+1}(x_{M+1}; \mathbf{x}_1^M) \\ I_K(y_L; \mathbf{y}_{L-M-1}^{L-1}) &= I_K(y_{M+1}; \mathbf{y}_1^M). \end{aligned}$$

Then, we have

$$\begin{aligned} \mathcal{H}_K(y_L|\mathbf{y}_{L-M-1}^{L-1}) &= \mathcal{H}_K(y_L|\mathbf{y}_1^{L-1}) \\ &- \sum_{y_1, \dots, y_L} P\{y_1; \dots; y_L\} \ln P\{y_L|\mathbf{y}_1^{L-1}\} \end{aligned} \quad (31)$$

$$\begin{aligned} \mathcal{H}_{K+1}(x_L|\mathbf{x}_{L-M-1}^{L-1}) &= \mathcal{H}_{K+1}(x_L|\mathbf{x}_1^{L-1}) \\ &- \sum_{x_1, \dots, x_L} P\{x_1; \dots; x_L\} \ln P\{x_L|\mathbf{x}_1^{L-1}\} \end{aligned} \quad (32)$$

where $P\{y_1; \dots; y_L\}$ and $P\{x_1; \dots; x_L\}$ are joint probabilities of quantized fading samples when there are K and $K+1$ partitioned intervals, respectively. Since

$$\begin{aligned} \mathcal{H}_{K+1} - \mathcal{H}_K p_{i,1} \ln \left(\frac{p_{i,1} + p_{i,2}}{p_{i,1}} \right) \\ + p_{i,2} \ln \left(\frac{p_{i,1} + p_{i,2}}{p_{i,2}} \right) > 0 \end{aligned} \quad (33)$$

(30) will be positive if

$$-\mathcal{H}_{K+1}(x_L|\mathbf{x}_1^{L-1}) + \mathcal{H}_K(y_L|\mathbf{y}_1^{L-1}) < \mathcal{H}_{K+1} - \mathcal{H}_K. \quad (34)$$

From [11], we have

$$\mathcal{H}_K(y_L|\mathbf{y}_1^{L-1}) \leq \frac{\mathcal{H}_K(y_1, \dots, y_L)}{L} \quad (35)$$

where

$$\mathcal{H}_K(y_1, \dots, y_L) = - \sum_{y_1, \dots, y_L} P\{y_1, \dots, y_L\} \times \ln P\{y_1, \dots, y_L\} \quad (36)$$

and where $\mathcal{H}_K(y_1, \dots, y_L)$ is the joint entropy among quantized fading samples y_1, \dots, y_L .

When L is sufficiently large, the joint probability can be expressed as [11]

$$P\{y_1, \dots, y_L\} = e^{-Lh_K} \quad (37)$$

where h_K is the entropy rate of the quantized fading process, which is a Markov process with K partitioned intervals. The entropy rate h_K can be further expressed as [11]

$$h_K = \lim_{L \rightarrow \infty} \mathcal{H}_K(y_1, \dots, y_L) = \mathcal{H}_K(y_L|\mathbf{y}_1^{L-1}) \leq \mathcal{H}_K. \quad (38)$$

Therefore, from (37) and (38), we obtain

$$P\{y_1, \dots, y_L\} \geq e^{-L\mathcal{H}_K}. \quad (39)$$

With (35), (36), and (39), the bound of $\mathcal{H}_K(y_L|\mathbf{y}_1^{L-1})$ can be derived as

$$0 \leq \mathcal{H}_K(y_L|\mathbf{y}_1^{L-1}) \leq \sum_{y_1, \dots, y_L} e^{-L\mathcal{H}_K} \mathcal{H}_K \quad (40)$$

or

$$0 \leq \mathcal{H}_K(y_L|\mathbf{y}_1^{L-1}) \leq K^L e^{-L\mathcal{H}_K} \mathcal{H}_K. \quad (41)$$

Therefore, we get

$$\begin{aligned} -K^L e^{-L\mathcal{H}_K} \mathcal{H}_K &\leq \mathcal{H}_{K+1}(x_L|\mathbf{x}_1^{L-1}) - \mathcal{H}_K(y_L|\mathbf{y}_1^{L-1}) \\ &\leq (K+1)^L e^{-L\mathcal{H}_{K+1}} \mathcal{H}_{K+1}. \end{aligned} \quad (42)$$

With (34) and (42), the information will be positive if $\mathcal{H}_{K+1} - \mathcal{H}_K$ is greater than the upper bound of (34) or

$$\mathcal{H}_{K+1} - \mathcal{H}_K \geq (K+1)^L e^{-L\mathcal{H}_{K+1}} \mathcal{H}_{K+1}. \quad (43)$$

Thus, we conclude that

$$\begin{aligned} p_{i,1} \ln \left(\frac{p_{i,1} + p_{i,2}}{p_{i,1}} \right) + p_{i,2} \ln \left(\frac{p_{i,1} + p_{i,2}}{p_{i,2}} \right) \\ \geq (K+1)^L e^{-L\mathcal{H}_{K+1}} \mathcal{H}_{K+1} \end{aligned} \quad (44)$$

which proves the theorem. ■

Proof of Theorem 4: Here, we follow the same notations as given in Theorem 3. Let y_i and x_i be quantized fading samples i obtained after and before merging the partitioned intervals, respectively. Consider the difference between the mutual information values after and before merging the partitioned intervals, which can be written as

$$\begin{aligned} \mathcal{I}_K(y_{M+1}; \mathbf{y}_1^M) - \mathcal{I}_{K+1}(x_{M+1}; \mathbf{x}_1^M) &= \mathcal{H}_K \\ &- \mathcal{H}_{K+1} - \mathcal{H}_K(y_{M+1}|\mathbf{y}_1^M) + \mathcal{H}_{K+1}(x_{M+1}|\mathbf{x}_1^M). \end{aligned} \quad (45)$$

From the proof of Theorem 3, we get

$$\begin{aligned} -K^L \mathcal{H}_K e^{-L\mathcal{H}_K} &\leq \mathcal{H}_{K+1}(x_{M+1}|\mathbf{x}_1^M) \\ &- \mathcal{H}_K(y_{M+1}|\mathbf{y}_1^M) \leq (K+1)^L \mathcal{H}_{K+1} e^{-L\mathcal{H}_{K+1}}. \end{aligned} \quad (46)$$

Moreover, we have

$$\mathcal{H}_K - \mathcal{H}_{K+1} - p_i \ln \frac{p_{i,i+1}}{p_i} - p_{i+1} \ln \frac{p_{i,i+1}}{p_{i+1}} < 0. \quad (47)$$

From (45)–(47), we conclude that the mutual information will be decreased at most by the following amount:

$$p_i \ln \frac{p_{i,i+1}}{p_i} + p_{i+1} \ln \frac{p_{i,i+1}}{p_{i+1}} + K^L \mathcal{H}_K e^{-L\mathcal{H}_K}$$

which concludes the theorem. \blacksquare

Proof of Theorem 5: For the sake of conciseness, we will shorten the expression during the proof. For example, we will express $\mathcal{H}(x_{M+1}u_0|\mathbf{x}_1^M = \omega_{M-1}u)$ as $\mathcal{H}(u_0|\omega_{M-1}u)$ and $P\{x_{M+1}u_0|\mathbf{x}_1^M = \omega_{M-1}u\}$ as $P\{u_0|\omega_{M-1}u\}$, i.e.,

$$\begin{aligned} & \mathcal{H}(u_0|\omega_{M-1}u) - \mathcal{H}(u_0|\omega_{M-1}) \\ &= \sum_{\forall u_0} \sum_{\forall \omega_{M-1}} \sum_{\forall u} P\{u_0; \omega_{M-1}u\} \ln P\{u_0|\omega_{M-1}u\} \\ & \quad - \sum_{\forall u_0} \sum_{\forall \omega_{M-1}} P\{u_0; \omega_{M-1}\} \ln P\{u_0|\omega_{M-1}\}, \end{aligned} \quad (48)$$

$$\begin{aligned} &= \sum_{\forall u_0} \sum_{\forall \omega_{M-1}} \sum_{\forall u} P\{u_0; \omega_{M-1}u\} \\ & \quad \times \ln \frac{P\{u_0|\omega_{M-1}u\}P\{u_0|\omega_{M-1}\}}{P\{u_0|\omega_{M-1}\}} \\ & \quad - \sum_{\forall u_0} \sum_{\forall \omega_{M-1}} P\{u_0; \omega_{M-1}\} \ln P\{u_0|\omega_{M-1}\}, \end{aligned} \quad (49)$$

$$\begin{aligned} &= D_{\text{avg}}(P\{u_0|\omega_{M-1}u\}||P\{u_0|\omega_{M-1}\}) \\ & \quad + \sum_{\forall u_0} \sum_{\forall \omega_{M-1}} \sum_{\forall u} P\{u_0; \omega_{M-1}u\} \ln P\{u_0|\omega_{M-1}\} \\ & \quad - \sum_{\forall u_0} \sum_{\forall \omega_{M-1}} P\{u_0; \omega_{M-1}\} \ln P\{u_0|\omega_{M-1}\}, \end{aligned} \quad (50)$$

$$\begin{aligned} &= D_{\text{avg}}(P\{u_0|\omega_{M-1}u\}||P\{u_0|\omega_{M-1}\}) \\ & \quad - \mathcal{H}_p(P\{u_0|\omega_{M-1}u\}||P\{u_0|\omega_{M-1}\}) \\ & \quad + \mathcal{H}(u_0|\omega_{M-1}). \end{aligned} \quad (51)$$

Since the mutual information can be written as

$$\mathcal{I}(u_0; \omega_{M-1}) = \mathcal{H}(u_0) - \mathcal{H}(u_0|\omega_{M-1}) \quad (52)$$

we obtain from (51) and (52) that

$$\begin{aligned} & D_{\text{avg}}(P\{u_0|\omega_{M-1}u\}||P\{u_0|\omega_{M-1}\}) \\ &= \mathcal{I}(u_0; \omega_{M-1}) - \mathcal{I}(u_0; \omega_{M-1}u) \\ & \quad + \mathcal{H}_p(P\{u_0|\omega_{M-1}u\}||P\{u_0|\omega_{M-1}\}) \\ & \quad - \mathcal{H}(u_0|\omega_{M-1}) \end{aligned} \quad (53)$$

which concludes the proof. \blacksquare

REFERENCES

- [1] T. Aulin, "A modified model for the fading signal at a mobile radio channel," *IEEE Trans. Veh. Technol.*, vol. VT-28, no. 3, pp. 182–203, Aug. 1979.
- [2] T. Aulin, "Characteristics of a digital mobile radio channel," *IEEE Trans. Veh. Technol.*, vol. VT-30, no. 2, pp. 45–53, May 1980.
- [3] F. Babich and G. Lombardi, "A Markov model for the mobile propagation channel," *IEEE Trans. Veh. Technol.*, vol. 49, no. 1, pp. 63–73, Jan. 2000.
- [4] F. Babich, O. E. Kelly, and G. Lombardi, "A context-tree based model for quantized fading," *IEEE Trans. Commun.*, vol. 3, no. 2, pp. 46–48, Feb. 1999.
- [5] P. Buhlmann and A. J. Wyner, "Variable length Markov chains," *Ann. Stat.*, vol. 27, no. 2, pp. 480–513, Apr. 1999.
- [6] P. Buhlmann, "Model selection for variable length Markov chains and tuning the context algorithm," *Ann. Inst. Stat. Math.*, vol. 52, no. 2, pp. 287–315, Jun. 2000.
- [7] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [8] C. D. Charalambous and N. Menemenlis, "General non-stationary models for short-term and long-term fading channels," in *Proc. EURACOM*, 2000, pp. 142–149.
- [9] L. F. Chang, "Throughput estimation of ARQ protocols for Rayleigh fading channel using fade- and interfade-duration statistics," *IEEE Trans. Veh. Technol.*, vol. 40, no. 1, pp. 223–229, Feb. 1991.
- [10] M. Chu and W. Stark, "Effect of mobile velocity on communications in fading channels," *IEEE Trans. Veh. Technol.*, vol. 49, no. 1, pp. 202–210, Jan. 2000.
- [11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [12] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol. 42, no. 9, pp. 1977–1997, Sep. 1963.
- [13] T. Eyceoz, A. Duel-Hallen, and H. Hallen, "Prediction of fast fading parameters by resolving the interference pattern," in *31st Asilomar Conf. Rec.*, 1997, vol. 1, pp. 167–171.
- [14] B. D. Fritchman, "A binary channel characterization using partitioned Markov chains," *IEEE Trans. Veh. Technol.*, vol. VT-13, no. 2, pp. 221–227, Apr. 1967.
- [15] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell Syst. Tech. J.*, vol. 39, pp. 1253–1266, 1960.
- [16] A. Goldsmith and S. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218–1230, Oct. 1997.
- [17] R. M. Gray, *Entropy and Information Theory*. New York: Springer-Verlag, 1990.
- [18] M. Gudmundson, "Correlation model for shadowing fading in mobile radio systems," *Electron. Lett.*, vol. 27, no. 23, pp. 2145–2146, 1999.
- [19] S. Hara, A. Ogino, M. Araki, M. Okada, and N. Morinaga, "Throughput performance of SAW-ARQ protocol with adaptive packet length in mobile packet data transmission," *IEEE Trans. Veh. Technol.*, vol. 45, no. 3, pp. 561–569, Aug. 1996.
- [20] R. A. Howard, *Dynamic Probabilistic Systems*. New York: Wiley, 1971.
- [21] C.-Y. Hsu, A. Ortega, and M. Khansari, "Rate control for robust video transmission over burst-error wireless channels," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 5, pp. 756–773, May 1999.
- [22] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principle and Applications to Speech and Videos*. New York: Prentice-Hall, 1984.
- [23] S. Kallel, "Analysis of memory and incremental redundancy ARQ schemes over a nonstationary channel," *IEEE Trans. Commun.*, vol. 40, no. 9, pp. 1474–1480, Sep. 1992.
- [24] A. Konrad, B. Y. Zhao, A. Joseph, and R. Ludwig, "A Markov-based channel model algorithm for wireless networks," in *Proc. 4th ACM Int. Workshop MSWiM*, Jul. 2001, pp. 28–36.
- [25] L. Kleninrok, *Queueing Systems*, vol. I. New York: Wiley, 1975.
- [26] W. Kumwilaisak, Y. T. Hou, Q. Zhang, W. Zhu, C.-C. J. Kuo, and Y.-Q. Zhang, "A cross-layer quality-of-service mapping architecture for video delivery in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 10, pp. 1685–1699, Dec. 2003.
- [27] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Stat.*, vol. 22, no. 1, pp. 79–86, Mar. 1951.
- [28] Y. Li and X. Huang, "The simulation of independent Rayleigh faders," *IEEE Trans. Commun.*, vol. 50, no. 9, pp. 1503–1514, Sep. 2002.
- [29] C. H. C. Leung, Y. Kikumoto, and S. A. Sorensen, "The throughput efficiency of the go-back-N ARQ scheme under Markov and related error structures," *IEEE Trans. Commun.*, vol. 36, no. 2, pp. 881–887, Feb. 1998.
- [30] M. J. Mason, G. C. Hess, and S. S. Gilbert, "Shadowing variability in an urban land mobile environment at 900 MHz," *Electron. Lett.*, vol. 26, no. 10, pp. 646–648, May 1990.
- [31] A. Ortega, "Optimization techniques for adaptive quantization of image and video under delay constraints," Ph.D. dissertation, Dept. Elect. Eng., Columbia Univ., New York, 1994.
- [32] J. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.
- [33] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1999.

- [34] 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, 1996, pp. 282–286.
- [35] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL: Univ. Illinois Press, 1949.
- [36] C. K. Siew and D. J. Goodman, "Packet data transmission over mobile radio channels," *IEEE Trans. Veh. Technol.*, vol. 38, no. 2, pp. 95–101, May 1989.
- [37] S. Sivaprakasam and K. S. Shanmugan, "An equivalent Markov model for burst errors in digital channels," *IEEE Trans. Commun.*, vol. 43, no. 2–4, pp. 1347–1355, Feb.–Apr. 1995.
- [38] C. Sgraja and J. Lindner, "Pilot/data-aided estimation of Rayleigh flat-fading channels," in *Proc. 7th Int. OFDM-Workshop*, 2002.
- [39] G. L. Stuber, *Principles of Mobile Communication*. Boston, MA: Kluwer, 1996.
- [40] C. Tan and N. Beaulieu, "On first-order Markov modeling for the Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2032–2040, Dec. 2000.
- [41] W. Turin, *Digital Transmission Systems: Performance Analysis and Modeling*. New York: McGraw-Hill, 1998.
- [42] W. Turin and R. van Nobelen, "Hidden Markov modeling of flat fading channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 9, pp. 1809–1817, Dec. 1998.
- [43] W. Turin, "Throughput analysis of the go-back-N protocol in fading radio channels," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 5, pp. 881–887, May 1999.
- [44] H. S. Wang and N. Moayeri, "Finite-state Markov channel: A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 45, no. 1, pp. 163–171, Feb. 1995.
- [45] 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, no. 2, pp. 163–171, May 1995.
- [46] M. P. Wand and M. C. Jones, *Kernel Smoothing*. London, U.K.: Chapman & Hall, 1995.
- [47] M. J. Weinberger, J. J. Rissanen, and G. Lombardi, "A universal finite memory source," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 643–652, May 1995.
- [48] D. Wu, "QoS provisioning in wireless networks: Research articles," *Wirel. Commun. Mob. Comput.*, vol. 5, no. 8, pp. 957–969, Dec. 2005.
- [49] C. Xiao, Y. R. Zheng, and N. C. Beaulieu, "Second-order statistical properties of the WSS Jakes' fading channel simulation," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 888–891, Jun. 2002.
- [50] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999.
- [51] M. Zorzi, R. R. Rao, and L. B. Milstein, "ARQ error control for fading mobile radio channels," *IEEE Trans. Veh. Technol.*, vol. 46, no. 2, pp. 445–455, May 1997.
- [52] M. Zorzi, "Some results on error control for burst-error channels under delay constraints," *IEEE Trans. Veh. Technol.*, vol. 50, no. 1, pp. 12–24, Jan. 2001.



Wuttipong Kumwilaisak (S'02–M'03) received the B.E. degree from the Chulalongkorn University, Bangkok, Thailand, in 1995 and both the M.S. and Ph.D. degrees from the University of Southern California, Los Angeles, in 2003, all in electrical engineering with support from a Thai Government Scholarship.

He was a Research Intern with Ericsson Eurolab, Aachen, Germany, in 2001 and with Microsoft Research Asia, Beijing, China, in 2002. From April 2003 to August 2004, he was a Senior Engineer and the Project Leader of the mobile platform solution team and multimedia laboratory with Samsung Electronics, Suwon, Korea. From March to November 2006, he was a Postdoctoral Fellow with Thomson Research Laboratory, Princeton, NJ. Since 2007, he has been an Assistant Professor with the Department of Electronics and Telecommunication Engineering, King Mongkut's University of Technology Thonburi, Bangkok. His research interests include the optimization and algorithmic design for wireless communications, multimedia processing, and multimedia communication systems.

Dr. Kumwilaisak was the recipient of the Best Paper Award from the Multimedia Communication Technical Committee of the IEEE Communications Society in 2005.



C.-C. Jay Kuo (S'83–M'86–SM'92–F'99) received the B.S. degree from the National Taiwan University, Taipei, Taiwan, R.O.C., in 1980 and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1985 and 1987, respectively, all in electrical engineering.

He is the Director of the Signal and Image Processing Institute and a Professor of electrical engineering, computer science, and mathematics with the University of Southern California (USC), Los Angeles. He has guided about 80 students to their Ph.D. degrees and supervised 20 postdoctoral research fellows. Currently, his research group at USC has around 35 Ph.D. students (see <http://viola.usc.edu>), which is one of the largest academic research groups in multimedia technologies. He is the coauthor of about 130 journal papers, 720 conference papers, and eight books. He has delivered more than 300 invited lectures at conferences, research institutes, universities, and companies. His research interests include the areas of digital image/video analysis and modeling, multimedia data compression, communication and networking, and biological signal/image processing.

Dr. Kuo is a Fellow of The International Society for Optical Engineers and a member of the Association for Computing Machinery. He is the Editor-in-Chief of the *Journal of Visual Communication and Image Representation* (Elsevier) and has been the Editor of eight other international journals. He was the recipient of the National Science Foundation Young Investigator Award in 1992, the Presidential Faculty Fellow Award in 1993, the Northrop Junior Faculty Research Award from the USC Viterbi School of Engineering in 1994, and the best paper awards from the Multimedia Communication Technical Committee of the IEEE Communication Society in 2005, the IEEE Vehicular Technology Fall Conference (VTC-Fall) in 2006, and the IEEE Conference on Intelligent Information Hiding and Multimedia Signal Processing in 2006. He was an IEEE Signal Processing Society Distinguished Lecturer in 2006.



Dapeng Wu (S'98–M'04–SM'06) received the B.E. degree in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 1990, the M.E. degree in electrical engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 1997, and the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, in 2003.

Since August 2003, he has been with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, as an Assistant Professor. His research interests include the areas of networking, communications, multimedia, signal processing, and information and network security.

Dr. Wu is a member of the Best Paper Award Committee, Technical Committee on Multimedia Communications, IEEE Communications Society. He was the recipient of the National Science Foundation CAREER award in 2007, the Best Paper Award from the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY in 2001, and the Best Paper Award in International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QShine) in 2006. He is currently the Editor-in-Chief of the *Journal of Advances in Multimedia* and an Associate Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and the *International Journal of Ad Hoc and Ubiquitous Computing*. He is also a Guest-Editor of IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, Special Issue on Cross-Layer Optimized Wireless Multimedia Communications. He is the Program Chair for the 2008 International Conference on Communications Signal Processing for Communications Symposium and has been a member of the executive committee and/or technical program committees of over 50 conferences. He is the Vice Chair of Mobile and Wireless Multimedia Interest Group, Technical Committee on Multimedia Communications, IEEE Communications Society.