Optimal Variable Length Markov Chain (VLMC) Modeling of Fading Channels

Wuttipong Kumwilaisak †, C.-C. Jay Kuo ‡, and Dapeng Wu §
†King Mongkut's University of Technology, Thonburi, Email: wuttipong.kum@kmutt.ac.th,
‡University of Southern California, Email: cckuo@sipi.usc.edu,
§University of Florida, Email: wu@ece.ufl.edu

***Abstract*— Channel characterization and modeling is 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 work. 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. 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 Kullback-Liebler (K-L) distance as the optimization criterion. Finally, we show simulation results that demonstrate the accuracy of the proposed fading channel representation in modeling the Rayleigh fading as well as the log-normal fading.**

I. Introduction

Discrete Markov model is widely used to model the fading channel [4]–[6]. These include both the first order Markov model [4], [6] and the variable order Markov model [5]. However, there are several issues needed to be addressed in order to make Markov model more accurate and useful in representing a wireless fading channel. First, the exact fading envelope or the fading channel SNR distribution is unknown in practice. Hence, the fading channel model with a discrete Markov under the assumption of known fading SNR distribution may not match well with the actual measurements [4], [6]. Second, the first-order Markov model, which is the most popular discrete Markov model in representing fading channel [4], [6], may not accurately model fading channel in every environment well [4]. Third, given the number of parameters characterizing discrete Markov model, the optimal structure of discrete Markov model, including its order and the number of Markov states, should be derived to best model a fading channel. Last but not the least, the transmission policy (*e.g.,* adaptive modulation [11]) used by a communication system, should be considered in discrete Markov modeling.

To address the above issues, we present a scheme that optimally represents a fading channel with variable length Markov chain (VLMC). VLMC is a general class of Markov chain, in which its structure has variable order and parsimonious number of transition probabilities. The proposed modeling consists of two main components: 1) optimal fading partitioning under the constraint of a transmission policy and 2) derivation of the best VLMC representation. The fading partition is to discretize/quantize a continuous fading channel gain. The optimality criterion for the quantization is based on a cost function of fading channel statistics and the transmission policy used in the system. Once a continuous fading channel gain is quantized, variable length Markov chain is used to model the channel. To obtain the optimal VLMC representation, we use a Kullback-Liebler (K-L) distance as the optimality criterion.

The rest of this paper is organized as follows. The fading partition under a transmission policy using the relaxation method is presented in Section II. In Section III, the concept of VLMC is reviewed. The optimal fading modeling with VLMC based on the K-L distance and the Lagrangian optimization technique is presented in Section IV. Simulation results for the proposed channel modeling are shown in Section V. Finally, concluding remarks are given in Section VI.

II. Optimal Fading Partition Under A Transmission Policy

Let us first consider a set of fading sample SNRs obtained from a fading process as $\Gamma = \{\gamma_1, \ldots, \gamma_L\}$, where γ_i is SNR sample i of the fading process and L is the total number of SNR samples in Γ. The probability density function (pdf) of fading SNRs can be derived from SNR samples in Γ. In this paper, we use the kernel density estimation (KDE) [3] to estimate the pdf of fading SNR. The estimated pdf of fading SNR based on the KDE can be written as

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

where $K(\cdot)$ is an arbitrary pdf (*e.g.*, the Gaussian) and h is the bandwidth of the kernel function $K(\cdot)$. To compute the optimal h in order to give the best closeness between the estimated and actual pdf of the fading process, an iterative approach called the solved equation method (STE) [3] is used.

Define a set of fading partition, which partitions the whole fading SNR range of the estimated pdf to be U intervals, as $\Delta = \{\alpha_1, \ldots, \alpha_{U-1}\}$ (*i.e.,* $\gamma_{lower} \leq \alpha_1 \ldots \leq \alpha_{U-1} \leq \gamma_{upper}$), where $[\alpha_{i-1}, \alpha_i)$ is interval i bounded by α_{i-1} and α_i, and γ_{lower} and γ_{upper} are the end points on the left and on the right of the estimated pdf, respectively. Each fading partition is viewed as one channel state and has its

1-4244-0355-3/06/$20.00 (c) 2006 IEEE

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

own fading channel parameters for characterizing the channel condition. To compute the optimal fading partition, we jointly consider two criterions: 1) the accuracy of fading channel parameters characterizing fading partition, and 2) the objective of communication system design. For the first criterion, we need to have the fading channel parameters characterizing fading partition to be as close to the fading channel parameters of the actual fading as possible. For the second criterion, we need to achieve the objective of communication system design, when we apply the specific transmission policy to transmit data.

Let cost functions, which are the function of Δ, concerning the accuracy of fading channel parameters and the objective of communication system design be $C_a(\Delta)$ and $C_b(\Delta)$, respectively. In this paper, we consider the probability of bit error as a fading channel parameter, whereas the objective of communication system design is to limit the averaged bit error per transmission symbol. We first define a cost function $C_a(\Delta)$, which measures the closeness between the probabilities of bit error characterizing fading partition and that of the actual fading channel condition to be approximated, as

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

where $\alpha_0 = \gamma_{lower}$, $\alpha_U = \gamma_{upper}$, $\hat{f}(\gamma)$ is the estimated fading SNR distribution, and $e_i\{\pi_i\}$ is the probability of bit error characterizing the fading partition $[\alpha_{i-1}, \alpha_i)$ with transmission policy π_i. The transmission policy π_i is defined as a transmission scheme (*e.g.,* modulation type), when the fading channel condition is characterized by the fading partition $[\alpha_{i-1}, \alpha_i)$. $P_b\{\gamma, \pi_i\}$ is the probability of bit error of the actual fading channel with the transmission policy π_i under the environment with the SNR equal to γ. For the transmission policy, we adopt an adaptive modulation scheme into our proposed partition algorithm. The adaptive modulation algorithm [11] uses possibly different modulation schemes under different fading channel conditions. The cost function corresponding to the adaptive modulation with the objective to limit the averaged bit error per symbol can be written as

$$C_b(\Delta) = \sum_{i=1}^{U} \int_{\alpha_{i-1}}^{\alpha_i} B_i \times P_b\{\gamma, \pi_i\} \hat{f}(\gamma) d\gamma, \quad (3)$$

where B_i is a number of bits per symbol of the modulation used in SNR interval i (*e.g.,* if the transmission policies π_i and π_{i+1} of fading interval i and $i{+}1$ are BPSK and QPSK, bits per symbol B_i and B_{i+1} are one and two bits, respectively.). Based on the defined cost functions, our problem can be formulated as

Formulation 1: **(Optimal fading partition under the transmission policy)**
Find the optimal set of fading partition Δ^* such that

$$\Delta^* = \min_{\forall \Delta} C_a(\Delta), \quad (4)$$

subject to

$$C_b(\Delta) \leq C_{b,target}, \quad (5)$$

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

The solution of Formulation 1 can be obtained through the minimization of the Lagrangian cost function. The cost function of the Lagrangian optimization technique derived from Formulation 1 can be written as

$$J_p(\lambda_p, \Delta) = C_a(\Delta) + \lambda_p C_b(\Delta), \quad (6)$$

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 of Formulation 1. Because there are two sets of variables in the cost function from (6) (*i.e.,* a set of fading partition $\Delta = \{\alpha_1, \ldots, \alpha_{U-1}\}$ and a set of the probability of bit error characterizing the fading partition $E = \{e_1(\pi_1), \ldots, e_{U-1}(\pi_{U-1})\}$), we propose to use the relaxation method to obtain the optimal solution. The algorithm can be described as follows.

Algorithm 1: **(Relaxation method to obtain the optimal fading partition)**

- Step 0: First, we take the partial derivation of the cost function in (6) with respective to $e_i\{\pi_i\}$ and set it to zero. We obtain

$$e_i\{\pi_i\} = \frac{\int_{\alpha_{i-1}}^{\alpha_i} P_b\{\gamma, \pi_i\} \hat{f}(\gamma) d\gamma}{\int_{\alpha_{i-1}}^{\alpha_i} \hat{f}(\gamma) d\gamma}, \qquad \forall i = 1, 2, \ldots, U. \quad (7)$$

  Then, we take the partial derivation of the cost function in (6) with respective to α_i and set it to zero, which leads to

$$\begin{aligned} (e_i\{\pi_i\} - P_b\{\alpha_i, \pi_i\})^2 \\ -(e_{i+1}\{\pi_{i+1}\} - P_b\{\alpha_i, \pi_{i+1}\})^2 = \\ \lambda_p[B_{i+1}(1 - P_b\{\alpha_i, \pi_{i+1}\}) - \\ B_i(1 - P_b\{\alpha_i, \pi_i\})], \qquad \forall i = 1, 2, \ldots, U. \end{aligned} \quad (8)$$

- Step 1: Initialize a set of fading partition. Select a value of λ_p.
- Step 2: Compute the value of $e_i\{\pi_i\}$, $\forall i = 1, \ldots, U-1$, from a set of fading partition using (7). Then, use the value of $e_i\{\pi_i\}$, $\forall i = 1, \ldots, U-1$ to compute a set of fading partition α_i, $\forall i = 1, \ldots, U-1$, using (8). We iterate between (7) and (8) until a set of fading partition is converged. This converged fading partition will be considered as the optimal fading partition with the considering value of λ_p.
- Step 3: Go to Step 1 until obtaining the value of λ_p, which meets the constraint of the averaged bit error per symbol. The final solution will serve as the solution of Formulation 1.

III. Variable Length Markov Chain: Some Background

Variable length Markov chain (VLMC) [7] is a Markovian process whose structure has a sparse memory with some states

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

lumped together. The structure of VLMC can be characterized by variable order and parsimonious number of transition probabilities. The underlying structure of VLMC generally leads to the simplification of Markov channel model comparing to the full Markov chain.

To describe the concept of VLMC in details, we review the definition of a categorial space and a context function [7]. First, the finite categorial space χ is an arbitrary set of a letter representing all possible data values generated from Markov process. For example, in Fig. 1, the finite categorial space is $\chi = \{a, b, c, d\}$. Then, concepts of the context function and the variable length memory are introduced to serve as tools in VLMC construction [7].

Definition 1: Let $(\mathbf{x}_t)_{t\in Z}$ be a sequel time series, where $\mathbf{x}^{\mathbf{t}}_{-\mathbf{l}+\mathbf{1}+\mathbf{t}} = x_{-l+1+t}, \ldots, x_t$, with value $x_t \in \chi$. We define a function denoted by Ψ that maps an infinite sequence (an infinite past) to a possibly shorter string as $\Psi : \mathbf{x}^{\mathbf{t}}_{-\infty} \rightarrow \mathbf{x}^{\mathbf{t}}_{-\mathbf{l}+\mathbf{1}+\mathbf{t}}, l = l(\mathbf{x}^{\mathbf{t}}_{-\infty}) = \min\{k; P\{x_{t+1}|\mathbf{x}^{\mathbf{t}}_{-\infty}\} = P\{x_{t+1}|\mathbf{x}^{\mathbf{t}}_{-\mathbf{k}+\mathbf{1}+\mathbf{t}}\}, \forall x_{t+1} \in \chi\}$, where $\Psi(\cdot)$ is called a context function and $\Psi(\mathbf{x}^{\mathbf{t}}_{-\infty}) = \mathbf{x}^{\mathbf{t}}_{-\mathbf{k}+\mathbf{1}+\mathbf{t}}$ is called a context of the process time t with length k. Let $0 \leq \zeta \leq \infty$ be the smallest integer such that $l(\mathbf{x}^{\mathbf{t}}_{-\infty}) \leq \zeta, \quad \forall \mathbf{x}^{\mathbf{t}}_{-\infty} \in \chi^{\infty}$. The number ζ is called the order of the context function $\Psi(\cdot)$, and if $\zeta \leq \infty$, $(\mathbf{x}_\mathbf{t})_{t\in Z}$ is called a variable length Markov chain of order ζ.

Obviously, the VLMC of order ζ is a Markov chain with the maximum order equaling ζ. VLMC can have a variable order $1 \leq l \leq \zeta$ inside its structure. The context represents how previous data values of Markov process contribute to the knowledge of the incoming data values with the length of the context corresponding to a number of the past data values. To give a concrete example in the context function concept, consider the transition probability of the VLMC in Fig. 1 with $\chi = \{a, b, c, d\}$ as

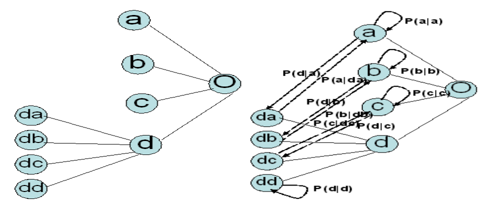
$$P\{x_t|x_{t-1},\ldots\} = \begin{cases} P\{x_t|x_{t-1} = a\}, \\ P\{x_t|x_{t-1} = b\}, \\ P\{x_t|x_{t-1} = c\}, \\ P\{x_t|x_{t-1} = d, x_{t-2} \in \{a, b, c, d\}\}. \end{cases}$$

The contexts used to characterize VLMC corresponding to the above transition probabilities in this example are

$$\Psi(\mathbf{x}^{\mathbf{t-1}}_{\infty}) = \begin{cases} a, & x_{t-1} = a, \mathbf{x}^{\mathbf{t-2}}_{\infty} arbitrary \\ b, & x_{t-1} = b, \mathbf{x}^{\mathbf{t-2}}_{\infty} arbitrary \\ c, & x_{t-1} = c, \mathbf{x}^{\mathbf{t-2}}_{\infty} arbitrary \\ da, db, dc, dd, & x_{t-1} = d, x_{t-2} \in \{a, b, c, d\}, \\ & \mathbf{x}^{\mathbf{t-3}}_{-\infty} arbitrary. \end{cases}$$

With the contexts of the underlying VLMC, it is convenient to represent VLMC as a *context tree*, which can be defined as follows.

Definition 2: Let $\Psi(\cdot)$ be a context function of VLMC and its context tree τ be defined as $\tau := \{\omega_l : \omega_l = \Psi(\mathbf{x}^{\mathbf{t}}_{-\infty}), \omega_l u \notin \tau, \forall u \in \chi, l \leq \zeta\}$, where ω_l is a context of length l corresponding to nodes of the tree and $\omega_l u$ is the concatenation of context ω_l and alphabet u.



(a) (b)

Fig. 1. (a) The context tree structure of the variable length Markov chain (b) The corresponding finite state machine of the context tree of (a).

IV. VLMC Modeling of Fading Channels

A. Problem Formulation

As seen from Section III, VLMC can be represented with a context tree to represent fading channel. A context tree can be built from a set of measured fading SNR, Γ, and a finite categorial space, χ. Each letter in χ represents fading partition obtained from Section II. Consider a set of measured fading SNR $\Gamma = \{\gamma_1, \gamma_2, \ldots, \gamma_L\}$. Define $\mathcal{A} = \{x_1, x_2, \ldots, x_L\}$ as a letter sequence obtained from assigning a letter in χ to every measured fading sample in Γ. We call letters in $\mathcal{A}$ as quantized fading samples. The assignment is depended on the range of fading partition that the value of a measured fading SNR is in. For example, if the value of γ_1 is between $[\alpha_0, \alpha_1]$, we assign letter "a" to x_1.

To characterize a context tree or VLMC, we need to compute the probability and the transition probability of a context. The probability and the transition probability of a context are defined as

$$P\{\mathbf{x}^{\mathbf{k+l-1}}_{\mathbf{k}} = \omega_l\} = \frac{N(\omega_l)}{L - l + 1}, \tag{9}$$

$$P\{x_{k+l} = u_0|\mathbf{x}^{\mathbf{k+l-1}}_{\mathbf{k}} = \omega_l\} = \frac{N(\omega_l u_0)}{N(\omega_l)}, \tag{10}$$

where ω_l is a context with length l existing in $\mathcal{A}$, $N(\omega_l)$ is a number of context ω_l obtained from $\mathcal{A}$, $P\{\mathbf{x}^{\mathbf{k+l-1}}_{\mathbf{k}} = \omega_l\}$ is the probability of context ω_l, $N(\omega_l u_0)$ is a number of context $\omega_l u_0$, which is the concatenated context between context ω_l and letter u_0, and $P\{x_{k+l} = u_0|\mathbf{x}^{\mathbf{k+l-1}}_{\mathbf{k}} = \omega_l\}$ is a transition probability given the past sequel of letters equal to context ω_l, where $u_0 \in \chi$. 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 existing in $\mathcal{A}$. Note that $N(\omega_l u_0)$ can be computed by looking ahead over the window by one step. The closeness measurement of transition probability between specific context ω_l and $\omega_l u$ can be achieved through the Kullback-Liebler (K-L) distance [10]. It can be expressed as

$$\begin{aligned} &\rho(P\{x_{k+l} = u_0|\mathbf{x}^{\mathbf{k+l-1}}_{\mathbf{k-1}} = \omega_l u\}||P\{x_{k+l} = u_0| \\ &\mathbf{x}^{\mathbf{k+l-1}}_{\mathbf{k-2}} = \omega_l\}) = P\{x_{k+l} = u_0|\mathbf{x}^{\mathbf{k+l-1}}_{\mathbf{k-1}} = \omega_l u\} \\ &\qquad \ln \frac{P\{x_{k+l} = u_0|\mathbf{x}^{\mathbf{k+l-1}}_{\mathbf{k-1}} = \omega_l u\}}{P\{x_{k+l} = u_0|\mathbf{x}^{\mathbf{k+l-1}}_{\mathbf{k-2}} = \omega_l\}}, \end{aligned} \tag{11}$$

where $\rho(P\{x_{k+l} = u_0|\mathbf{x}^{\mathbf{k+l-1}}_{\mathbf{k-1}} = \omega_l u\}||P\{x_{k+l} = u_0|\mathbf{x}^{\mathbf{k+l-1}}_{\mathbf{k-2}} = \omega_l\})$ is the K-L distance of transition probability between context ω_l and $\omega_l u$. The smaller the K-L distance, the closer the values of transition probability of contexts ω_l and $\omega_l u$. In its physical meaning, there is little contribution in the knowledge improvement of the incoming letter, u_0, from the previous context, when concatenating letter u with context ω_l comparing with context ω_l alone.

There may be more than one possible VLMCs to represent the same fading channel. To select the optimal one, the trade-off between the complexity of VLMC and its accuracy in reflecting the stationary process of fading channel is considered. The complexity of VLMC structure can be defined in terms of a number of transition probability and a number of states. Define the summation of K-L distance of terminal nodes of a context tree corresponding to VLMC channel model Υ as

$$D_\Upsilon(P\{x_{k+l} = u_0|\mathbf{x_{k-1}^{k+l-1}} = \omega_l u\}||P\{x_{k+l} = u_0|\mathbf{x_{k-2}^{k+l-1}} = \omega_l\}) = \sum_{\forall \omega_l u \in \tau_t} \rho(\{P\{x_{k+l} = u_0|\mathbf{x_{k-1}^{k+l-1}} = \omega_l u\}||P\{x_{k+l} = u_0|\mathbf{x_{k-2}^{k+l-1}} = \omega_l\}), \quad (12)$$

where τ_t is a set of terminal nodes of a context tree corresponding to VLMC Υ, and l is the length of a context corresponding to the terminal nodes of a context tree and less than or equal to ζ. Based on the meaning of K-L distance provided above, the less the summation of K-L distance of terminal nodes, the better VLMC represents the statistical property of fading process represented by the letters in χ. $D_\Upsilon(P\{x_{k+l} = u_0|\mathbf{x_{k-1}^{k+l-1}} = \omega_l u\}||P\{x_{k+l} = u_0|\mathbf{x_{k-2}^{k+l-1}} = \omega_l\})$ is a convex function in a set of transition probabilities, which can be proven as follows.

Theorem 4.1: **(Convexity of the summed K-L distance of terminal nodes):** $D_\Upsilon(P\{x_{k+l} = u_0|\mathbf{x_{k-1}^{k+l-1}} = \omega_l u\}||P\{x_{k+l} = u_0|\mathbf{x_{k-2}^{k+l-1}} = \omega_l\})$ is a convex function in a set of transition probabilities of VLMC terminal nodes.

Proof: From Theorem 2.7.2 in [10], $\rho(P\{x_{k+l} = u_0|\mathbf{x_{k-1}^{k+l-1}} = \omega_l u\}||P\{x_{k+l} = u_0|\mathbf{x_{k-2}^{k+l-1}} = \omega_l\})$ is convex in a pair of transition probabilities. We know from [9] that the summation of convex functions is convex. Therefore, the summed K-L distance of the VLMC terminal nodes is convex in a set of transition probabilities of the VLMC terminal nodes. $\diamond$

The problem formulation to search for the optimal VLMC channel representation can be stated as follows.

Formulation 2: **(Optimal VLMC channel representation)**

Find the optimal VLMC channel representation Υ^* such that

$$\Upsilon^* = \min_{\Upsilon} D_\Upsilon(P\{x_{k+l} = u_0|\mathbf{x_{k-1}^{k+l-1}} = \omega_l u\}|| P\{x_{k+l} = u_0|\mathbf{x_{k-2}^{k+l-1}} = \omega_l\}), \quad (13)$$

subject to

$$N_\Upsilon \leq N_{\Upsilon,target}, \quad (14)$$

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

The cost function equivalent to Formulation 2 can be written in terms of the Lagrangian optimization as

$$J_c(\lambda_c) = D_\Upsilon(P\{x_{k+l} = u_0|\mathbf{x_{k-1}^{k+l-1}} = \omega_l u\}|| P\{x_{k+l} = u_0|\mathbf{x_{k-2}^{k+l-1}} = \omega_l\}) + \lambda_c N_\Upsilon, \quad (15)$$

where $J_c(\lambda_c)$ is a cost function of the Lagrangian optimization and λ_c is the Lagrange multiplier corresponding to the cost function.

B. Solution of the Optimization Problem

Due to the convexity of the summed K-L distance of the VLMC terminal nodes, we use the Lagrangian optimization technique in searching for the optimal VLMC model, which best represents the statistics of fading channel. To solve the underlying problem, we consider two steps: 1) constructing the summed K-L distance characteristic curve, and 2) searching for the optimal VLMC representation.

1) Constructing the Summed K-L Characteristic Curve: The summed K-L distance characteristic curve provides the relation between the summed K-L distance of terminal nodes from a context tree and the number of transition probability characterizing VLMC. It is built by considering several possible context trees representing the same quantized fading process, $\mathcal{A}$. We limit the maximum depth of a context tree or the maximum order of VLMC to ξ_{max} in order to constrain the size of a context tree. The algorithm for constructing the summed K-L characteristic curve can be described as follows.

Algorithm 2: **(Construction of the summed K-L distance characteristic curve)**

- Step 0: Initialize 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 $P\{\mathbf{x_{k-1}^{k+l}} = \omega_l\}$, $P\{x_{k+l+1} = u_0|\mathbf{x_{k-1}^{k+l}} = \omega_l\}$, and $P\{x_{k+l+1} = u_0|\mathbf{x_{k}^{k+l}} = \omega_l u\}$ 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 existing 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 $\rho(P\{x_{k+l} = u_0|\mathbf{x_{k-1}^{k+l-1}} = \omega_l u\}||P\{x_{k+l} = u_0|\mathbf{x_{k-2}^{k+l-1}} = \omega_l\})$, which is the K-L distance between $P\{x_{k+l} = u_0|\mathbf{x_{k-1}^{k+l-1}} = \omega_l u\}$ and $P\{x_{k+l} = u_0|\mathbf{x_{k-2}^{k+l-1}} = \omega_l\}$ for specific context $\omega_l u$ and ω_l from parameters obtained from Step 1.
- Step 4: Select threshold T_{th}. Prune nodes corresponding to context $\omega_l u$ if

$$|\rho(P\{x_{k+l} = u_0|\mathbf{x_{k-1}^{k+l-1}} = \omega_l u\}|| P\{x_{k+l} = u_0|\mathbf{x_{k-2}^{k+l-1}} = \omega_l\})| < T_{th}. \quad (16)$$

Repeat the pruning process for every node in τ_0, until the condition in (16) is not satisfied for all left nodes.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

- Step 5: From the pruned tree, compute the summed K-L distance of terminal nodes of the tree. Repeat Step 4 and 5 for various values of threshold by starting with τ_0.

The pairs of the summed K-L distance and a set of transition probability with different T_{th} characterize the summed K-L distance characteristic curve.

2) Searching for the Optimal VLMC Representation: In this section, we use the bisection algorithm [8] to search for λ_c^*, which provides the optimal VLMC representation based on the cost function $J_c(\lambda_c)$, from the summed K-L distance characteristic curve. To be concise in presentation, let $D_\Upsilon(\lambda_c)$ and $N_\Upsilon(\lambda_c)$ be the summed K-L distance defined in (12) and the number of transition probability of VLMC channel model, which is the optimized solution, when the Lagrange multiplier is equal to λ_c. The algorithm to search for the optimal VLMC representation of fading channel can be stated step-by-step as follows.

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

- Step 0: Start with two values λ_u and λ_l such that $N_\Upsilon(\lambda_u) \leq N_{\Upsilon,target} \leq N_\Upsilon(\lambda_l)$.
- Step 1: Set $\lambda_{next} = \frac{D_\Upsilon(\lambda_l) - D_\Upsilon(\lambda_u)}{N_\Upsilon(\lambda_l) - N_\Upsilon(\lambda_u)} + \epsilon$, where ϵ is an arbitrarily small positive number added to make sure that the smallest number of channel parameters is picked, if λ_{next} is a singular slope value.
- Step 2: Repeat the optimization process of (15) for $\lambda_c = \lambda_{next}$. If $N_\Upsilon(\lambda_{next}) = N_{\Upsilon,target}$, then stop. Else if $N_\Upsilon(\lambda_{next}) \geq N_{\Upsilon,target}$, set $\lambda_l = \lambda_{next}$, or else if $N_\Upsilon(\lambda_{next}) \leq N_{\Upsilon,target}$, set $\lambda_u = \lambda_{next}$. Go to Step 1.

V. Experimental Results

In this section, we present a sequence of computer simulation results to demonstrate the main concepts discussed in previous sections. First, we study the fading SNR distribution estimation using the proposed method as described in Section II. The closeness between the estimated and the actual fading SNR distributions can be measured objectively using the following K-L distance

$$D(\hat{f}(\gamma)||f(\gamma)) = \int_{-\infty}^{\infty} \hat{f}(\gamma) \ln \frac{\hat{f}(\gamma)}{f(\gamma)}, \quad (17)$$

where $\hat{f}(\gamma)$ and $f(\gamma)$ denote the estimated and the actual fading SNR distributions, respectively. Fig. 2 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 fast fading environment (with a larger normalized Doppler frequency), it demands fewer samples to obtain the same K-L distance.

Next, we consider the complete cost function as given in (6). 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 α_1. The transmission policy is adaptive modulation with two choices (BPSK and QPSK),

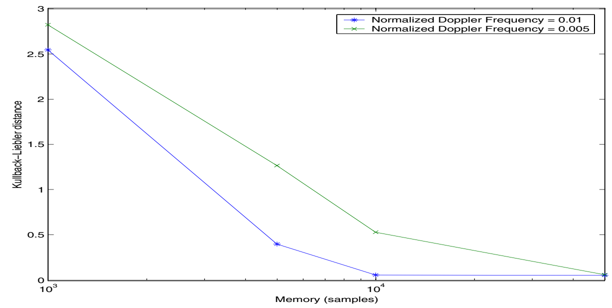


Fig. 2. The K-L distance between the Rayleigh fading channel and the estimate using the proposed fading partition scheme as a function of the number of fading samples.

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 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 trade-off between $C_a(\Delta^*)$ and $C_b(\Delta^*)$) but concentrate on $C_a(\Delta^*)$ alone in this simple case. However, when there are more intervals and more modulation choices, the optimal solution may occur in some $\lambda_p > 0$.

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

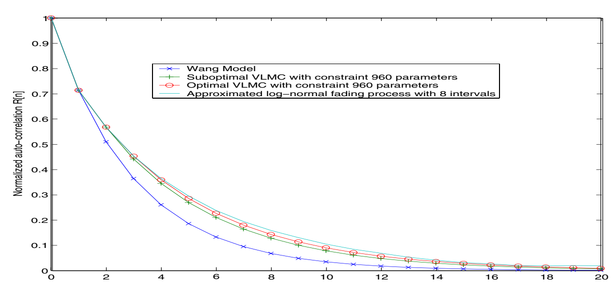


Fig. 3. The comparison of the normalized auto-correlation among different VLMC channel models and the discretized fading process under the log-normal fading process in sub-urban environments with $\rho = 0.82$ and $\sigma = 7.5$ dB with 8 intervals

Next, we study the performance of the optimal VLMC channel model in approximating the statistical property of fading channels. 50,000 samples of each fading channel are generated in the simulation. We investigate the optimal VLMC channel model in both log-normal long-term and the Rayleigh

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.

short-term fading. Fig. 3 shows the performance of the optimal VLMC to approximate the discretized fading process of the log-normal long term fading with eight intervals in an urban environment [1]. The performance can be measured in terms of the normalized auto-correlation. We compare our results with popular Wang's model [6] and sub-optimal VLMC under the constraint on the number of transition probabilities. As shown in the simulation, the optimal VLMC channel 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 auto-correlation. The average order of the optimal VLMC with 8 intervals and a constraint of 960 parameters is 3.18.

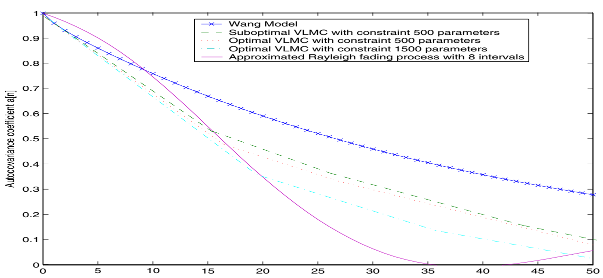


Fig. 4. Comparison of the auto-covariance coefficient among different VLMC channel models and the discretized fading process with 8 intervals in a Rayleigh fading process with normalized Doppler frequency 0.01

For the Rayleigh fading environment, we conduct the simulation in the same way as the log-normal fading environment. The auto-covariance coefficient is used to assess the statistical property of VLMC and discretized Rayleigh fading channel. Fig.4 shows how well the optimal VLMC with a different number of transition probabilities can model a discretized fading channel. We consider a discretized fading channel eight intervals with the Rayleigh fading. The optimal VLMC with a larger number of transition probabilities models the discretized fading channel better than that with a smaller number of transition probabilities. This can be seen in terms of the closeness of the auto-covariance coefficient. Comparing with Wang's model, the optimal VLMC offers a better model in fading channel modeling. The average orders of the optimal VLMC with the constraints of 500 and 1500 parameters are equal to 2.7 and 3.8, respectively, which are low comparing to the full Markov chain with the same order of VLMC. Fig. 5 compares the average fade duration obtained from the optimal VLMC and that of the Rayleigh fading channel. The average fade duration is measured in terms of the number of symbols. The constraint on the number of parameters is set to be equal to 1500. We can see that the optimal VLMC approximates the fading duration of the Rayleigh fading channel well.

VI. Conclusion

In this paper, we proposed an optimal fading channel modeling with VLMC. The optimization criterion is the K-L

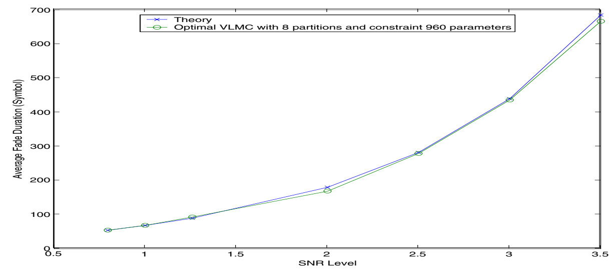


Fig. 5. Comparison of the average fade duration of Rayleigh fading obtained from theory and optimal VLMC when normalized Doppler frequency is equal to 0.01

distance. The proposed scheme consists of two main components. One is the fading partition mechanism, which takes the transmission policy into account. The other is the derivation of the optimal VLMC to represent fading channel under the constraint on the number of the transition probabilities. The optimal VLMC model can be obtained by constructing the summed K-L characteristic curve and selecting the optimal VLMC using the Lagrangian optimization. From the experimental results, the optimal VLMC provides a better approximation of the fading channel than a benchmark, which is sub-optimal. Moreover, we show that to represent fading channel well, VLMC may require different order and different number of transition probabilities, depending on fading channel environments (*e.g.,* the Doppler frequency).

VII. Acknowledgement

This research is partly supported by King Mongkut's University, Thonburi Research Fund and partly supported by Electronics and Telecommunication Department, King Mongkut's University of Technology, Thonburi.

References

[1] M. Gudmundson, "Correlation model for shadowing fading in mobile radio systems," *Electronics Letters*, vol. 27, no. 23, pp. 2145 -2146, 1999.

[2] G.L. Stuber, *Principles of Mobile Communication*, Boston, MA, Kluwer, 1996.

[3] M. P. Wand and M. C. Jones, *Kernel Smoothing*, Chapman and Hall, London, 1995.

[4] 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, 2000.

[5] 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, 1999.

[6] 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, 1995.

[7] P. Buhlmann and A.J. Wyner, "Variable length Markov chains," *Annals of Statistics*, vol. 27, pp. 480-513, 1999.

[8] A. Ortega, *Optimization Techniques for Adaptive Quantization of Image and Video under Delay Constraints.* Ph.D. thesis, Dept. of Electrical Engineering, Columbia University, New York, NY, 1994.

[9] S. Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[10] T. M. Cover and J.A. Thomas, *Elements of Information Theory*, John Wiley and Sons INC., New York, 1991.

[11] A. Goldsmith and S. Chua, "Variable-Rate variable power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218-1230, 1997.

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2006 proceedings.