# A Variable-Order Discrete Model for the Fading Channel

Fulvio Babich[†],Owen E. Kelly[§]and Giancarlo Lombardi[†]

[†]Dipartimento di Elettrotecnica, Elettronica ed Informatica
Università di Trieste, Via A. Valerio 10, I-34127 Trieste, Italy
babich, lombardi@univ.trieste.it

[§]Wireless Information Network Laboratory (WINLAB)
Rutgers University, Piscataway, NJ, USA 08855-0909
oekelly@winlab.rutgers.edu

**Abstract.** Fading on the mobile propagation channel is modeled as a Markov chain with unspecified state space and order. State space and memory are selected via the context tree pruning (CTP) algorithm that has provable optimality in fitting Markov models of unknown order. CTP provides a hierarchy of models of increasing dimension each of which may be considered the "best" approximation of the fading channel among Markov models of that dimension. A particular model may then be selected by a criterion appropriate for the task that the discrete model will perform. An example herein considers the model's ability to faithfully mimic statistics of sojourn times in each channel state. That criterion choice is relevant for simulation of channel access and use schemes that exploit channel memory.

## 1 Introduction

Wireless communications are experiencing continued growth due to their intrinsic features of flexibility. That growth brings pressure to make better use of channel capacity, either by allowing a greater number of users to access services or by increasing the effective bandwidth of services themselves. The main limitation is the mobile fading channel, and although coding and interleaving may limit the effects of fading by making the data stream *appear* independent, a capacity penalty is introduced by that strategy. Communication systems that achieve capacity *must* consider channel memory [1]. To this purpose and because they are amenable to analysis and efficient simulation, finite state Markov models have been considered by several authors [2, 3, 4, 5, 6, 7]. In particular, first-order models have been built as approximations to the classic Clarke's model [8] and statistical analysis and simulation support the accuracy of the first order model for very slow fading or fast fading [9]. Between those regimes, however, the available results suggest that the first order model can be an oversimplification.

In this paper a modeling technique developed originally for data compression is adapted to the task of estimating a Markov chain model of the fading channel [10]. A *context tree* is a particular minimal parameterization of a Markov chain and the context tree pruning algorithm described in [10] has provable optimality in estimating the Markov model when the order of dependence is unknown. CTP provides a hierarchy of models of increasing dimension each of which may be considered the "best" approximation of the fading channel among models of that dimension. As the models become more complex, the degree to which they approximate the fading channel improves. The designer chooses the smallest model that meets their approximation requirements as defined by a problem specific fidelity criterion. For access schemes, we consider the case that the goodness of the model is defined with respect to its fidelity in reproducing the *times* that the channel is in a particular state.

We unify the comparison of finite state models of different memory length by using the context tree pruning (CTP) algorithm together with a fidelity criterion (e.g. Kolmogorov distance applied to sojourn time distribution in each channel state). Context tree data sources and CTP are summarized in §2. §3 describes the fading simulator. The application of context tree pruning to simulated fading data is presented in §4 and further discussion in §5.

## 2 A Variable Order Markov Model

Figure 1 illustrates the spirit of much recent work in information theory that highlights the importance of describing models using a *minimal* number of parameters [11, 12]. A *context tree* is a particular minimal parameterization of the $k$-th order Markov source $P(x_n|x_{n-1}\cdots x_{n-k})$ and also the name of the data structure by which such sources may be estimated (e.g. [10]). Figures 2 and 3 compare the representations for an example with memory $k=2$. The estimation task is to discover the conditional probability mass function of a finite alphabet Markov process based on an observed *training* sequence $\boldsymbol{x} = x_1x_2\cdots x_N$ where each $x_n \in \mathcal{A}$ and $\mathcal{A}$ is a finite set. The probability mass function for a symbol, $x_{n+1}$, resides on the tree leaf specified by the recent history or *context* of the process: $x_n, x_{n-1}, \ldots$. The source is defined by its set of contexts, $\mathcal{C}$, and a parameter vector, $\theta$, that specifies transition probabilities. The context tree source has fewer defining parameters ($\dim\theta$) because, where possible, contexts that provide no useful distinction are merged. In figure 2 uses alphabet $\mathcal{A} = \{a, b, c\}$ and the context set (defined over $\mathcal{A}^*$) is $\mathcal{C} = \{aa, ba, ca, b, c\}$.

Estimating a context tree source consists of counting the number of occurrences in the training sequence of all possible subsequences of length less-than or equal to some maximum order. Counts are arranged in a tree analogous to Figure 2 but with histograms also at interior nodes. The second estimation stage consists of examining the accumulated tree and selecting a particular

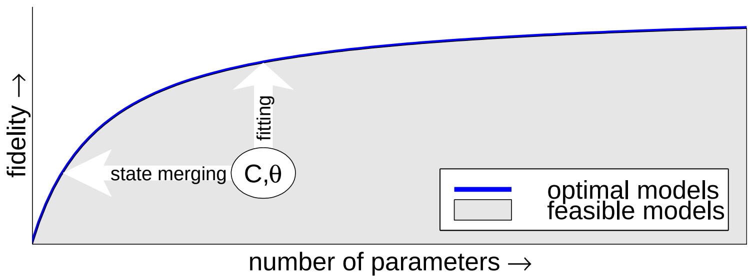


**Fig. 1.** Model optimality is attained by parameter fitting and parameter reduction.

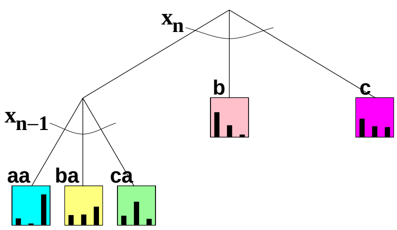


**Figure 2**: Context tree source. Tree leaves show conditional distributions of $x_{n+1}$ where the conditioning sequence, $x_n\ x_{n-1}$, is represented by the path that leads to the leaf.

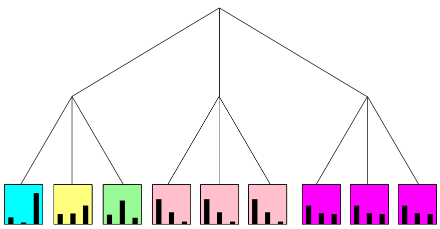

**Figure 3**: The traditional parameterization specifies the same memory for every context (here 2). The number of parameters required to represent the *same* source as Figure 2 increases from 10 to 18.

best tree by merging contexts or *pruning*. Context tree pruning [10] is optimal in that *if the true data source is a context tree source* then the negative-log-probability (or compressed size of $\mathbf{x}$ in bits) that CTP assigns to its own training sequence is within $\mathcal{O}(1)$ of the best-achievable lower bound [11]. For sufficiently long training, CTP is able to discover the correct set of contexts $\mathcal{C}$ with probability one and to estimate $\theta$ with high accuracy.

For every context tree source $(\mathcal{C}, \theta)$, there exists an equivalent first-order Markov chain $(\mathcal{S}, P)$ whose states $s \in \mathcal{S}$ also correspond to short sequences in $\mathcal{A}^*$ and transition probabilities $(P)_{ij}$ are determined by $\theta$. For example, if $x_n$ is a binary sequence ($\mathcal{A} = \{0, 1\}$) from a context tree source then the Markov chain state space might be $\mathcal{S} = \{11, 10, 0\}$ and we say the second Markov state "occurs" at time $n$ if the channel states satisfy $x_n = 1, x_{n-1} = 0$.

## 3 Fading Model

The classical Clarke model is adopted in simulating the baseband fading process $\tilde{f}(t)$. The in-phase and quadrature fading processes $\tilde{f}_c(t), \tilde{f}_s(t)$ are both assumed to be Gaussian, stationary and uncorrelated. The expression of power spectrum for both of them is the following:

$$S(f) = \begin{cases} \frac{\sigma^2}{2\pi f_D} \frac{1}{\sqrt{1-\left(\frac{f}{f_D}\right)^2}}, & \text{for } |f| < f_D \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $\sigma^2 = 1$ is process power, $f_D = \frac{v}{\lambda}$ is the maximum Doppler shift, $v$ is the mobile speed, and $\lambda$ is the carrier wavelength. The corresponding autocorrelation function is $R(\tau) = \sigma^2 J_0(2\pi f_D \tau)$, where $J_0(\cdot)$ is zeroth-order Bessel function of the first-kind.

We assume in the example of §4 that the in-phase and quadrature processes are zero mean, so that the resulting fading process is *Rayleigh type*, i.e. the real envelope $f(t) = \left|\tilde{f}(t)\right|$ follows the Rayleigh distribution:

$$p_f(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right). \tag{2}$$

## 4 Context Tree Approximation for Rayleigh Fading

We begin by sampling the fading envelope $\boldsymbol{f} = \{f(kT)\}_{k=-\infty}^{+\infty}$ with sampling period $T$ to construct a vector, $\boldsymbol{f}$, of measurements. Let the sampled fading envelope $\boldsymbol{f}$ be quantized on $L-1$ levels $\{A_1, A_2, \ldots, A_{L-1}\}$, forming an $L$-valued discrete-time process:

$$x_i = \sum_{k=1}^{L-1} \mathbf{1}[f_i > A_k] \tag{3}$$

where $\mathbf{1}[cond]$ is the indicator function, that equals 1 if *cond* is true, else 0. Now $x_k$ defines the *channel state* at time $k$ and lies in the alphabet $\mathcal{A} = \{0, 1, \ldots, L-1\}$. For example, we might define a binary process by choosing $A_1$ to be the minimum acceptable channel gain; then $x_i = 0$ implies a fading state and $x_i = 1$ implies an acceptable channel state, as in the Gilbert–Elliott model [2, 3].

A long sequence of real or simulated fading samples $\{f_i\}_{i=1}^N$ allows us to construct a quantized sequence $\mathbf{x} = \{x_i\}_{i=1}^N$ and then to estimate a context tree source, $(\widehat{\mathcal{C}}(\mathbf{x}), \widehat{\theta}(\mathbf{x}))$ from $\mathbf{x}$. If there exists a finite $\mathcal{C}$ that describes the source of $\mathbf{x}$, then $\widehat{\mathcal{C}}(\mathbf{x})$ will grow in size as $N$ increases, and finally stop at the correct context set $\mathcal{C}$. But if $\mathbf{x}$ is not a finite order Markov process, the models will continue to grow in complexity as the length of the training sequence grows. An arbitrary degree of approximation can be therefore achieved by using an arbitrarily long training sequence.

The data compression optimality of CTP may be rephrased that CTP discovers the simplest possible model that describes the data with a given fidelity. Let us select some very large $N$ and carry out the simulation and estimation for process $\mathbf{x}$. Before $\widehat{\mathcal{C}}(\mathbf{x})$, there exists an increasing list of context sets $\emptyset = \mathcal{C}_0 \subset \mathcal{C}_1 \subset \widehat{\mathcal{C}}(\mathbf{x})$ and their corresponding parameters $\theta_0, \theta_1, \ldots, \theta_{\mathbf{x}}$. For each $k$, $(\mathcal{C}_k, \theta_k)$ is the best $(\dim \theta_k)$–parameter Markov model — best in the probability assignment sense (information distance). But the assignment of probabilities to sequences is the *fundamental* task of a random process, therefore a model that approximates well in this sense, approximates well in *every* sense (heuristically).

However, information distance may not be appropriate for our approximation task, we therefore re-evaluate the models based on a task-specific distance measure. Let $a \in \mathcal{A}$ be a fading state of interest, we define the sojourn time $\tau_a$ as the number of consecutive steps in which process $\mathbf{x}$ remains in state $a$ having just entered $a$. For discovering the interactions between protocol design and channel memory, the distribution of sojourn times in each channel state obtained from the model must match that of the actual channel. Let $D(\cdot,\cdot)$ be the distance between these two distributions. Let $G_a^k(n)$ be $\Pr[\tau_a \le n]$ for a context tree source $(\mathcal{C}_k, \theta_k)$. We would like to measure the fidelity of the $k$-th model by how closely the cumulative distribution function $G_a^k(n)$ approximates $\Pr[\tau_a \le n]$, i.e. $D(G_a^k(n), \Pr[\tau_a \le n])$. But we have only the sequence $\mathbf{x}$ and therefore we use $D(G_a^k, H_a)$ where $H_a(n)$ is the empirical estimate of $\Pr[\tau_a \le n]$ from sequence $\mathbf{x}$. ($N$ is chosen very large so that $H_a(n) \approx \Pr[\tau_a \le n]$.) The Kolmogorov distance

$$D(G_a^k, H_a) = \sup_n \left| G_a^k(n) - H_a(n) \right| \quad ,$$

is here adopted as a measure of deviation between the two distributions.

As the channel state $x_i = a$ may be a union of states of the Markov chain represented by the context tree, the sojourn time in $a$ is not geometrically distributed (as would happen if $a$ were a state of Markov chain), but has a more complicated distribution, that can be algebraically evaluated. This property enables context tree sources to approximate sources that have non-trivial sojourn time distributions [13, 14].

Figure 5 shows the context tree source fitted to $N = 10^5$ samples of a fading sequence whose two levels are defined by fade margin $F = 8\,\mathrm{dB}$. The fading rate is $f_D T = 0.1$. Figure 6 shows the number of parameters of context trees fitted to $10^5$ samples of fading sequences from slow, intermediate, and fast fading regimes and using a range of fade margins. It can be observed that the CTP algorithm finds first-order markovian behavior of the fading for the lowest value $f_D T = 0.0125$, corresponding to slow fading, and the substantial memoryless behavior for the highest value $f_D T = 0.3$, corresponding to fast fading. Intermediate fading with $f_D T = 0.1$ gives rise to more complex models, but the complexity decreases with fade margin.

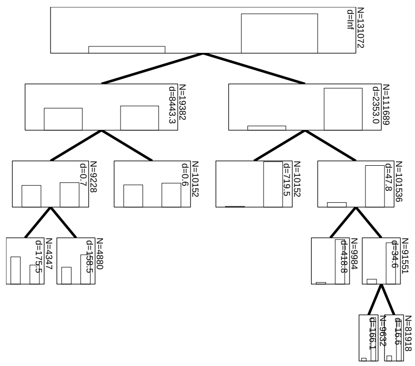

**Figure 5**: Context tree fitted to $10^5$ samples of quantized fading with $f_D T = 0.1$ and fade margin $F = 8$ dB.

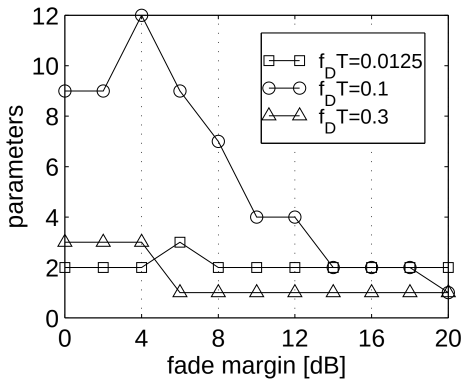


**Figure 6**: Number of parameters of the "best" Markov model fitted to data with CTP versus the fade margin for different values of $f_D T$.

Figure 7 shows the empirical sojourn time distribution for being in channel state 0 or channel state 1. Figure 8 shows how the Kolmogorov distance varies with model size. It can be noted that the reduction of model complexity from 7 parameters to 6 parameters does not introduce substantial degradation to model behavior. If only channel state 1 is of interest, then even a 4–parameter model appears to provide satisfactory approximation.

## 5 Discussion

A method and tools have been introduced that allow straightforward construction of finite state models for the mobile channel fading and objective comparison between models. Using context tree pruning on a large sample obtained from the target random process provides a hierarchy of models with increasing numbers of parameters. Each model may be considered "best" among models with that degree of complexity. The choice of a particular model leads to state definitions and transition probabilities for the approximating Markov chain. To select a particular model, the designer must provide a measure of fidelity and make their own choice of what is an acceptable

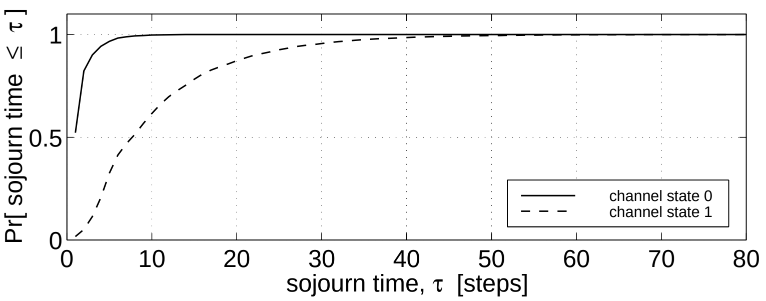


**Fig. 7.** Cumulative distributions $(H_0(n), H_1(n))$ of sojourn time in channel state 0 and 1 estimated from quantized fading with $f_D T = 0.1$ and $F = 8$ dB.

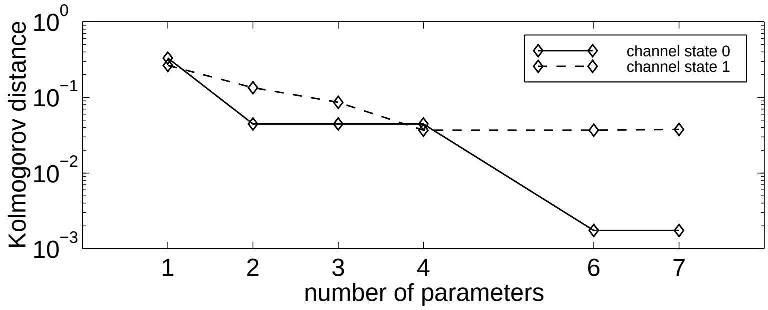


**Fig. 8.** Kolmogorov distance between the estimated distributions $(H_0(n), H_1(n))$ and the distributions obtained from context tree models that are subsets of the 7-parameter model shown in Figure 5 (versus $\dim \theta$).

compromise between fidelity and complexity. The method was demonstrated using Rayleigh fading quantized to two values and as a fidelity measure the Kolmogorov distance between sojourn time distributions for the approximating model and an estimate of the "true" model. But the method extends readily to other fading models, finer quantization, and any distance measure that the designer deems appropriate. The best finite order Markov models that represent various fading regimes of interest have still to be cataloged. In future work we will consider the possibility of using significance level as a fidelity criterion [15].

The high speed associated with future wireless access systems will increase the effects of channel memory. Systems that ignore memory by accommodating the worst case will be conservative designs that lose significant capacity as a result. Systems that are not conservative in that way will nonetheless become more fragile in operating regimes where channel memory is evident.

In much work with Markov models for fading, the model is chosen for its simplicity and later justified. We have provided a tool that allows designers to *measure* the channel memory. Such a tool may prove very useful for the design, optimization, and performance assessment of communication systems over Markov channels that have unknown memory.

## References


1. Andrea J. Goldsmith and Pravin P. Varaiya. Capacity, mutual information, and coding for finite-state Markov channels. *IEEE Trans. Info. Theory*, 42(3):868–886, May 1996.
2. E. N. Gilbert. Capacity of a Burst-Noise Channel. *Bell System Technical Journal*, 39:1253 – 1266, Sep 1960.
3. E. O. Elliott. Estimates of Error Rates for Codes on Burst-Noise Channels. *Bell System Technical Journal*, 42:1977 – 1997, Sep 1963.
4. H. S. Wang. On verifying the first-order Markovian assumption for a Rayleigh fading channel model. In *Proc. ICUPC '94*, pages 160 – 164, San Diego, CA, Sep. 1994.
5. H. S. Wang and N. Moayeri. Finite-state Markov channel: A useful model for radio communication channels. *IEEE Trans. Veh. Technol.*, 44(1):163–171, Feb 1995.
6. M. Zorzi, R. R. Rao, and L. B. Milstein. On the accuracy of a first-order Markov model for data transmission on fading channels. In *Proc. IEEE ICUPC '95*, pages 211 – 215, Tokyo, Japan, Nov. 1995.
7. M. Zorzi, R. R. Rao, and L. B. Milstein. A Markov Model for Block Erorrs on Fading Channel. In *Proc. PIMRC '96*, Taiwan, Oct. 1996.
8. R. H. Clarke. A statistical theory of mobile-radio reception. *Bell System Technical Journal*, 47:957–1000, 1968.
9. F. Babich and G. Lombardi. A Measurement Based Markov Model for the Indoor Propagation Channel. In *Proceedings of IEEE VTC '97*, Phoenix, AZ, May 5 – 7, 1997.
10. M. J. Weinberger, J. J. Rissanen, and M. J. Feder. A universal finite memory source. *IEEE Trans. Info. Theory*, 41(3):643–652, May 1995.
11. Jorma Rissanen. *Stochastic Complexity in Statistical Inquiry*, volume 15 of *Series in Computer Science*. World Scientific, 1989.
12. Bertrand S. Clarke and Andrew R. Barron. Jeffreys' prior is asymptotically least favorable under entropy risk. *J. Statist. Plann. Inference*, 41(1):37–60, 1994.
13. Gerardo Rubino and Bruno Sericola. Sojourn times in finite Markov processes. *J. Appl. Prob.*, 27:744–756, 1989.
14. Attila Csenki. The joint distribution of sojourn times in finite Markov processes. *Adv. in Appl. Probab.*, 24(1):141–160, 1992.
15. Patrick Billingsley. Statistical methods in Markov chains. *Ann. Math. Stat.*, 32:12–40, 1961.