Causal-CSIT Rate Adaptation for Block-Fading Channels

Khoa D. Nguyen
Institute for Telecommunications Research
University of South Australia
Mawson Lakes Blvd, SA5095
khoa.nguyen@unisa.edu.au

Roy Timo
Institute for Communications Engineering
Technische Universität München
Munich, Germany
roy.timo@tum.de

Lars. K. Rasmussen
ACCESS Linnaeus Center
KTH Royal Institute of Technology
Stockholm, Sweden
lars.rasmussen@ieee.org

***Abstract*—We propose a rate-adaptive error-correction coding framework for transmission over finite-length block-fading channels, with causal channel state information at the transmitter. A dynamic programming optimization approach can be use to obtain the optimal rate adaptation strategy. A suboptimal rate adaptation strategy is developed for sequential random codes. Numerical results show that significant performance gains are achievable with the proposed adaptation scheme.**

I. Introduction

The block-fading channel is a mathematical model for delay-constrained transmission over slowly varying fading channels. In this model, each codeword is transmitted over a finite number of blocks, and each block experiences a constant fading realization [1]. Even when each block consists of an infinite number of channel uses, the channel is non-ergodic. Consequently the transmission rate supported by the channel is a random variable, depending on the channel realizations. Performance measures of the channel include the outage probability and the throughput. The *outage probability* is the probability that the transmission rate is not supported by the channel, while *throughput* is the average information rate that can be successfully conveyed over the channel. With channel state information at the transmitter (CSIT), power and/or rate adaptive transmission techniques can be employed to optimize the performance, see for example [2]–[5]. For systems with full CSIT, water-filling or mercury/water filling power allocation is optimal [3], [6]. For systems with causal CSIT where only CSI up to and including the current block is available, the optimal power allocation strategies are solved in [4] using the dynamic programming approach. Meanwhile, the optimal rate adaptation is straightforward: block-wise transmission at a rate equal to the instantaneous capacity.

For the more realistic finite block length case, rate adaptation is not as trivial. In this scenario, there exists a tradeoff between transmission rate/throughput, block length and error probability [7], [8]. For systems with full CSIT, a suitable backoff from capacity is needed to achieve the error probability requirement [8]. When CSIT is causally available, a trivial scheme is block-wise adaptive transmission. However, block-wise transmission uses codes with smaller block length than the delay constraint, and thus larger backoff from capacity is necessary [7]. Optimal rate adaptation requires a coding scheme whose code rate can be adapted causally across the fading blocks. One such attempt is introduced in [9] for automatic-repeat-request systems with delayed CSIT; however, the optimal rate adaptation strategy was not studied.

In this paper, we study rate-adaptive transmission for finite-length block-fading channels with causal CSIT. We propose a sequential coding scheme for causal rate adaptation, which generalizes the *code with expandable message space* introduced in [9]. We propose a dynamic programming approach to optimize the system throughput, subject to an average error probability constraint. A suboptimal optimization approach is introduced for the sequential random coding scheme. Numerical results show that the proposed adaptation rule provides significant gains compared to non-adaptive transmission and the trivial block-wise adaptive transmission scheme. Furthermore, the performance of the suboptimal scheme is close to the upper bound benchmark achieved by systems with full CSIT.

Unless otherwise defined, the following notational rules are used. Lowercase and uppercase correspondingly denote deterministic and random variables. A bold symbol $\mathbf{x}$ denotes a vector or a matrix, whose dimension can be deduced from the context. $\mathbf{x}^i$ denotes a vector with the first i elements of $\mathbf{x}$, while $\mathbf{x}_i^j$ denotes the concatenation $\mathbf{x}_i \ldots \mathbf{x}_j$.

II. System Model

Consider transmission over a channel with B fading blocks, where each block consists of n channel uses. The input-output transition probabilities of block $b \in \{1, \ldots, B\}$ are

$$\mathbb{P}[\mathbf{Y}_b = \mathbf{y}_b | \mathbf{X}_b = \mathbf{x}_b, S_b = s_b] = \prod_{i=1}^{n} Q_{s_b}(y_{b,i}|x_{b,i}), \qquad (1)$$

where $\mathbf{X}_b = [x_{b,1}, \ldots, x_{b,n}] \in \mathcal{X}^n$ and $\mathbf{Y}_b = [y_{b,1}, \ldots, y_{b,n}] \in \mathcal{Y}^n$ are the transmit and receive signals in block b respectively. The channel state s_b is assumed to be causally known at the transmitter; for example, $\mathbf{s}^b \triangleq [s_1, \ldots, s_b]$ is available at the transmitter at the start of block b. The channel state sequence $\mathbf{S}^B = [S_1, \ldots, S_B]$ is i.i.d. on $\mathcal{S}$ according to the

This work was supported by the Australian Research Council Grant DE12010016, the Alexander von Humboldt Foundation, and the Swedish Research Council.

law $\mathbb{P}[S_b = s_b] = f_S(s)$. For a given channel realization s^B, the channel experienced by each codeword is

$$\mathbb{P}\left[\mathbf{Y}_1^B = \mathbf{y}_1^B | \mathbf{X}_1^B = \mathbf{x}_1^B, \mathbf{S} = s^B\right] = \prod_{b=1}^{B} \mathbb{P}[\mathbf{Y}_b = \mathbf{y}_b | \mathbf{X}_b = \mathbf{x}_b, S_b = s_b]. \quad (2)$$

The transmitter is required to reliably communicate the output of a uniform Bernoulli source $\mathbf{W} = W_1, W_2, \ldots$ to the receiver. In the following, we propose a rate-adaptive scheme that exploits causal CSIT to maximize throughput.

A. Encoding and Decoding

Causal rate adaptation can be realized by a transmission scheme consisting of the following mappings:

- A *rate-adaptation strategy* $\boldsymbol{\ell} = [\ell_1, \ell_2, \ldots, \ell_B]$, where

$$\ell_b : \mathcal{S}^b \to \mathcal{L} \subseteq \mathbb{N}, \quad (3)$$

  where $\mathcal{L}$ is the set of strategies in each block.
- A sequential encoder for each $s^b \in \mathcal{S}^B$ and $b = 1, \ldots, B$,

$$f_{\mathbf{s}^b} : \{0,1\}^{L_b(\mathbf{s}^b)} \to \mathcal{X}^n \quad (4)$$

  where $L_b(\mathbf{s}^b) \triangleq \sum_{i=1}^{b} \ell_i(\mathbf{s}^i)$.
- A decoder for each channel realization $\mathbf{s}^B \in \mathcal{S}^B$,

$$g_{\mathbf{s}^B} : \mathcal{Y}^{Bn} \times \mathcal{S}^B \to \{0,1\}^{L_B(\mathbf{s}^B)} \quad (5)$$

The transmitter sends $\mathbf{X}_1 \triangleq f_{\mathbf{S}^1}(\mathbf{U}_1)$ with n channel uses in block 1, where $\mathbf{U}_1 \triangleq (W_1, W_2, \ldots, W_{\ell_1(\mathbf{S}^1)})$ denotes the first $\ell_1(\mathbf{S}^1)$ bits of the sequence $\mathbf{W}$. In block 2, the transmitter sends $\mathbf{X}_2 \triangleq f_{\mathbf{S}^2}(\mathbf{U}_1, \mathbf{U}_2)$ with n channel uses, where $\mathbf{U}_2 \triangleq (W_{\ell_1(\mathbf{S}^1)+1}, \ldots, W_{\ell_1(\mathbf{S}^1)+\ell_2(\mathbf{S}^2)})$ denotes the next ℓ_2 bits from $\mathbf{W}$. The process repeats: in block $b = 2, 3, \ldots, B$, the transmitter sends $\mathbf{X}_b \triangleq f_{\mathbf{S}^b}(\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_b)$, where

$$\mathbf{U}_b \triangleq (W_{L_{b-1}+1}, W_{L_{b-1}+2}, \ldots, W_{L_b}).$$

The key idea here is that the transmitter adapts the number ℓ_b of "fresh" bits it encodes in block b, depending on the current and previous channel states $S_1, S_2, \ldots, S_b$.

At the end of block B, the receiver estimates the first L_B bits of $\mathbf{W}$ via

$$\widehat{\mathbf{U}}_1^B \triangleq g_{\mathbf{S}^B}(\mathbf{Y}_1^B, \mathbf{S}^B), \quad (6)$$

where $\widehat{\mathbf{U}}_1^B$ denotes the receiver's estimate of $(\mathbf{U}_1, \ldots, \mathbf{U}_B)$.

B. Performance Metrics

The performance of the above scheme is characterised, in part, by the average block error probability

$$P_e(\boldsymbol{\ell}) \triangleq \mathbb{P}\left[\widehat{\mathbf{U}}_1^B \neq \mathbf{U}_1^B\right] \quad (7)$$

and the average throughput

$$T(\boldsymbol{\ell}) \triangleq \frac{1}{nB} \mathbb{E}_{\mathbf{S}^B}\left[L_B(\mathbf{S}^B)\big(1 - P_e(\mathbf{S}^B, \boldsymbol{\ell}(\mathbf{S}^B))\big)\right], \quad (8)$$

where

$$P_e(\mathbf{s}^B, \boldsymbol{\ell}(\mathbf{s}^B)) \;\triangleq\; \mathbb{P}\left[\widehat{\mathbf{U}}_1^B \neq \mathbf{U}_1^B | \mathbf{S}^B = \mathbf{s}^B\right] \quad (9)$$

denotes the average probability of error conditioned on a given state sequence $\mathbf{s}^B \in \mathcal{S}^B$. The throughput $T(\boldsymbol{\ell})$ is the average information rate (in bits per channel use) that is successfully conveyed over the channel. We are interested in finding the optimal adaptive strategy $\boldsymbol{\ell}^\star$ that maximizes the throughput $T(\boldsymbol{\ell})$, subject to an average error probability constraint:

$$\begin{cases} \text{Maximize} & T(\boldsymbol{\ell}) \\ \text{Subject to} & P_e(\boldsymbol{\ell}) \leq \varepsilon. \end{cases} \quad (10)$$

To solve (10) we form the Lagrangian

$$M(\lambda, \boldsymbol{\ell}) \triangleq \mathbb{E}_{\mathbf{S}^B}\big[L_B(\mathbf{S}^B)(1 - P_e(\mathbf{S}^B, \boldsymbol{\ell}(\mathbf{s}^B))) - \lambda P_e(\mathbf{S}^B, \boldsymbol{\ell}(\mathbf{S}^B))\big] \quad (11)$$

and denote

$$\boldsymbol{\ell}_\lambda^\star \triangleq \arg\max_{\boldsymbol{\ell}} M(\lambda, \boldsymbol{\ell}). \quad (12)$$

Following [10, Proposition 3.3.4], a solution to (10) is $\boldsymbol{\ell}_{\lambda^\star}^\star$ where $\lambda^\star \geq 0$ is chosen such that

$$\lambda^\star(P_e(\boldsymbol{\ell}_{\lambda^\star}^\star) - \varepsilon) = 0. \quad (13)$$

Therefore solving (10) reduces to searching for a $\lambda^\star$ satisfying (13). This requires efficiently solving (12) for any $\lambda \geq 0$. In the sequel, we propose the dynamic programming approach to solve (12). Dependencies on λ are made implicit for notational convenience.

C. Dynamic optimization approach

For a given λ, (12) can be solved sequentially using dynamic programming. Specifically, since the rate-adaptation strategy $\boldsymbol{\ell}$ causally depends on the channel states, maximizing (11) is equivalent to (14) at the top of the next page.

Defining $(\boldsymbol{\ell}^b, \mathbf{s}^b)$ as the *state* of the transmission at the end of block b, the optimal $\widehat{\ell}_{b+1}$ given each $(\boldsymbol{\ell}^b, \mathbf{s}^b)$ and s_{b+1} can be solved sequentially for $b = B-1, \ldots, 0$ as

$$\widehat{\ell}_{b+1}(\boldsymbol{\ell}^b, \mathbf{s}^{b+1}) = \arg\max_{\ell_{b+1}} \left\{\widehat{M}_{b+1}(\boldsymbol{\ell}^{b+1}, \mathbf{s}^{b+1}))\right\}, \quad (15)$$

where

$$\widehat{M}_B(\boldsymbol{\ell}^B, \mathbf{s}^B) = (1 - P_e(\mathbf{s}^B, \boldsymbol{\ell})) \sum_{b=1}^{B} \ell_b - \lambda P_e(\mathbf{s}^B, \boldsymbol{\ell}) \quad (16)$$

$$\widehat{M}_b(\boldsymbol{\ell}^b, \mathbf{s}^b) = \sum_{s_{b+1}} f_S(s_{b+1}) \widehat{M}_{b+1}([\boldsymbol{\ell}^b \; \widehat{\ell}_{b+1}], \mathbf{s}^{b+1}). \quad (17)$$

Then the optimal strategy is

$$\ell_b^\star(\mathbf{s}^b) = \widehat{\ell}_b(\boldsymbol{\ell}^{\star b-1}(\mathbf{s}^{b-1}), \mathbf{s}^b) \quad (18)$$

for $b = 1, \ldots, B$ where $\boldsymbol{\ell}^{\star b}(\mathbf{s}^b) \triangleq [\ell_1^\star(\mathbf{s}^1) \ldots \ell_b^\star(\mathbf{s}^b)]$.

The dynamic programming approach reduces the multivariate optimization problem in (12) to a series of one-dimensional searches. However, the size of the state space $\mathcal{S}^b\mathcal{L}^b$, and thus the number of searches in block b, grows as $(|\mathcal{S}||\mathcal{L}|)^b$. The complexity can be prohibitive even for relatively small B. In the next section, we consider optimising the rate-adaptation strategy $\boldsymbol{\ell}$ for sequential random codes and illustrate a technique to reduce the complexity of solving (12).

$$\max_{\boldsymbol{\ell}} M(\lambda,\boldsymbol{\ell}) = \sum_{s_1} f_S(s_1) \max_{\ell_1} \left(\sum_{s_2} f_S(s_2) \max_{\ell_2} \left(\dots \sum_{s_B} f_S(s_B) \max_{\ell_B} \left(\left(\sum_{b=1}^{B} \ell_b \right) (1 - P_e(\mathbf{s}^B,\boldsymbol{\ell})) - \lambda P_e(\mathbf{s}^B,\boldsymbol{\ell}) \right) \dots \right) \right). \quad (14)$$

III. Rate Adaptation for Sequential Random Codes

A. Sequential Random Codes

We concentrate on the achievable performance of the proposed rate adaptation scheme via rate-adaptive random sequential tree codes. For achievable results, we focus on random sequential codes [11] with the encoding function f_b at block b defined as

$$f_b : \{0,1\}^{b\times \ell_m} \to \mathcal{X}^n \quad (19)$$

$$f_b(\underline{\mathbf{U}}_1^b) = \mathbf{X}_b, \quad (20)$$

where the components of $\mathbf{X}_b$ are i.i.d. drawn from a distribution $Q_X(x)$ and ℓ_m is the maximum number of new bits that can be included in block b. Without loss of generality we can choose $\ell_m = \lfloor nB \max_{s\in\mathcal{S}} I(Q_s) \rfloor$, where $I(Q_s)$ is the input-output mutual information (in bits per channel use) of a channel with transition probabilities Q_s and input distribution Q_X [12].

To realize a rate-adaptation strategy $\boldsymbol{\ell}$ the transmitter forms $\underline{\mathbf{U}}_b = [\mathbf{U}_b, 0, \dots, 0] \in \{0,1\}^{\ell_m}$ at block b, where $\mathbf{U}_b$ is defined in Section II-A, and sends $\mathbf{X}_b = f_b(\underline{\mathbf{U}}_1^b)$. For a channel realisation $\mathbf{s}^B$ and received signal $\mathbf{y}^B$, the maximum likelihood decoder outputs[1]

$$\widehat{\underline{\mathbf{u}}}^B = \mathop{\arg\max}_{\substack{\underline{\mathbf{u}}_b \in \{[\mathbf{u},0,\dots,0]:\mathbf{u}\in\{0,1\}^{\ell_b(\mathbf{s}^b)}\} \\ b=1,\dots,B}} \mathbb{P}\left[\mathbf{y}^B | f_1(\underline{\mathbf{u}}_1), \dots, f_B(\underline{\mathbf{u}}_1^B), \mathbf{s}^B\right].$$

and recovers an estimate $\widehat{\mathbf{u}}^B$ of $\mathbf{U}^B$.

Note that the mechanism of constructing the code (random coding) does not depend on the channel state, while the message set and the transmitted codewords depend on the channel realization through the rate-adaptation strategy. Thus, we have partly separated the code design from the rate-adaptation strategy. This approach is significantly simpler than optimising the general scheme described in Section II-A.

Given a channel realization $\mathbf{s}^B$ and a transmission strategy $\boldsymbol{\ell}$, the error performance of the sequential random codes can be bounded as follows [11]

$$\begin{aligned} P_e(\mathbf{s}^B,\boldsymbol{\ell}) &= \mathbb{P}\left[\widehat{\underline{\mathbf{U}}}_1^B \neq \underline{\mathbf{U}}_1^B | \mathbf{S}^B = \mathbf{s}^B\right] \\ &\le \sum_{b=1}^{B} \mathbb{P}\left[\widehat{\underline{\mathbf{U}}}_b \neq \underline{\mathbf{U}}_b | \widehat{\underline{\mathbf{U}}}_1^{b-1} = \underline{\mathbf{U}}_1^{b-1}, \mathbf{S}^B = \mathbf{s}^B\right] \\ &\triangleq \sum_{b=1}^{B} P_b(\mathbf{s}^B,\boldsymbol{\ell}). \end{aligned} \quad (21)$$

For two messages $\underline{\mathbf{U}}_1^B$ and $\widehat{\underline{\mathbf{U}}}_1^B$ such that $\underline{\mathbf{U}}_1^{b-1} = \widehat{\underline{\mathbf{U}}}_1^{b-1}$ and $\underline{\mathbf{U}}_b \neq \widehat{\underline{\mathbf{U}}}_b$, the corresponding codewords satisfies $\mathbf{X}_1^{b-1} = \widehat{\mathbf{X}}_1^{b-1}$ while $\mathbf{X}_b^B$ and $\widehat{\mathbf{X}}_b^B$ are mutually independent. Therefore the error probability $P_b(\mathbf{s}^B, \boldsymbol{\ell}(\mathbf{s}^B))$ is the same as the error probability of transmission over a $(B-b+1)$–block channel with states $[s_b, \dots, s_B]$ using random block codes with rate $\frac{\sum_{i=b}^{B} \ell_i}{(B-b+1)n}$. Numerous bounds on the achievable performance of random block codes exist in the literature (see [7], [13], [14] and references therein). The results in [13], [14] are readily extended to our case as follows.

Lemma 1 (Saddle-point approximation [14]): The error probability $P_b(\mathbf{s}^B,\boldsymbol{\ell})$ can be approximated by

$$P_b(\mathbf{s}^B,\boldsymbol{\ell}) \approx \alpha(c_1,c_2) \exp\left(-n\sum_{i=b}^{B} E_0(s_i,\widehat{\rho}) - \widehat{\rho}\ell_i\right), \quad (22)$$

where

$$E_0(s_i,\rho) = -\ln \sum_{y\in\mathcal{Y}} \left(\sum_{x\in\mathcal{X}} Q_X(x) Q_{s_i}(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho} \quad (23)$$

$$\widehat{\rho} = \mathop{\arg\max}_{\rho\in[0,1]} \left\{ n\sum_{i=b}^{B} E_0(s_i,\rho) - \rho\ell_i \right\}$$

$$c_1 = \sum_{i=b}^{B} \ell_i - n \frac{\partial E_0(s_i,\rho)}{\partial\rho}\bigg|_{\rho=\widehat{\rho}}$$

$$c_2 = -n\sum_{i=b}^{B} \frac{\partial^2 E_0(s_i,\rho)}{\partial\rho^2}\bigg|_{\rho=\widehat{\rho}}$$

$$\alpha(c_1,c_2) = \int_0^{\infty} e^{-\widehat{\rho}z}\phi(z;c_1,c_2)dz + \int_{-\infty}^{0} e^{(1-\widehat{\rho})z}\phi(z;c_1,c_2)dz$$

and $\phi(z;\mu,\sigma^2)$ is the pdf of a Gaussian random variable with mean μ and variance σ^2.

Lemma 1 gives a tight approximation for the error probability $P_b(\mathbf{s}^B,\boldsymbol{\ell})$ for a wide range of channel states, strategies $\boldsymbol{\ell}$ and block-length n. Meanwhile, letting $\alpha(c_1,c_2) = 1$ in (22) returns the simpler Gallager's upper bound on error probability [13]. Gallager's bound is not tight when the transmission rate is close to capacity. We will exploit the simplicity of Gallager's bound to solve the optimziation problem in (10), and use Lemma 1 to evaluate the achievable performance of the obtained strategy.

B. Dynamic Programming with Reduced State Space

Directly exploiting the bounds (21) and (22) in solving (10) requires the same complexity as the general case. We first propose an approximation for the error probability $P_e(\mathbf{s}^B,\boldsymbol{\ell})$, aiming at reducing the complexity of solving (11).

Following (21), (22) with $\alpha(c_1,c_2) = 1$, we can write

$P_e(\mathbf{s}^B,\boldsymbol{\ell})$

[1] When multiple messages achieve the maximum, the decoder randomly chooses one of the messages.

$$\leq \sum_{b=1}^{B} \min_{\rho_i \in [0,1]} \exp\left(-n\sum_{i=b}^{B} E_0(s_i,\rho_i) + \rho_i \ell_i\right)$$
$$\leq B \max_{b\in\{1,\dots,B\}} \exp\left(-\max_{\rho\in[0,1]} \left\{ n\sum_{i=b}^{B} E_0(s_i,\rho) - \rho\ell_i \right\}\right)$$
$$\stackrel{(a)}{\leq} B \exp\left(-\max_{j\in\mathcal{J}} \min_{b\in\{1,\dots,B\}} \left\{ n\sum_{i=b}^{B} E_0(s_i,\rho_j) - \rho_j\ell_i \right\}\right)$$
$$\triangleq B\tilde{P}_e(s^B,\boldsymbol{\ell}) \tag{24}$$

for some $\mathcal{J} \triangleq \{1,\dots,|\mathcal{J}|\}$ and $\rho_j \in [0,1]$. In (a) we have simplified the maximization over $\rho \in [0,1]$ to maximizing over $j \in \mathcal{J}$ for tractability. Numerical results show that $\mathcal{J} = \{1,2\}$ is sufficient to obtain good adaptation strategies. A suitable choice of $\{\rho_1,\dots,\rho_{|\mathcal{J}|}\}$ will depend on parameters such as the fading distribution $f_S(s)$, the block length n and the target average error probability ε.

For $b \in \{1,\dots,B\}$ and $j \in \mathcal{J}$, define

$$a_{b,j} = \min_{k\in\{1,\dots,b\}} \left\{ \sum_{i=k}^{b} nE_0(s_i,\rho_j) - \rho_j\ell_i \right\}. \tag{25}$$

We can then easily verify that

$$a_{b+1,j} = \min\{a_{b,j},0\} + nE_0(s_{b+1},\rho_j) - \rho_j\ell_{b+1}. \tag{26}$$

Therefore, substituting $\tilde{P}_e(\mathbf{s}^B,\boldsymbol{\ell})$ for $P_e(\mathbf{s}^B,\boldsymbol{\ell})$, the Lagragian function in (11) is approximated by

$$\tilde{M}(\lambda,\boldsymbol{\ell}) \triangleq \mathbb{E}_{\mathbf{S}^B}\left[L_B(\mathbf{S}^B)\left(1-\exp\left(-\max_{j\in\mathcal{J}} a_{B,j}\right)\right) -\lambda\exp\left(-\max_{j\in\mathcal{J}} a_{B,j}\right)\right]$$

Similar to the approach in Section II-C, define the state of the transmission at the end of block b as $\mathbf{a}_b = (a_{b,0},a_{b,1},\dots,a_{b,|\mathcal{J}|}) \in \mathbb{R}^{|\mathcal{J}|+1}$ where $a_{b,0} \triangleq \sum_{i=1}^{b}\ell_i$ and $a_{b,j}, j\in\mathcal{J}$ are defined as in (25). Then we can sequentially solve for the optimal $\tilde{\ell}_{b+1}$ given each state $\mathbf{a}_b$ as

$$\tilde{\ell}_{b+1}(\mathbf{a}_b,s_{b+1}) = \arg\max_{\ell_{b+1}} \left\{\tilde{M}_{b+1}(\mathbf{a}_{b+1}(\mathbf{a}_b,s_{b+1},\ell_{b+1}))\right\}$$

where

$$\tilde{M}_B(\mathbf{a}_B) \triangleq a_{b,0}\left(1-\exp\left(-\max_{j\in\mathcal{J}} a_{B,j}\right)\right) - \lambda\exp\left(-\max_{j\in\mathcal{J}} a_{B,j}\right) \tag{27}$$
$$\tilde{M}_b(\mathbf{a}_b) \triangleq \sum_{s_{b+1}} f_S(s_{b+1})\tilde{M}_{b+1}(\mathbf{a}_{b+1}(\mathbf{a}_b,s_{b+1},\tilde{\ell}_{b+1})) \tag{28}$$

and $\mathbf{a}_{b+1}(\mathbf{a}_b,s_{b+1},\ell_{b+1})$ is given by

$$a_{b+1,0} = a_{b,0} + \ell_{b+1},$$
$$a_{b+1,j} = \min\{a_{b,j},0\} + E_0(s_{b+1},\rho_j) - \rho_j\ell_{b+1}, j\in\mathcal{J}$$

By reasonably bracketing and quantizing $\mathbf{a}_b$, the state space grows polynomially with B and linearly with $|\mathcal{L}|$. The suboptimal strategy is significantly simpler than the optimal one in Section II-C, especially when either B or $|\mathcal{L}|$ grows large.

The suboptimal solution for (11) is therefore

$$\mathbf{a}_0 = (0,0,0) \tag{29}$$
$$\tilde{\ell}^\star_b(s^b) = \tilde{\ell}_{b-1}(\mathbf{a}^\star_{b-1}(s^{b-1}),s_b) \tag{30}$$
$$\mathbf{a}^\star_b(s^b) = \mathbf{a}_b(\mathbf{a}^\star_{b-1}(s^{b-1}),s_b,\tilde{\ell}^\star_b(s^b)). \tag{31}$$

IV. Numerical Results

For benchmarking purposes, we consider three trivial adaptation schemes:

- *Full non-causal CSIT*: the channel states of all B blocks are available prior to transmission. The optimal adaptation strategy is obtained by solving

$$\text{Maximize} \sum_{s^B} f_S(\mathbf{s}^B)\left(\sum_{\ell=1}^{B}\ell_b(\mathbf{s}^B)\right)(1-P_e(\mathbf{s}^B,\boldsymbol{\ell})) - \lambda^\star P_e(\mathbf{s}^B,\boldsymbol{\ell}) \tag{32}$$

  where $\lambda^\star$ satisfies (13). It can be verified that the optimal solution satisfies $\ell_b(\mathbf{s}^B) = 0, b = 2,\dots,B$. Thus for each $\lambda > 0$, the optimal solution is $\overline{\boldsymbol{\ell}}(\ell^\star_1(\mathbf{s}^B))$ where $\overline{\boldsymbol{\ell}}(\ell_1) \triangleq [\ell_1,0,\dots,0]$ and $\ell^\star_1(\mathbf{s^B})$ solves

$$\max_{\ell_1} f_S(\mathbf{s}^B)\ell_1(1-P_e(\mathbf{s}^B,\overline{\boldsymbol{\ell}}(\ell_1)) - \lambda P_e(\mathbf{s}^B,\overline{\boldsymbol{\ell}}(\ell_1))$$

  The throughput for non-causal CSIT serves as an upper bound on the performance of the proposed scheme.
- *Block-wise coding*: A straightforward and trivial adaptation rule is to transmit independent codewords for each fading block. The system can be treated as a special case of the proposed scheme for $B = 1$. The problem is much simpler than the proposed scheme, however the significantly shorter code leads to much worse throughput error probability tradeoff [7].
- *Non adaptive transmission:* ignore the channel state information and choose the transmission rule $\overline{\boldsymbol{\ell}}(\ell^\star_1)$ where

$$\ell^\star_1 = \arg\max_{\ell} \sum_{s^B} f_S(\mathbf{s}^B)\ell_1(1-P_e(\mathbf{s}^B,\overline{\boldsymbol{\ell}}(\ell_1)) - \lambda^\star P_e(\mathbf{s}^B,\overline{\boldsymbol{\ell}}(\ell_1)) \tag{33}$$

  and $\lambda^\star$ satisfies (13).

Recall that the suboptimal causal rate adaptation scheme is solved using the Gallager's bound (setting $\alpha(c_1,c_2) = 1$ in (22)). However we use Lemma 1 to derive the benchmarking schemes as well as evaluating the performance of all strategies.

To illustrate the performance of the proposed scheme, we consider a binary-symmetric block-fading channel, where block $b \in \{1,\dots,B\}$ is a BSC with crossover probability s_b. Specifically, $\mathcal{X} = \mathcal{Y} = \{0,1\}$ and the transition probability in (1) can be written as

$$Q_{s_b}(y_{b,i}|x_{b,i}) = \begin{cases} 1-s_b, & y_{b,i} = x_{b,i} \\ s_b, & \text{otherwise.} \end{cases} \tag{34}$$

We consider a block-fading channel with $B = 3$ and the following fading distribution

$$\mathbb{P}[S = 0.1/3] = \mathbb{P}[S = 0.2/3] = \mathbb{P}[S = 0.3/3] = 1/3. \tag{35}$$

The ergodic capacity of the channel is 0.656 bits per channel use. We consider sequential random codes, as described in Section III-A and the causal adaptation strategies are obtained from the algorithm in Section III-B with $\mathcal{J} = \{1, 2\}$.

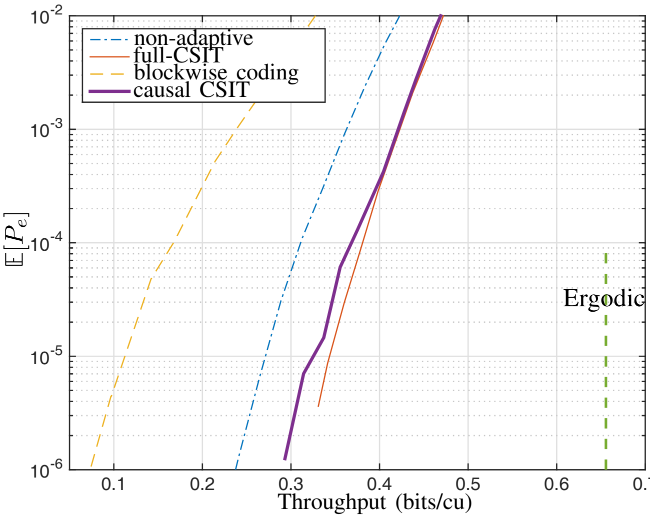

Fig. 1. Throughput versus error probability achieved by various transmission rules over a binary symmetric channel with fading distribution in (35). The channel has $B = 3$ blocks with block length $n = 50$.

Figure 1 illustrates the performance of various transmission schemes for block length $n = 50$. The causal adaptive rule is obtained by choosing $(\rho_1, \rho_2) = (1, 0.5)$. Despite that only causal CSIT is available, the proposed suboptimal scheme operates quite close to the adaptation rule with full CSIT. Compared to non-adaptive transmission, the proposed adaptation rule for causal CSIT provides a gain of 0.05 bits per channel use at error probability 10^{-4}. In contrast, the adaptation rule with block-wise coding performs worse than non-adaptive transmission. For this particular setting, especially the small block length, the gain from adaptation is insignificant compared to the penalty due to short codes.

The performances of the transmission schemes for block-length $n = 10^3$ are illustrated in Figure 2. The causal adaptive rule is obtained by choosing $(\rho_1, \rho_2) = (0.15, 0.05)$. In this case, the throughputs achieved by all transmission scheme are closer to the ergodic capacity compared to the case $n = 50$, as expected. Note that the performance of all schemes, except for the non-adaptive one, should approach the ergodic capacity when $n \to \infty$. The penalty of using block-wise coding is not as severe as in the previous case. Thus the block-wise adaptive transmission achieve significant gains compared to non-adaptive transmission. The block-wise scheme is still substantially inferior to the proposed scheme, which performs close to the full CSIT benchmark.

V. Conclusions

In this paper we have proposed a rate adaptive coding scheme for the block-fading channel with causal CSIT. We aim at maximizing the throughput subject to an average error probability constraint. The optimization problem can be solved

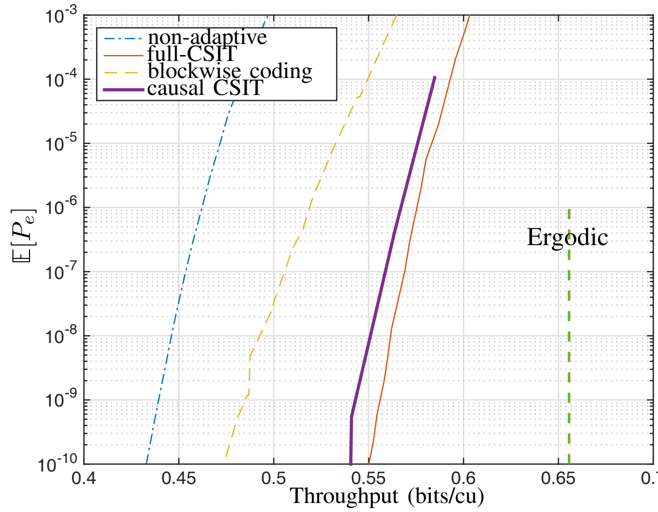

Fig. 2. Throughput versus error probability achieved by various transmission rules over a binary symmetric channel with fading distribution in (35). The channel has $B = 3$ blocks with block length $n = 1000$.

via dynamic programming. A suboptimal scheme has been derived for a sequential random coding scheme. The proposed scheme achieves significant throughput gains compared to existing trivial adaptation schemes. This motivates future development of practical causal rate adaptive codes, as well as developing strategies for rate adaptation with practical codes.

References

[1] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Informatic-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[2] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[3] G. Caire, G. Taricco, and E. Biglieri, "Optimal power control over fading channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, Jul. 1999.

[4] R. Negi and J. M. Cioffi, "Delay-constrained capacity with causal feedback," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, , Sep. 2002.

[5] K. D. Nguyen, A. Guillén i Fàbregas, and L. K. Rasmussen, "Power allocation for block-fading channels with arbitrary input constellations," *IEEE Trans. on Wireless Comm.*, vol. 8, no. 5, May 2009.

[6] A. Lozano, A. M. Tulino, and S. Verdú, "Opitmum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, Jul. 2006.

[7] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel Coding Rate in the Finite Blocklength Regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[8] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Diversity versus Channel Knowledge at Finite Block-Length," in *Inf. Theory Workshop*, 2013, pp. 572–576.

[9] K. F. Trillingsgaard and P. Popovski, "Block-fading channels with delayed CSIT at finite blocklength," *CoRR*, vol. abs/1401.7134, 2014.

[10] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Athena Scientific, Belmont, Massachusetts, 1999.

[11] A. Sahai, "Anytime information theory," Ph.D. dissertation, Massachusetts Institute of Technology, 2001.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley and Sons, 2006.

[13] R. G. Gallager, *Information theory and reliable communication.* John Wiley and Sons, 1968.

[14] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "The saddlepoint approximation: A unification of exponents, dispersions and moderate deviations," *CoRR*, vol. abs/1402.3941, 2014.