Optimizing Pilot Length for a Go/No-Go Decision in Two-State Block Fading Channels with Feedback

Chung-Yu Lou, Babak Daneshrad, and Richard D. Wesel
Department of Electrical Engineering, University of California, Los Angeles, California 90095–1594
Email: chungyulou@ucla.edu, {babak, wesel}@ee.ucla.edu

***Abstract*—We propose an approach where each user independently seeks to minimize the amount of time that they occupy the channel. Essentially, we seek to minimize the number of transmitted symbols required to communicate a packet assuming variable-length coding with feedback. Users send a pilot sequence to estimate the channel quality and decide whether to proceed with a transmission or wait for the next opportunity. Thus a user may choose to leave the channel even though it has already gained access, in order to increase the network throughput and also save its own energy resources. This paper optimizes the number of pilots and the channel identification threshold to minimize the total number of transmitted symbols (including pilots) required to communicate the packet. We prove a sufficient condition for the optimal pilot length and the channel identification threshold. This optimal parameter pair is solved numerically and the reduction in channel occupancy is shown for various channel settings.**

I. Introduction

This paper considers a multi-user wireless network in which each user communicates packets using variable-length coding with feedback. This setting is general in that it includes a range of techniques from simply repeating packets (with or without Chase combining [1]) based on ACKs and NACKs to schemes with decoding attempts after every received symbol with potentially frequent feedback to the transmitter. Our interest lies in a distributed network where global channel state information (CSI) is not available to the users and each user must learn their own channel state. Furthermore, we consider applications such as a sensor network where resource management may take precedence over fairness concerns.

Techniques using feedback such as the incremental redundancy (IR) scheme proposed in [2] have long been used to enhance link robustness. Recently, IR combined with hybrid automatic repeat request (HARQ) has gained significant attention [3]–[7]. In addition to improving the robustness of a link, feedback can also be used to inform the transmitter that the channel is bad enough that the communication attempt should be discontinued until a fade abates. Rather than struggling in a deep fade or extremely noisy channel, postponing transmission until the channel improves releases the channel for other users in a multi-user setting.

In a distributed multi-user wireless network, the multiple access protocol is in general based on fairness, such as slotted ALOHA or carrier sense multiple access with collision avoidance (CSMA/CA). In this case, the network throughput is reduced by a pair of transmitter (TX) and receiver (RX) with low signal-to-noise ratio (SNR) accessing the channel. Overall network utility can be improved by sacrificing some fairness to include the realities of the channel quality by prioritizing the users in part based on their SNRs.

However, the SNR is usually not known before sending the pilots. Therefore, we consider a mechanism whereby a user can choose to leave the channel once it believes its SNR is below a certain threshold. When a TX starts to access the channel, it first sends pilots to estimate the SNR, and the RX will respond whether the data transmission should be attempted. If the user decides to leave the channel, it will wait for at least a channel coherence time. By doing so, the user can reduce its energy consumption while also increasing the total network throughput.

An accurate channel estimation relies on sufficient training pilots. However, sending too many pilots unacceptably lowers the rate of that transmission. This trade-off is discussed in [8], [9], where the total duration of training and data transmission is bounded by the channel coherence time. In [10], [11], non-ideal CSI feedback is considered and the duration of the RX sending the CSI back to the TX is also taken into account. In this case, the total duration of training, CSI feedback, and data transmission is limited by the channel coherence time.

In this work, we consider a two-state block fading channel where the channel coefficient remains constant during the training and data transmission. Because of the IR scheme, even though the message length is constant, the transmitted packet has a variable length because the duration required to transmit successfully depends on the SNR. Therefore, we will minimize the number of transmitted symbols (the channel usage) required by a packet transmission, which includes both the initial and any subsequent pilot sequences and the variable length of the data transmission itself. Thus by optimizing the pilot length and the threshold to leave the channel, the user is capable of smart channel selection with a relatively low cost of training. We prove that the optimal parameter pair can be solved numerically and provide a procedure to compute it. The channel usage reduction of avoiding deep fades is shown for various channel settings.

The paper is organized as follows: Section II describes the system model. Section III proves the optimality of our

This material is based upon work supported by the National Science Foundation under Grant 1162501. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

procedure for computing the optimal parameters for the pilot length and the threshold to leave the channel. Section IV shows the reduction in channel occupancy achieved by avoiding bad channels through training. Section V concludes this paper.

II. System Model

When the TX starts to access the channel, it first sends n training symbols to estimate the SNR of the block fading channel for this block. In a block fading channel, the training symbols as well as the data symbols in the same block experience the same fading coefficient. This coefficient is independent of all the previous blocks. In this paper, we consider a two-state Gaussian channel, i.e. the fading coefficient has two possible realizations, and the received symbols are impacted by additive white Gaussian noise (AWGN).

We assume that the RX knows the noise power by monitoring the background noise while no signal is present, and the received signal is properly scaled such that the noise has unit power. Let the i^{th} received training symbol be

$$Y_i = X + N_i, \tag{1}$$

where X is a positive random variable which depends on the channel fading coefficient and $N_i \sim \mathcal{N}(0,1)$ is the noise. Let the two possible realizations of X be x_{G} and x_{B} with probabilities p_{G} and p_{B}, where $p_{\text{G}} + p_{\text{B}} = 1$, representing the good and bad states of the channel, respectively. In other words, the SNRs of the good and bad states are x_{G}^2 and x_{B}^2, where $x_{\text{G}} > x_{\text{B}}$. The statistics of the channel $(p_{\text{G}}, p_{\text{B}}, x_{\text{G}}, x_{\text{B}})$ are known, and the realization of X is to be estimated. Note that, if either $p_{\text{G}} = 0$, $p_{\text{B}} = 0$, or $x_{\text{G}} = x_{\text{B}}$, then this two-state channel is essentially a static channel and needs no training at all. Therefore, we only consider the non-trivial cases.

Having n measures of the random variable X affected by the Gaussian noise, the RX then uses the sufficient statistic

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{2}$$

to determine whether the data transmission should begin by comparing it with a threshold T. If $\bar{Y} \geq T$, the RX believes that the channel is in the good state and allows the TX to initiate the data transmission. Otherwise, the RX asks the TX to leave the channel and wait for the next block. On the next block that the TX gains access to the channel, the training process starts over again because the channel realization in the new block is independent of the blocks observed in previous attempts.

In this study, we assume for simplicity that the feedback is noiseless and instantaneous, but note that this feedback is of a single bit of information and is required only once per transmission attempt so that this assumption neglects a small fixed overhead. Since $\bar{Y} \sim \mathcal{N}\left(X, \frac{1}{n}\right)$, the conditional probabilities of initiating the data transmission while the channel is in the good and bad states are given by

$$\begin{aligned}
\mathrm{P}\big(\bar{Y} \geq T \big| X = x_{\text{G}}\big) &= \mathrm{Q}\big(\sqrt{n}\,(T - x_{\text{G}})\big) \\
\mathrm{P}\big(\bar{Y} \geq T \big| X = x_{\text{B}}\big) &= \mathrm{Q}\big(\sqrt{n}\,(T - x_{\text{B}})\big),
\end{aligned}$$

where $\mathrm{Q}(\cdot)$ is the Gaussian Q-function

$$\mathrm{Q}(x) = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{t^2}{-2}\right) \mathrm{d}t.$$

During the data transmission, the TX sends a packet containing k bits. With the IR scheme, the TX keeps sending data symbols until the RX decodes the packet successfully and sends an acknowledgement. Let the expected channel usage of the data transmission given the channel being in the good and bad states be τ_{G} and τ_{B}, respectively, where $0 < \tau_{\text{G}} < \tau_{\text{B}}$. We denote $\bar{\tau}$ as the total expected channel usage to deliver a packet. This includes one or more length-n pilot sequences. Hence, $\bar{\tau}$ is calculated as

$$\begin{aligned}
\bar{\tau} = n &+ \mathrm{P}\big(\bar{Y} \geq T \big| X = x_{\text{G}}\big)\, p_{\text{G}} \tau_{\text{G}} \\
&+ \mathrm{P}\big(\bar{Y} \geq T \big| X = x_{\text{B}}\big)\, p_{\text{B}} \tau_{\text{B}} + \mathrm{P}\big(\bar{Y} < T\big)\, \bar{\tau}.
\end{aligned} \tag{3}$$

Subtracting the last term of (3) from both sides produces

$$\bar{\tau} = \frac{n + p_{\text{B}}\, \mathrm{Q}(\sqrt{n}\,(T - x_{\text{B}}))\, \tau_{\text{diff}}}{p_{\text{G}}\, \mathrm{Q}(\sqrt{n}\,(T - x_{\text{G}})) + p_{\text{B}}\, \mathrm{Q}(\sqrt{n}\,(T - x_{\text{B}}))} + \tau_{\text{G}}, \tag{4}$$

where $\tau_{\text{diff}} = \tau_{\text{B}} - \tau_{\text{G}} > 0$.

Our goal is to minimize $\bar{\tau}$ by picking the optimal parameter pair $(n^\star, T^\star)$. In Section IV, we will compare $\bar{\tau}\big|_{n=n^\star, T=T^\star}$ with $\bar{\tau}_{\text{ref}}$. The referenced channel usage $\bar{\tau}_{\text{ref}}$ is defined as the expected channel usage to deliver a packet without training and given by

$$\bar{\tau}_{\text{ref}} = \bar{\tau}\big|_{n=0} = p_{\text{G}}\tau_{\text{G}} + p_{\text{B}}\tau_{\text{B}}. \tag{5}$$

In (4), it is clear that τ_{G} is a lower bound for $\bar{\tau}$. Achieving this lower bound requires identifying the low-SNR channel perfectly $\left(\mathrm{P}\big(\bar{Y} \geq T \big| X = x_{\text{B}}\big) = 0\right)$ with absolutely no cost $(n = 0)$. Given the lower bound for $\bar{\tau} = \tau_{\text{G}}$, we further define a corresponding upper bound on the expected reduction in channel occupancy as

$$\Delta\bar{\tau} = \bar{\tau}_{\text{ref}} - \tau_{\text{G}} = p_{\text{B}}\tau_{\text{diff}}. \tag{6}$$

III. Optimal Pilot Length and Threshold

Although our target n is a non-negative integer, (4) can be evaluated for any non-negative real n. Thus, to simplify the derivation, our procedure identifies the optimal integer $n^\star$ by examining the integers neighboring the real number $\hat{n}$ that minimizes $\bar{\tau}$. The selected integer is the optimal integer $n^\star$ because of the continuity of (4).

The discussion proceeds as follows: A sufficient condition for a local minimum facilitates a proof that there is only one local minimum, which is the global minimum. Then, solving numerically for the optimal nonnegative real $\hat{n}$ leads to the optimal integer $n^\star$ and optimal pair $(n^\star, T^\star)$.

A. Sufficient Condition for a Local Minimum

Considering the following extreme points establishes that there are no minima on the boundaries:

- As $n \to \infty$, $\bar{\tau} \to \infty$ and cannot be a local minimum.
- As $n > 0$ and $T \to \infty$, we have $\bar{\tau} \to \infty$ and it cannot be a local minimum.

- When $n > 0$ and $T \to -\infty$, we can shown that $\bar{\tau} = n + \bar{\tau}_{\text{ref}}$, which is not a local minimum since n can be reduced.
- When $n = 0$, we know $\bar{\tau} = \bar{\tau}_{\text{ref}}$. If T is finite, then

$$\frac{\partial}{\partial n}\bar{\tau}\bigg|_{n=0} = 2p_{\text{G}}p_{\text{B}}\tau_{\text{diff}}\left(1 - \frac{x_{\text{G}} - x_{\text{B}}}{2\sqrt{2\pi n}}\right)\bigg|_{n=0} < 0.$$

  So local minima do not appear on $n = 0$.

Thus the only possible local minima occur in the interior of the domain. Let the gradient and Hessian of $\bar{\tau}$ with respect to (T, n) be

$$\nabla\bar{\tau} = \begin{bmatrix} \frac{\partial}{\partial T}\bar{\tau} \\ \frac{\partial}{\partial n}\bar{\tau} \end{bmatrix} \tag{7}$$

and

$$\nabla^2\bar{\tau} = \begin{bmatrix} \frac{\partial^2}{\partial T^2}\bar{\tau} & \frac{\partial^2}{\partial T \partial n}\bar{\tau} \\ \frac{\partial}{\partial n \partial T}\bar{\tau} & \frac{\partial^2}{\partial n^2}\bar{\tau} \end{bmatrix}. \tag{8}$$

Because the gradient and Hessian are continuous, a local minimum satisfies that its gradient is a zero vector, and its Hessian is positive definite.

Let $\hat{n}$ and $\hat{T}$ be a solution of $\nabla\bar{\tau} = 0$. We will derive a relationship between $\hat{n}$ and $\hat{T}$ that shows for any such $\hat{n}$ and $\hat{T}$, $\nabla^2\bar{\tau}\big|_{n=\hat{n},T=\hat{T}} > 0$.

By evaluating $\frac{\partial}{\partial T}\bar{\tau}\big|_{n=\hat{n},T=\hat{T}} = 0$, we obtain

$$\begin{aligned}(p_{\text{G}}Q_{\text{G}} + p_{\text{B}}Q_{\text{B}})\left(p_{\text{B}}\frac{\partial}{\partial T}Q_{\text{B}}\tau_{\text{diff}}\right)\bigg|_{n=\hat{n},T=\hat{T}} \\ = (n + p_{\text{B}}Q_{\text{B}}\tau_{\text{diff}})\left(p_{\text{G}}\frac{\partial}{\partial T}Q_{\text{G}} + p_{\text{B}}\frac{\partial}{\partial T}Q_{\text{B}}\right)\bigg|_{n=\hat{n},T=\hat{T}},\end{aligned} \tag{9}$$

where we define $Q_{\text{G}} = \text{Q}(\sqrt{n}\,(T - x_{\text{G}}))$, and $Q_{\text{B}} = \text{Q}(\sqrt{n}\,(T - x_{\text{B}}))$. Similarly, from $\frac{\partial}{\partial n}\bar{\tau}\big|_{n=\hat{n},T=\hat{T}} = 0$, we get

$$\begin{aligned}(p_{\text{G}}Q_{\text{G}} + p_{\text{B}}Q_{\text{B}})\left(1 + p_{\text{B}}\frac{\partial}{\partial n}Q_{\text{B}}\tau_{\text{diff}}\right)\bigg|_{n=\hat{n},T=\hat{T}} \\ = (n + p_{\text{B}}Q_{\text{B}}\tau_{\text{diff}})\left(p_{\text{G}}\frac{\partial}{\partial n}Q_{\text{G}} + p_{\text{B}}\frac{\partial}{\partial n}Q_{\text{B}}\right)\bigg|_{n=\hat{n},T=\hat{T}}.\end{aligned} \tag{10}$$

In order to show that $\nabla^2\bar{\tau}\big|_{n=\hat{n},T=\hat{T}}$ is positive definite, we need to prove the two inequalities

$$\frac{\partial^2}{\partial T^2}\hat{\tau}\bigg|_{n=\hat{n},T=\hat{T}} > 0 \tag{11}$$

and

$$\det\nabla^2\bar{\tau}\big|_{n=\hat{n},T=\hat{T}} > 0. \tag{12}$$

Using (9) and the fact that $n + p_{\text{B}}Q_{\text{B}}\tau_{\text{diff}}\big|_{n=\hat{n},T=\hat{T}} > 0$ and $p_{\text{B}}\frac{\partial}{\partial T}Q_{\text{B}}\tau_{\text{diff}}\big|_{n=\hat{n},T=\hat{T}} < 0$, the first inequality (11) is equivalent to

$$\begin{aligned}\left(p_{\text{G}}\frac{\partial}{\partial T}Q_{\text{G}} + p_{\text{B}}\frac{\partial}{\partial T}Q_{\text{B}}\right)\left(p_{\text{B}}\frac{\partial^2}{\partial T^2}Q_{\text{B}}\tau_{\text{diff}}\right)\bigg|_{n=\hat{n},T=\hat{T}} \\ < \left(p_{\text{B}}\frac{\partial}{\partial T}Q_{\text{B}}\tau_{\text{diff}}\right)\left(p_{\text{G}}\frac{\partial^2}{\partial T^2}Q_{\text{G}} + p_{\text{B}}\frac{\partial^2}{\partial T^2}Q_{\text{B}}\right)\bigg|_{n=\hat{n},T=\hat{T}}.\end{aligned}$$

After evaluating the partial derivatives of the above expression, (11) is also equivalent to

$$\hat{n}^{\frac{3}{2}}p_{\text{G}}p_{\text{B}}\exp\left(\frac{\hat{G}^2 + \hat{B}^2}{-2}\right)(\hat{B} - \hat{G}) > 0, \tag{13}$$

where

$$\hat{G} = \sqrt{\hat{n}}\,(\hat{T} - x_{\text{G}}) \tag{14}$$

$$\hat{B} = \sqrt{\hat{n}}\,(\hat{T} - x_{\text{B}}). \tag{15}$$

Furthermore, (13) is always true because every term on the left-hand side (LHS) is positive, including $\hat{B} - \hat{G} = \sqrt{n}\,(x_{\text{G}} - x_{\text{B}}) > 0$.

Regarding the second inequality (12), we use (9) and (10) multiple times and the fact that both $n + p_{\text{B}}Q_{\text{B}}\tau_{\text{diff}}\big|_{n=\hat{n},T=\hat{T}}$ and $p_{\text{B}}\frac{\partial}{\partial T}Q_{\text{B}}\tau_{\text{diff}}\big|_{n=\hat{n},T=\hat{T}}$ are non-zero to change the Q-functions and their first order partial derivatives with respect to n into the first order partial derivatives with respect to T. The inequality (12) simplifies to

$$\begin{aligned}&\left(\frac{\partial}{\partial T}Q_{\text{G}}\frac{\partial^2}{\partial T^2}Q_{\text{B}} - \frac{\partial}{\partial T}Q_{\text{B}}\frac{\partial^2}{\partial T^2}Q_{\text{G}}\right)\bigg|_{n=\hat{n},T=\hat{T}} \\ &\cdot\left(\frac{\partial}{\partial T}Q_{\text{G}}\frac{\partial^2}{\partial n^2}Q_{\text{B}} - \frac{\partial}{\partial T}Q_{\text{B}}\frac{\partial^2}{\partial n^2}Q_{\text{G}}\right)\bigg|_{n=\hat{n},T=\hat{T}} \\ &\quad > \left(\frac{\partial}{\partial T}Q_{\text{G}}\frac{\partial^2}{\partial n \partial T}Q_{\text{B}} - \frac{\partial}{\partial T}Q_{\text{B}}\frac{\partial^2}{\partial n \partial T}Q_{\text{G}}\right)^2\bigg|_{n=\hat{n},T=\hat{T}}.\end{aligned}$$

After evaluating the partial derivatives, the above inequality is equivalent to

$$(\hat{B} - \hat{G})\left[\hat{B}\left(1 + \hat{B}^2\right) - \hat{G}\left(1 + \hat{G}^2\right)\right] > \left(\hat{B}^2 - \hat{G}^2\right)^2,$$

which is equivalent to

$$1 - \hat{G}\hat{B} > 0. \tag{16}$$

To prove (16), we must rely on a relationship between $\hat{n}$ and $\hat{T}$. Specifically, from (9), we get

$$\hat{n}\left(p_{\text{G}}\,e^{\frac{\hat{G}^2}{-2}} + p_{\text{B}}\,e^{\frac{\hat{B}^2}{-2}}\right) = p_{\text{G}}p_{\text{B}}\tau_{\text{diff}}\left[\text{Q}(\hat{G})\,e^{\frac{\hat{B}^2}{-2}} - \text{Q}(\hat{B})\,e^{\frac{\hat{G}^2}{-2}}\right]. \tag{17}$$

Multiplying (10) with $\left(p_{\text{B}}\frac{\partial}{\partial T}Q_{\text{B}}\tau_{\text{diff}}\right)/\left(n + p_{\text{B}}Q_{\text{B}}\tau_{\text{diff}}\right)\big|_{n=\hat{n},T=\hat{T}}$ and then using (9), we know

$$\hat{n}\left(p_{\text{G}}\,e^{\frac{\hat{G}^2}{-2}} + p_{\text{B}}\,e^{\frac{\hat{B}^2}{-2}}\right) = p_{\text{G}}p_{\text{B}}\tau_{\text{diff}}\,e^{\frac{\hat{G}^2 + \hat{B}^2}{-2}}\,\frac{\hat{G} - \hat{B}}{-2\sqrt{2\pi}}. \tag{18}$$

Since (17) and (18) have the same LHS, we obtain the equality

$$\hat{G} + 2\sqrt{2\pi}\,\text{Q}(\hat{G})\,e^{\frac{\hat{G}^2}{2}} = \hat{B} + 2\sqrt{2\pi}\,\text{Q}(\hat{B})\,e^{\frac{\hat{B}^2}{2}}, \tag{19}$$

where $\hat{G}$ only appears on the LHS and $\hat{B}$ only appears on the right-hand side (RHS). By introducing the function

$$f(x) = x + 2\sqrt{2\pi}\,\text{Q}(x)\exp\left(\frac{x^2}{2}\right), \tag{20}$$

we get $f(\hat{G}) = f(\hat{B})$. The second derivative of this function is given by

$$f''(x) = 2\sqrt{2\pi}\,\text{Q}(x)\exp\left(\frac{x^2}{2}\right)\left(x^2 + 1\right) - 2x. \tag{21}$$

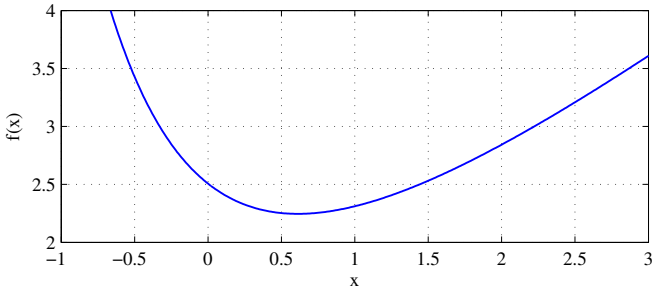

Fig. 1. The function $f(x)$ given in (20).

From (21), $f''(x) > 0$ when $x \leq 0$. When $x > 0$, we can also show $f''(x) > 0$ through the Gaussian Q-function lower bound [12]

$$\mathrm{Q}(x) > \frac{x}{x^2+1}\frac{1}{\sqrt{2\pi}}\exp\left(\frac{x^2}{-2}\right) \qquad x > 0.$$

Hence $f(x)$ is strictly convex. Having $f(x)$ plotted in Fig. 1, we can properly draw a horizontal line and create two intersections such that the x-coordinates of the left and right intersections correspond to $\hat{G}$ and $\hat{B}$ respectively, because $f(\hat{G}) = f(\hat{B})$ and $\hat{G} < \hat{B}$. Since the minimum of $f(x)$ happens at $x = 0.6120$, we have $\hat{G} < 0.6120 < \hat{B}$. To prove (16), we shall discuss the two cases of $\hat{G}$:

- If $\hat{G} \leq 0$, it is straightforward that $1 - \hat{G}\hat{B} > 0$ because $\hat{B} > 0$.
- If $0 < \hat{G} < 0.6120$, then $2.5066 < f(\hat{G}) < 2.2460$, which yields $1.4536 > \hat{B} > 0.6120$. Combining the ranges of $\hat{G}$ and $\hat{B}$, we have $1 - \hat{G}\hat{B} > 0.1104 > 0$.

Thus, (16) is valid, and it leads to $\det \nabla^2 \bar{\tau}\big|_{n=\hat{n},T=\hat{T}} > 0$.

Since both (11) and (12) are true, $\bar{\tau}$ has a local minimum at any point $(\hat{n}, \hat{T})$ for which $\nabla\bar{\tau}|_{n=\hat{n},T=\hat{T}} = 0$. We now show that there is exactly one such point.

B. Existence and Uniqueness of the Local Minimum

From the definition of $\hat{G}$ and $\hat{B}$ in (14) and (15), the corresponding $(\hat{n}, \hat{T})$ can be obtained from $(\hat{G}, \hat{B})$ through

$$\hat{n} = \left(\frac{\hat{G} - \hat{B}}{x_\mathrm{G} - x_\mathrm{B}}\right)^2 \tag{22}$$

and

$$\hat{T} = \frac{\hat{G}x_\mathrm{B} - \hat{B}x_\mathrm{G}}{\hat{G} - \hat{B}}. \tag{23}$$

Thus, it is sufficient to prove the existence and uniqueness of $(\hat{G}, \hat{B})$. Using (22) to substitute for $\hat{n}$ in (18) yields

$$(\hat{B} - \hat{G})\left(p_\mathrm{G}\, e^{\frac{\hat{B}^2}{2}} + p_\mathrm{B}\, e^{\frac{\hat{G}^2}{2}}\right) = p_\mathrm{G}p_\mathrm{B}\tau_\mathrm{diff}\frac{(x_\mathrm{G} - x_\mathrm{B})^2}{2\sqrt{2\pi}}. \tag{24}$$

Note that the LHS of (24) is a function of $(\hat{G}, \hat{B})$ while the RHS is a positive constant. We can further simplify the LHS through the function $f(x)$ defined in (20).

Let $y = f(x)$, and the two inverse functions $x = f_\mathrm{G}^{-1}(y)$ and $x = f_\mathrm{B}^{-1}(y)$ for the ranges $x < 0.6120$ and $x > 0.6120$, respectively. Then $\hat{G}$ and $\hat{B}$ can be expressed in terms of y and the LHS of (24) is $h(y)$ which written as

$$h(y) = \left[f_\mathrm{B}^{-1}(y) - f_\mathrm{G}^{-1}(y)\right]\left[p_\mathrm{G}\, e^{\frac{f_\mathrm{B}^{-1}(y)^2}{2}} + p_\mathrm{B}\, e^{\frac{f_\mathrm{G}^{-1}(y)^2}{2}}\right], \tag{25}$$

which is a function of a single variable $y > 2.2460$. When y approaches its minimum, $h(y)$ approaches zero. As y goes to infinity, $h(y)$ goes to infinity as well. Therefore, by continuity, $h(y)$ meets the RHS of (24) at least once. This proves the existence of $(\hat{G}, \hat{B})$.

To prove the uniqueness, we show that $h(y)$ is strictly increasing. Consider its first derivative $h'(y)$, in which we are able to show

$$\frac{\mathrm{d}}{\mathrm{d}y}\left[f_\mathrm{B}^{-1}(y) - f_\mathrm{G}^{-1}(y)\right] p_\mathrm{G} \exp\left(\frac{f_\mathrm{B}^{-1}(y)^2}{2}\right) > 0$$

using $\frac{\mathrm{d}}{\mathrm{d}y}f_\mathrm{B}^{-1}(y) > 0 > \frac{\mathrm{d}}{\mathrm{d}y}f_\mathrm{G}^{-1}(y)$, $f_\mathrm{B}^{-1}(y) > f_\mathrm{G}^{-1}(y)$, and $f_\mathrm{B}^{-1}(y) > 0$. Therefore, $h'(y)$ is lower bounded by

$$\begin{aligned} h'(y) &> \frac{\mathrm{d}}{\mathrm{d}y}\left[f_\mathrm{B}^{-1}(y) - f_\mathrm{G}^{-1}(y)\right] p_\mathrm{B} \exp\left(\frac{f_\mathrm{G}^{-1}(y)^2}{2}\right) \\ &= p_\mathrm{B} \exp\left(\frac{f_\mathrm{G}^{-1}(y)^2}{2}\right)\left\{\frac{\mathrm{d}}{\mathrm{d}y}f_\mathrm{B}^{-1}(y)\right. \\ &\quad \left. + \left[f_\mathrm{G}^{-1}(y)\, f_\mathrm{B}^{-1}(y) - 1 - f_\mathrm{G}^{-1}(y)^2\right]\frac{\mathrm{d}}{\mathrm{d}y}f_\mathrm{G}^{-1}(y)\right\}. \end{aligned} \tag{26}$$

Since $\frac{\mathrm{d}}{\mathrm{d}y}f_\mathrm{B}^{-1}(y) > 0$, $\frac{\mathrm{d}}{\mathrm{d}y}f_\mathrm{G}^{-1}(y) < 0$ and $f_\mathrm{G}^{-1}(y)\, f_\mathrm{B}^{-1}(y) < 1$, the RHS of (26) is positive. Hence, $h(y)$ is strictly increasing, and the uniqueness of $(\hat{G}, \hat{B})$ is proven.

To summarize, we have shown that there is a unique solution to $\nabla\bar{\tau} = 0$, which implies a local minimum. In other words, the solution to $\nabla\bar{\tau} = 0$ yields the global minimum. This guarantees that Newton's method will always converge to the global minimum.

C. Numerical Method

Given all the parameters $(p_\mathrm{G}, p_\mathrm{B}, x_\mathrm{G}, x_\mathrm{B}, \tau_\mathrm{G}, \tau_\mathrm{B})$, the point $(\hat{n}, \hat{T})$ that yields the global minimum of $\bar{\tau}$ can be computed using Newton's method. Since $\hat{n}$ may not be an integer, we consider its neighbor integers $\hat{n}_1 = \lfloor\hat{n}\rfloor$ and $\hat{n}_2 = \lceil\hat{n}\rceil$. For $\hat{n}_1$ and $\hat{n}_2$, the corresponding optimal thresholds $\hat{T}_1$ and $\hat{T}_2$ are computed. The final optimal parameter pair $(n^\star, T^\star)$ is then chosen between $(\hat{n}_1, \hat{T}_1)$ and $(\hat{n}_2, \hat{T}_2)$, whichever achieves the smaller total expected channel usage $\bar{\tau}$. The computation of $\hat{T}_i$, $i \in \{1, 2\}$ is discussed below.

With the integer $n = \hat{n}_i$ fixed, $\bar{\tau}$ is a function only of T. Denote this function as $\bar{\tau}_i(T)$, and let its first and second derivatives be $\bar{\tau}_i'(T)$ and $\bar{\tau}_i''(T)$, respectively. When $\hat{n}_i = 0$, the value of $\hat{T}_i$ is not important because $\bar{\tau}_i(T) = \bar{\tau}_\mathrm{ref}$ regardless of T. For $\hat{n}_i > 0$, for any finite $\hat{T}_i$ satisfying $\bar{\tau}_i'(\hat{T}_i) = 0$, we are able to show $\bar{\tau}_i''(\hat{T}_i) > 0$, i.e. $\hat{T}_i$ is a local minimum, by following the proof of (11). Furthermore, we know $\bar{\tau}_i'(T) > 0$ for $T \gg 0$ because $\lim_{T\to\infty}\bar{\tau}_i(T) = \infty$. These two facts and the continuity of $\bar{\tau}_i(T)$ and $\bar{\tau}_i'(T)$ imply that, if it exists, $\hat{T}_i$ is unique and thus it yields the global minimum.

We now discuss the two cases where this finite solution to $\bar{\tau}_i'(T) = 0$ does and does not exist. Consider

$$\lim_{T\to-\infty} \bar{\tau}_i'(T) = \lim_{T\to-\infty} \frac{\sqrt{\hat{n}_i}}{\sqrt{2\pi}} e^{\frac{\hat{n}_i(T-x_{\mathrm{B}})^2}{-2}} p_{\mathrm{B}}\left(\hat{n}_i - p_{\mathrm{B}}\tau_{\mathrm{diff}}\right). \quad (27)$$

Case 1: If $0 < \hat{n}_i < p_{\mathrm{B}}\tau_{\mathrm{diff}}$, then $\bar{\tau}_i'(T) < 0$ for $T \ll 0$ since (27) approaches zero from below. In addition, we know $\bar{\tau}_i'(T) > 0$ for $T \gg 0$. Thus, because of the intermediate value theorem of the continuous function $\bar{\tau}_i'(T)$, the finite solution to $\bar{\tau}_i'(T) = 0$ exists and it is $\hat{T}_i$. Moreover, since the local minimum $\hat{T}_i$ is unique, Newton's method is guaranteed to converge to it.

Case 2: If $\hat{n}_i > p_{\mathrm{B}}\tau_{\mathrm{diff}}$, then $\bar{\tau}_i'(T) > 0$ for $T \ll 0$ because (27) approaches zero from above. This implies that there must be an even number of finite roots of $\bar{\tau}_i'(T) = 0$. Since $\hat{T}_i$ must be unique if it exists, there is no such root[1]. In this case, $\hat{\tau}(T)$ is strictly increasing and the global minimum happens at $T \to -\infty$. Having such threshold means that the user will always initiate the data transmission after the training. This is reasonable because the price of training $\hat{n}_i$ is greater than the upper bound for the expected reduction in channel occupancy $\Delta\bar{\tau} = p_{\mathrm{B}}\tau_{\mathrm{diff}}$ defined in (6). Furthermore, if $\hat{n}_i = p_{\mathrm{B}}\tau_{\mathrm{diff}}$, the same threshold should be used for the same reason.

The procedure to compute the optimal parameter pair $(n^\star, T^\star)$ is summarized as follows:

1) Find the optimal parameter pair $(\hat{n}, \hat{T})$ that yields the global minimum of $\bar{\tau}$ using Newton's method, where $\hat{n}$ is a positive real number.
2) Let $\hat{n}_1 = \lfloor \hat{n} \rfloor$ and $\hat{n}_2 = \lceil \hat{n} \rceil$. Calculate $\hat{T}_i$ for $i \in \{1, 2\}$ according to:
   - If $\hat{n}_i = 0$, set $\hat{T}_i = 0$.
   - If $0 < \hat{n}_i < \Delta\bar{\tau}$, find $\hat{T}_i$ that yields the global minimum of $\bar{\tau}|_{n=\hat{n}_i}$ using Newton's method.
   - If $\hat{n}_i \geq \Delta\bar{\tau}$, set $\hat{T}_i = -\infty$. This can only happen when $i = 2$, and implies that $\hat{n}_1$ will be optimal.
3) Pick the smaller total expected channel usage between $\bar{\tau}|_{n=\hat{n}_1, T=\hat{T}_1}$ and $\bar{\tau}|_{n=\hat{n}_2, T=\hat{T}_2}$, and return the corresponding parameter pair.

IV. Numerical Results

The approach described above to find the optimal parameter pair $(n^\star, T^\star)$ is general. It can be applied to any transmission scheme that has the essential elements of feedback to let the transmitter know whether to proceed with the communication and also variable-length coding and feedback that would permit transmission with expected lengths of τ_{G} on the good channels and τ_{B} on the bad channels.

This section provides numerical results when the optimal parameter pairs $(n^\star, T^\star)$ are used with a specific scheme for variable-length coding with feedback. We expect that the qualitative behavior would be similar for many other such schemes.

[1]Repeated roots are not allowed because $\bar{\tau}_i''(\hat{T}_i) > 0$.

We consider a message length of $k = 128$ bits, which are encoded using a rate-compatible code into a variable-length packet whose bits are modulated using binary phase-shifted keying (BPSK). Our example uses a feedback scheme in which the RX tries to decode the variable-length packet after receiving every BPSK symbol and will reply with an acknowledgement once it decodes successfully.

This IR scheme can be analyzed through rate-compatible sphere-packing (RCSP) [7]. The probability of decoding error after receiving m symbols is

$$\mathrm{P}_m(\gamma) = 1 - \mathrm{F}_{\chi^2(m)}\left(\frac{m\,(1+\gamma)}{2^{\,2k/m}}\right), \quad (28)$$

where γ is the SNR and $\mathrm{F}_{\chi^2(m)}(\cdot)$ is the cumulative distribution function (CDF) of the chi-square distribution with m degrees of freedom. To guarantee a successful data transmission, we do not put any limit on the number of available data symbols, and $\lim_{m\to\infty} \mathrm{P}_m(\gamma) = 0$. With the probability upper bound given in [13], the expected channel usage of the data transmission is upper bounded by

$$\tau_{\mathrm{data}}(\gamma) \leq 1 + \sum_{m=1}^{\infty} \mathrm{P}_m(\gamma).$$

While computing τ_{G} and τ_{B}, only finite terms are evaluated. For instance,

$$\tau_{\mathrm{G}} \approx 1 + \sum_{m=1}^{\hat{m}_{\mathrm{G}}} \mathrm{P}_m\big(x_{\mathrm{G}}^2\big),$$

where $\hat{m}_{\mathrm{G}}$ is the smallest positive integer that satisfies the following two criteria. The first criterion is $\left(1 + x_{\mathrm{G}}^2\right) 2^{\,-2k/\hat{m}_{\mathrm{G}}} > 1 + 0.5x_{\mathrm{G}}^2$. Thus the sum of the omitted terms is upper bounded by

$$\begin{aligned}
\sum_{m=\hat{m}_{\mathrm{G}}+1}^{\infty} \mathrm{P}_m\left(x_{\mathrm{G}}^2\right) &< \sum_{m=\hat{m}_{\mathrm{G}}+1}^{\infty} 1 - \mathrm{F}_{\chi^2(m)}\left(\frac{m\left(1+x_{\mathrm{G}}^2\right)}{2^{\,2k/\hat{m}_{\mathrm{G}}}}\right) \\
&< \sum_{m=\hat{m}_{\mathrm{G}}+1}^{\infty} 1 - \mathrm{F}_{\chi^2(m)}\left[m\left(1+0.5x_{\mathrm{G}}^2\right)\right] \\
&\leq \sum_{m=\hat{m}_{\mathrm{G}}+1}^{\infty} \left[\left(1+0.5x_{\mathrm{G}}^2\right) e^{-0.5x_{\mathrm{G}}^2}\right]^{m/2}, \quad (29)
\end{aligned}$$

where (29) follows from Chernoff bound for the chi-square tail probability. Since $0 < \left(1 + 0.5x_{\mathrm{G}}^2\right) e^{-0.5x_{\mathrm{G}}^2} < 1$, the sum of the geometric series in (29) is finite. The second criterion of $\hat{m}_{\mathrm{G}}$ is that (29) is upper bounded by a small $\varepsilon = 10^{-2}$. Hence the error of τ_{G} incurred by computing only the first $\hat{m}_{\mathrm{G}}$ terms is bounded by ε. The computation of τ_{B} is carried out in the same manner as τ_{G}, but $\gamma = x_{\mathrm{B}}^2$ is used instead.

Next, we show the benefit of channel selection using pilots by comparing two schemes. The optimal selective scheme sends pilots to estimate the channel with the optimal parameter pair $(n^\star, T^\star)$ given in Section III. If the data transmission begins, the IR scheme mentioned earlier is used. The non-selective scheme always initiates the data transmission without estimating the channel, and the same IR scheme is used.

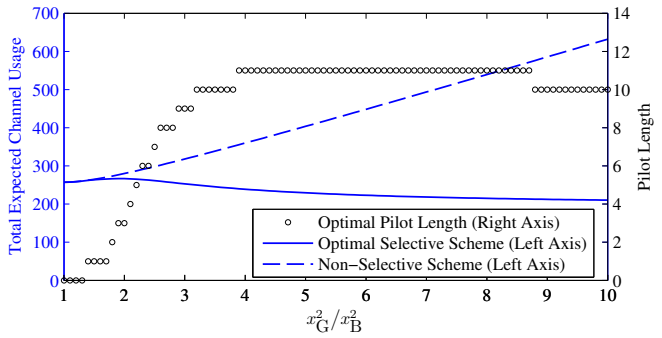

Fig. 2. A comparison of $\bar{\tau}|_{n=n^\star,T=T^\star}$ and $\bar{\tau}_{\text{ref}}$. The good and bad states are equally probable $p_G = p_B = 0.5$, and the ratio of the associated SNRs varies while the averaged SNR is fixed to $p_G x_G^2 + p_B x_B^2 = 1$ (0 dB). The packet size is $k = 128$ bits.

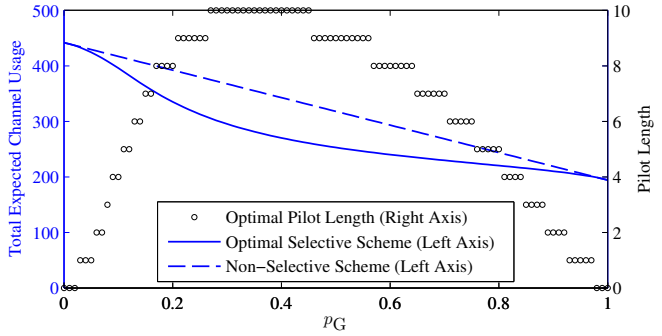

Fig. 3. A comparison of $\bar{\tau}|_{n=n^\star,T=T^\star}$ and $\bar{\tau}_{\text{ref}}$. The SNRs of the good and bad states are fixed to $x_G^2 = 1.5$ (1.76 dB) and $x_B^2 = 0.5$ (−3.01 dB), and the associated probabilities vary. The packet size is $k = 128$ bits.

In Fig. 2, we plot the total expected channel usage (including training) of the optimal selective scheme and the non-selective scheme. The two channel states are equally probable and the averaged SNR is one. As the difference between the good and bad states increases, there is more room for the non-selective scheme to improve and smart channel selection can bring more reduction in channel usage. If the two states are similar to each other (x_G^2 is close to x_B^2), it requires more training symbols to distinguish the two states and the potential expected reduction $\Delta\bar{\tau}$ is small. In this case, the best choice of the selective scheme is "no selection" ($n^\star = 0$) and the two schemes have the same performance.

A similar comparison is illustrated in Fig. 3. Now the probabilities of the good and bad states vary while the associated SNRs are fixed. The largest reduction appears at moderate probability. When p_G approaches zero or one, the two-state channel acts like a static channel and requires almost no training. In this case, the optimal pilot length goes to zero and the channel usage reduction go to zero.

V. Conclusion

In this paper, we present a channel selection scheme based on training. After receiving several training symbols, the receiver determines whether the data transmission should be initiated or postponed. The optimal pilot length and the threshold to leave the channel balance the cost of training and the reliability of channel estimation such that the total expected channel usage is minimized. We prove that the optimal parameter pair can be achieved using Newton's method, and describe a procedure to compute it. The reduction in channel occupancy is shown for various channel settings and is evident when the two channel states are very distinct and occur with comparable probabilities.

Note that the derivation is independent of the specific transmission scheme used in our numerical results. For example, channel selection based on training could be applied to incremental redundancy with less frequent decoding attempts [6] or repeated redundancy using Chase combining [1]. Although this paper focuses on two-state channels, we are currently working on generalizing this concept to multiple-state channels and even continuous-coefficient block-fading channels.

References

[1] D. Chase, "Code combining–a maximum-likelihood decoding approach for combining an arbitrary number of noisy packets," vol. 33, no. 5, pp. 385–393, May 1985.

[2] D. Mandelbaum, "An adaptive-feedback coding scheme using incremental redundancy," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 388–389, May 1974.

[3] P. Frenger, S. Parkvall, and E. Dahlman, "Performance comparison of HARQ with Chase combining and incremental redundancy for HSDPA," in *Proc. 2001 IEEE 54th Veh. Technology Conf. (VTC Fall)*, vol. 3, Oct. 2001, pp. 1829–1833.

[4] J. Kim, W. Hur, A. Ramamoorthy, and S. McLaughlin, "Design of rate-compatible irregular LDPC codes for incremental redundancy hybrid ARQ systems," in *Proc. 2006 IEEE Int. Symp. Inform. Theory (ISIT)*, Jul. 2006, pp. 1139–1143.

[5] L. Szczecinski, P. Duhamel, and M. Rahman, "Adaptive incremental redundancy for HARQ transmission with outdated CSI," in *Proc. 2 IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2011, pp. 1–6.

[6] A. Williamson, T.-Y. Chen, and R. Wesel, "A rate-compatible sphere-packing analysis of feedback coding with limited retransmissions," in *Proc. 2012 IEEE Int. Symp. Inform. Theory (ISIT)*, Jul. 2012, pp. 2924–2928.

[7] T.-Y. Chen, N. Seshadri, and R. Wesel, "A sphere-packing analysis of incremental redundancy with feedback," in *Proc. 2011 IEEE Int. Conf. Commun. (ICC)*, Jun. 2011, pp. 1–5.

[8] V. Pohl, P. H. Nguyen, V. Jungnickel, and C. Von Helmolt, "Continuous flat-fading MIMO channels: achievable rate and optimal length of the training and data phases," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1889–1900, Jul. 2005.

[9] S. Savazzi and U. Spagnolini, "Optimizing training lengths and training intervals in time-varying fading channels," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1098–1112, Mar. 2009.

[10] W. Santipach and M. Honig, "Optimization of training and feedback overhead for beamforming over block fading channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6103–6115, Dec. 2010.

[11] M. Kobayashi, N. Jindal, and G. Caire, "Training and feedback optimization for multiuser MIMO downlink," *IEEE Trans. Commun.*, vol. 59, no. 8, pp. 2228–2240, Aug. 2011.

[12] R. D. Gordon, "Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument," *The Ann. of Math. Stat.*, vol. 12, no. 3, pp. 364–366, Sep. 1941.

[13] T.-Y. Chen, D. Divsalar, and R. Wesel, "Chernoff bounds for analysis of rate-compatible sphere-packing with numerous transmissions," in *Proc. 2012 IEEE Inform. Theory Workshop (ITW)*, Sep. 2012, pp. 477–481.