On the Use of Sounding in Wideband Channels

Sheng Jing, Lizhong Zheng, and Muriel Médard

Abstract— **In wideband channels, the energy per degree of freedom does not suffice to provide an accurate measurement of the channel over the entire spectrum. However, in the presence of feedback, we may garner information at the sender about some aspects of the channel quality over certain portion of the spectrum. In this paper, we investigate a scheme to capture the effect of such information. We consider channel sounding with a finite amount of energy over a block-fading channel in both time and frequency. The quality of each subchannel is assessed as being the cross-over probability in a BSC. In order to characterize a judicious policy for allocating energy to different subchannels in view of establishing their usefulness for transmission, we use a multi-armed bandit approach. This approach provides us with a cohesive framework to consider the relative costs and benefits of allotting energy for sounding versus transmission, and for repeated sounding of a single channel versus sounding of different channels. In particular, we are able to give a characterization of the number of subchannels that should be probed for capacity maximization in terms of the available transmission energy, the available bandwidth and the fading characteristics of the channel.**

I. INTRODUCTION

For communications in wideband channels, the energy per degree of freedom is usually not sufficient to measure accurately the channel over the entire spectrum. However, with a feedback channel, we are able to collect some information at the sender about some aspects of the channel quality over certain portions of the spectrum. We need to balance the energy between sounding the channels and transmitting information. We model the independently fading channels as a set of independent Binary Symmetric Channels (BSC) and the quality of each channel is characterized as the BSC's cross-over probability. A multi-armed bandit approach is used to study the tradeoff between allotting energy for sounding channels and transmitting information. Specifically, we are able to characterize the number of subchannels that should be probed for capacity maximization in terms of the available transmission energy, the available bandwidth and the fading characteristics.

A. Physical Channel

A wideband channel with a large bandwidth, B is assumed and divided into subchannels. Each subchannel has the same bandwidth equal to the coherence bandwidth, W_c. They are centered at $\{f_i, i = 1, 2, \ldots, L, L \gg 1\}$. The i^{th} subchannel occupies the frequency band of $(f_i - \frac{W_c}{2}, f_i + \frac{W_c}{2})$. The subchannels are mutually independent.

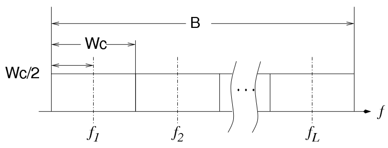

Fig. 1. Wideband Channel Divided into Subchannels

B. System Model

The i^{th} subchannel is modeled as a BSC (Figure 2), with the cross-over probability P_i characterizing its quality.

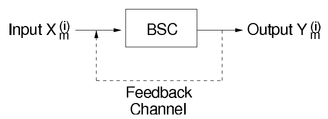

Fig. 2. System Model: i^{th} Subchannel

The subchannel is assumed to experience block fading: the cross-over probability P_i keeps constant in the k^{th} coherence block, $kT_cW_c \leq m \leq (k+1)T_cW_c$; P_i in different coherence blocks are mutually independent. For example, communication on a block fading channel using BPSK perturbed by AWGN is equivalent to transmission through a BSC, in which the cross-over probability P_i is determined by the amplitude of the fading coefficient. The feedback channel only provides noiseless duplicates of the testings. The energy consumption of each input symbol is assumed to be 1 unit. Given E units of energy over all subchannels in one coherence block T_c, which is equivalent to an average power constraint $\frac{E}{T_c}$ (units per second) over all subchannels, we are allowed to input at most E symbols into all the subchannels in one coherence block. The channel state information P_i is unknown to either the sender or the receiver.

II. COMMUNICATION SCHEME

Each coherence block is divided into two parts: the channel testing phase and the data transmission phase. We assume

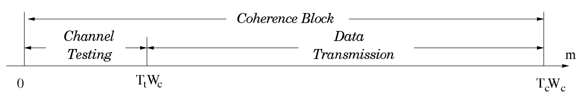

Fig. 3. Coherence Block

the existence of a noiseless feedback channel but only for

This work was supported by the National Science Foundation under grant nos. CCR-0093349, ANI-0335256 and CNS-0434974, the National Science Foundation CAREER Award no. 0347395 and the ARO under grant no. S0176938

Sheng Jing, Lizhong Zheng and Muriel Médard are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA (email: sjing@mit.edu, lizhong@mit.edu and medard@mit.edu)

the channel testing phase. Thus, in the data transmission phase, each subchannel is treated as an ordinary BSC without feedback. Let E_{te}, E_{tr} be the energy consumed in the channel testing phase and the data transmission phase, respectively. The testing and transmission scheme in a coherence block is as follows:

1) At the beginning of each coherence block, we choose a subset of M subchannels from the L available subchannels, $(L \gg 1)$;
2) We use E_{te} units of energy to test the M subchannels with a channel testing algorithm;
3) At the end of the channel testing phase, we choose one subchannel according to the testing results;
4) We spend the remaining E_{tr} units of energy in transmitting information on the chosen subchannel;
5) We go to Step 1 at the beginning of the next block.

We assume that only one subchannel is chosen at the end of the channel testing phase. This assumption can be better understood through the example in Figure 4.

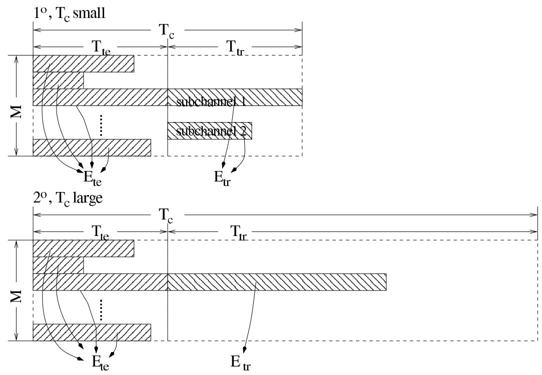

Fig. 4. How T_c affects channel selection

We assume that subchannel 1 is the best and subchannel 2 is the second best according to the testing results. In the upper figure where T_c is small, we cannot consume all the E_{tr} units of energy by the end of T_c if we only choose to transmit on subchannel 1. Thus, we can send more data by choosing more than one subchannel, i.e. subchannel 2. In the lower figure where T_c is large, we can use the entire E_{tr} units of energy by the end of T_c on subchannel 1. Since subchannel 1 has a better quality than any other subchannel, it is desirable to transmit only on subchannel 1. We assume a relationship between E and T_c, such that, for every coherence block, the transmission energy E_{tr} can be used entirely on a single subchannel by the end of that coherence block. This relationship guarantees no loss of average rate per block if we only choose the best subchannel for transmission, rather than a set of subchannels.

The channel testing algorithm allocates the E_{te} units of energy over the M subchannels and choose the subchannel that has the best testing result. This is similar to the setup of Multi-Armed Bandit Problem (MABP).

III. Introduction to MABP

MABP is a classical problem in the decision theory. We play with M machines, each of which is associated with a stochastic reward. The objective is to design a scheme of pulling machines to maximize the total rewards. "In spite of MABP's simplicity, it includes the tradeoff between exploration over more machines and exploitation on a particular machine" [3]. MABP has been studied in various settings [1]. In [3], the authors considered MABP from a new perspective, i.e. the Probably Approximately Correct (PAC) model. When the i^{th} machine is pulled, a stochastic reward $R(i)$ is received, which is assumed to be distributed in $\{0,1\}$ with probability $\{1-P_i, P_i\}$ respectively. The values of $\{P_i, i=1,2,\dots,M\}$ are unknown. The optimal machine, denoted as machine $I^\star$, is the one with the biggest expected reward, denoted as $r^\star$. We define the ε-optimal machine to be those whose expected rewards are no less than $r^\star - \varepsilon$. As stated in [3], the objective in PAC model is to find out a near optimal machine with high probability, i.e. an ε-optimal machine with probability at least $1-\delta$. We consider the complexity to be the number of machine uses before the decision is made. We denote machine $I^{\star\star}$ to be the selected machine. The objective in the PAC model is to design a strategy to minimize the number of machine uses such that

$$Pr\{P_{I^{\star\star}} > P_{I^\star} - \varepsilon\} > 1 - \delta \qquad (1)$$

where δ and ε are positive parameters. The algorithm which satisfies (1) is called a (ε, δ)-PAC algorithm. In [3], the Median Elimination algorithm is proposed as a (ε, δ)-PAC algorithm and its complexity is proved to be $O\left(\frac{\log\frac{1}{\delta}}{\varepsilon^2}M\right)$. Here is the general form of the Median Elimination algorithm (ε_ls and δ_ls are determined from ε and δ):

1) Let S_1 be the set of initial M sub-channels;
2) ε_1, δ_1, $l=1$;
3) Test every subchannel k in S_l for $\frac{\ln(\frac{3}{\delta_l})}{(\frac{\varepsilon_l}{2})^2}$ times, and let $\hat{p}_k^l$ denote the empirical crossover probability of subchannel k up to the l^{th} stage;
4) Find the median of $\{\hat{p}_k^l, k \in S\}$, denoted by m_l;
5) $S_{l+1} = S_l \setminus \{k : \hat{p}_k^l < m_l\}$;
6) If $|S_l| = 1$, then algorithm terminates and output S_l as the chosen sub-channel; Otherwise, $\varepsilon_{l+1} \Leftarrow \varepsilon_l$, $\delta_{l+1} \Leftarrow \delta_l$, $l = l+1$, go to step 3.

In [4], The complexity of the (ε, δ)-PAC algorithm is shown to be lower bounded by $\Theta\left(\frac{\log\frac{1}{\delta}}{\varepsilon^2}M\right)$. We adopt the Median Elimination algorithm as the channel testing algorithm, since it achieves the lower bound of complexity.

IV. Performance Expression

The performance of the i^{th} subchannel is a monotonically non-increasing function of the cross-over probability P_i. An example is the BSC's capacity $C_i = 1 - H_b(P_i)$. We consider the simple case where P_i takes two possible values, p_G and p_B. The probability density function (PDF) of P_i, $f_{P_i}(p)$, is composed of two impulses, one at p_G and the other at p_B. Figure 5 is a typical example of $f_P(p)$ and $f_C(c)$, where

$0 \le p_G < p_B \le \frac{1}{2}$.

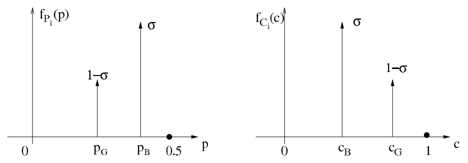

Fig. 5. Typical Probability Distribution of P_i, C_i

When constructing the communication scheme, an (ε, δ) Median Elimination algorithm is generated ($\varepsilon = p_B - p_G$) and applied. The testing results satisfy

$$Pr\{P_{I^{\star\star}} = p_G \,|P_{I^\star} = p_G\} > 1 - \delta \quad (2)$$

The energy consumption of the channel testing, which is equal to the complexity of the testing algorithm, is bounded as follows

$$\begin{aligned} A_1 M \le E_{te} &\le A_2 M \quad (3) \\ \text{where} \quad A_i &= \frac{C_i}{\varepsilon^2} \log \frac{1}{\delta} \qquad i = 1, 2 \end{aligned}$$

We examine the influence of M and σ (parameter of $f_{P_i}(p)$) on the average number of bits that can be reliably sent during the transmission phase, $R(E, M, \sigma)$. This equals to the ergodic capacity for the given transmission energy, if we allow for coding over many coherence blocks. An upper bound and a lower bound of $R(E, M, \sigma)$ are derived in the following lemma.

Lemma 1: $R(E, M, \sigma)$ is bounded as follows

$$R_l(M, E, \sigma) \le R(E, M, \sigma) \le R_u(M, E, \sigma) \quad (4)$$

$$\begin{aligned} \text{where } R_u(M, E, \sigma) &= \max\{0, (E - A_1 M)\,[c_B \\ &\quad + (c_G - c_B)(1 - \sigma^M)]\} \quad (5) \\ R_l(M, E, \sigma) &= \max\{0, (E - A_2 M)\,[c_B \\ &\quad + (c_G - c_B)(1 - \delta)(1 - \sigma^M)]\} \quad (6) \end{aligned}$$

Proof:

Let $I^{\star\star} = \arg\min_{i=1,2,\ldots,M}\{\widehat{P}_i\}$, where $\widehat{P}_i$ is the i^{th} subchannel's empirical crossover probability. Then we have the expression:

$$\begin{aligned} & E\{\text{bits reliably sent in one coherence block}\} \\ = \; & (E - E_{te}) \cdot (c_B + (c_G - c_B) Pr[P_{I^{\star\star}} = p_G]) \quad (7) \end{aligned}$$

where $\Pr[\{P_{I^{\star\star}} = p_G\}]$ is derived as follows:

$$\begin{aligned} & \Pr[\{P_{I^{\star\star}} = p_G\}] \\ = \; & 1 - \Pr[\{P_{I^{\star\star}} = p_B\}] \\ = \; & 1 - \Pr[\{\{P_{I^\star} = p_G\} \cap \{P_{I^{\star\star}} = p_B\}\} \cup \{P_{I^\star} = p_B\}] \\ = \; & 1 - (1 - \sigma^M) \Pr[P_{I^{\star\star}} - P_{I^\star} > \varepsilon \,|\, P_{I^\star} = p_G] - \sigma^M \\ > \; & (1 - \delta) \cdot (1 - \sigma^M) \quad (8) \end{aligned}$$

(8) comes from the definition of the (ε, δ)-PAC algorithm. On the other hand,

$$\begin{aligned} & \Pr[P_{I^{\star\star}} = p_G] \\ = \; & 1 - (1 - \sigma^M) \Pr[P_{I^{\star\star}} - P_{I^\star} > \varepsilon \,|\, P_{I^\star} = p_G] - \sigma^M \\ < \; & 1 - \sigma^M \quad (9) \end{aligned}$$

Thus, we have a lower bound and an upper bound:

$$(1 - \delta) \cdot (1 - \sigma^M) < Pr[P_{I^\star} = p_G] < 1 - \sigma^M \quad (10)$$

Combining (3), (7) and (10), we get desired bounds for $R(M, E, \sigma)$. ■

V. Design of the Testing Scheme

We use lemma 1 to characterize the desirable range of M. The maximum points $M_u^\star(\sigma, E)$ and $M_l^\star(\sigma, E)$ for $R_u(M, E, \sigma)$ and $R_l(M, E, \sigma)$ are derived in the appendix. It is reasonable to choose $M^\star$, the desirable number of subchannels we should start testing, between $M_u^\star(\sigma, E)$ and $M_l^\star(\sigma, E)$. For a fixed testing algorithm, we have $A = A_1 = A_2$, and $M^\star$, $M_u^\star$ and $M_l^\star$ converge as $\delta \to 0$. In practice, our computation shows the difference to be negligibly small (Figure 7 and 6) The following results show that $M_u^\star(\sigma, E)$ and $M_l^\star(\sigma, E)$ grow as $O(\log E)$ when E is large.

When $E \gg \frac{A c_B}{(c_G - c_B) \ln \frac{1}{\sigma}}$,

$$\begin{aligned} & M_u^\star(\sigma, E) \\ = \; & \log_{\frac{1}{\sigma}} \left(1 + \left(\frac{\ln \frac{1}{\sigma}}{A(1 + \ln B_1)}\right) E\right) + \log_{\frac{1}{\sigma}} \left(\frac{1 + \ln B_1}{\ln B_1}\right) \end{aligned}$$

where $B_1 = \frac{c_G}{(c_G - c_B)}$.

When $E \gg \frac{A c_B}{(c_G - c_B)(1 - \delta) \ln \frac{1}{\sigma}}$,

$$\begin{aligned} & M_l^\star(\sigma, E) \\ = \; & \log_{\frac{1}{\sigma}} \left(1 + \left(\frac{\ln \frac{1}{\sigma}}{A(1 + \ln B_2)}\right) E\right) + \log_{\frac{1}{\sigma}} \left(\frac{1 + \ln B_2}{\ln B_2}\right) \end{aligned}$$

where $B_2 = 1 + \frac{c_B}{(1 - \delta)(c_G - c_B)}$.

As for the influence of σ, since we have

$$\frac{dM_u^\star(\sigma, E)}{d\sigma} = \frac{1 + \ln B_1}{\ln B_1} \cdot \frac{1}{\sigma (\ln \frac{1}{\sigma})^2} > 0 \quad (11)$$

$$\frac{dM_l^\star(\sigma, E)}{d\sigma} = \frac{1 + \ln B_2}{\ln B_2} \cdot \frac{1}{\sigma (\ln \frac{1}{\sigma})^2} > 0 \quad (12)$$

so $M_u^\star(\sigma, E)$ and $M_l^\star(\sigma, E)$ increase with σ. This indicates that, for large E, when the subchannel grows worse, we spend a larger portion of energy in channel testing.

For the integer constraint on M, when $M^\star$ is not an integer, we can find $M_0 < M^\star < M_1$, where M_0 and M_1 are integers and $M_1 - M_0 = 1$. $M^\star$ can be expressed as a linear combination of M_0 and M_1, i.e. $M^\star = q_0 M_0 + q_1 M_1$, where $q_0 + q_1 = 1, q_0 > 0, q_1 > 0$. Then at the beginning of each coherence block, we flip a coin which gives "head" with probability q_0 and "tail" with probability q_1. Then we set $M = M_0$ if we get a head and set $M = M_1$ if we get a tail. Due to the concavity of $R_l(M, E, \sigma)$, the scheme described above gives the best achievable $R_l(M, E, \sigma)$ under the integer constraint on M. The problem can be then solved without considering the integer constraint.

The case where $p_G = 0.1, p_B = 0.45, \delta = 0.01$ is studied.

1) Fix $\sigma = 0.95$, plot $M^\star$ for various $E \in (0, 4 \times 10^6)$

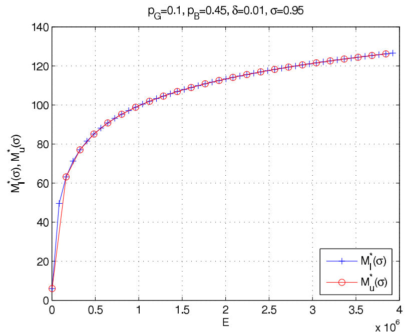

Fig. 6. $M^\star$ vs. E (σ fixed)

2) Fix $E = 5 \times 10^4$, plot $M^\star$ for various $\sigma \in (0, 1)$

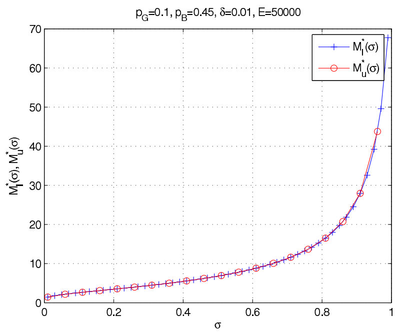

Fig. 7. $M_l^\star$ and $M_u^\star$ vs. σ (for a fixed large E)

VI. Conclusions

In order to maximize the rate per block for a wideband block fading channel, with energy constraint E per block, we select a portion of the frequency band to test and transmit over one subchannel. The desirable size of this portion of the frequency band, $M^\star$ has the following properties:

- $M^\star$ grows asymptotically with $O(\log E)$, when E is large;
- For large E, when each subchannel grows worse, we spend a larger portion of E in testing subchannels ($M^\star$ increases).

Acknowledgment

The authors would like to thank David Tse and R. Srikant for helpful discussions and suggestions.

References

[1] D.A. Berry and B. Fristedt. *Bandit Problems*. Chapman and Hall, 1985.

[2] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. *Gambling in a rigged casino: The adversarial multi-armed bandit problem.* In Proc. 36th Annual Symposium on Foundations of Computer Science, pages 322-331. IEEE Computer Society Press, 1995.

[3] E. Even-Dar, S. Mannor and Y. Mansour, *PAC bounds for multi-armed bandit and Markov decision processes*, Fifteenth Annual Conference on Computational Learning Theory, pages 255-270. Springer, 2002.

[4] Shie Mannor and John Tsitsiklis, *The Sample Complexity of Exploration in the Multi-Armed Bandit Problem*, Journal of Machine Learning Research 5 (2004) 623-648. s

Appendix

Since $c_G > c_B$, $0 < \delta \ll 1$, $0 < \sigma < 1$, we have that

$$\frac{d^2 R_l(M,E,\sigma)}{dM^2} \quad < \quad 0 \qquad \text{for } M \in \left[0, \frac{E}{A}\right] \tag{13}$$

$R_l(M,E,\sigma)$ is a concave function of M, so the maximum point $R_l^\star(M,E,\sigma)$ exists either within $[0, \frac{E}{A}]$ when $\frac{dR_l(M,E,\sigma)}{dM} = 0$ or on the boundary.

$$\begin{aligned} & \frac{dR_l}{dM}(M=0,E,\sigma) \\ = \; & -Ac_B + E(c_G - c_B)(1-\delta)\ln\frac{1}{\sigma} \end{aligned} \tag{14}$$

$$\begin{aligned} & \frac{dR_l}{dM}(M=\frac{E}{A},E,\sigma) \\ = \; & -A\left(c_B + (c_G - c_B)(1-\delta)(1-\sigma^{\frac{E}{A^2}})\right) \end{aligned} \tag{15}$$

Since $\frac{dR_l}{dM}(M=\frac{E}{A},E,\sigma) < 0$, we have that

$$\begin{aligned} & M_l^\star(\sigma,E) \\ = \; & \begin{cases} \frac{E}{A} - \frac{B_2(\frac{1}{\sigma},E)-1}{\ln(\frac{1}{\sigma})} & \text{if } E \geq \frac{Ac_B}{(c_G-c_B)(1-\delta)\ln\frac{1}{\sigma}} \\ 0 & \text{if } E < \frac{Ac_B}{(c_G-c_B)(1-\delta)\ln\frac{1}{\sigma}} \end{cases} \end{aligned} \tag{16}$$

$$\begin{aligned} \text{where} \quad & B_2(\frac{1}{\sigma},E) \\ = \; & LambertW\left(\frac{(C_G - \delta(C_G - C_B))\cdot e^{\left(1+\frac{\ln(\frac{1}{\sigma})E}{A}\right)}}{(1-\delta)(C_G - C_B)}\right) \end{aligned}$$

Since the he *Lambert* function $W(z)$ has the following asymptotic expansion when z is large:

$$W(z) \quad = \quad \ln z - \ln\ln z + O\left[\left(\frac{\ln\ln z}{\ln z}\right)\right]$$

Then, when $E \gg \frac{Ac_B}{(c_G-c_B)(1-\delta)\ln\frac{1}{\sigma}}$

$$\begin{aligned} & M_l^\star(\sigma,E) \\ = \; & \log_{\frac{1}{\sigma}}\left(1 + \left(\frac{\ln\frac{1}{\sigma}}{A(1+\ln B_2)}\right)E\right) + \log_{\frac{1}{\sigma}}\left(\frac{1+\ln B_2}{\ln B_2}\right) \end{aligned}$$

where $B_2 = 1 + \frac{c_B}{(1-\delta)(c_G-c_B)}$.

Similarly, for $R_u(M=0,E,\sigma)$, when $E \gg \frac{Ac_B}{(c_G-c_B)\ln\frac{1}{\sigma}}$

$$\begin{aligned} & M_u^\star(\sigma,E) \\ = \; & \log_{\frac{1}{\sigma}}\left(1 + \left(\frac{\ln\frac{1}{\sigma}}{A(1+\ln B_1)}\right)E\right) + \log_{\frac{1}{\sigma}}\left(\frac{1+\ln B_1}{\ln B_1}\right) \end{aligned}$$

where $B_1 = \frac{c_G}{(c_G-c_B)}$.