

Capacity Definitions of General Channels with Receiver Side Information

Michelle Effros
Department of Electrical Engineering
California Institute of Technology, Pasadena, CA 91125
effros@caltech.edu

Andrea Goldsmith and Yifan Liang
Department of Electrical Engineering
Stanford University, Stanford, CA 94305
{andrea,yfl}@wsl.stanford.edu

Abstract—We consider three capacity definitions for general channels with channel side information at the receiver, where the channel is modeled as a sequence of finite dimensional conditional distributions not necessarily stationary, ergodic, or information stable. The Shannon capacity is the highest rate asymptotically achievable with arbitrarily small error probability. The outage capacity is the highest rate asymptotically achievable with a given probability of decoder-recognized outage. The expected capacity is the highest *expected* rate asymptotically achievable with a single encoder and multiple decoders, where the channel side information determines the decoder in use. Expected capacity equals Shannon capacity for channels governed by a stationary ergodic random process but is typically greater for general channels. These alternative definitions essentially relax the constraint that all transmitted information must be decoded at the receiver. We derive equations for these capacity definitions through information density. Examples are also provided to demonstrate their implications.

I. INTRODUCTION

The well-known capacity formula

$$C = \lim_{n \rightarrow \infty} \sup_{X^n} \frac{1}{n} I(X^n; Y^n)$$

does not hold in full generality. Dobrushin proved it for the class of information stable channels in [1]. Verdú and Han showed the formula

$$C = \sup_{\mathbf{X}} \underline{I}(\mathbf{X}; \mathbf{Y})$$

for general channels in [2], where $\underline{I}(\mathbf{X}; \mathbf{Y})$ is the liminf in probability of the normalized information densities. This formula highlights the pessimistic nature of the Shannon capacity definition – which forces the use of a single code with arbitrarily small error probability. For example, consider a collection of channels $\{W_s : s \in \mathcal{S}\}$ parameterized by s . Suppose the random variable S is chosen according to some distribution $p(s)$ at the beginning of transmission and then held fixed. This is called a *composite channel* [3] or *averaged channel* [4], and the capacity is dominated by the performance of the “worst” channel, no matter how small its probability.

To deal with channels such as the composite channel above, we relax the constraint that all transmitted information has to be correctly decoded. Although the transmitter is forced to use a single code in the absence of channel side information, the

receiver may decide how much information can be reliably decoded based on channel state information at the receiver (CSIR). The outage capacity [5] illustrates this idea. Here we design a coding scheme that works well most of the time, but, with some maximal probability q , the decoder sees a bad channel and declares an outage, and the transmitted information is lost. The encoding scheme is designed to maximize the capacity for non-outage states. Previously examined in [5], outage capacity is a common criterion used in wireless fading channels.

Another method for dealing with channels with variable quality is to apply the broadcast strategy [6]. The transmitter views the composite channel as a broadcast channel with a collection of virtual receivers indexed by channel realization S . The encoder uses a broadcast code to transmit to the virtual receivers. The receiver uses side information S to choose the appropriate decoder. The goal is to identify the point in the broadcast rate region that maximizes the *expected rate*, where the expectation is taken with respect to the distribution $p(s)$ on \mathcal{S} . In [7] Shamai derives the expected capacity for the Gaussian slowly fading channel. The broadcast strategy is also used in [8] to minimize expected distortion. Here we bound the expected capacity for a broad class of channels.

The alternative capacity definitions are of particular interest for applications where it is desirable to maximize average received rate and acceptable for the encoder not to know the exact delivered rate. Examples include systems with some acceptable outage probability, feedback channels where the receiver tells the transmitter which symbols to resend, or communication systems using multiple resolution source codes.

These alternative definitions and main capacity results were proposed in our previous work [3]. Here we revisit capacity theorems for general channels with CSIR and give details of the proofs in Section II. In Section III we compare capacity definitions and their implications through examples of the Gilbert-Elliott channel and the binary symmetric channel (BSC) with random crossover probabilities. Conclusions are given in Section IV.

II. CAPACITY UNDER DIFFERENT DEFINITIONS

Consider a sequence of n -dimensional channels $\mathbf{W} = \{W^n = P_{Z^n|X^n}\}_{n=1}^{\infty}$ where W^n is the conditional distribution from the input space \mathcal{X}^n to the output space \mathcal{Z}^n . Let \mathbf{X} and \mathbf{Z} denote the input and output processes respectively, where

each process is specified by a sequence of finite-dimensional distributions, e.g. $\mathbf{X} = \{X^n = (X_1^{(n)}, \dots, X_n^{(n)})\}_{n=1}^\infty$.

For the special case where the decoder has CSIR, we represent this information as an output of the channel. Specifically, we let $Z^n = (S, Y^n)$, where S is the channel side information and Y^n is the output. We assume that S is independent of \mathbf{X} and unknown to the encoder. Thus $W^n(z^n|x^n) = P_S(s)P_{Y^n|X^n,S}(y^n|x^n, s)$. The information density is defined similarly as in [2], i.e.

$$i_{X^n W^n}(x^n; y^n|s) = \log \frac{P_{Y^n|X^n,S}(y^n|x^n, s)}{P_{Y^n|S}(y^n|s)}. \quad (1)$$

The logarithms in this paper are base 2 if not stated otherwise. Throughout we will consider sequences of $(2^{nR}, n)$ codes for channel \mathbf{W} , where an $(2^{nR}, n)$ code is a collection of 2^{nR} blocklength- n codewords and the associated decoding regions.

A. Shannon Capacity

The achievability and converse theorems for the Shannon capacity C of a general channel

$$\begin{aligned} C &= \sup_{\mathbf{X}} \underline{I}(\mathbf{X}; \mathbf{Z}) = \sup_{\mathbf{X}} \underline{I}(\mathbf{X}; \mathbf{Y}|S) \\ &= \sup_{\mathbf{X}} \sup \left\{ \alpha : \lim_{n \rightarrow \infty} \Pr \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n|S) \leq \alpha \right] = 0 \right\} \end{aligned}$$

are proved by Theorems 2 and 5 of [2]. We here provide another proof of achievability based on the notion of typical sets. In this proof we omit explicitly denoting the conditioning on the side information S for ease of notation.

Encoding: Generate the codebook by choosing $X^n(1), \dots, X^n(2^{nR})$ i.i.d. according to some distribution $P_{X^n}(x^n)$.

Decoding: Define, for any $\epsilon > 0$, the *typical set* $A_\epsilon^{(n)}$ as

$$A_\epsilon^{(n)} = \left\{ (x^n, y^n) : \frac{1}{n} i_{X^n W^n}(x^n; y^n) \geq \underline{I}(\mathbf{X}; \mathbf{Y}) - \epsilon \right\}. \quad (2)$$

Channel output Y^n is decoded to $X^n(i)$ where i is the unique index for which $(X^n(i), Y^n) \in A_\epsilon^{(n)}$.

Error Analysis: Assuming equiprobable inputs, the expected probability of error is:

$$\begin{aligned} P_e^{(n)} &= \Pr(\text{error}|1 \text{ sent}) \\ &\leq \Pr((X^n(1), Y^n) \notin A_\epsilon^{(n)}) + \sum_{j=2}^{2^{nR}} \Pr((X^n(j), Y^n) \in A_\epsilon^{(n)}) \\ &\leq \Pr \left[\frac{1}{n} i_{X^n W^n}(X^n(1); Y^n) < \underline{I}(\mathbf{X}; \mathbf{Y}) - \epsilon \right] \\ &\quad + 2^{nR} \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} P_{X^n}(x^n) P_{Y^n}(y^n) \\ &\leq \epsilon + 2^{n(R - \underline{I}(\mathbf{X}; \mathbf{Y}) + \epsilon)} \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} P_{X^n W^n}(x^n, y^n) \quad (3) \end{aligned}$$

for n large enough. The final inequality uses (1), (2), and the fact that $(x^n, y^n) \in A_\epsilon^{(n)}$ implies

$$P_{Y^n}(y^n) \leq 2^{-n(\underline{I}(\mathbf{X}; \mathbf{Y}) - \epsilon)} P_{W^n}(y^n|x^n).$$

All probabilities above are conditioned on codeword 1 being sent: we omit explicit conditioning for conciseness. From (3)

$$P_e^{(n)} \leq \epsilon + 2^{n(R - \underline{I}(\mathbf{X}; \mathbf{Y}) + \epsilon)} \rightarrow \epsilon$$

for all $R \leq \underline{I}(\mathbf{X}; \mathbf{Y}) - \epsilon$ and arbitrary $\epsilon > 0$, which completes our proof.

B. Outage Capacity

Consider a sequence of $(2^{nR}, n)$ codes. Let $P_o^{(n)}$ be the probability that the decoder declares an outage. Let $P_e^{(n)}$ be the probability that the receiver decodes improperly given that an outage is not declared. We say that a rate $(1-q)R$ is outage- q achievable if there exists a sequence of $(2^{nR}, n)$ channel codes such that $\lim_{n \rightarrow \infty} P_o^{(n)} \leq q$ and $\lim_{n \rightarrow \infty} P_e^{(n)} = 0$. The outage- q capacity of the above described channel is defined to be the supremum over all outage- q achievable rates, i.e.

$$\begin{aligned} C_q^o &= (1-q) \sup_{\mathbf{X}} \underline{I}_q(\mathbf{X}; \mathbf{Y}|S) \\ &= (1-q) \sup_{\mathbf{X}} \sup \left\{ \alpha : \lim_{n \rightarrow \infty} \Pr \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n|S) \leq \alpha \right] \leq q \right\}. \quad (4) \end{aligned}$$

Notice that $C_0^o = C$, so the outage- q capacity is a generalization of the Shannon capacity. The achievability proof follows the same typical-set argument given above. The converse result likewise follows [2]. A closely-related concept of ϵ -capacity was defined in [2], where the error probability ϵ consists of decoding errors which the receiver is unaware of. Whereas in outage capacity the receiver declares an outage based on CSIR when it cannot decode with vanishing error probability. As a consequence no decoding is performed for outage states.

In the example of a composite channel, the encoder uses a single code book and sends information at rate $C_q^o/(1-q)$. The receiver correctly decodes the information proportion $(1-q)$ of the time and declares an outage proportion q of the time. Thus the average rate is C_q^o . The value q can be chosen to maximize the outage capacity C_q^o . When an outage occurs, the receiver may notify the sender for retransmission or approximate the unreliable information from surrounding samples.

C. Expected Capacity

Another strategy for increasing reliably-received rate is to use a single encoder and a collection of decoders, each parameterized by s and decoding at a rate R_s . The transmitter is forced to use a single encoder without channel side information. The receiver, on the other hand, can choose the appropriate decoder based on CSIR. Denote by $P_e^{(n,s)}$ the probability of error associated with channel s . We define the expected capacity C^e as the supremum of all achievable rates $\mathbb{E}_S R_S$ of any code sequence that satisfies $\mathbb{E}_S P_e^{(n,S)} \rightarrow 0$. We show that C^e satisfies the lower bound $C^e \geq \sup_q C_q^o$ and the upper bound

$$C^e \leq \sup_{\mathbf{X}} \limsup_{n \rightarrow \infty} \mathbb{E}_S \mathbb{E}_{X^n Y^n | S} \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n|S) \middle| S \right]. \quad (5)$$

The lower bound is achieved using the code used for outage- q capacity. For the upper bound, we assume channel side

information is provided to the transmitter (CSIT) so it can adapt the transmission rate to the channel state. In this case we also obtain those rates $\{R_s\}_{s \in \mathcal{S}}$ achievable without CSIT. Denote by $X_s^n(1), \dots, X_s^n(2^{nR_s})$ and $D_s(1), \dots, D_s(2^{nR_s})$ the set of codewords and decoding regions corresponding to channel s . We fix $\gamma > 0$ and define for each $s \in \mathcal{S}$ and $1 \leq i \leq 2^{nR_s}$

$$\begin{aligned} B_s(i) &= \{Y^n \in \mathcal{Y}^n : \frac{1}{n} i_{X^n W^n}(X^n(i); Y^n | s) \leq R_s - \gamma\} \\ &= \{Y^n \in \mathcal{Y}^n : P_{X^n | Y^n, S}(X^n(i) | Y^n, s) \leq 2^{-n\gamma}\} \end{aligned} \quad (6)$$

where (6) follows from (1). Notice that for any s with $R_s > 0$

$$\begin{aligned} &P_{X^n Y^n | S} \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n | s) \leq R_s - \gamma \middle| S \right] \\ &\leq \sum_{i=1}^{2^{nR_s}} \left[2^{-nR_s} P_{Y^n | X^n, S}(D_s(i)^c | X^n(i), s) \right. \\ &\quad \left. + \sum_{y^n \in B_s(i) \cap D_s(i)} P_{X^n Y^n | S}(X^n(i), y^n | s) \right] \\ &\leq P_e^{(n, s)} + \sum_{i=1}^{2^{nR_s}} \sum_{y^n \in B_s(i) \cap D_s(i)} 2^{-n\gamma} P_{Y^n | S}(y^n | s) \\ &\leq P_e^{(n, s)} + 2^{-n\gamma}. \end{aligned} \quad (7)$$

Furthermore we have

$$\begin{aligned} &\mathbb{E}_S \liminf_{n \rightarrow \infty} P_{X^n Y^n | S} \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n | S) \leq R_S - \gamma \middle| S \right] \\ &\leq \lim_{n \rightarrow \infty} \mathbb{E}_S P_{X^n Y^n | S} \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n | S) \leq R_S - \gamma \middle| S \right] \\ &\leq \lim_{n \rightarrow \infty} [\mathbb{E}_S P_e^{(n, S)} + 2^{-n\gamma}] = 0, \end{aligned}$$

where the chain of inequalities follows from Fatou's lemma, (7), and the code constraint $\mathbb{E}_S P_e^{(n, S)} \rightarrow 0$. Since the probability must be non-negative, we conclude

$$\liminf_{n \rightarrow \infty} P_{X^n Y^n | S} \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n | S) \leq R_S - \gamma \middle| S \right] = 0$$

almost surely (a.s.) in S . Thus for any $\epsilon > 0$,

$$P_{X^n Y^n | S} \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n | S) \leq R_S - \gamma \middle| S \right] < \epsilon$$

occurs infinitely often a.s. Assuming $i_{X^n W^n}(X^n; Y^n | S)$ is bounded by M , we then have

$$\mathbb{E}_{X^n Y^n | S} \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n | S) \middle| S \right] > (R_S - \gamma)(1 - \epsilon) - \epsilon M$$

also occurs infinitely often a.s. Since ϵ is arbitrary, we see that

$$\mathbb{E}_S \mathbb{E}_{X^n Y^n | S} \left[\frac{1}{n} i_{X^n W^n}(X^n; Y^n | S) \middle| S \right] \geq \mathbb{E}_S R_S - \gamma$$

infinitely often for arbitrary γ , which gives us the upper bound (5) for expected capacity. Note that the expectation in the upper bound (5) is indeed $\frac{1}{n} I(X^n; Y^n | S)$.

III. EXAMPLES

A. Gilbert-Elliott Channel

The Gilbert-Elliott channel [9] is a two-state Markov chain, where each state is a BSC. The transition probabilities between channel states are g and b respectively, the initial state distribution is given by π_G and π_B for states G and B . The crossover probabilities for the ‘‘good’’ and ‘‘bad’’ BSCs satisfy $0 \leq p_G < p_B \leq 1/2$. We let $x_n \in \{0, 1\}$, $y_n \in \{0, 1\}$, and $z_n = x_n \oplus y_n$ denote the channel input, output, and error on the n th transmission. Whether or not the channel is stationary or ergodic depends on the channel parameters. We study the capacity definitions under a variety of scenarios.

Example 1: Ergodic Case, Stationary or Non-Stationary

Consider the case with g and b nonzero, $\pi_G = g/(g+b)$ and $\pi_B = b/(g+b)$, so the Gilbert-Elliott channel is stationary and ergodic. In this case the information density $\frac{1}{n} i_{X^n W^n}(X^n; Y^n)$ converges to a delta function at the average mutual information, so capacity equals average mutual information as usual. The Shannon capacity C is equal to the expected capacity $\pi_G C_G + \pi_B C_B$, where $C_G = 1 - h(p_G)$, $C_B = 1 - h(p_B)$ and $h(p) = -p \log p - (1-p) \log(1-p)$ is the binary entropy function.

This is not a composite channel. Since any transmission intrinsically experiences both good and bad channel states, the receiver has no basis of choosing certain transmissions over the others to declare outage. Outage capacity is not well defined for stationary and ergodic channels, but becomes more interesting for non-ergodic channels as in Example 2.

If $\pi_G \neq g/(g+b)$ but b and g are nonzero, then the Gilbert-Elliott channel is ergodic but not stationary. However, the distribution on the states G and B converges to a stationary distribution. Thus the channel is asymptotically mean stationary, and our four definitions of capacity have the same values as in the stationary case.

Example 2: Stationary and Nonergodic Case

We now set $g = b = 0$. So the initial channel state is chosen according to probabilities $\{\pi_G, \pi_B\}$ and then remains fixed for all time. The Shannon capacity equals that of the bad channel. The outage capacity $C_q^o = (1-q)C_B$ if the outage probability $q < \pi_B$ and $C_q^o = (1-q)C_G$ otherwise. The loss incurred from the lack of side information at the encoder is that the expected capacity is strictly less than the average of individual capacities $\pi_B C_B + \pi_G C_G$ and is equal to [6]

$$\max_{0 \leq r \leq 1/2} 1 - h(r * p_B) + \pi_G [h(r * p_G) - h(p_G)], \quad (8)$$

where $\alpha * \beta = \alpha(1-\beta) + (1-\alpha)\beta$. The interpretation here is that the broadcast code achieves rate $1 - h(r * p_B)$ for the bad channel and an additional rate $h(r * p_G) - h(p_G)$ for the good channel, so the average rate is the expected capacity.

Using the *tangent match* approach developed in [10] we can obtain r^* which maximizes (8). Namely if we define

$$A = \frac{1 - 2p_B}{1 - 2p_G}, \quad f(p_1, p_2) = \frac{\log(1/p_1 - 1)}{\log(1/p_2 - 1)},$$

then $r^* = 0$ if $\pi_G \leq Af(p_B, p_G)$; $r^* = 1/2$ if $\pi_G \geq A^2$ and r^* solves $f(r^*p_G, r^*p_B) = A/\pi_G$ otherwise. A more general case will be considered in the next example.

B. BSC with random crossover probabilities

We consider a BSC with random crossover probability $0 \leq p \leq 1/2$. At the beginning of transmission p is chosen according to some distribution $f(p)$ and then held fixed. We also denote $F(p) = \int_0^p f(s)ds$ as the cumulative distribution function. This is an example of a composite channel. The Shannon capacity is $C = 1 - h(p^*)$ where $p^* = \sup\{p : f(p) > 0\} = \inf\{p : F(p) = 1\}$, and the outage- q capacity is $C_q^o = (1 - q)[1 - h(p_q)]$ where $p_q = \inf\{p : F(p) \geq 1 - q\}$.

We then consider a broadcast approach on this channel. The receiver is equivalent to a continuum of ordered users, each indexed by the BSC crossover probability p and occurring with probability $f(p)dp$. The transmitter sends an infinite number of layers of coded information and each user decodes an incremental rate $|dR(p)|$ corresponding to its own layer. Since the BSC broadcast channel is degraded, a user with crossover probability p can also decode layers indexed by larger crossover probabilities, hence we achieve a rate of

$$R(p) = - \int_p^{1/2} dR(p). \quad (9)$$

The problem of determining the expected capacity then boils down to the characterization of the broadcast rate region.

In the discrete case with N users, assuming $0 \leq p_1 \leq \dots \leq p_N \leq (1/2)$, the capacity region is shown to be [11]

$$\{\mathbf{R} : R_i = h(r_i * p_i) - h(r_{i-1} * p_i), 1 \leq i \leq N\} \quad (10)$$

where $0 = r_0 \leq r_1 \leq \dots \leq r_N = 1/2$. Since the original broadcast channel is stochastically degraded it has the same capacity region as a cascade of N BSC's. The capacity region boundary is traced out by augmenting $(N - 1)$ auxiliary channels [11] and varying the crossover probabilities of each. For each i , r_i equals the overall crossover probability for auxiliary channels 1 up to i . See Figure 1 for an illustration.

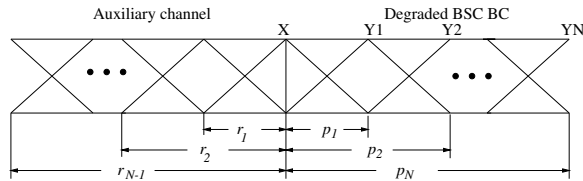


Fig. 1. BSC broadcast channel with auxiliary channels for random coding

We extend the above result to the continuous case with an infinite number of auxiliary channels. In this case we define a monotonically increasing function $r(p)$ equal to the overall crossover probability of auxiliary channels up to that indexed by p . In the following we use both $r(p)$ and r_p interchangeably. For the layer indexed by p , the incremental rate is

$$-dR(p) = h(p * r_p) - h(p * r_{p-dp}) \approx \delta \log(1/x - 1),$$

where $x = p * r_p$, $\delta = (1 - 2p)r'_p dp$, the first order approximation $r_{p-dp} \approx r_p - r'_p dp$ and $h(x - \delta) \approx h(x) - h'(x)\delta$. Here δ is a small variation, and we do not explicitly address the problematic limiting case $h'(x) \rightarrow \infty$ as $x \rightarrow 0$.

Overall the expected rate is

$$\begin{aligned} C^e &= \int_0^{1/2} f(p)R(p)dp = - \int_0^{1/2} F(p)dR(p) \\ &= \int_0^{1/2} F(p) \log\left(\frac{1}{p * r_p} - 1\right) (1 - 2p)r'_p dp. \end{aligned} \quad (11)$$

The optimal $r(p)$ maximizing the expected rate can be solved through calculus of functional variation. Let us represent

$$S(p, r_p, r'_p) = F(p) \log\left(\frac{1}{p * r_p} - 1\right) (1 - 2p)r'_p. \quad (12)$$

The optimal $r(p)$ should satisfy the Euler equation [12]

$$S_r - \frac{d}{dp} S_{r'} = 0 \quad (13)$$

where $S_r = \partial S / \partial r$ and $S'_r = \partial S / \partial r'$. After some algebra (13) simplifies to

$$\frac{(p * r_p)^{-1} - (1 - p * r_p)^{-1}}{\log_e(1 - p * r_p) - \log_e(p * r_p)} = \frac{(1 - 2p)f(p) - 2F(p)}{F(p)}. \quad (14)$$

In general (14) has no closed-form solution but there exist obvious numerical approaches.

As an example we examine capacity under different definitions for a crossover probability uniformly distributed on $[0, 1/2]$. The Shannon capacity $C = 0$. The outage- q capacity $C_q^o = (1 - q)[1 - h(\frac{1-q}{2})]$. To evaluate the expected capacity we solve $r(p)$ for each p in (14). It is seen that $0 \leq r_p \leq 1/2$ only for $p_l \leq p \leq p_u$ with the two cutoff probabilities $r(p_l) = 0$ and $r(p_u) = 1/2$. For the uniform distribution case, $p_l = 0.136$ and $p_u = 1/6$, which illustrates it is unnecessary to use the channel all the time to achieve the expected capacity. In fact no information is sent when $p \geq 1/6$.

In Figure 2 we plot the capacity under different definitions. We observe that, although the achievable rate in non-outage states, $C_q^o/(1 - q)$, could exceed the expected capacity C^e , the outage- q capacity C_q^o is always dominated by the expected capacity C^e . We further define cutoff outage probabilities $q_l = 1 - 2p_l$ and $q_u = 1 - 2p_u$. In Figure 3 we plot the achievable rates in each state for various capacity definitions. We see that the code for outage- q capacity achieves a constant rate for non-outage states and a rate 0 otherwise. For this example, the incremental rates $|dR(p)|$ are nonzero only for $p_l \leq p \leq p_u$. Therefore the code for expected capacity achieves a rate 0 when $p > p_u$. When p decreases from p_u to p_l , the rate gradually increases from 0 to 0.38 bits per channel use, and stays at this constant level for $p < p_l$. Since all channels are equally probable, the area under each curve is the average rate of that strategy. The area under the expected capacity curve is the largest. The expected capacity curve is, in some places, lower than the curve for outage- q_l capacity. Although the outage- q_l code achieves a rate higher than the broadcast

code for expected capacity when $p < p_l$, the same code has decoding rate 0 for all other channel states $p > p_l$, giving a lower area under the total curve.

In summary, when optimizing $r(p)$ for the expected capacity we first identify the cutoff probabilities (p_l, p_u) and then solve (14) to obtain $r(p)$ for each p . We want to emphasize that the correct cutoff range, although seemingly a very coarse characterization of the optimal solution, is crucial to the expected rate. Consider some alternative approaches:

- Optimal cutoff $[p_l, p_u]$: $r(p) = 0$ for $p \leq p_l$, $r(p) = \frac{(p-p_l)^\gamma}{2(p_u-p_l)^\gamma}$ for $p_l \leq p \leq p_u$, and $r(p) = 1/2$ for $p \geq p_u$;
- Entire range $[0, 1/2]$: $r(p) = (1/2)(2p)^\gamma$.

The choice of γ makes $r(p)$ convex ($\gamma > 1$), linear ($\gamma = 1$) or concave ($\gamma < 1$) in both approaches. The achievable rates are plotted for 20 trials in each case together with C^e in Figure 4. In each trial γ is randomly chosen between 0 and 4. We observe that the correct cutoff range yields an achievable expected rate very close to C^e . But the expected rate varies dramatically if we naively spread $r(p)$ over $[0, 1/2]$.

IV. CONCLUSION

In view of the pessimistic nature of Shannon capacity for composite channels with CSIR, we propose alternative capacity definitions including outage capacity and expected capacity. These definitions lend insights to applications where side information at the receiver combined with appropriate source coding strategies can exploit these more flexible notions of capacity. We prove capacity theorems or bounds under each definition, and illustrate how average achievable rates can be improved through examples of Gilbert-Elliot channels and a BSC with random crossover probabilities.

REFERENCES

- [1] R. Dobrushin. General formulation of Shannon's main theorem in information theory. *Amer. Math. Soc. Trans.*, 33:323–438, 1963.
- [2] S. Verdú and T. Han. A general formula for channel capacity. *IEEE Trans. Inform. Theory*, 40(4):1147–1157, July 1994.
- [3] M. Effros and A. Goldsmith. Capacity definitions and coding strategies for general channels with receiver side information. In *IEEE ISIT*, page 39, Cambridge MA, August 1998.
- [4] R. Ahlswede. The weak capacity of averaged channels. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 11:61–73, 1968.
- [5] L. Ozarow, S. Shamai and A. Wyner. Information theoretical considerations for cellular mobile radio. *IEEE Trans. Veh. Tech.*, 43(2):359–378, May 1994.
- [6] T. Cover. Broadcast channels. *IEEE Trans. Inform. Theory*, 18:2–14, Jan. 1972.
- [7] S. Shamai and A. Steiner. A broadcast approach for a single-user slowly fading MIMO channel. *IEEE Trans. Inform. Theory*, 49(10):2617–2635, Oct. 2003.
- [8] D. Gündüz and E. Erkip. Joint source channel codes for MIMO block fading channels. Submitted to *IEEE Trans. Inform. Theory*, March 2006.
- [9] M. Mushkin and I. Bar-David. Capacity and coding for the Gilbert-Elliot channels. *IEEE Trans. Inform. Theory*, 35(6):1277–1290, Nov. 1989.
- [10] Y. Liang and A. Goldsmith. Rate regions and optimal power allocation for TD fading broadcast channels without CSIT. In *Allerton*, Monticello IL, Sept. 2005.
- [11] P. Bergmans. Random coding theorem for broadcast channels with degraded components. *IEEE Trans. Inform. Theory*, 19(2):197–207, March 1973.
- [12] D. Luenberger. *Optimization by vector space methods*. John Wiley & Sons, Inc., New York NY, 1969.

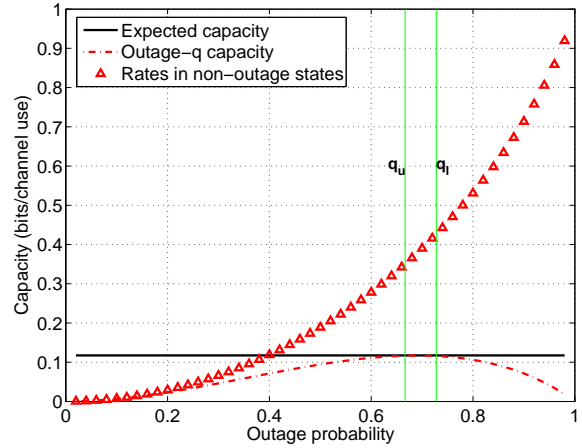


Fig. 2. Capacity under different definitions

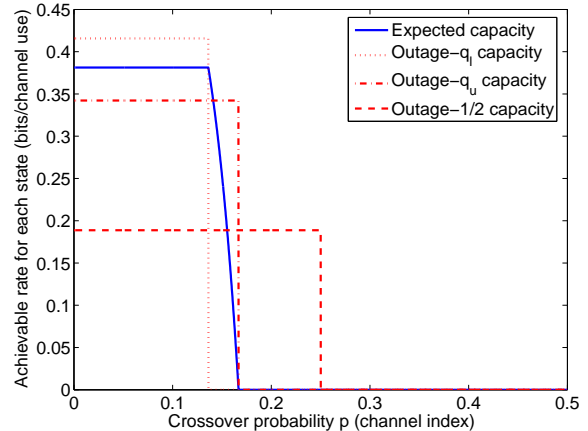


Fig. 3. Achievable rate for each channel state

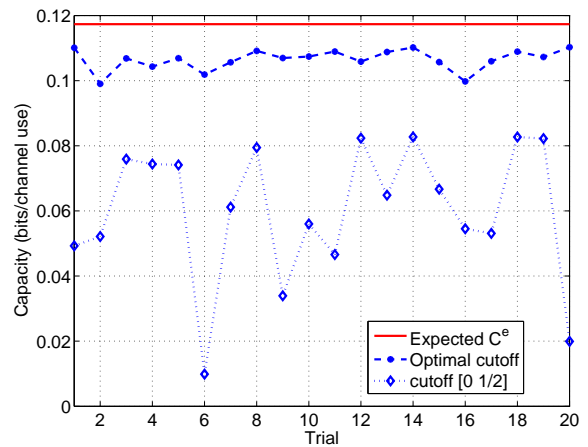


Fig. 4. Effect of cutoff range