Comments on unknown channels

Kristen Woyach, Kate Harrison, Gireeja Ranade and Anant Sahai
Wireless Foundations, EECS, UC Berkeley
{kwoyach, harriska, gireeja, sahai}@eecs.berkeley.edu

***Abstract*—The idea of modeling an unknown channel using a broadcast channel was first introduced by Cover[1] in 1972. This paper builds on his line of thought to consider priority encoding of communication over unknown channels without feedback, using fixed-length codes and from a single-shot, individual channel perspective. A ratio-regret metric is used to understand how well we can perform with respect to the actual channel realization.**

I. Introduction

How does one communicate without knowledge of the quality of the communication channel? This practical problem has been extensively considered from a wide range of perspectives. This paper considers single-shot, fixed-blocklength transmission without feedback, with a priority ordering over the message bits.

One of the first approaches for the unknown channel problem was put forth by Cover in [1], [2], where he suggested modeling an unknown (compound) channel using a family of degraded broadcast channels. This approach explicitly considers different performances over a range of channel realizations, and contrasts with Blackwell's arbitrarily varying channel (AVC) approach, which provides hard guarantees for adversarial scenarios [3]. A third approach is to use rateless codes [4].

Communication over unknown channels naturally parallels problems in universal source coding. Universal *lossless* source coding is like the AVC approach in one sense, in that both provide hard guarantees on the quantities of received information and no errors are tolerated. Variable-length rateless channel coding is another error-free approach. On the other hand, the degraded broadcast approach to transmission over unknown channels corresponds to the soft source coding guarantees that are given by successive refinement for *lossy* source-coding [5], [6], where the message bits are assigned a priority.

Priority encoding [7] has mostly been looked at in information theory through the lens of error-exponent analysis [8]. Here, we aim to provide a simplified[2] view of the problem in the channel coding case, with the fundamental goal of understanding the tradeoff between outage and graceful performance degradation. The idea is similar to the variable-blocklength-encoding discussed in [9], with message priority being used to determine the encoding.

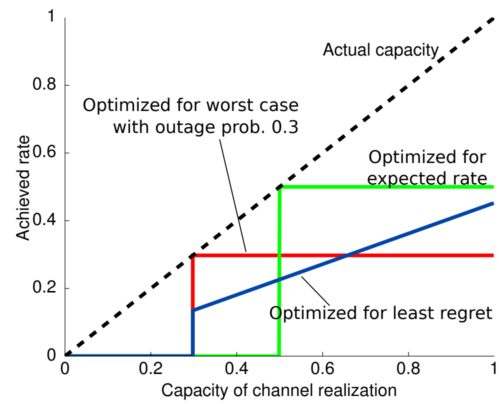

Fig. 1. Rate achieved given a channel realization in the range between 0.3 and 1 using timesharing schemes. Scheme-1 (optimized for worst case) with outage probability 0.3 codes all information at rate 0.3. Scheme-2 (optimized for expected rate) maximizes expected rate, assuming a uniform channel realization between 0.3 and 1. Notice some targeted users (realizations better than 0.3) get no rate, and both schemes allocate all resources to one realization. Scheme-3 (optimized for least regret) allows higher rate for better realizations, and has smooth improvement above the outage probability, 0.3.

Much of the work that follows up on Cover's idea is based on the idea of having a prior distribution over a range of channel parameters, and using this prior to maximize expected performance [10], [9], [11]. When a prior over the fades is available, successive refinement ideas have even been combined with the degraded broadcast approach to optimize the expected distortion of a Gaussian source [12], [13].

But what if the transmitter has no prior? If they were allowed, rateless codes [4] would be almost perfect: they guarantee decodability and are largely agnostic to any prior over the channel parameters[3]. The individual sequence approach also assumes no prior information about the channel, and tries to achieve a good rate as compared to the actual channel realization [14], [15] using a scheme that relies on a small amount of feedback[4]. This feedback is also practically essential to allow the use of rateless codes.

[1]The title of this paper is a tribute to Cover's "Comments on broadcast channels".

[2]Error-exponent analysis at its heart is an analysis of channel atypicality. We simplify this view by directly considering variable channel behaviors without the underlying tail distributions.

[3]Even rateless codes might be suboptimal when dealing with a family of channels that are not all optimized by the same input distribution (e.g. the Z channel). A rateless code suffers a penalty for choosing the wrong input distribution for the realized channel.

[4]Implicit in the individual channel framework are two categories of uncertainty: uncertainty over the channel distribution, and uncertainty over the channel realization. Here, we focus on the former category. For the latter category, tools from [16], [17], [14], [15] use feedback to implement rateless codes and learn the channel realization. Without feedback, common-randomness (i.e. interleaving) can be used to turn an individual "noise" sequence into essentially an unknown channel noise distribution.

Communication over channels in heterogeneous networks often involve unknown recipients, and in these cases, not only is the channel unknown, but there is also no informative prior available or any hope of feedback[5]. We extend the ideas of successive refinement and a broadcast approach for one-shot transmissions over channels without a prior distribution.

Our objective is to do well with respect to a reference: the realized channel capacity. We use *regret* — the ratio of the realized channel capacity to the achieved rate — as a performance metric, and minimize the worst-case regret. Universal source coding deals with a similar notion — the minmax redundancy [19]. Such comparisons are also commonly used as metrics in the machine-learning literature [20].

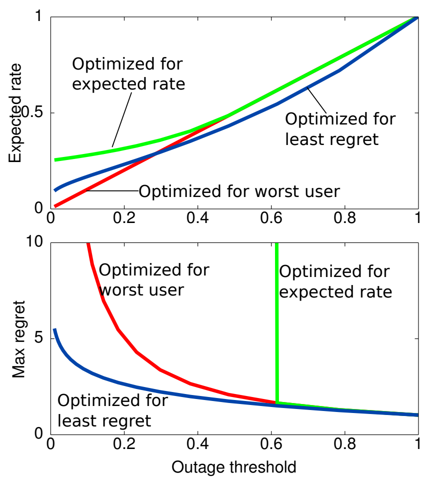

Fig. 2. Performance of rate allocations in Fig. 1, given different metrics. The expected rate allocation trades a higher outage probability for higher long-term rate. The minmax regret allocation trades away some long-term rate for better short-term performance.

Consider Fig. 1, which compares time-sharing schemes optimized for different metrics assuming a uniform prior over channel realizations with capacities between 0 and 1. In this figure, we assume 0.3 is the outage level — any lower realization is not deemed worth devoting resources to. Scheme-1 (optimized for worst case) transmits all information at rate 0.3 and ensures non-zero rate for every targeted realization. Scheme-2 (optimized for expected rate) optimizes the expected rate given a uniform prior over $[0.3, 1]$. Scheme-3 (optimized for least regret) minimizes the maximum regret over realizations between 0.3 and 1.

Notice in Fig. 1, Schemes-1 and -2 are qualitatively similar: both allocate all available resources to a single channel realization. Any realization below this threshold will be in outage, and any above the threshold will receive information at the threshold rate. On the other hand, Scheme-3 has a graceful improvement in achieved rate as the channel improves above the chosen outage realization.

[5]For example, in cognitive radio networks, a primary may need to signal its priority status to an unknown secondary with delay constraints. Secondary radios may be required to transmit a prioritized identity to the primary along with the main transmission intended for its own receiver. In cases of harmful interference, the identity of the interferer needs to be resolved to the best possible resolution even if the main transmission cannot be decoded [18].

Fig. 2 shows how the three schemes compare under the metrics of expected rate and worst regret. As we will see in Sec. III-B, *every* timesharing scheme requires some outage to maintain finite regret. Hence, the x-axis shows the lowest targeted realization (in Fig. 1, this was 0.3). With the expected rate metric (Fig. 2, top), Scheme-2 will do better than serving the worst user as long as it is beneficial to make the actual outage threshold higher than the predetermined threshold. Scheme-3 will always have a worse expected rate, since it sacrifices some long-term average rate for better one-shot performance.

We note that neither Scheme-1 or -2 does particularly well for one-shot transmissions (given by the max regret metric, Fig. 2, bottom). The worst regret using Scheme-2 can be infinite for a large chunk of targeted but low-SNR users. Scheme-2 has maximum regret that decreases as the inverse of the outage capacity. Compare this to the minmax regret scheme (Scheme-3) which has maximum regret decreasing as the natural log of the outage capacity. In this paper, we will see that a similar regret curve is achievable for many channels with unknown parameter.

II. Problem Setup

Our problem involves an encoder and decoder communicating over an unknown channel. The channel, $P_\pi(Y|X)$, $X \in \mathcal{X}$, $Y \in \mathcal{Y}$, is drawn from a known set $\mathcal{S}$ of possible memoryless channels. The message is an infinitely long bitstring $M = m_0 m_1 m_2 \cdots \in \{0,1\}^\infty$. The encoder $\mathcal{E} : \{0,1\}^\infty \to \mathcal{X}^n$ is unaware of the channel. The decoder $\mathcal{D}_\pi : \mathcal{Y}^n \to \{0,1\}^{k(\pi)}\ \forall\, \pi$ is aware of the channel π and commits to $k(\pi)$. The probability of error, $P(\mathcal{D}_\pi(Y^n) \neq m_0 m_1 \cdots m_{k(\pi)-1})$, must go to zero as $n \to \infty$. Let $R_n(\pi, \mathcal{E}, \mathcal{D}_\pi) = \frac{k(\pi)}{n}$ be the rate achieved for the channel π, with encoding $\mathcal{E}$ and decoder $\mathcal{D}_\pi$. Then, we define $R(\pi, \mathcal{E}, \mathcal{D}_\pi) = \lim_{n\to\infty} \frac{k(\pi)}{n}$.

A special case of this problem is where $\mathcal{S} = \mathcal{S}_\beta$ is a degraded family parametrized by a single parameter, β, the capacity and $\beta_{min} \leq \beta \leq \beta_{max}$. A degraded family is a family of channels where $\forall\ \beta_1 > \beta_2$ and corresponding channels $P_{\beta_1}(Y_1|X)$, $P_{\beta_2}(Y_2|X)$ in $\mathcal{S}_\beta$, $\exists\ P_{\beta_1,\beta_2}(Y_2|Y_1)$ such that $X - Y_1 - Y_2$ form a Markov chain.

We will now show that the problem of transmitting over an unknown channel in $\mathcal{S}_\beta$ is equivalent to the problem of transmitting nested messages over a degraded broadcast channel to users with channels from $\mathcal{S}_\beta$. User U_β has channel $P_\beta(Y|X)$. Let $\mathcal{M}$ be the set of all messages to all users. The messages come from a degraded message set also parametrized by β. R_β is the rate of the message to U_β, i.e. $M_\beta \in \mathcal{M}_\beta = \{1, 2 \cdots 2^{\lfloor nR_\beta \rfloor}\}$. A degraded message set is one where $\forall \beta_1 > \beta_2, M_{\beta_2}$ is a prefix of M_{β_1}. $\mathcal{E}_{deg} : \mathcal{M} \to \mathcal{X}^n$, and decoder $\mathcal{D}_{deg,\beta} : Y^n \to \mathcal{M}_\beta$. The total rate to U_β is $\widetilde{R}_\beta$.

Theorem 2.1: $\widetilde{R}_\beta$ is achievable at U_β under the degraded broadcast setup with a degraded message set $\forall\ \beta_{min} \leq \beta \leq \beta_{max}$ iff $\exists\ \mathcal{E}, \mathcal{D}_\beta$ such that $R(\beta, \mathcal{E}, \mathcal{D}_\beta) = \widetilde{R}_\beta$.

Proof: ($\Rightarrow$) $\forall n$ note that the infinite string M can be easily parsed as a degraded message set, and consider the

$R(\beta)$s achieved by $\mathcal{E}_{deg}, \mathcal{D}_{deg,\beta}$. With the re-encoded string as message sets, $R(\beta, \mathcal{E}_{deg}, \mathcal{D}_{deg,\beta}) = \widetilde{R}(\beta)\ \forall\ \beta$.
($\Leftarrow$) $\forall n$ consider $M_{\beta_{max}} \in \{1, 2, \cdots 2^{\lfloor nR_{\beta_{max}} \rfloor}\}$, and represent it as a binary string, $\widetilde{M} \in \{0,1\}^{n\lfloor R_{\beta_{max}} \rfloor}$. Notice that, $\forall \beta \leq \beta_{max}$, M_β is a prefix of $M_{\beta_{max}}$. Now, fill in all bits of $\widetilde{M}$ beyond $\lfloor nR_{\beta_{max}} \rfloor$ randomly to get $M \in \{0,1\}^\infty$. Clearly, $\widetilde{M}$ is a prefix of M. Now, using $\mathcal{E}$ and decoders $\mathcal{D}_\beta$, U_β can decode $R(\beta, \mathcal{E}, \mathcal{D}_\beta)$. But this is the desired rate, $\widetilde{R}_\beta$. ■

Clearly, the two formulations are equivalent. With this in mind, we will interchangeably think of channel realizations as virtual receivers of the degraded broadcast channel [12].

Let $R'(\beta) = \lim_{\epsilon \to 0} \frac{\lfloor R_{\beta+\epsilon} - R_\beta \rfloor}{\epsilon}$ be the differential amount of information sent at rate β. The total decodable information, $R(\beta)$, for $P_\beta(y|x)$ is the integral of the incremental rates achieved at all weaker channels, i.e. $R(\beta) = \int_0^\beta R'(\beta)\ d\beta$.

Definition 2.1: The *regret* for allocation R' at β, $\xi(R', \beta)$,

$$\xi(R', \beta) = \frac{\beta}{\int_0^\beta R'(\alpha)\ d\alpha}, \tag{1}$$

is the ratio of the actual capacity to the achieved rate.

We aim to minimize the worst-case regret $\min_{R'} \max_\beta \xi(R', \beta)$. However, there is a tension between the range of users (β's) supported and the minmax regret achieved[6]. Hence, we define C_o:

Definition 2.2: The outage level, C_o, is the minimum capacity supported by an achievable scheme, i.e. $R(\beta) = 0$ for $\beta < C_o$ and $R(\beta) > 0$ for $\beta \geq C_o$.

We define the metric for performance as the worst case regret for a fixed outage level:

$$\xi_m(C_o) = \min_{R'} \max_{\beta \geq C_o}\ \xi(R', \beta). \tag{2}$$

As we will see in Sec. III-B on binary erasure channels, $C_o > 0$ is necessary to be able to achieve finite minmax regret, $\xi_m(C_o)$. For the BEC, $\xi_m(0) = \infty$.

III. Examples

A. Linear deterministic channels

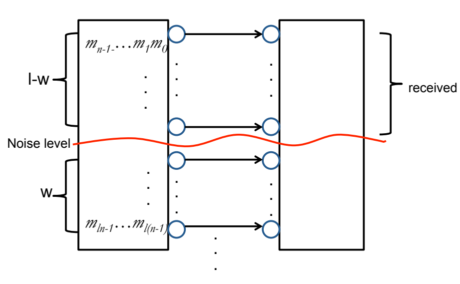

Fig. 3. Achieved rate for the deterministic model is equal to the capacity of the realized channel.

To illustrate the regret metric and understand the best possible case, we explore linear deterministic[7] channels [23]. For this channel, $\xi(\mathbf{R}, \beta)$, is always 1, regardless of C_o. Even though the realized channel capacity is unknown, it can always be achieved.

Consider a linear deterministic channel, $Y = X + Z$, such that the signal, X has power $P = 2^l$, and the noise Z has maximum power $P_{noise} = 2^w$. Thus, $y_i = x_i$ for $l \geq i \geq 1$, and all lower bits are lost to the noise. The parameter w, is unknown and indexes a family of degraded channels. It is clear that the capacity of any member of this family is given by $\beta = l - w$.

Theorem 3.1: For the deterministic channel described above $\xi_m(C_o) = 1\ \forall\ C_o$.

Proof: Achievability: Let the message $M = m_0 m_1 m_2 ...$, ordered so m_0 is the highest priority bit. Then, $\forall$ blocklength n send $x_l = m_0 m_1 \cdots m_{n-1}, x_{l-1} = m_n, \cdots m_{2n-1}$ and so on. Given noise realization $\widetilde{w}$, the receiver can decode all n bits from the first $l - \widetilde{w}$ bit levels . The achieved rate is $R(\beta) = l - \widetilde{w} = \beta$, with regret $\xi(\mathbf{R}, \beta) = 1$ for every β. *Converse:* The regret is trivially bounded below by 1. ■

The deterministic channel achieves a perfect regret of 1 because the bit-levels, or "virtual channels" [12], [13], do not interact. So, information sent at one level does not interfere with the transmissions on any other level. As we will see, other channels cannot perform so well.

B. Binary erasure channels

Consider a family of degraded BECs parametrized by their erasure probability $0 \leq \epsilon \leq 1$, with capacity parameter $\beta = 1 - \epsilon$. Any achievable rate allocation $R'(\beta)$ is contained within the timesharing region [24], [7] as below[8]:

Theorem 3.2: The rate region for a degraded family of BECs $P_\beta(y|x)$ is given by $R'(\beta)$ satisfying

$$\int_0^1 \frac{R'(\beta)}{\beta} d\beta \leq 1. \tag{3}$$

Thus, the best performance under any metric achieved using time-sharing as in Fig. 2 is in fact the optimal performance possible for the family of BECs.

Lemma 3.3: The minmax regret for the BEC for outage level C_o, $\xi_m(C_o)$, is achieved by an R' such that $\xi(R', \beta) = \theta$ for some constant θ, and $R'(\beta) = \rho$, a constant, $\forall\ \beta > C_o$. The only atom in the rate allocation is at C_o.

Proof: The proof follows from a standard argument using calculus of variations. We first argue that the rate allocation must be continuous across the parameter space except for at C_o, and then use this to show that the minmax is achieved when the regret is constant across the parameter range. ■

Theorem 3.4: The minmax regret $\xi_m(C_o)$ for the degraded BEC family is $\xi_m(C_o) = 1 - \ln(C_o) = 1 + \ln(1/C_o)$.

[6][21] notes a tension between reliability and utility in the cognitive radio setting with regards to choosing whitespace device transmission power limits. The authors find that sacrificing the use of a few receivers yields a substantial increase in utility for the remaining transmitters.

[7]A parallel analysis will give regret 1 for a truncation channel [22] as well.

[8]Although traditionally rate regions are hosted in finite dimensions, our minmax regret formulation is naturally considered over a continuous family. So, we adopt the appropriate differential rate regions whenever possible.

Proof: The equal regret $\xi(R', \beta) = \theta$ and equal rate $R'(\beta) = \rho \;\forall\; \beta > C_o$ constraints from lemma 3.3 imply

$$\theta = \frac{C_o}{R'(C_o)} = \frac{\beta}{R'(C_o) + \rho \cdot (\beta - C_o)} \;\forall\; \beta. \tag{4}$$

Hence, $\rho = \frac{R'(C_o)}{C_o}$. Plugging into the rate region (3), and noting that the minmax point must exist on the boundary of the rate region, we have

$$\frac{R'(C_o)}{C_o} + \int_{C_o^+}^{1} \frac{R'(\beta)}{\beta} d\beta = 1 \tag{5}$$

$$\rho + \int_{C_o^+}^{1} \frac{\rho}{\beta} d\beta = 1, \text{ since } R'(\beta) = \rho,\; C(\beta) = \beta \tag{6}$$

$$\rho = (1 - \ln C_o)^{-1} \tag{7}$$

Hence, $\xi_m(C_o) = \theta = \frac{1}{\rho} = 1 - \ln(C_o) = 1 + \ln(\frac{1}{C_o})$. ■

The above minmax regret computation holds for any time-sharing based scheme, since the BEC rate region is equivalent to the timesharing region over the same range of capacities.

As $C_o \to 0$, $\theta \to \infty$. This is not surprising. Since the rate region of the BEC can be achieved using timesharing, we know that any achievable regret vector $\xi(R', \beta)$, and hence min-regret $\xi_m(C_o)$, can also be achieved using a timesharing scheme. A regret of 1 is unattainable since as soon as a block of time $l(\beta)$ is allocated for a certain erasure probability $\epsilon = 1 - \beta$:

- weaker realizations with a higher erasure rate will be unable to decode the block, and hence, will not be able to achieve their capacity and face regret.
- stronger realizations with a lower erasure rate will face a regret due to the lower transmission rate.

In this way, the BEC behaves in a fundamentally different way from the linear deterministic channels.

C. Gaussian channels

Consider a degraded family of complex Gaussian channels with power constraint P, parametrized by their capacity, $C_o \leq \beta < \infty$, in nats. We know that $\beta = \ln(1 + P/N(\beta))$, so $N(\beta) = P/(e^{\beta} - 1)$. Any achievable rate allocation lies within the Gaussian degraded broadcast rate region[9] given by superposition parametrized by a power allocation with an atom at P_{C_o} and otherwise continuous over the range of β:

$$P = P_{C_o} + \int_{C_o}^{\infty} P(\beta) d\beta \tag{8}$$

$$R(C_o) = \ln\left(1 + \frac{P_{C_o}}{P - P_{C_o} + N(C_o)}\right) \tag{9}$$

$$R(\beta) = R(C_o) + \int_{C_o}^{\beta} \frac{P(\alpha)}{\int_{\alpha}^{\infty} P(y)dy + N(\alpha)} d\alpha \tag{10}$$

for all $\beta \geq C_o$.

We believe that the minmax regret for the Gaussian is achieved by an R' such that $\xi(R', \beta) = \theta$ for some constant θ. Based on this conjecture, the minmax regret $\xi_m(C_o)$ for the degraded Gaussian family is the solution to

$$\begin{aligned} &\min \;\; \theta \\ &\text{s.t. } P(\beta, \theta) \geq 0 \;\forall\; \beta \geq C_o \end{aligned} \tag{11}$$

where $P(\beta, \theta) = P(C_o, \theta) \cdot e^{-\frac{\beta - C_o}{\theta}} - P \cdot \sum_{m=1}^{\infty} \frac{me^{-mC_o}}{1 - m\theta} \cdot \left(e^{-m(\beta - C_o)} - e^{-\frac{\beta - C_o}{\theta}}\right)$, and $P(C_o, \theta) = \frac{P}{\theta}\left(\frac{e^{C_o - C_o/\theta}}{e^{C_o} - 1}\right)$.

To see this, note that if $\xi(R', \beta) = \theta$ as above, then $R(C_o) = C_o/\theta$. Hence, $R'(\beta) = \theta^{-1}$ for all β, where θ is the optimal regret. The optimizing power allocation, $P(\beta, \theta)$ above, is the solution to the first order integral equation $P(\beta, \theta) = \theta^{-1}(\int_{\beta}^{\infty} P(y, \theta)dy + N(\beta))$, with initial condition $P(C_o, \theta)$. $P(C_o, \theta)$ is calculated using the power constraint equation, and is feasible if $P(\beta, \theta) \geq 0 \;\forall\; \beta \geq C_0$. $P(\beta, \theta)$ has at most three extrema (over β), which makes the feasibility constraint easy to check.

Numerical search leads to the same solution as above. The Gaussian outage-regret curve is in Fig. 4 for $\beta_{max} = \infty$. We also include curves for $\beta_{max} = \{.5, 1, 5\}$ for comparison.

D. Binary symmetric channels

It is difficult to express a differential form of the rate region (found in [27]). So, we numerically calculate an achievable minmax regret by finding an achievable rate allocation that produces an equal regret for all channel realizations above outage. The result is shown in Fig. 4.

IV. Discussion and future work

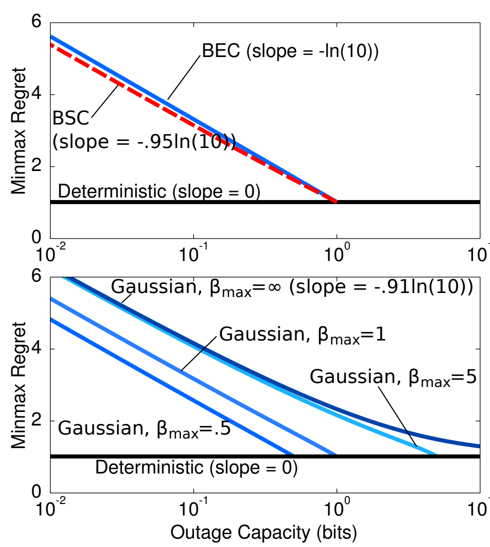

Fig. 4. Regret vs. outage curve for different channel models

Fig. 4 numerically compares the regret versus outage curves for the BEC, BSC and Gaussian channels. First, let us note that the curve for the BEC is a straight line, $1 + \ln(C_o)$, and both the BSC and Gaussian curves have linear regimes where the minmax regret falls as $\gamma \ln(C_o)$ for some $\gamma \geq 0$. This is significantly better than the γ/C_o regret achieved in the

[9] To find this differential form, take the discrete region in [25], let the power at each rate get very small, and linearize using $\ln(1 + x) \approx x$ for small x. More details are given in [26].

introduction by optimizing our transmission for the weakest user at rate C_o. We believe that the slope γ of the curves represents the ability of the channel to support superposition coding. The deterministic model curve has slope $\gamma = 0$, and represents the best possible situation. Each bit-level above noise can be independently used for transmission. The BEC, on the other hand, has slope $-1 \cdot \ln(10)$, (corrected since the x-axis is in base 10). This is the timesharing curve: the resources allocated to the individual realizations interact such that no superposition is possible as discussed earlier. These two slopes bound what is possible for *any* degraded family of channels because slope better than 0 is not possible, and a timesharing scheme is always achievable.

Observe that in Fig. 4, as the Gaussian SNR $\to \infty$, the slope, $\gamma \to 0$. This matches the traditional intuition that the Gaussian at high SNR is well represented by the deterministic channel, and essentially perfect superposition coding is possible. The slopes of the Gaussian model in low SNR scenarios are tricker to interpret. Since it is difficult to obtain an explicit expression for γ in the Gaussian case, we rely on numerical calculations to understand it. Preliminary calculations indicate that they approach the slope of the BEC, indicating that superposition gains, while technically possible, are very small for low SNR Gaussian broadcast channels. However, numerical estimates for the slope of the BSC give $\gamma \approx -.95 \ln(10)$. This is better than the slope of the BEC, and suggests that perhaps gains from superposition coding are possible for the BSC broadcast channel.

An interesting question to consider is: could we construct a channel with a regret-outage slope of any desired value between 0 and -1? Such a construction could provide insight into the transitional behavior of the Gaussian — what exactly is going on in between the low and high SNR regimes?

These ideas can be extended beyond even different families of channels. For example if a cognitive radio is sending an identity signal through an unknown channel to a primary [18], both the channel quality and the possible blocklength may be unknown. Does there exist a strategy similar to the approach here to allow identity rate to gracefully improve with both a better channel and more time? At what cost?

Finally, what is the lossy source coding analog of this channel coding regret formulation? There is obviously a connection between the minmax regret results here and the ideas of successive refinement in [6] which remains to be explored.

V. Acknowledgements

We thank the U.S. NSF (CNS-403427, CNS-0932410 and Graduate Fellowships) for their generous funding support and the reviewers for their comments.

References

[1] T. Cover, "Broadcast channels," *Information Theory, IEEE Transactions on*, vol. 18, no. 1, pp. 2–14, 1972.

[2] ——, "Comments on broadcast channels," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2524–2530, 1998.

[3] D. Blackwell, L. Breiman, and A. Thomasian, "The capacities of certain channel classes under random coding," *The Annals of Mathematical Statistics*, vol. 31, no. 3, pp. 558–567, 1960.

[4] M. Luby, "LT codes," in *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*. IEEE, 2002, pp. 271–280.

[5] W. Equitz and T. Cover, "Successive refinement of information," *Information Theory, IEEE Transactions on*, vol. 37, no. 2, pp. 269–275, 1991.

[6] L. Lastras and T. Berger, "All sources are nearly successively refinable," *Information Theory, IEEE Transactions on*, vol. 47, no. 3, pp. 918–926, 2001.

[7] S. Boucheron and M. Salamatian, "About priority encoding transmission," *Information Theory, IEEE Transactions on*, vol. 46, no. 2, pp. 699–705, 2000.

[8] S. Borade, B. Nakiboglu, and L. Zheng, "Unequal error protection: An information-theoretic perspective," *Information Theory, IEEE Transactions on*, vol. 55, no. 12, pp. 5511–5539, 2009.

[9] S. Verdú and S. Shamai, "Variable-rate channel capacity," *Information Theory, IEEE Transactions on*, vol. 56, no. 6, pp. 2651–2667, 2010.

[10] S. Shamai, "A broadcast strategy for the Gaussian slowly fading channel," in *Information Theory. 1997. Proceedings., 1997 IEEE International Symposium on*. IEEE, 1997, p. 150.

[11] L. Varshney, S. Mitter, and V. Goyal, "Channels that die," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*. IEEE, 2009, pp. 566–573.

[12] C. Tian, A. Steiner, S. Shamai, and S. Diggavi, "Successive refinement via broadcast: Optimizing expected distortion of a Gaussian source over a Gaussian fading channel," *Information Theory, IEEE Transactions on*, vol. 54, no. 7, pp. 2903–2918, 2008.

[13] C. Ng, D. Gunduz, A. Goldsmith, and E. Erkip, "Minimum expected distortion in Gaussian layered broadcast coding with successive refinement," in *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*. IEEE, 2007, pp. 2226–2230.

[14] Y. Lomnitz and M. Feder, "Communication over individual channels," *Information Theory, IEEE Transactions on*, vol. 57, no. 11, pp. 7333–7358, 2011.

[15] V. Misra and T. Weissman, "The porosity of additive noise sequences," *Information Theory, 2012. ISIT 2012. Proceedings. International Symposium on*, 2012.

[16] O. Shayevitz and M. Feder, "Communicating using feedback over a binary channel with arbitrary noise sequence," in *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on*. IEEE, 2005, pp. 1516–1520.

[17] K. Eswaran, A. Sarwate, A. Sahai, and M. Gastpar, "Zero-rate feedback can achieve the empirical capacity," *Information Theory, IEEE Transactions on*, vol. 56, no. 1, pp. 25–39, 2010.

[18] A. Sahai, K. Woyach, G. Atia, and V. Saligrama, "A technical framework for light-handed regulation of cognitive radios," *Communications Magazine, IEEE*, vol. 47, no. 3, pp. 96–102, 2009.

[19] N. Merhav and M. Feder, "Universal prediction," *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2124–2147, 1998.

[20] N. Cesa-Bianchi and G. Lugosi, *Prediction, learning, and games*. Cambridge University Press, 2006.

[21] K. Harrison and A. Sahai, "Potential collapse of whitespaces and the prospect for a universal power rule," in *New Frontiers in Dynamic Spectrum Access Networks (DySPAN), 2011 IEEE Symposium on*. IEEE, 2011, pp. 316–327.

[22] N. Martins, M. Dahleh, and N. Elia, "Feedback stabilization of uncertain systems in the presence of a direct link," *Automatic Control, IEEE Transactions on*, vol. 51, no. 3, pp. 438–447, 2006.

[23] A. Avestimehr, S. Diggavi, and D. Tse, "Wireless network information flow: A deterministic approach," *Information Theory, IEEE Transactions on*, vol. 57, no. 4, pp. 1872–1905, 2011.

[24] S. Borade, L. Zheng, and M. Trott, "Multilevel broadcast networks," in *Information Theory, 2007. ISIT 2007. IEEE International Symposium on*. IEEE, 2007, pp. 1151–1155.

[25] P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise (corresp.)," *Information Theory, IEEE Transactions on*, vol. 20, no. 2, pp. 279–280, 1974.

[26] http://www.eecs.berkeley.edu/~kwoyach/papers/itw2012Extended.pdf.

[27] T. Cover and J. Thomas, *Elements of information theory*, 2nd ed. Wiley, 2006.