Information Embedding and Related Problems: Recent Results and Applications

Brian Chen Stark C. Draper Gregory W. Wornell

Chinook Communications, Inc.
Lexington, MA 02421
{bchen,gww}@chinook.com

Massachusetts Institute of Technology
Cambridge, MA 02139
{bchen,scd,gww}@allegro.mit.edu

Abstract

Some applications and generalizations of the problem of coding with side information are developed. In the context of Gaussian broadcast channels, we show how embedding codes, based on information embedding techniques, can have significant advantages over superposition codes in several scenarios of interest, particularly when upgrading legacy digital systems.

Performance limits of noisy generalizations of the coding with side information problem are also developed for finite-alphabet and Gaussian-quadratic cases. The information embedding version—writing on dirty paper wearing bad glasses—may have applications to the multiple-access channel. The source coding version, which generalizes the Wyner-Ziv problem, has applications ranging from constrained relay channels to forms of sequential estimation that arise in distributed sensor networks. For the latter, distortion-rate behavior is characterized in the Gaussian-quadratic case for a sensor pipeline.

1 Introduction

While some of the earliest applications of information embedding techniques have been in watermarking and data hiding areas, the array of applications, richness of the theory, and interconnections to other problems continue to grow. This paper establishes some new results in this realm.

We first explore aspects of the role information embedding techniques have to play in communication over the classical broadcast channel. In particular, after summarizing how codes based on information embedding can be used as scalable alternatives to superposition codes for the broadcast channel, we develop the somewhat special role it has to play in some backward-compatible upgrading of digital communication infrastructure in the broadcast context.

Information embedding also has potential applications to multiple-access channel problems. As a foundation for investigations in this area, we analyze the problem of writing on dirty paper while wearing bad glasses, i.e., information embedding where the encoder's knowledge of the host signal is imperfect.

As has been explored recently by several groups, there is an important information theoretic duality between information embedding and the problem of source coding with

This work has been supported in part by the National Science Foundation under Grant No. CCR-0073520, and by a grant from Microsoft Research.

side information, whereby good information embedding encoders make good Wyner-Ziv decoders [13], and vice-versa. Motivated by this relationship, we investigate the the impact of imperfect encoder knowledge on the Wyner-Ziv problem. We term this "noisy source coding with side information."

Single-letter characterizations of the performance limits for the noisy variants of both source and channel coding problems are developed in the general finite-alphabet context, as well as in the Gaussian-quadratic case.

Finally, an application of these results to a sequential estimation problem that arises in distributed sensor networks is developed. The results leads to some efficient structures for "sensor pipelines."

2 Embedding Codes for Broadcast Channels

In this section we develop aspects of the role information embedding techniques have to play in broadcast channel applications. We first review the Gaussian broadcast channel and describe how information embedding techniques can be used to achieve channel capacity. We then consider a broadcast scenario with two classes of receivers: legacy receivers and new receivers. We discuss scenarios where information embedding techniques allow us to upgrade this digital communications architecture in a more efficient backwards-compatible manner than can be accomplished with previous techniques.

First, consider the two-receiver broadcast channel. Recall that when the transmitter is constrained to use power P and the two receivers experience noises with variances N_1 and N_2, respectively, with $N_1 \geq N_2$, then the achievable rate region for this channel is [8]

$$R_1 \leq \frac{1}{2}\log\left[1 + \frac{\alpha P}{(1-\alpha)P + N_1}\right] \quad \text{and} \quad R_2 \leq \frac{1}{2}\log\left[1 + \frac{(1-\alpha)P}{N_2}\right], \tag{1}$$

where α may be arbitrarily chosen ($0 \leq \alpha \leq 1$) to trade off rate between the receivers. The classical and familiar method for achieving these rates is superposition coding combined with successive cancellation decoding [8]. In particular, the encoder maps the messages for each receiver, m_1 and m_2 respectively, into the corresponding channel codewords, $\mathbf{c}_1$ and $\mathbf{c}_2$, and transmits the superposition $\mathbf{c}_1+\mathbf{c}_2$. The weak receiver (experiencing noise N_1) effectively treats $\mathbf{c}_2$ as extra channel noise in its decoding of m_1. The strong receiver (experiencing noise N_2) also decode m_1, subtracts off the corresponding codeword $\mathbf{c}_1$ from its received signal, then decodes m_2 from the residual.

In the general M-receiver broadcast channel, each stronger receivers must cancel all the messages intended for the weaker ones, so that ultimately the Mth receiver must know all M codebooks and decode all M messages. From this perspective, superposition codes, which are based on an "interference cancellation" principle, generally do not lead to scalable broadcast protocols.

However, as described in [4] in the symmetric case, the broadcast channel frontier (1) can also be achieved in a fundamentally different way via information embedding, and we term the result an "embedding code." In the two-receiver case, as before, the encoder uses codeword $\mathbf{c}_1$ to transmit information to the weak receiver at average power αP. However, with embedding coding the encoding of the message for the strong receiver is chosen as a function of that for the weak receiver. In particular, the strong receiver's message is embedded into the weak receiver's message using information embedding, corresponding to an embedding distortion $(1-\alpha)P$. With capacity-achieving information embedding techniques, the host and embedding signals are independent and the composite signal has power P. The weak receiver decodes m_1 just as in the superposition coding case, ignoring the distortion caused by the embedding, achieving rate R_1 in (1). However, the strong receiver extracts m_2 using an information embedding decoder, which requires no knowledge of the (Gaussian) host signal $\mathbf{c}_1$. Since, the strong receiver's distortion-to-noise ratio is $(1-\alpha)P/N_2$, rate R_2 in (1) achievable, consistent with the Costa's result [6].

In the general M-receiver broadcast channel, the receivers are ordered by strength, and successively embedded into the weakest to achieve capacity.[1] Thus, in contrast to superposition coding, there is an ordering of receivers that is important at the encoder. However, the benefit of such embedding codes is a decoupling at the receivers: each receiver need only know its own codebook to extract its message. In particular, it treats the messages for all weaker receivers as its "host" and messages for all stronger receivers as "noise" in its decoding. In this way embedding codes, which are based on an "interference avoidance" principle, lead to an inherently more scalable broadcast protocol.

One other property of embedding codes is that the messages can be embedded in an arbitrary order at the encoder, with the decoders adjusted accordingly, although capacity region frontier is only achievable by the weak-to-strong ordering. However, with superposition codes, the decoders cannot be arbitrarily ordered: only the stronger receivers can strip messages of weaker ones.

However, even beyond scalability issues, there are other reasons why embedding codes can be preferable to superposition codes. Indeed, as we develop next, they yield strictly better performance in some broadcasting contexts with legacy infrastructure.

2.1 Upgrading Legacy Digital Broadcast Infrastructure

Consider a deployed broadcast system whose legacy receivers can reliably decode codebooks of rate

$$R_{\text{legacy}} = \frac{1}{2}\log[1+\text{SNR}_{\text{legacy}}] \quad \text{if SNR} \geq \text{SNR}_{\text{legacy}}, \tag{2}$$

where SNR is the (common) realized channel SNR. Some time after deployment of large numbers of legacy receivers, the system is upgraded. The upgrade results in improved communications channel quality (due to, e.g., higher power transmitters or lower noise amplifiers) so that the channel SNR is now $\text{SNR}_{\text{upgrade}} > \text{SNR}_{\text{legacy}}$. With this upgrade, a new class of receivers is also introduced that can reliably decode codebooks of better rate

$$R_{\text{upgrade}} = \frac{1}{2}\log[1+\text{SNR}_{\text{upgrade}}] \quad \text{if SNR} \geq \text{SNR}_{\text{upgrade}}. \tag{3}$$

As we now develop, if the legacy receivers must be kept in-service, we must constrain the design of the new receivers with this in mind. To simplify the exposition, we restrict our attention to a two-receiver (legacy and new) scenario, to which independent messages are to be sent. The M-receiver case follows analogously.

Referring to Fig. 1, the lower dotted and upper dashed lines delineate the capacity region for the original and upgraded channels, respectively. The legacy receiver is constrained to operate along the solid line connecting (a) to (b).

We consider solutions based on resource (e.g., time) sharing, embedding codes, and superposition codes. With resource sharing some fraction of, e.g., time slots in a TDMA system (or frequency bins in an OFDM or FDMA system) are allocated for transmission

[1]Important generalizations are developed in [3].

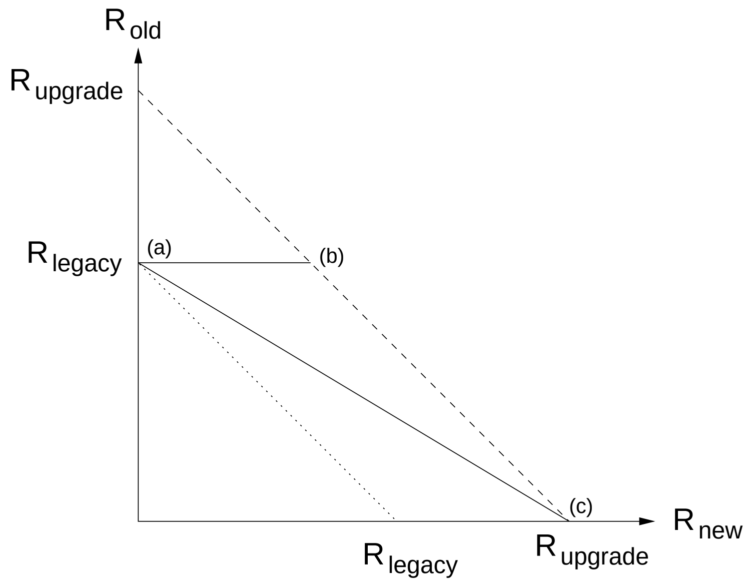

Figure 1: Achievable region for Gaussian broadcast channels illustrating the advantages of an information embedding approach over resource sharing.

at rate R_{legacy} to the legacy receiver and the remaining fraction are allocated for transmission at rate R_{upgrade} to the new receiver. The overall transmission rate therefore lies on the line connecting points (a) and (c) in Fig. 1. By contrast, using either superposition codes or embedding codes we can achieve any rate on the solid line connecting (a) and (b), and in turn by subsequent time-sharing, any rate on the line between (b) and (c).

For the embedding code solution, the legacy codeword serves as the host signal and has average power P. A distortion signal, which is the host-dependent encoding of the embedded information, of power D is added to it giving a composite signal power $P+D$. To keep the transmitted power constant, a gain of $\sqrt{P/(P+D)}$ is applied before transmission; we assume that legacy receivers have automatic gain control mechanisms that will apply an inverse gain of $\sqrt{(P+D)/P}$. If the noise at the receivers has variance N, we can consider an equivalent model without the gains, but with a noise variance of $N(1+D/P)$. Define the signal-to-noise and distortion-to-noise ratios to be $\text{SNR}_{\text{upgrade}} = P/N$ and $\text{DNR} = D/N$, respectively. Then, taking into account the automatic gain control, we define the effective DNR and SNR as follows:

$$\begin{aligned} \text{SNR}_{\text{eff}} &= \frac{P}{N(1+D/P)} = \frac{\text{SNR}_{\text{upgrade}}}{1+D/P}, \\ \text{DNR}_{\text{eff}} &= \frac{D}{N(1+D/P)} = \frac{\text{DNR}}{1+D/P}. \end{aligned}$$

The legacy receiver sees an SNR of

$$\text{SNR}_{\text{old}} = \frac{P}{D+N(1+D/P)} = \frac{\text{SNR}_{\text{upgrade}}/(1+D/P)}{1+\text{DNR}/(1+D/P)} = \frac{\text{SNR}_{\text{eff}}}{1+\text{DNR}_{\text{eff}}}, \tag{4}$$

while the new receiver sees a distortion-noise-ratio of DNR_{eff}. The achievable rates are

$$R_{\text{old}} \leq \frac{1}{2}\log\left[1+\frac{\text{SNR}_{\text{eff}}}{1+\text{DNR}_{\text{eff}}}\right] \quad \text{and} \quad R_{\text{new}} \leq \frac{1}{2}\log[1+\text{DNR}_{\text{eff}}].$$

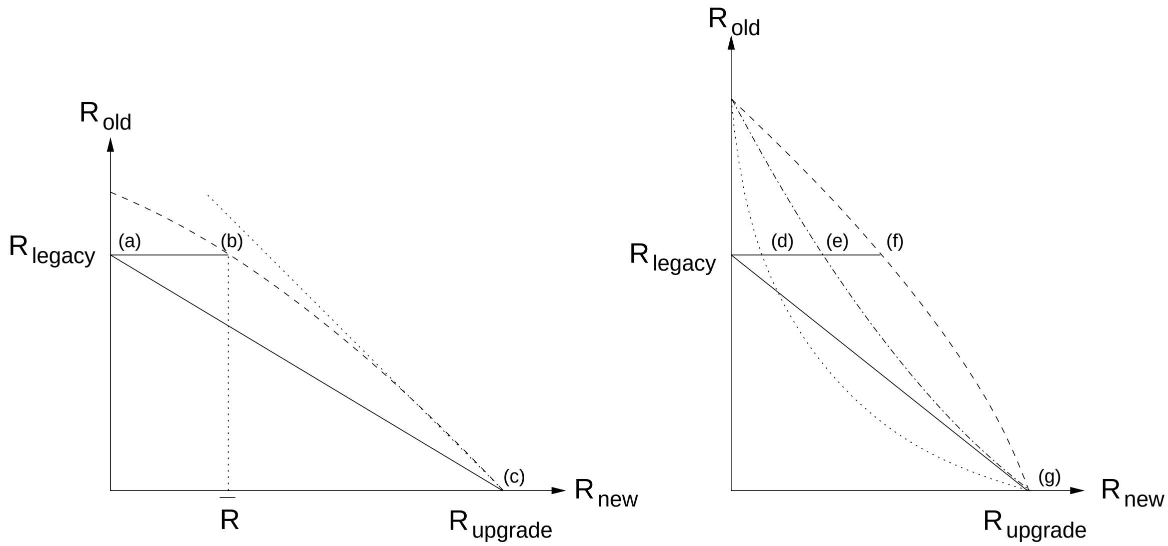

Figure 2: Achievable region for Gaussian broadcast channels with different receiver SNR's. Left ($N_{\text{legacy}} > N_{\text{upgrade}}$): point (b) can be achieved with either embedding or superposition codes. Right ($N_{\text{legacy}} < N_{\text{upgrade}}$): point (e) can be achieved with embedding but not superposition codes, and point (f) cannot be reached by either code.

The sum rate

$$\begin{aligned} R_{\text{sum}} &= R_{\text{old}} + R_{\text{new}} \\ &\le \frac{1}{2}\log\left[1+\frac{\text{SNR}_{\text{eff}}}{1+\text{DNR}_{\text{eff}}}\right] + \frac{1}{2}\log[1+\text{DNR}_{\text{eff}}] = \frac{1}{2}\log\left[1+\text{DNR}_{\text{eff}}+\text{SNR}_{\text{eff}}\right] \\ &= \frac{1}{2}\log\left[1+\frac{\text{DNR}+\text{SNR}_{\text{upgrade}}}{1+D/P}\right] = \frac{1}{2}\log[1+\text{SNR}_{\text{upgrade}}] = R_{\text{upgrade}}. \end{aligned} \quad (5)$$

If $\text{SNR}_{\text{old}} = \text{SNR}_{\text{legacy}}$ then $R_{\text{old}} = R_{\text{legacy}}$, the legacy rate we are committed to supporting. Since $R_{\text{sum}} = R_{\text{upgrade}}$, the maximum possible sum rate, it follows that fulfilling the design requirement of a system upgrade that continues to support the legacy receiver does not result in a sum-rate inefficiency. Furthermore, as Fig. 1 reflects, this solution leads to greater sum rate than does the resource-sharing solution. That is embedding codes enable legacy receivers to operate always at rate R_{legacy} and to push the rate pair out to (b), which can be then used to time-share with (c). Consistent with the discussion in Sec. 2, superposition codes can also be used to achieve point (b), so that both embedding and superposition codes are equally good in this setting, and preferable to resource-sharing.

Interestingly, however, a somewhat different conclusion can emerge when the two receivers experience different noise levels N_{legacy} and N_{upgrade}. In general, their relative strengths can be arbitrary in practice. For example, the case $N_{\text{legacy}} > N_{\text{upgrade}}$ models a situation in which the new receivers also have quieter electronics or better equalization, while the case $N_{\text{legacy}} < N_{\text{upgrade}}$ models a scenario in which the new receivers are further away from the transmitter. The left-hand and right-hand plots in Fig. 2 model these two scenarios, respectively. In this figure, R_{legacy} is the legacy rate that we must continue to support, and the dashed curve in both plots is the frontier of achievable rates at the new channel SNRs.

First consider the left-hand plot of Fig. 2, for which $N_{\text{legacy}} > N_{\text{upgrade}}$. As before, time-sharing between R_{legacy} and R_{upgrade} achieves rates along the solid line connecting points (a) and (c). Embedding and superposition codes get us to point (b), and we can time-share between this point and (c). At (b) the achievable rate pair is $(R_{\text{legacy}}, \bar{R})$. However, now $R_{\text{legacy}} + \bar{R} < R_{\text{upgrade}}$. That is, the sum rate is strictly less than if we stopped transmitting to legacy receivers and transmitted exclusively to new receivers. This is because the now-concave frontier of the achievable region lies below the (dotted) constant sum-rate line emanating from $(R_{\text{upgrade}}, 0)$. Thus, in this case although we attain the frontier of the capacity region, we pay a price in overall throughput (sum rate) by having to accommodate legacy receivers.

This behavior contrasts with that exhibited in the the right-hand plot of Fig. 2, for which $N_{\text{legacy}} < N_{\text{upgrade}}$. Because the new receiver has the lower SNR, it cannot decode the legacy receiver's message m_1 and strip its codeword. However, the legacy receiver does not have successive cancellation decoding as part of its functionality, so it cannot decode the new receiver's message m_2 and strip its codeword. Therefore, both receivers suffer interference when superposition coding is used. With embedding codes, however, we can partially circumvent the interference. As discussed in section 2, either codeword can serve as the host signal, and rates (1) remain achievable. But unless the embedding is done in the codeword destined for the lower-SNR receiver, these rates no longer lie on the achievable rate frontier. The rates pairs achievable using superposition coding (sc) and embedding coding (ec) are, respectively,

$$R_{\text{old,sc}} = \frac{1}{2}\log\left[1 + \frac{\alpha P}{(1-\alpha)P + N_{\text{legacy}}}\right], \qquad R_{\text{new,sc}} = \frac{1}{2}\log\left[1 + \frac{(1-\alpha)P}{\alpha P + N_{\text{upgrade}}}\right], \quad (6)$$

$$R_{\text{old,ec}} = \frac{1}{2}\log\left[1 + \frac{\alpha P}{(1-\alpha)P + N_{\text{legacy}}}\right], \qquad R_{\text{new,ec}} = \frac{1}{2}\log\left[1 + \frac{(1-\alpha)P}{N_{\text{upgrade}}}\right], \quad (7)$$

where α may be arbitrarily chosen $(0 \leq \alpha \leq 1)$. The rate pair of (6) gives us the dotted curve in the right-hand plot. Time-sharing between (d) and (g) gives us the set of achievable rates. The rate pair of (7) gives us the dash-dot curve in the right-hand plot. Time-sharing between (e) and (g) yields the set of achievable rates. Hence, in this case embedding coding is not able to avoid all interference (which would get us to (f)), but does strictly better than superposition coding, which cannot suppress any interference.

3 Noisy Coding with Side Information

Generalizations of the basic information embedding problem also have a potentially interesting role to play in, e.g., complementary applications to the multiple-access channel. Likewise, the corresponding generalizations of its dual, Wyner-Ziv coding, also have interesting applications to sequential estimation.

In this section we begin by characterizing the performance limits of the noisy information embedding and noisy source coding with side information problems. In particular, we develop single-letter capacity expressions for both scenarios in the finite-alphabet case, and develop the associated capacities for the Gaussian-quadratic case [9].

3.1 Writing on Dirty Paper Wearing Bad Glasses

Fig. 3 depicts the generalized information embedding scenario of interest. The n-dimensional signal $\mathbf{x}$ is the host signal, and is assumed to be a vector of independent identically-

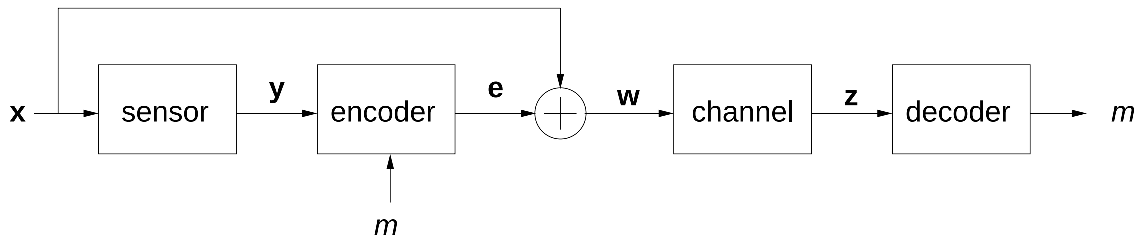

Figure 3: Information embedding with noisy host information. The signals $\mathbf{x}$, $\mathbf{y}$, m, $\mathbf{e}$, $\mathbf{w}$ and $\mathbf{z}$ are, respectively, the host, host observation, message, embedding signal, composite signal, and channel output.

distributed (i.i.d.) random variables, $x_i \sim p(x)$. The encoder observes $\mathbf{y}$, which is related to the host via the memoryless transition density $p(y|x)$. The message to be embedded is m, and the output of the encoder is the embedding signal $\mathbf{e}$. The embedding signal $\mathbf{e}$ is added to the host $\mathbf{x}$, producing the composite signal $\mathbf{w}$. A distortion constraint is placed between the composite signal $\mathbf{w}$ and the host signal $\mathbf{x}$. Finally, the decoder measures $\mathbf{z}$, which is an observation of $\mathbf{w}$ via the memoryless channel with transition density $p(z|w)$.

The information embedding capacity with "imperfect" host information is denoted $C_{\mathrm{I}}^{\mathrm{IE}}(d)$.. The derivation [9] can be viewed as generalizing results in [5] to accommodate a distortion constraint, just as [1] generalizes [10] to accommodate such distortion constraints.

Given noisy host observations, $C_{\mathrm{I}}^{\mathrm{IE}}(d)$ is the maximum achievable rate for communicating a message m such that $\Pr(\hat{m} \neq m)$ is arbitrarily small and $E\left[\frac{1}{n}\sum_{k=1}^{n} D(x_k, w_k)\right]$ is arbitrarily close to d for sufficiently large n.

Theorem 1 *For general distortion measures $D(\cdot,\cdot)$, the capacity $C_{\mathrm{I}}^{\mathrm{IE}}(d)$ can be expressed in the form $C_{\mathrm{I}}^{\mathrm{IE}}(d) = \max[I(z;u) - I(u;y)]$, where the maximum is taken over all distributions $p_{u|y}(u|y)$ and functions $f : \mathcal{U} \times \mathcal{Y} \to \mathcal{E}$ satisfying $E[D(x;w)] \leq d$ where $e = f(u,y)$ and u is an auxiliary random variable.*

3.1.1 Gaussian-Quadratic Case

In this case, $\mathbf{x}$ is an n-length sequence of i.i.d. zero-mean, variance σ_x^2 Gaussian random variables, i.e., $x_i \sim \mathcal{N}(0, \sigma_x^2)$. The encoder observes $\mathbf{y}$, a noise-corrupted version of $\mathbf{x}$ according to $y_i = x_i + v_{1,i}$ where $v_{1,i} \sim \mathcal{N}(0, N_1)$. As a function of $\mathbf{y}$ and m, the encoder produces the embedding signal $\mathbf{e}$, and the channel input is $\mathbf{w} = \mathbf{x} + \mathbf{e}$ where $E[(x-w)^2] \leq d$. The communication channel is an additive white Gaussian noise channel, $z_i = w_i + v_{2,i}$ where $v_{2,i} \sim \mathcal{N}(0, N_2)$. For this channel, the capacity is

$$C_{\mathrm{I}}^{\mathrm{IE}}(d) = \frac{1}{2}\log\left[1 + \frac{d}{\sigma_{x|y}^2 + N_2}\right]. \tag{8}$$

The two terms in the denominator of (8) correspond to the two sources of uncertainty. The first term is the minimum mean-squared error (MMSE) estimation error in the host estimate $\hat{\mathbf{x}} = E[\mathbf{x}|\mathbf{y}]$. The second term is the uncertainty caused by the channel noise. In the limit of small estimation error ($N_1 \to 0$), we get Costa's familiar capacity [6].

Eq. (8) also tells us that, in the case of Gaussian measurements, a separation theorem applies to noisy information embedding: without loss in performance the encoder can be realized as the cascade of MMSE estimation (of $\mathbf{x}$ from $\mathbf{y}$) followed by information

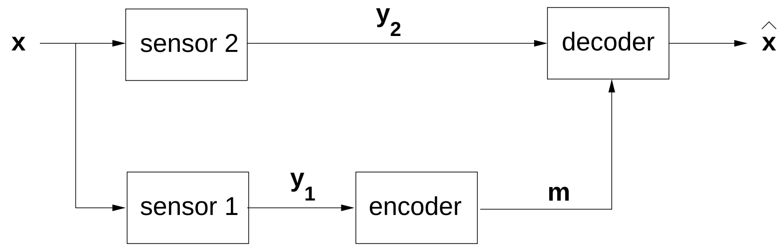

Figure 4: Noisy source coding with side information. The signals $\mathbf{x}$, $\mathbf{y}_1$, $\mathbf{y}_2$ and m are, respectively, the source, remote source observation, side information, and message

embedding (as if the estimate were the true state). Since the MMSE estimation error acts as extra channel noise, this must be taken into account by the distortion compensation in the (QIM) information embedder [4]. A geometric view is developed in [9].

The test channel used to determine capacity is similar to the channel used in [6]. The auxiliary random variable $u = \alpha y + e$, where and e is the encoder output, a zero-mean Gaussian random variable of variance d that is independent of y. The input to the channel is $w = x + e$. Optimization over α yields the capacity given in (8).[2]

3.2 Noisy Source Coding with Side Information

Fig. 4 depicts the source coding with side information scenario of interest. The n-dimensional i.i.d. source vector $\mathbf{x}$ is observed via two memoryless transition probabilities $p(y_1|x)$ and $p(y_2|x)$ at the encoder and decoder, respectively. The encoder converts its observation of $\mathbf{y}_1$ into a message m that is communicated over a rate-constrained channel to the decoder. The decoder produces $\hat{\mathbf{x}}$, an estimate of the source $\mathbf{x}$, as a function of $\mathbf{y}_2$ and m. This scenario differs from Wyner-Ziv source coding because $\mathbf{x}$ is not uniquely determinable from $\mathbf{y}_1$. This is a useful model for a variety of practical scenarios, such as multiple-microphone problems in acoustic applications and other constrained relaying problems.[3]

The rate-distortion function for source-coding with side-information and 'imperfect' encoder observations is denoted $R_{\mathrm{I}}^{\mathrm{WZ}}(d)$. This is the minimum achievable rate so that $E\left[D(x^n, \hat{x}^n)\right]$ is arbitrarily close to d for sufficiently large n. In [9] we show the following theorem:

Theorem 2 *Let (x, y_1, y_2) be drawn i.i.d. $\sim$ $p(y_1|x)p(y_2|x)p(x)$ and let $D(x^n, \hat{x}^n) = \frac{1}{n}\sum_{i=1}^{n} D(x_i, \hat{x}_i)$ be given. The rate distortion function for noisy source coding with side information is $R_{\mathrm{I}}^{\mathrm{WZ}}(d) = \min[I(u; y_1) - I(u; y_2)]$, where the minimization is over all functions $f : \mathcal{Y}_2 \times \mathcal{U} \to \hat{\mathcal{X}}$ and conditional distributions $p(u|y_1)$, such that*

[2]When the encoder observes a noisy version of the host, but must also embed in some function of that noisy observation, the capacity changes to $C(d) = \frac{1}{2}\log[1 + (d - \sigma^2_{x|y})/N_2]$, where $d - \sigma^2_{x|y} \geq 0$. Capacity is achieved by embedding in the MMSE estimate.

[3]For the general Gaussian relay channel [7], our results suggest that the relay sends the message that would be most useful for the decoder to estimate the codeword. A combination of stripping and noisy source coding with side information allows the following rate to be achieved,

$$R_{\mathrm{relay}} = \frac{1}{2}\log\left[(P_1 + P_2 + N_2)/\left[N_2\left(1 + \frac{P_2 N_1}{P_1(N_1 + N_2) + N_1 N_2}\right)\right]\right].$$

As the relay power grows ($P_2 \to \infty$) this strategy achieves capacity. But as the channel to the relay becomes perfect ($N_1 \to 0$) this approach suffers because cooperation is not exploited [9].

$\sum_x \sum_{y_1} \sum_{y_2} \sum_u p(x, y_1, y_2) p(u|y_1) D(x, f(y_2, u)) \leq d$.

3.2.1 Gaussian-Quadratic Case

In this case, $\mathbf{y}_1$ and $\mathbf{y}_2$ are observations of $\mathbf{x}$ through Gaussian channels: $y_{1,i} = x_i + v_{1,i}$, $y_{2,i} = x_i + v_{2,i}$, where $v_{j,i} \sim \mathcal{N}(0, N_j)$ and the two noise sources are independent of each other and of the source.

$$R_{\mathrm{I}}^{\mathrm{WZ}}(d) = \frac{1}{2} \log \left[\frac{\sigma^2_{x|y_2} - \sigma^2_{x|y_1,y_2}}{d - \sigma^2_{x|y_1,y_2}} \right], \tag{9}$$

where $\sigma^2_{x|y_1,y_2} \leq d \leq \sigma^2_{x|y_2}$. Note that as $N_1 \to 0$, the encoder measures the source perfectly, and (9) is the Wyner-Ziv rate developed in [12]. Likewise, as $N_2 \to \infty$ there is no side information at the decoder. This is the situation of noisy quantization, and the rate converges to that developed in [11].

Note again that (9) implies that a separation theorem applies in the Gaussian case: without loss of performance the optimal encoder can be factored into the cascade of MMSE estimation followed by Wyner-Ziv coding, with the decoder modified accordingly.

The test channel used to determine capacity is developed in[9]. In brief, the auxiliary random variable $u = \alpha y_1 + e$ where $e \sim \mathcal{N}(0, \alpha d^*)$ is independent of y_1 and y_2. Optimal choices of α and d^* are $\alpha = \sigma^2_{x|y_2} - d/(\sigma^2_{x|y_2} + N_1)$ and $d^* = d - \sigma^2_{x|y_1,y_2}$. In addition, the decoder (reconstruction) function f is

$$f(y_2, u) = \hat{x} = \frac{d}{N_2} y_2 + \left[1 + \frac{N_1}{\sigma^2_{x|y_2}} \right] u. \tag{10}$$

3.3 Sequential Estimation and Sensor Pipelines

In this section we apply noisy source coding with side information to a problem of sequential estimation that arises in, e.g., distributed sensor networks. Consider the scenario depicted in Fig. 5. A number of sensors nodes are linked together serially via rate-constrained channels of rates $R_2, R_3, \ldots$. Node n observes $\mathbf{y}_n = \mathbf{x} + \mathbf{v}_n$ where $\mathbf{x} \sim \mathcal{N}(0, \sigma_x^2 \mathbf{I})$, $\mathbf{v}_n \sim \mathcal{N}(0, N\mathbf{I})$, and receives a data stream from node $(n-1)$ at rate R_n. Each node acts in sequence, sending on the information that will most help the next node in its estimation of $\mathbf{x}$. For example, node n fuses together its observation with whatever data it received from the nodes earlier in the chain (nodes 1, 2, …, $n-1$) to produce its estimate $\hat{\mathbf{x}}_n$, with a mean-squared distortion error d_n. Knowing that node n makes an observation of the source, node $(n-1)$ sends on the most useful bits to assist in making that estimate. The rate-distortion region for this problem can be derived using the results on noisy source coding with side information from Section 3.

In order to illustrate the results we restrict $R_2 = R_3 = \ldots = R$, and $\mathbf{v}_1 \sim \mathbf{v}_2 \sim \mathbf{v}_3 \sim \ldots \sim \mathcal{N}(0, N\mathbf{I})$. Under these conditions, the distortion at the nth sensor is

$$d_n = \frac{d_{n-1} N}{d_{n-1} + N} + \sigma^2_{x|y} \frac{\left(1 - \frac{d_{n-1}}{\sigma_x^2}\right)}{\left(1 + \frac{d_{n-1}}{N}\right)} 2^{-2R}, \tag{11}$$

where $\sigma^2_{x|y}$ is the MMSE in the estimate of x given any single observation $\mathbf{y}_i = \mathbf{x} + \mathbf{v}_i$. The first term in (11) dictates how the MMSE $\sigma^2_{x|y_1,\ldots,y_n}$ decreases as a function of n. As $R \to \infty$ the method presented achieves this lower bound.

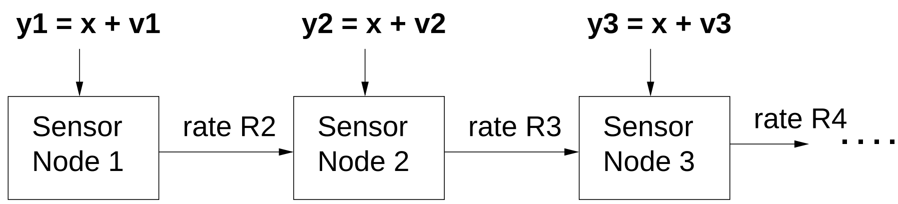

Figure 5: Sequential estimation in a serial sensor network with rate-constrained connections. Noisy source coding with side information results give the solution to this problem.

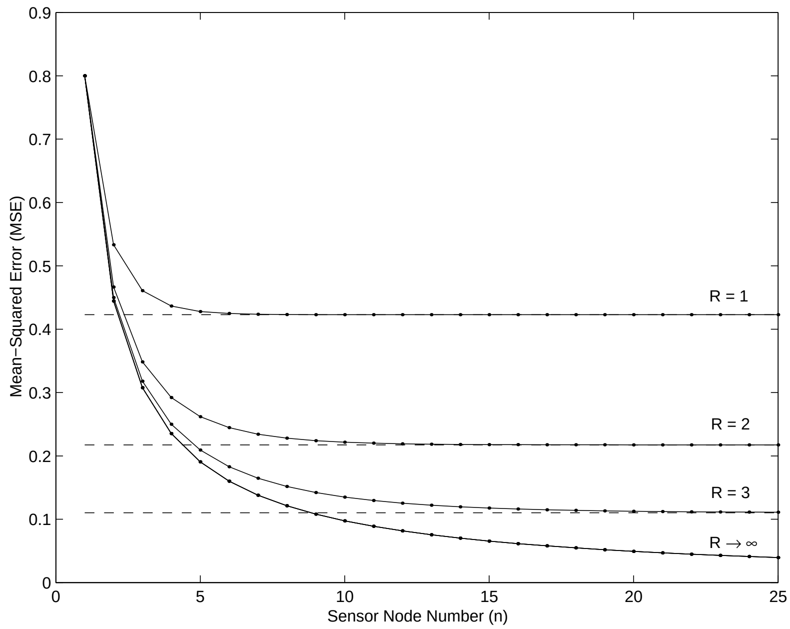

Figure 6: Successive estimation performance declines with node number, but is lower-bounded for finite rate ($R < \infty$). Here $\sigma_x^2 = 4$, $N = 1$, and $R = 1, 2, 3, \infty$. Solid lines plot the evolution of the estimation error d_n for the different rates. Dashed lines plot the lower-bound on performance for each finite rate.

The distortion in (11) decreases monotonically with n. In the limit as $n \to \infty$ $d_n \to d$ which can be derived by setting $d_n = d_{n-1} = d$ in (11), yielding $d \sim \Theta(2^{-R})$; specifically,

$$d = \frac{\sigma_{x|y}^2}{2\sigma_x^2}\left[\sqrt{N^2 2^{-2R} + 4\sigma_x^2(\sigma_x^2 + N)} - N2^{-R}\right] 2^{-R}. \qquad (12)$$

In Fig. (6) we plot the decrease in mean-squared estimation error versus node number for $\sigma_x^2 = 4$, $N = 1$. Node 1 has only its own observation, so its error is $\sigma_{x|y}^2$. Node 2 has its own observation plus whatever information it gets at rate R from node 1, etc. We plot the results for $R = 1, 2, 3, \infty$. The data points that correspond to each node's estimation error are connected by solid lines for convenience. The limit given by (12) is plotted as a dashed line for each case.

Eq. (12) provides a means for determining the bit pipe size required to achieve a target steady state distortion. However, using this bit pipe size between all pairs of nodes is inefficient: since earlier nodes do not have reliable estimates to convey, their bit pipes can be smaller. We complete our analysis of the sequential analysis problem by examining the asymptotically optimal rate of growth of bit pipes that should be used.

In particular, we find that if we allow the the transmission rate to increase at each stage (i.e. $R_2 < R_3 < \cdots$), the distortion can be made to stay within a finite multiple of $\sigma^2_{x|y_1,\ldots,y_n}$, i.e. $d_n \leq \kappa\sigma^2_{x|y_1,\ldots,y_n}$, $\kappa \geq 1$. As n gets large, to meet the constraint this constraint R_n must grow logarithmically as

$$R_n \geq \log\left[\frac{\mathrm{SNR}}{1+\mathrm{SNR}}\right] + \log\left[\frac{n}{\kappa(\kappa-1)}\right]$$

where $\mathrm{SNR} = \sigma_x^2/N$ is an individual sensor's SNR.

These sequential estimation results can be compared with the corresponding results for the closely related non-sequential CEO problem [2].

References

[1] R. J. Barron, B. Chen, and G. W. Wornell. The duality between information embedding and source coding with side information and its implications and applications. *Submitted Jan. 2000 to IEEE Trans. Info. Theory; also in ISIT-2001.*

[2] T. Berger, Z. Zhang, and H. Viswanathan. The CEO problem. *IEEE Trans. Info. Theory*, 42:887–902, May 1996.

[3] G. Caire and S. Shamai. On achievable rates in a Gaussian broacast channel with multiple transmit antennas. In *Proc. Int. Symp. Inform. Theory*, Washington, DC, June 2001.

[4] B. Chen and G. W. Wornell. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. *IEEE Trans. Info. Theory*, 47:1423–1443, May 2001.

[5] M. Chiang and T. M. Cover. Unified duality of channel capacity and rate distortion with state information. In *Proc. Int. Symp. Inform. Theory*, Washington, DC, June 2001.

[6] M. H. Costa. Writing on dirty paper. *IEEE Trans. Info. Theory*, 29:439–441, May 1983.

[7] T. M. Cover and A. A. El Gamal. Capacity theorems for the relay channel. *IEEE Trans. Info. Theory*, 25:572–584, September 1979.

[8] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley and Sons, 1991.

[9] S. C. Draper and G. W. Wornell. Writing on dirty paper with bad glasses and related problems, 2001. in preparation.

[10] S. I. Gel'fand and M. S. Pinsker. Coding for channel with random parameters. *Problems of Control and Information Theory*, 9(1):19–31, 1980.

[11] J. K. Wolf and J. Ziv. Transmission of noisy information to a noisy receiver with minimum distortion. *IEEE Trans. Info. Theory*, 16:406–411, July 1970.

[12] A. D. Wyner. The rate-distortion function for source coding with side information at the decoder–II: General sources. *Information and Control*, 38:60–80, 1978.

[13] A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. Info. Theory*, 22:1–10, January 1976.