

Dirty Paper Writing and Watermarking Applications

G.RaviKiran

February 10, 2003

1 Introduction

Following an underlying theme in Communication is the duo of Dirty Paper Writing and Watermarking. In 1983 utilising some previous results by Gelfand, Pinsker, El Gamal and Heegard, Costa gave a method for achieving capacity in the presence of a known interference. Costa, in his paper, doesn't mention about *any* applications of his result, let alone Watermarking. In 1999 Cox *et al* recognised the similarity between the two.

The summary will be arranged in the following order. Dirty Paper writing will be dealt with in Section 2, an introduction to watermarking and some of the issues related to it in Section 3. The relation between Watermarking and Dirty Paper Coding, and the concept of Quantisation Index Modulation will be discussed in Section 4.

2 Writing on Dirty Paper

2.1 Problem:

To send a message $M = m; m \in \{1, \dots, |M|\}$ to the receiver in n uses of the channel. Where $M = \lceil e^{nR} \rceil$ where R is the rate in nats per channel use. There is an interference source \mathbf{S} which is assumed to be corrupting the transmission. The encoder(transmitter) has full knowledge of the value of the source at any instant of time. Further, the source is assumed to be a sequence of independent identically distributed (i.i.d) $N(0, Q)$ random variables. The value of \mathbf{S} is known to the transmitter but not the receiver. Based on m and \mathbf{S} , the encoder sends a codeword \mathbf{X} which must satisfy the power constraint $(1/n) \sum_{i=1}^n X_i^2 \leq P$.

In addition to the interference \mathbf{S} there is the usual channel noise \mathbf{Z} , So the channel output is given by $\mathbf{Y} = \mathbf{X} + \mathbf{S} + \mathbf{Z}$, where the channel noise \mathbf{Z} is distributed according to $N(0, N)$. Upon receipt of \mathbf{Y} the decoder creates an estimate \hat{m} of the index m . Probability of error in making the estimate is given

by

$$P_e = \frac{1}{|M|} \sum_{k=1}^{|M|} \text{Prob}\{\hat{m} \neq k | m = k\} \quad (1)$$

Index M is assumed to be uniform over the interval $\{1, \dots, |M|\}$. The decoder, however, will have to base his estimate solely on the channel output \mathbf{Y} . The problem is surmised in Figure 1.

2.2 The naive solution

The naive encoding scheme to communicate over the channel of figure 1 is to try and cancel out \mathbf{S} using some of the power available for transmission. i.e if $P > Q$, Q being the variance (power) of the interfering signal, the encoder could use $P-Q$ for transmitting \mathbf{M} and use Q for transmitting $-\mathbf{S}$. The maximum rate of information transmission under this scheme would be $C((P-Q)/N)$ where N is the variance of \mathbf{Z} , and $C(x) = \frac{1}{2} \log(1+x)$, the capacity, in nats per channel use, of a AWGN channel with SNR of x . More generally if the encoder uses a fraction α of the available power for cancelling \mathbf{S} , the maximum achievable rate would be $C(\frac{(1-\alpha)P}{N+(\sqrt{Q}-\sqrt{\alpha P})^2})$.

This scheme, definitely, cannot achieve the rate which would have been possible in the absence of \mathbf{S} . The optimal encoding scheme would be one which uses codewords in the direction of \mathbf{S} . It looks at the space surrounding \mathbf{S} and chooses codewords that are distinguishable when viewed from the channel output. It adapts its signal to \mathbf{S} instead of trying to erase it.

A sheet of paper covered with independent dirt spots of normally distributed intensity. Problem is to write a message on this sheet of paper so that the reader can read it. The writer knows the original location of the dirt marks, but the reader can see only the sum of the message and the dirt. The optimum writer would somehow use the dirt spots, link them in ways, to convey maximum information. Hence the name Dirty Paper Writing.

2.3 The Optimum Solution

(I) The capacity of a DMC with \mathbf{S} known to encoder is

$$C = \max_{p(\mathbf{u}, \mathbf{x} | \mathbf{s})} \{I(U; Y) - I(U; S)\} \quad (2)$$

Argument: We need to transmit m , which is an index. First we will assign a codeword \mathbf{u} to it. Then depending on \mathbf{u} and \mathbf{s} we will send an appropriate \mathbf{x} . The first assignment can be done easily. Generate $e^{n(I(U; Y) - \epsilon)}$ i.i.d sequences \mathbf{u} , selecting uniformly from typical \mathbf{u} . Distribute these sequences over e^{nR} bins. Let $i(\mathbf{u})$ be the index of the bin which contains \mathbf{u} . Given \mathbf{s} , look in bin m for a \mathbf{u} such that (\mathbf{u}, \mathbf{s}) are jointly typical. If the number of sequences in the bin m is greater than $e^{n(I(U; S) - \delta)}$, its highly probable that we will find such a \mathbf{u} . The codeword \mathbf{u} has now been assigned, we have to find an appropriate \mathbf{x} . Choose

\mathbf{x} such that $(\mathbf{x}, \mathbf{u}, \mathbf{s})$ are jointly typical. This \mathbf{x} is now sent through the channel. At the output the decoder finds a \mathbf{u} such that (\mathbf{u}, \mathbf{y}) is jointly typical. Hence we know the assigned codeword, the decoded message index $\hat{m} = i(\mathbf{u})$. There were two constraints we met in the process. There have to be $e^{n(I(U;S)-\delta)}$ sequences in each bin so $|M|.e^{n(I(U;S)-\delta)}$ sequences in total. Where $nR = \ln(|M|)$, but we had generated only $e^{n(I(U;Y)-\epsilon)}$. So $|M|.e^{n(I(U;S)-\delta)} < e^{n(I(U;Y)-\epsilon)}$ or

$$R < I(U;Y) - I(U;S) - \epsilon - \delta \quad (3)$$

the result follows. The problem now reduces to finding such an \mathbf{X} .

(II) \mathbf{X} of the form $X = U - \alpha S$ achieves capacity.

Argument:

As $Y = X + S + Z$

$$\begin{aligned} I(U;Y) &= H(X + S + Z) - H(X + S + Z|U) \\ &= H(X + S + Z) - H(X + S + Z|(X + \alpha S)) \\ &= H(X + S + Z) + H(X + \alpha S) - H(X + S + Z; X + \alpha S) \\ &= \frac{1}{2} \ln((2\pi e)^2 (P + Q + N)(P + \alpha^2 Q)) \\ &\quad - \frac{1}{2} \ln((2\pi e)^2 ((P + Q + N)(P + \alpha^2 Q) - (P + \alpha Q)^2)) \end{aligned}$$

$$I(U;Y) = \frac{1}{2} \ln \left(\frac{(P + Q + N)(P + \alpha^2 Q)}{PQ(1 - \alpha)^2 + N(P + \alpha^2 Q)} \right) \quad (4)$$

and

$$I(U;S) = \frac{1}{2} \ln \left(\frac{P + \alpha^2 Q}{P} \right) \quad (5)$$

From (3)

$$R < I(U;Y) - I(U;S) \quad (6)$$

$$R < \frac{1}{2} \ln \left(\frac{(P + Q + N)P}{PQ(1 - \alpha)^2 + N(P + \alpha^2 Q)} \right) \quad (7)$$

Maximising with respect to α we get

$$\max(R) = \frac{1}{2} \ln \left(1 + \frac{P}{N} \right) \quad (8)$$

which is achieved for $\alpha = \frac{P}{P+N}$.

This is the maximum rate for the optimum encoding scheme. Why is it optimum? Because its achieving the capacity of a channel with a transmit power of P and noise power of N . Its as if the interference doesn't exist, or as if the interference signal \mathbf{S} is known both to the encoder and the decoder.

3 Watermarking

Watermarking is hiding a message signal or "watermark", within another signal, called a "host signal". The embedding must be done such that the embedded signal causes no serious distortion to its host. At the same time, the embedding

must be robust to common degradations of the watermarked signal, the watermark must survive whenever the host signal does. The most direct model for watermarking is given in figure 2.

3.1 Host Interference

There are two classes of watermarking (1) Host interference non-rejecting methods (2) Host interference rejecting methods.

(1) Host interference nonrejecting methods have the general property that the host signal is a source of interference in the system, and generally result from system designs that do not allow the encoder in figure 2 to sufficiently exploit knowledge of the host signal.

Such methods have embedding functions of the form

$$\mathbf{x}(\mathbf{s}, m) = \mathbf{s} + \mathbf{w}(m) \tag{9}$$

where $\mathbf{w}(m)$ is a noise sequence which is amplitude modulated by the message index m or a function of m , $a(m)$.

For this class of embedding methods, the host signal acts as interference that inhibits the decoders ability to estimate m . Consequently, even in the absence of any channel perturbations, one can usually embed only a small amount of information. Thus, these methods are useful only when the host signal is available at the decoder.

(2) Information embedding systems can achieve host-interference rejection when knowledge of the host signal at the encoder is adequately exploited in system design. For examples Least Bit Modulation (LBM). In general the embedding function of such methods are of the form

$$\mathbf{x} = \mathbf{q}(\mathbf{s}) + \mathbf{d}(m) \tag{10}$$

where $\mathbf{q}(\cdot)$ represents the coarse quantiser that determines the most significant bits and depends only on the host, $\mathbf{d}(\cdot)$ determines the least significant bits and depends only on the message index m . The QIM method proposed by Wornell *et al.* uses the same idea in a more general setting.

4 The Dirty Connection

A minor rearrangement of figure 2, shows how to view the watermarking problem as a special case of Dirty Paper Writing. The rearranged model of watermarking is shown in figure 3.

The similarities are striking and are not just skin deep. The encoding can be written as follows

$$\mathbf{x}(\mathbf{s}, m) = \mathbf{s} + \mathbf{e}(\mathbf{s}, m) \tag{11}$$

The distortion between the original host signal and the watermarked host signal is, therefore

$$\mathbf{e}(\mathbf{s}, m) = \mathbf{x}(\mathbf{s}, m) - \mathbf{s} \tag{12}$$

The condition that the host signal is not overly perturbed by the watermarking thus translates to an equivalent power constraint on the distortion, say $var(e) < P$. The connection is complete, \mathbf{e} in watermarking is equivalent to \mathbf{x} of dirty paper writing. The results will, intuitively, just carry over. The optimum scheme would be when

$$\mathbf{x} = \mathbf{u}(m) - \alpha \mathbf{s} \quad (13)$$

and capacity will be achieved when

$$\alpha = \frac{P}{P + N} \quad (14)$$

or when written differently

$$\alpha = \frac{DNR}{1 + DNR} \quad (15)$$

where $P/N =$ Distortion to Noise ratio(DNR)

Its easy to get lost in idealities, capacity is scheived when \mathbf{S} is randon with normal distribution. Which it definitely is not in the case of a host signal. The “optimum scheme” was optimum as it acheived capacity, when \mathbf{S} is not normal, it won’t acheive capacity and it may not even be optimum. A practical implementation of the “optimum” encoding scheme is QIM where \mathbf{u} is just a quantiser.

4.1 QIM

View the embedding function $\mathbf{X}(\mathbf{S}, m)$ as an ensemble of functions of \mathbf{S} , indexed by m . Figure 4 illustrates a possible QIM.

The minimum distance d_{min} measures the robustness to perturbations, and the sizes of the quantization cells, one of which is shown in the figure, determine the distortion. In this example, one bit is to be embedded so that $m \in \{1, 2\}$. Thus, we require two quantizers, and their corresponding sets of reconstruction points in R^N are represented in figure 4 with \times s and \circ s. If $m=1$, the host signal is quantized with the \times -quantizer, i.e., is chosen to be the closest to \times . If $m=2$, \mathbf{s} is quantized with the \circ -quantizer. As \mathbf{s} varies, the composite signal value varies from one \times point ($m=1$) to another or from one \circ point ($m=2$) to another, but it never varies between a \times point and a \circ point. Thus, even with an infinite energy host signal, one can determine if channel perturbations are not too severe. This scheme alone is akin to LBM and is not going to do any better in a significant manner. So Wornell talks about Distortion-Compensated QIM (DCQIM) where the Dirty Paper model of (13) is arrived at in a different and round-about manner.

4.2 DCQIM

For a fixed rate and a given quantizer ensemble, scaling all quantizers by α increases d_{min} by a factor of $1/\alpha^2$, thereby increasing the robustness of the embedding. However, the embedding- induced distortion also increases by a

factor of $1/\alpha^2$. Adding back a fraction $1 - \alpha$ of the quantization error to the quantization value removes, or compensates for, this additional distortion. The resulting embedding function is

$$\mathbf{x}(\mathbf{s}, m) = \mathbf{q}(\mathbf{s}, m, \Delta/\alpha) + (1 - \alpha)[\mathbf{s} - \mathbf{q}(\mathbf{s}, m, \Delta/\alpha)] \quad (16)$$

Where $\mathbf{q}(\mathbf{s}, m, \Delta/\alpha)$ is the m th quantiser and α is viewed as a factor scaling the original reconstruction points. The optimum α is found, not surprisingly, to be $\alpha = \frac{DNR}{1+DNR}$

4.3 Optimality

So is DCQIM optimum? Its not very clear that it is. Its following the same path as Dirty Paper Writing and so the same conditions apply. Its definitely optimum if the interference is gaussian in distribution, but when the host is not gaussian it may or it may not.

5 References

- [1] Max.H.Costa, “Writing on Dirty Paper”, *IEEE Transactions on Information Theory.*, vol IT-29, pp 439-441, May 1983.
- [2] Brian Chen, Gregory.W.Wornell, “Quantisation Index modulation:A class of provably good methods for Digital Watermarking and Information Embedding”, *IEEE Transactions on Information Theory.*, vol 47, pp 1423-1443, May 2001.
- [3] Richard .J.Barron, Brian Chen, Gregory.W.Wornell, “ The Duality between Information Embedding and Source coding with side information and Some applications”, *Submitted to IEEE Transactions on Information Theory.*

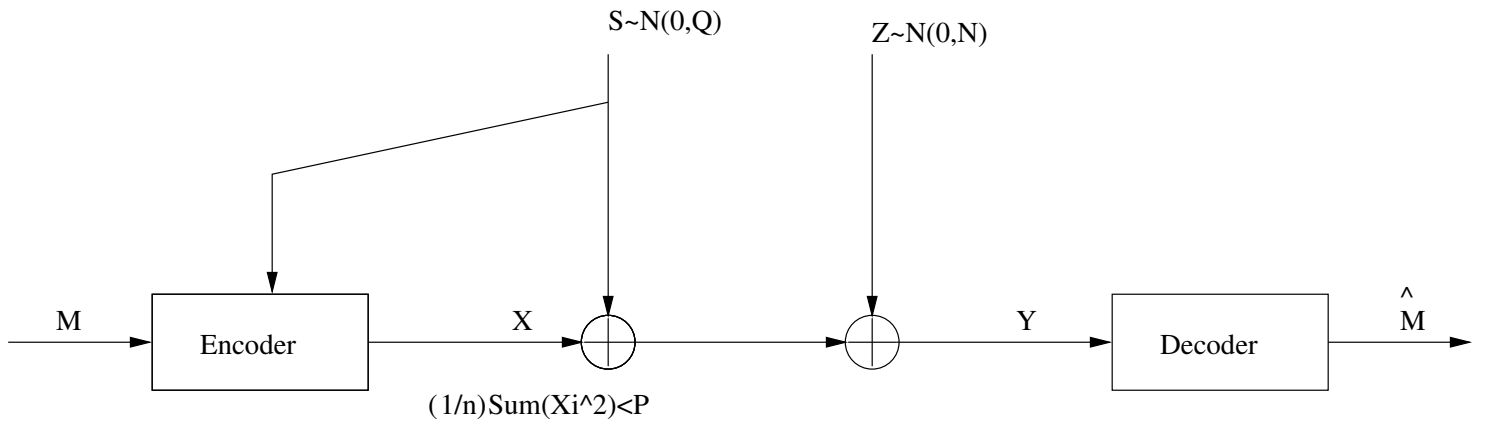


Figure 1: Model for Dirty Paper Writing

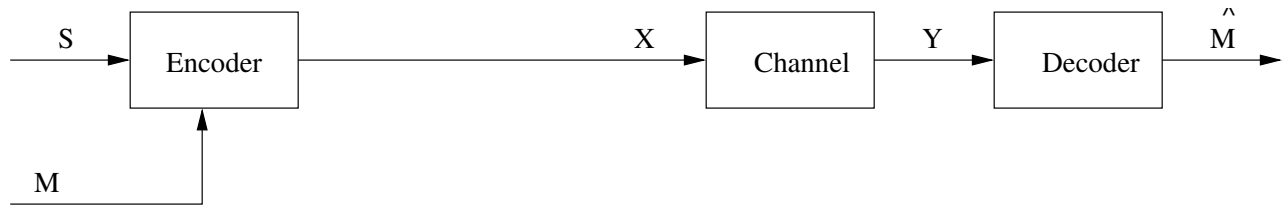


Figure 2: Model for Watermarking

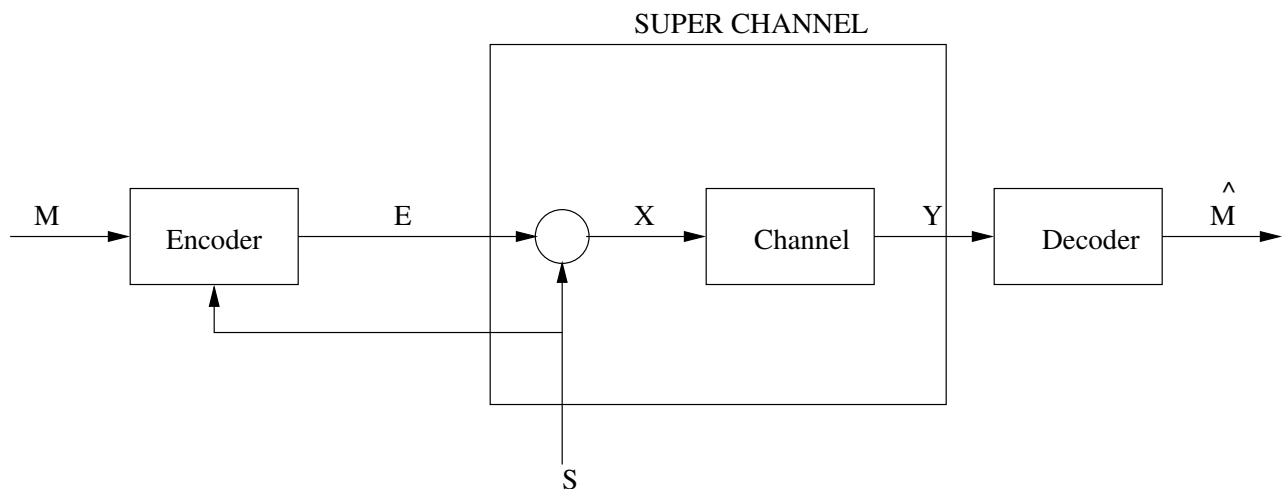


Figure 3: Rearranged Model for Watermarking

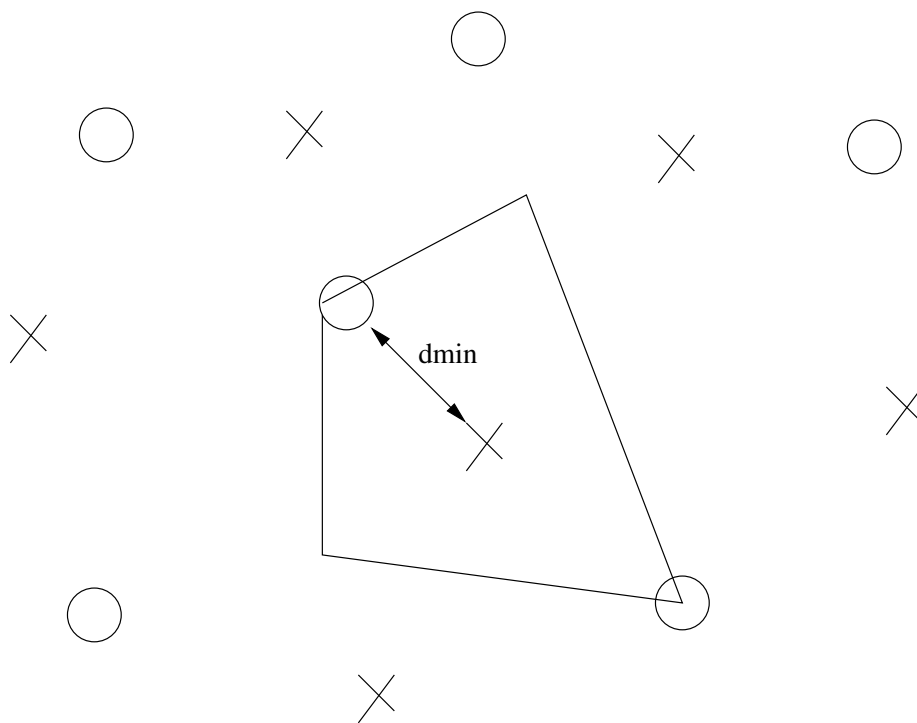


Figure 4: An example of QIM