Codes for Writing on Dirty Paper

Dirk van den Borne[1], Ton Kalker[12], and Frans Willems[12]

[1] Eindhoven University of Technology, Eindhoven, The Netherlands
[2] Philips Research Laboratories, Eindhoven, The Netherlands

Abstract. We have investigated coding methods that aim at achieving the Costa capacity. First we have considered the uncoded case to determine what can be gained from coding. After that we have designed codes with a trellis structure that are combined with a two-level quantizer at the transmitter side. Simulations are carried out to determine the coding gains of these methods. We have focussed on vector quantization instead of on scalar quantization to achieve a good performance.

1 System description

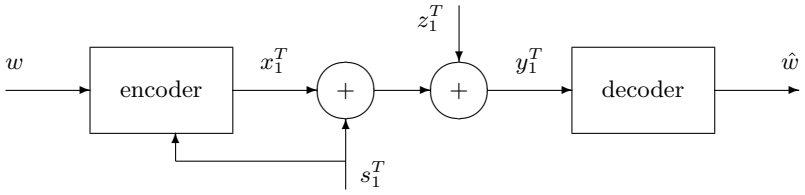

Fig. 1. Costa's Gaussian side-information model.

Costa [3] determined the capacity of the additive white Gaussian noise channel with side information at the transmitter (see figure 1). In this model the channel input vector $x_1^T = (x_1, x_2, \cdots, x_T)$ consisting of T real-valued components is modified by two additive memoryless Gaussian noise sources. The first noise source produces the vector $s_1^T = (s_1, s_2, \cdots, s_T)$ with components $\in \mathcal{N}(0, Q)$. This noise vector is observed by the encoder before the transmission process starts. It is what we call *side information* for the encoder. The encoder produces the code vector $x_1^T = e(w, s_1^T)$. Here $w \in \{1, 2, \cdots, M\}$ is the index of the message that must be conveyed. All messages are equally probable, i.e. $\Pr\{W = w\} = 1/M$ for all $w \in \{1, 2, \cdots, M\}$. All code vectors x_1^T should satisfy a power constraint, i.e. $\sum_{t=1,T} x_t^2 \leq TP$, where P denotes the signal power. The channel noise vector $z_1^T = (z_1, z_2, \cdots, z_T)$ has components $\in \mathcal{N}(0, N)$. The channel output vector $y_1^T = (y_1, y_2, \cdots, y_T) = x_1^T + s_1^T + z_1^T$ is processed by the decoder and leads to the message estimate $\hat{w} = d(y_1^T)$.

The system has two parameters that determine its performance. For given transmitter power P and noise variances Q and N the rate

$$R \stackrel{\Delta}{=} \frac{1}{T} \log_2(M) \tag{1}$$

should be as large as possible, while on the other hand the expected error-probability

$$P_{\mathcal{E}} \stackrel{\Delta}{=} \Pr\{\hat{W} \neq W\} \tag{2}$$

should be as small as possible. The largest rate that is achievable with arbitrarily small expected error-probability is called the capacity C. Costa showed that this capacity is determined only by the transmitter power P and the variance N of the channel noise, i.e. the noise that is unknown to the transmitter. It appears[3] that

$$C = \frac{1}{2}\log_2\left(1 + \frac{P}{N}\right) \text{ bits/transmission.} \tag{3}$$

Since the transmitter knows a part of the noise, this signaling method is called "writing on dirty paper". Chen and Wornell [2] have been the first to realize that Costa's result applies to digital watermarking and information embedding. In this paper we explore options for turn Costa's results into a practical watermarking system.

2 Baseline system, uncoded transmission

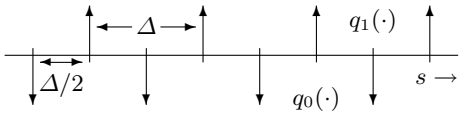

Fig. 2. Two uniform scalar quantizers.

Chen and Wornell [1] proposed a method that is based on quantization index modulation (QIM). They use two uniform quantizers, a quantizer $q_0(\cdot)$ and a quantizer $q_1(\cdot)$. The distance between the reconstruction points for each of these quantizers is Δ, their offset is $\Delta/2$ (see figure 2). We can now transmit a uniform binary digit $b \in \{0, 1\}$ by applying either the 0-quantizer or the 1-quantizer to the noise sample s. The input x is used to make $x + s$ equal to $q_b(s)$. In other words

$$x = q_b(s) - s \text{ for } b = 0, 1. \tag{4}$$

For variance $Q >> \Delta^2$ the power needed to modify s is the mean-square quantization error, i.e.

$$P = \int_{-\Delta/2}^{\Delta/2} x^2 dx = \frac{\Delta^2}{12}. \tag{5}$$

The receiver observes the noisy version $y = x + s + z = q_b(s) + z$ of the quantized s. It decides for $\hat{b} \in \{0, 1\}$ if y is closer to a reconstruction point $q_{\hat{b}}(s)$ for some

[3] If the side-information is ignored the maximum rate is $\frac{1}{2}\log_2(1 + P/(N + Q))$.

s than to any other reconstruction point $q_{1-\hat{b}}(s')$ for some s'. If the magnitude of the noise sample z is larger than $\Delta/4$ an error may occur. Therefore

$$P_{\mathcal{E}} \leq 2Q\left(\frac{\Delta}{4\sqrt{N}}\right) = 2Q\left(\sqrt{\frac{3P}{4N}}\right), \tag{6}$$

where $Q(t) \stackrel{\Delta}{=} \int_t^\infty \frac{1}{\sqrt{2\pi}} \exp(-\beta^2/2) d\beta$. Note that this uncoded method yields

$$R = 1 \text{ bit/transmission.} \tag{7}$$

If we are interested in achieving error probability $P_{\mathcal{E}} = 10^{-5}$ the argument of the Q-function should be 4.417. Therefore a signal-to-noise ratio $P/N = 26.02$ is required if we apply the baseline scheme.

3 Gap to capacity

From the capacity expression (3) it can be concluded that a signal-to-noise ratio

$$(P/N)_{\min} = 2^{2R} - 1. \tag{8}$$

suffices to obtain reliable communication. Therefore we define the gap to capacity as

$$g \stackrel{\Delta}{=} \frac{P/N}{(P/N)_{\min}} = \frac{P/N}{2^{2R}-1}. \tag{9}$$

This implies that the baseline system described in the previous section operates at $g = 26.02/3.0 = 9.38$ dB from capacity. It is our objective to study techniques and codes that close this gap.

4 Spreading techniques

In the previous section we have discussed a QIM-scheme that achieves rate $R = 1$ bit/transmission. By spreading, which is similar to spread-transform dither modulation (see e.g. [1]), we can construct from such a scheme a scheme with rate $R' = 1/k$ that requires transmitter power $P' = P/k$ for $k = 1, 2, \cdots$ without changing the (bit) error probability. To do so we consider a block of k transmissions and define

$$\overline{s} \stackrel{\Delta}{=} \frac{1}{\sqrt{k}} \sum_{i=1,k} s_k, \;\; \overline{x} \stackrel{\Delta}{=} \frac{1}{\sqrt{k}} \sum_{i=1,k} x_k, \text{ and } \overline{z} \stackrel{\Delta}{=} \frac{1}{\sqrt{k}} \sum_{i=1,k} z_k. \tag{10}$$

Now the transmitter chooses $\overline{x}$ such that

$$\overline{x} + \overline{s} = q_b(\overline{s}). \tag{11}$$

This requires transmitter power $P' = \Delta^2/(12k) = P/k$ by assuming that $x_i = \overline{x}/\sqrt{k}$ for $i = 1, k$. Note that the variance of the noise variable $\overline{Z}$ is equal to N. The receiver makes the decision $\hat{b}$ based on $\overline{y} = \overline{x} + \overline{s} + \overline{z}$.

Note that the gap to capacity increases by spreading since P/k is proportional to $1/k$ while $2^{2R/k} - 1$ is convex-$\cup$ in $1/k$. The gap approaches $26.02/\ln(4) = 12.73$ dB for $k \to \infty$.

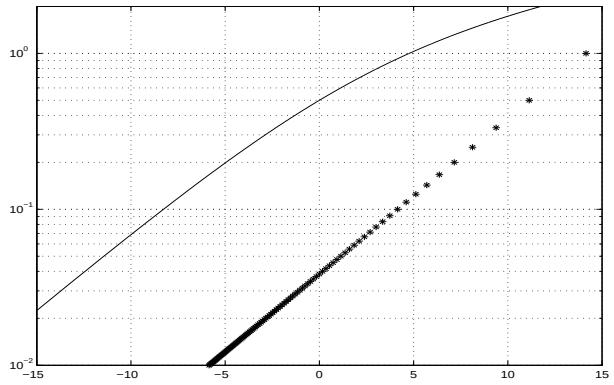

Fig. 3. Capacity in bits/transmission as a function of P/N in dB (horizontally). The *-curve denotes the R' versus P'/N behavior for $k = 1, 2, \cdots$ for the baseline system.

5 Compensation

Compensation was proposed by Chen and Wornell [2] to adapt QIM to Costa coding. Here compensation is described differently in order to keep the error-behavior of the system with and without compensation roughly the same.

Consider again the baseline system, with transmitter power P and channel noise variance N as a reference. Assume that $P/N > 1$. An alternative system is obtained by moving a noise sample s towards its reconstruction point $q_b(s)$, however not entirely. What we do is fix an $0 < \alpha \leq 1$ and take

$$x + s = q_b(s) + (1 - \alpha)(s - q_b(s)). \tag{12}$$

Then the now required transmitter power P_c is

$$P_c = E[X^2] = \alpha^2 E[q_b(S) - S] = \alpha^2 \Delta^2/12 = \alpha^2 P. \tag{13}$$

Since the signal in $y = x + s + n_c$ is $q_b(s)$, the variance of total noise in y is

$$(1 - \alpha)^2 E[q_b(S) - S] + E[Z_c^2] = (1 - \alpha)^2 \Delta^2/12 + N_c = (1 - \alpha)^2 P + N_c. \tag{14}$$

Suppose that, in order to keep the performance of the alternative system identical to that of the baseline system, we make the total noise variance[4] equal to N then

$$N_c = N - (1 - \alpha)^2 P. \tag{15}$$

The effective signal-to-noise ratio of the alternative system is

$$P_c/N_c = \frac{\alpha^2 P}{N - (1 - \alpha)^2 P}. \tag{16}$$

[4] We assume here that the compensation term is statistically independent of the signal. Observe however that the total noise is not Gaussian so it is not guaranteed that the performance remains constant.

It can be shown that this signal-to-noise ratio is maximized by taking

$$\alpha^* = \frac{P - N}{P}, \tag{17}$$

and moreover

$$P_c/N_c = P/N - 1. \tag{18}$$

It is important to note that

$$\alpha^* = \frac{P_c}{P_c + N_c}, \tag{19}$$

which is consistent with the observations of Costa [3].

If this compensation method is used for the baseline method and if we require an error probability of 10^{-5} the signal-to-noise ratio is decreased from 26.02 downto 25.02 which results in a gain of 0.17 dB.

6 Coding methods

6.1 Background

Eggers [4] investigated coding techniques based on scalar quantizing. He needed, for a $R = 1/3$ turbo-code, a signal-to-noise ratio of less than 1.5 dB for bit-error rate 10^{-5}. The minimal ratio is $2^{2/3} - 1 = -2.31$ dB, hence his method is roughly 3.8 dB from capacity. It was pointed out by Erez et al. [5] that by using vector instead of scalar quantization a gain of 1.53 dB could be achieved. A similar gain was obtained by the shaping methods proposed by Forney [6]. We apply Forney's techniques here to "collect" shaping gain. Instead of turbo-codes we focus on trellis codes however. A similar approach was suggested recently in [7].

The codes that we study here are based on the well-known rate-1/2 convolutional code with generator polynomials $(x^6+x^4+x^3+x+1, x^6+x^5+x^4+x^3+1)$. The number of states of this convolutional code is 64, its free (Hamming) distance $d_{H,\text{free}} = 10$. This code, in feed-forward mode, is combined with the two-level quantizer that is described as part of the baseline system in section 2. First we will investigate a method based on scalar quantization. After that we will modify the setup and show that applying vector quantization (slightly) outperforms scalar quantization.

6.2 Coding with scalar quantization

In the scalar case the message stream enters the convolutional encoder, and produces a code stream (see figure 4). For each binary message digit d two binary code digits b' and $b"$ are generated. With these two code digits the two-level quantizer modifies the noise samples s' and $s"$. We can use this method without ($\alpha = 1$) and with compensation. The receiver is equipped with a Viterbi detector which is processing using the squared Euclidean distance metric.

It will be clear that the required transmitter power of this method is equal to that of the baseline system, both in the uncompensated as as well as in the

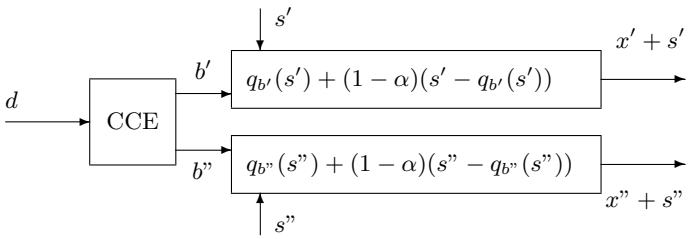

Fig. 4. Encoder that performs for scalar quantization.

compensated case. What is different however is the distance between the signal sequences $x_1^T + s_1^T$. The squared Euclidean distance between two such signal sequences is at least

$$d^2_{E,\text{free}} = d_{H,\text{free}} \cdot (\Delta/2)^2 = 5\Delta^2/2. \tag{20}$$

This is an increase by a factor 10 over the baseline case, achieved at the cost a smaller rate however. In figure 6 we see that we achieve a bit-error probability of 10^{-5} at $P/N \approx 7.0$ dB in the uncompensated case and at $P/N \approx 6.0$ dB with compensation. Since the $(P/N)_{\min} = 0$ dB for $R = 1/2$, the gap is 6.0 resp. 7.0 dB. Spreading the baseline system is roughly 11.14 dB from the Costa capacity for $R = 1/2$, so we have gained 5.1 resp. 4.1 dB relative to spreading.

6.3 Coding with vector quantization

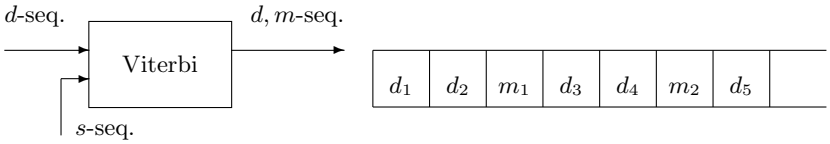

Fig. 5. Encoder for the vector-quantization case.

In the vector case not all the digits that enter the convolutional encoder are message digits. Part of the digits (see figure 5), the m (match-) digits, are used to obtain a signal sequence $x_1^T + s_1^T$ close to the sequence s_1^T (the side information sequence). In the figure, of every three digits that will enter the convolutional encoder, the last digit is a match digit. A Viterbi algorithm chooses, given the message (d-) digits, these match digits in such a way that the squared Euclidean distance between the signal sequence and the side information sequence is minimized. Since part of the convolutional code input digits are applied for matching, we have to face a rate-loss. We considered a match digit for every one, two, and three message digits. The corresponding rates are $R = 1/4, 1/3$, and $3/8$. We

have determined the bit-error probability for these setups, both for the uncompensated and compensated case (see the three rightmost plots in figure 6). It can be seen from this figure that the signal-to-noise ratios that are needed to achieve bit-error probability 10^{-5} are approximately 3.0 dB, 4.0 dB and 4.2 dB.

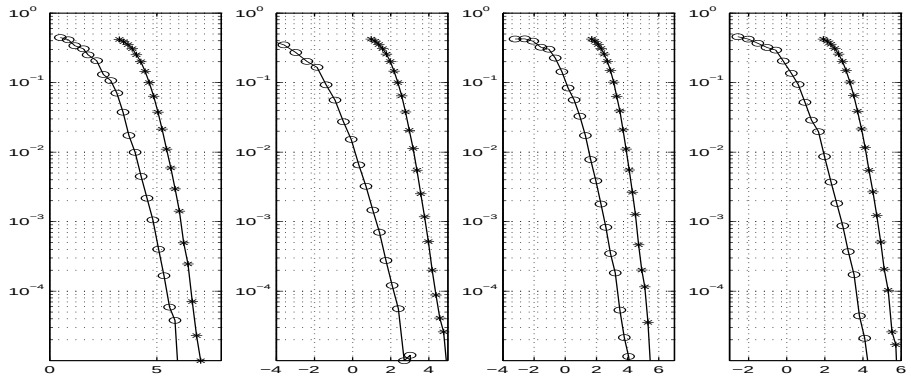

Fig. 6. Bit-error probability as a function of P/N (horizontally) for a $R = 1/2$ scalar code (left) and three vector codes (from left to right $R = 1/4, 1/3, 3/8$). For each point 10^6 message digits were processed.

All our simulations are carried out with a fixed Δ. We consider a range of values of the channel noise variance N. First, in the uncompensated case the required transmitter power P is measured together with the bit-error rate. After that, to apply compensation, we take $\alpha^* = (P-N)/P$, and $N_c = N-(1-\alpha^*)^2 P$. Then we measure P_c and the bit-error rate (which does not change very much).

6.4 Comparison

To compare the results of the scalar and vector case we consider the normalized signal-to-noise ratio defined as

$$(P/N)_{\text{norm}} = \frac{P/N}{2^{2R}-1}. \tag{21}$$

For the scalar $R = 1/2$ setup and the vector technique with $R = 3/8$ we have plotted the bit-error rate in figure 7 as a function of the *normalized* signal-to-noise ratio. It can be concluded from this figure that vector quantization gives an improvement of roughly 0.2 dB.

7 Conclusion

Although we have found a small improvement we are still 6 dB away from capacity. This is 2.2 dB worse than the turbo codes used by Eggers [4].

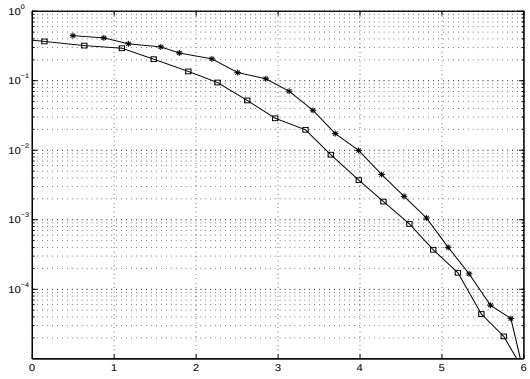

Fig. 7. Bit-error rate versus normalized signal-to-noise ratio in dB. The *-curve denotes the $R = 1/2$ scalar technique, the other curve the $R = 3/8$ vector method.

Note that our vector code has to compensate the rate loss between $R = 1/2$ of the scalar code and $R = 3/8$ which is $(2^1 - 1)/(2^{3/4} - 1) = 1.66$ dB. That we still gain 0.2 dB shows the potential of vector quantization.

In this paper we have always compared our results to capacity. If instead we would compare our simulation results to the spreading results for the baseline system our conclusion would be more positive. Then the rate loss counts for $4/3 = 1.25$ dB only, and the gain resulting from vector quantization would be roughly 0.6 dB.

References

1. B. Chen and G.W. Wornell, "Preprocessed and Postprocessed Quantization Index Modulation Methods for Digital Watermarking," *Proc. of SPIE: Security and Watermarking of Multimedia Contents II,* vol. 3971, pp. 48 - 59, San Jose, CA, January 2000.
2. B. Chen and G.W. Wornell, "Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding," *IEEE Trans. Inform. Theory,* vol. IT-47, pp. 1426-1443, May 2001.
3. M.H.M. Costa, "Writing on Dirty Paper, *IEEE Trans. Inform. Theory,* vol. IT-29, pp. 439-441, May 1983.
4. J. Eggers and B. Girod, *Informed Watermarking*, preprint, 2002.
5. U. Erez, S. Shamai, and R. Zamir, "Capacity and Lattice Strategies for Cancelling Known Interference," *ISITA,* Nov. 2000.
6. G.D. Forney Jr., "Trellis Shaping," *IEEE Trans. Inform. Theory,* vol. IT-38, pp. 281- 300, Jan. 1992.
7. W. Yu and J.M. Cioffi, "Trellis Precoding for the Broadcast Channel," *Proc. IEEE GLOBECOM 2001,* San Antonio, TX, Nov. 25-29, 2001.