Principles of Coding and Detection in Communication: Capacity of the AWGN Channel

Uri Erez

1 Problem Statement

Consider an AWGN channel,

$$\mathbf{y} = \mathbf{x} + \mathbf{z}, \tag{1}$$

where all vectors are of length N. The input is subject to a power constraint

$$\frac{1}{N}\|\mathbf{x}\|^2 \leq P \tag{2}$$

and the noise is i.i.d. Gaussian with variance σ_z^2. We shall prove the following theorem.

Shannon's (1948): For any rate satisfying $R < C = \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma_z^2}\right)$ and for any $\epsilon > 0$, for sufficiently large N there exists a codebook of size 2^{NR} and a decoder such that the average probability of error $\bar{P}_e < \epsilon$.

Remarks:

- The converse to the statement of the theorem is also true. That is, it can be shown that transmission over the AWGN channel at higher rates than $\frac{1}{2}\log_2\left(1 + \frac{P}{\sigma_z^2}\right)$ with arbitrarily small probability of error is not possible, and hence this value is the *capacity* of the AWGN channel.
- The direct part (i.e., achievablity of transmission at rates approaching $\frac{1}{2}\log_2(1+\frac{P}{\sigma_z^2})$) however does not rely on the noise being AWGN and the theorem holds for a much broader class of channels. In fact, the proof below requires only that the noise have average power σ_z^2 and that it is empirically *uncorrelated* with the input signal. Thus, the noise may be non-Gaussian, and does not even need to be statistically independent of the transmitted signal.

2 Random Codebook Generation

We use the notation $\mathcal{B}(\mathbf{v}, r)$ to denote a ball of radius r centered at $\mathbf{v}$. For any blocklength N, let us generate a codebook of size $M = 2^{NR}$

$$\mathcal{C} = \{\mathbf{X}^i\}_{i=1}^{M} \tag{3}$$

by drawing each of the codewords uniformly at random (and independently of each other) over a ball $\mathcal{B}(0, r_x)$ where

$$r_x = \sqrt{NP} \tag{4a}$$

$$\triangleq \sqrt{N}\rho_x. \tag{4b}$$

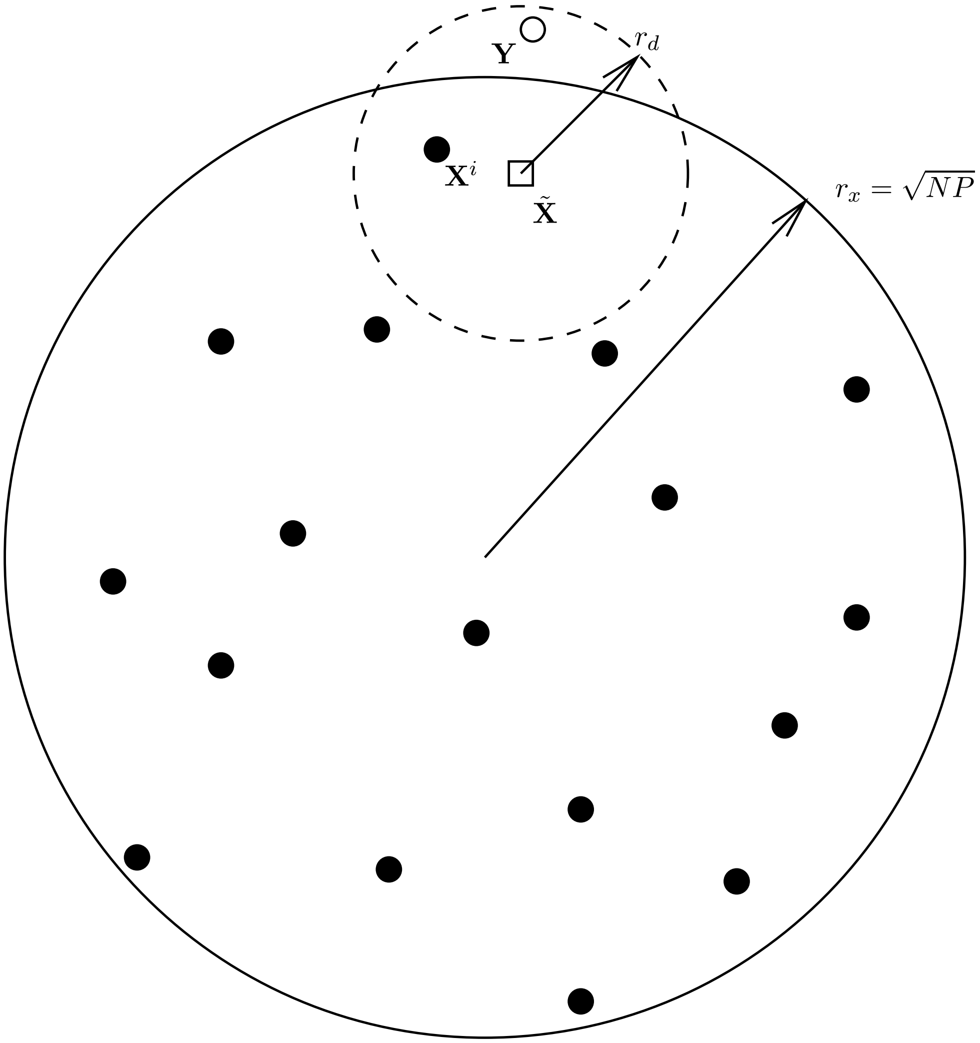
It follows that indeed for any $i = 1, \ldots, M$,

$$\frac{1}{N}\|\mathbf{X}^i\|^2 \leq P. \tag{5}$$

It remains to analyze the average (over the ensemble of codebooks and codewords) probability of error.

We know that the optimal decoder is an ML decoder which amounts in the case of an AWGN channel, to a minimum Euclidean distance (MED) decoder. Nonetheless, we shall consider a suboptimal decoder as it will make the analysis easier, while still allowing to approach capacity.

Decoding will consist of two operations. First, a *linear* estimate of the transmitted vector is formed. Next, if there exists one (and only one) codeword within a certain radius (slightly greater than the average estimation error) around this estimator, this codeword is the decoded codeword. Otherwise, we arbitrarily choose $\mathbf{x}^1$ (or any other codeword) as the decoded codeword.



3 Linear Estimator of Transmitted Vector

Denote the linear estimator of $\mathbf{X}$ from the channel output $\mathbf{Y}$ by $\tilde{\mathbf{X}} = \alpha \mathbf{Y}$. Then the estimation error is given by

$$\begin{aligned} \mathbf{e} &= \mathbf{X} - \tilde{\mathbf{X}} && \text{(6a)} \\ &= \mathbf{X} - \alpha(\mathbf{X} + \mathbf{Z}) && \text{(6b)} \\ &= (1-\alpha)\mathbf{X} - \alpha \mathbf{Z}. && \text{(6c)} \end{aligned}$$

The power of $\mathbf{e}$ satisfies

$$\begin{aligned} \sigma_e^2 &\stackrel{\triangle}{=} \frac{1}{N}\mathbb{E}[\|\mathbf{e}\|^2] && \text{(7a)} \\ &= \frac{1}{N}(1-\alpha)^2\mathbb{E}[\|\mathbf{X}\|^2] - \frac{2\alpha(1-\alpha)}{N}\sum_{n=0}^{N-1}\mathbb{E}[X_n Z_n] + \alpha^2\sigma_z^2 && \text{(7b)} \\ &= \frac{1}{N}(1-\alpha)^2\mathbb{E}[\|\mathbf{X}\|^2] + \alpha^2\sigma_z^2 && \text{(7c)} \\ &\leq (1-\alpha)^2 P + \alpha^2\sigma_z^2. && \text{(7d)} \end{aligned}$$

Let us choose α so as to minimize (7d). We choose α such that

$$\frac{d}{d\alpha}\left((1-\alpha)^2 P + \alpha^2\sigma_z^2\right) = -2(1-\alpha)P + 2\alpha\sigma_z^2 = 0. \qquad (8)$$

Thus the minimizing value of α satisfies

$$P - \alpha P = \alpha\sigma_z^2.$$

Rearranging terms we obtain

$$\alpha = \frac{P}{P+\sigma_z^2} = \frac{\text{SNR}}{1+\text{SNR}}. \qquad (9)$$

Substituting this value of α into (7d), the resulting estimation error satisfies

$$\begin{aligned} \frac{1}{N}\mathbb{E}[\|\mathbf{e}\|^2] &\leq \frac{P\sigma_z^2}{P+\sigma_z^2} && \text{(10a)} \\ &\stackrel{\triangle}{=} \rho_e^2. && \text{(10b)} \end{aligned}$$

4 Sphere Decoder

Consider the sphere $\mathcal{B}(\tilde{\mathbf{X}}, r_d)$ of radius $r_d = \sqrt{N}\rho_d$ centered at $\tilde{\mathbf{X}}$. Then, the decoded codeword index will be

$$\hat{i} = \begin{cases} j & \mathbf{X}^j \in \mathcal{B}(\tilde{\mathbf{X}}, r_d) \text{ and no other codeword is in } \mathcal{B}(\tilde{\mathbf{X}}, r_d) \\ 1 & \text{otherwise} \end{cases} \qquad (11)$$

5 Error Events

Define the following events:

1. The codeword that was transmitted $\mathbf{X}^i$ does not lie in $\mathcal{B}(\tilde{\mathbf{X}}, r_d)$. Denote this event by $\mathcal{E}_i^c$. We refer to this error event as *type I*:

$$\mathcal{E}_i^c \triangleq \left\{ \mathbf{X}^i \notin \mathcal{B}(\tilde{\mathbf{X}}, r_d) \right\} \tag{12}$$

2. The codeword $\mathbf{X}^j$, $j \neq i$, lies in $\mathcal{B}(\tilde{\mathbf{X}}, r_d)$. Denote this event by $\mathcal{E}_j$. We refer to such an error event as *type II*:

$$\mathcal{E}_j \triangleq \left\{ \mathbf{X}^j \in \mathcal{B}(\tilde{\mathbf{X}}, r_d) \right\} \tag{13}$$

Then, given that the ith codeword was transmitted, the error event is equal to

$$\mathcal{E} = \mathcal{E}_i^c \cup \bigcup_{j \neq i} \mathcal{E}_j. \tag{14}$$

By the union bound the expected (over the ensemble of codebooks) probability of error given that the message to be sent was i satisfies,

$$P_e[i] \leq \Pr(\mathcal{E}_i^c) + \sum_{j \neq i} \Pr(\mathcal{E}_j)$$

6 Type I Error Probability

We begin by bounding the probability of the event $\mathcal{E}_i^c$, *given that* $\mathbf{X}^i$ was sent. That is, the codeword that was transmitted, $\mathbf{X}^i$, does not lie within a radius r_d from $\tilde{\mathbf{X}}$. This occurs when, given $\mathbf{X}^i$ was transmitted,

$$\|\mathbf{X}^i - \tilde{\mathbf{X}}\|^2 > r_d^2$$

We may expand the l.h.s. of the last inequality as follows:

$$\begin{aligned}
\frac{1}{N}\|\mathbf{X}^i - \tilde{\mathbf{X}}\|^2 &= \frac{1}{N}\|(1-\alpha)\mathbf{X}^i + \alpha\mathbf{Z}\|^2 && (15a)\\
&= (1-\alpha)^2 \frac{1}{N}\sum_{n=0}^{N-1}(X_n^i)^2 + 2(1-\alpha)\alpha\frac{1}{N}\sum_{n=0}^{N-1} X_n^i Z_n + \alpha^2 \frac{1}{N}\sum_{n=0}^{N-1} Z_n^2. \\
&\triangleq A_n + B_n + C_n. && (15b)
\end{aligned}$$

Clearly we have for all N

$$A_n \triangleq (1-\alpha)^2 \frac{1}{N}\|\mathbf{X}\|^2 \leq (1-\alpha)^2 P.$$

We also observe that by the law of large numbers we have that $C_n \to \alpha^2\sigma_z^2$ in probability. Further, since B_n being a linear combination of i.i.d. random variables, we have that,

$$\begin{aligned}
\mathrm{Var}(B_n|\mathbf{X}^i) &= 4(1-\alpha)^2\alpha^2 \frac{1}{N^2}\sum_{n=0}^{N-1}(X_n^i)^2\sigma_z^2 && (16a)\\
&= \frac{1}{N}4(1-\alpha)^2\alpha^2\left(\frac{1}{N}\|\mathbf{X}^i\|^2\right)\sigma_z^2 && (16b)\\
&\leq \frac{1}{N}4(1-\alpha)^2\alpha^2\sigma_z^2 P. && (16c)
\end{aligned}$$

Therefore

$$\begin{aligned}\mathrm{Var}(B_n) &= \mathbb{E}_{\mathbf{X}^i}\left[\mathrm{Var}(B_n|\mathbf{X}^i)\right] &\text{(17a)}\\ &\leq \frac{1}{N}4(1-\alpha)^2\alpha^2\sigma_z^2 P. &\text{(17b)}\end{aligned}$$

Hence, the variance of B_n goes to zero as $N \to \infty$ and it follows that B_n tends to zero in probability. We note that as B_n is simply the empirical correlation between the transmitted signal and noise sequences and that for the theorem to hold, all that is needed is that this correlation goes to zero as $N \to \infty$.

We have thus established that

$$\begin{aligned}\frac{1}{N}\|\mathbf{X}^i - \tilde{\mathbf{X}}\|^2 &\leq (1-\alpha)^2 P + \alpha^2\sigma_z^2 &\text{(18a)}\\ &= \frac{P\sigma_z^2}{P+\sigma_z^2} &\text{(18b)}\\ &= \rho_e^2 &\text{(18c)}\end{aligned}$$

with probability going to one as $N \to \infty$. In order that the error probability will go to 0 as $N \to \infty$, we require (a sufficient condition) that

$$\rho_d > \rho_e. \tag{19}$$

With such a choice, we have

$$\lim_{N\to\infty} \Pr(\mathcal{E}_i^c) = \lim_{N\to\infty} \Pr\left(\|\mathbf{X}^i - \tilde{\mathbf{X}}\| > \sqrt{N}\rho_d\right) = 0. \tag{20}$$

7 Type II Error Probability

We turn to bound the probability that some other codeword $\mathbf{X}^j$, $j \neq i$, lies in $\mathcal{B}(\tilde{\mathbf{X}}, r_d)$ *given that* $\mathbf{X}^i$ was sent. Since the codewords are uniformly distributed over $\mathcal{B}(0, r_x)$, it follows that for any $j \neq i$

$$\begin{aligned}\Pr(\mathcal{E}_j) &\leq \frac{\mathrm{Volume}[\mathcal{B}(\tilde{\mathbf{X}}, r_d)]}{\mathrm{Volume}[\mathcal{B}(0, r_x)]} &\text{(21a)}\\ &= \left(\frac{r_d}{r_x}\right)^N. &\text{(21b)}\end{aligned}$$

Using the union bound we have

$$\begin{aligned}\Pr(\cup_{j\neq i}\mathcal{E}_j) &\leq (M-1)\Pr(\mathcal{E}_j) &\text{(22a)}\\ &\leq 2^{NR}\left(\frac{r_d}{r_x}\right)^N &\text{(22b)}\\ &= \left(2^R \cdot \frac{r_d}{r_x}\right)^N &\text{(22c)}\\ &= \left(2^R \cdot \frac{\rho_d}{\rho_x}\right)^N. &\text{(22d)}\end{aligned}$$

Thus, the probability of a type II error will go to zero exponentially in N if we choose ρ_d such that

$$2^R \cdot \frac{\rho_d}{\rho_x} < 1, \tag{23}$$

which can be rewritten as

$$\begin{aligned} \rho_d &< \rho_x \cdot 2^{-R} && \text{(24a)} \\ &\triangleq \rho_v. && \text{(24b)} \end{aligned}$$

8 Putting It All Together

Combining the results of Sections 6 and 7 we conclude that sufficient conditions for the error probability to go to 0 as $N \to \infty$ is

$$\rho_e < \rho_d < \rho_v. \tag{25}$$

By substituting (4b), (10a) and (24b), the inequality $\rho_e < \rho_v$ is equivalent to:

$$\sqrt{\frac{P\sigma_z^2}{P+\sigma_z^2}} < \sqrt{P} \cdot 2^{-R} \tag{26}$$

which is equivalent to

$$\begin{aligned} R &< \frac{1}{2}\log_2\left(\frac{P+\sigma_z^2}{\sigma_z^2}\right) && \text{(27a)} \\ &= \frac{1}{2}\log_2\left(1+\mathrm{SNR}\right). && \text{(27b)} \end{aligned}$$

Observe that this result holds for any message i due to the symmetrical construction of the ensemble. Thus, we also have

$$\lim_{N\to\infty} \bar{P}_e = \lim_{N\to\infty} \frac{1}{M}\sum_{i=1}^{M} P_e[i] = 0.$$

Finally, since the average probability of error over the ensemble of codewords goes to zero, it follows that there exists at least one codebook for which the same holds. This concludes the proof.

9 Error Probability Decay Rate

Different parameters of the sphere decoder lead to different decays of the the error probability. In this section we use the same α which minimized the effective noise $\mathbf{e}$. However, we carry out optimization over the radius of the decoding sphere, r_d. Further, we bound the Type I error probability more tightly.

As we shall see in the sequel, the error probability can be made to decay at an exponential rate. Since the probability of Type I error decreases (improves) with the increase of the decoding radius r_d, whereas the probability of Type II error — increases, our goal is to find the optimum value of r_d, which strikes a balance between the two.

9.1 Type I Error

Given that message i was chosen, from the definition of the type I error event (12)

$$\begin{aligned}\Pr(\mathcal{E}_i^c) &= \Pr(\mathbf{X}^i \notin \mathcal{B}(\tilde{\mathbf{X}}, r_d)) && \text{(28a)}\\ &= \Pr((1-\alpha)\mathbf{X}^i - \alpha\mathbf{Z} \notin \mathcal{B}(\mathbf{0}, r_d)) && \text{(28b)}\\ &= \Pr(\mathbf{e} \notin \mathcal{B}(\mathbf{0}, r_d))\,. && \text{(28c)}\end{aligned}$$

The effective noise $\mathbf{e}$ is a sum of a white Gaussian noise $\alpha\mathbf{Z}$ and a spherical "noise" $(1-\alpha)\mathbf{X}^i$ distributed uniformly over a ball. The following lemma, which will be proved in a home exercise, suggests that the resulting noise is no worse than a white Gaussian noise having the same second moment.

Lemma. *Let $\mathbf{Z}'$ be a WGN vector and $\mathbf{B}$ — a spherical noise vector, having the same second moment. Then, for any $\mathbf{x}$ inside the sphere*

$$\frac{1}{N}\ln\frac{f_B(\mathbf{x})}{f_{Z'}(\mathbf{x})} \le \epsilon_N, \tag{29a}$$

where

$$\epsilon_N = \frac{1}{2}\ln(2\pi e G_N^*), \tag{29b}$$

and

$$G_N^* = \frac{\Gamma^{2/N}\left(\frac{N}{2}+1\right)}{(N+2)\pi}. \tag{29c}$$

In addition

$$\lim_{N\to\infty} G_N^* = \frac{1}{2\pi e}. \tag{29d}$$

Therefore, $\epsilon_N \to 0$ as $N \to \infty$.

Corollary. *Let $\mathbf{Z}'$ be a WGN vector and $\mathbf{B}$ — a spherical noise vector having the same second moment. Then, for a white Gaussian vector $\mathbf{Z}$ which is statistically independent of $\mathbf{Z}'$ and $\mathbf{B}$,*

$$\Pr\left(\mathbf{B}+\mathbf{Z} \notin \mathcal{B}(\boldsymbol{0}, r_d)\right) \le e^{N\epsilon_N}\Pr\left(\mathbf{Z}'+\mathbf{Z} \notin \mathcal{B}(\boldsymbol{0}, r_d)\right)\,, \tag{30}$$

where $\epsilon_N \to 0$ as $N \to \infty$.

From this lemma one easily concludes that

$$\Pr(\mathbf{e} \notin \mathcal{B}(\mathbf{0}, r_d)) \le e^{N\cdot\epsilon_N}\cdot\Pr\left(\mathbf{Z}^* \notin \mathcal{B}(\mathbf{0}, r_d)\right), \tag{31}$$

where $\mathbf{Z}^* \sim \mathcal{N}(\mathbf{0}, I_N\cdot\rho_e^2)$, and $\epsilon_N \to 0$ as $N \to \infty$.

Using Poltyrev's bound (proved in the home assignment), leads to

$$\begin{aligned}\Pr\left(\mathbf{Z}^* \notin \mathcal{B}(\mathbf{0}, r_d)\right) &= \Pr\left(\frac{1}{N}\|\mathbf{Z}^*\|^2 \ge \rho_d^2\right) && \text{(32a)}\\ &\le \begin{cases}\exp\left[-N\cdot\frac{1}{2}\left(\frac{\rho_d^2}{\rho_e^2}-1-\ln\frac{\rho_d^2}{\rho_e^2}\right)\right], & \rho_d > \rho_e\\ 1, & \text{o.w.}\end{cases}\,, && \text{(32b)}\end{aligned}$$

where recall that $\rho_d = {}^{r_d}/\sqrt{N}$ is the normalized decoding radius and ρ_e is the effective noise power, defined in (7d).

Combining the results of (28c), (31) and (32) we see that the probability of the Type I error is upper bounded by

$$\Pr(\mathcal{E}_i^c) \le \exp\left[-N\frac{1}{2}\left(\frac{\rho_d^2}{\rho_e^2} - \ln\frac{\rho_d^2}{\rho_e^2} - 1 - \epsilon_N\right)\right] , \tag{33}$$

and the corresponding error exponent is

$$\lim_{N\to\infty} -\frac{1}{N}\ln\Pr(\mathcal{E}_i^c) \ge \frac{1}{2}\left(\frac{\rho_d^2}{\rho_e^2} - \ln\frac{\rho_d^2}{\rho_e^2} - 1\right) \tag{34a}$$

$$\triangleq E_1 \tag{34b}$$

9.2 Type II Error

Substituting (4a) in (21b), for any $j \neq i$, we have

$$\Pr(\mathcal{E}_j) \le \left(\frac{r_d}{r_x}\right)^N \tag{35a}$$

$$= \left(\frac{\rho_d^2}{P}\right)^{N/2} , \tag{35b}$$

which allows in turn, in conjunction with the union bound, to achieve the following upper bound on the type II error probability:

$$\Pr\left(\bigcup_{j\neq i}\mathcal{E}_j\right) \le (M-1)\Pr(\mathcal{E}_j) \tag{36a}$$

$$\le 2^{NR_{\text{bits}}}\left(\frac{\rho_d^2}{P}\right)^{N/2} \tag{36b}$$

$$= e^{NR_{\text{nats}}}\left(\frac{\rho_d^2}{P}\right)^{N/2} \tag{36c}$$

$$= e^{-E_2 N} , \tag{36d}$$

where

$$E_2 \triangleq \frac{1}{2}\ln\frac{P}{\rho_d^2} - R_{\text{nats}} . \tag{37}$$

9.3 Optimizing over r_d

Since the error probability is bounded from the above by the sum of the probabilities of the two error types, we have

$$\Pr(\mathcal{E}) \le e^{-N(E_1-\epsilon_N)} + e^{-NE_2} . \tag{38}$$

Or in terms of exponents:

$$\lim_{N\to\infty} -\frac{1}{N}\ln\Pr(\mathcal{E}) \geq \min(E_1, E_2) \stackrel{\triangle}{=} E. \tag{39}$$

In order to tighten the upper bound (r.h.s. of (38)), we recall that an increase in r_d improves E_1 but deteriorates E_2. Thus, we determine the optimum value of r_d by determining the value for which equality holds between the two exponents, i.e.,

$$E_1 = E_2\,. \tag{40}$$

By substituting (34b) and (37), we attain

$$\frac{1}{2}\left(\frac{\rho_d^2}{\rho_e^2} - \ln\frac{\rho_d^2}{\rho_e^2} - 1\right) = \frac{1}{2}\ln\frac{P}{\rho_d^2} - R_{\text{nats}}\,. \tag{41}$$

By substituting further $\rho_e^2 = \frac{P\sigma_z^2}{P+\sigma_z^2}$, and the capacity expression $C_{\text{nats}} = \frac{1}{2}\ln\left(1 + \frac{P}{\sigma_z^2}\right)$, we arrive at the following normalized-radius value

$$\rho_d^2 = \left[2\cdot(C_{\text{nats}} - R_{\text{nats}}) + 1\right]\rho_e^2 \tag{42a}$$

$$= \left[2\cdot(C_{\text{nats}} - R_{\text{nats}}) + 1\right]\frac{P\sigma_Z^2}{P+\sigma_Z^2}. \tag{42b}$$

Substituting (42a) in E_2 (37) (or in E_1 (34b)) we obtain the following (total) error exponent:

$$E = E_2 \tag{43a}$$

$$= \frac{1}{2}\ln\frac{P+\sigma_Z^2}{\sigma_Z^2} - \frac{1}{2}\ln\left[2\cdot(C_{\text{nats}} - R_{\text{nats}}) + 1\right] - R_{\text{nats}} \tag{43b}$$

$$= \Big(C_{\text{nats}} - R_{\text{nats}}\Big) - \frac{1}{2}\ln\Big(2(C_{\text{nats}} - R_{\text{nats}}) + 1\Big)\,. \tag{43c}$$

We see that indeed, the error exponent is positive for any $R < C$, which implies an exponential decay to zero with the block-length N, for such rates.

With this choice of ρ_d, using (38), we have the following bound:

$$\Pr(\mathcal{E}) \leq 2e^{-N(E-\epsilon_N)}\,. \tag{44}$$

We conclude by stating that even though we fixed α to be $\frac{P}{P+\sigma_z^2}$, it turns out that in general, in error-exponent sense, this is not the optimal choice. The optimal choice of α varies with the rate value R. By optimizing on both α and r_d, along with using a better decoding rule,[1] the optimal error exponent is achieved above the critical rate (see Fig. 1).

[1] ML decoding would be optimal of course, but perhaps not necessary in order to achieve the optimum exponent.

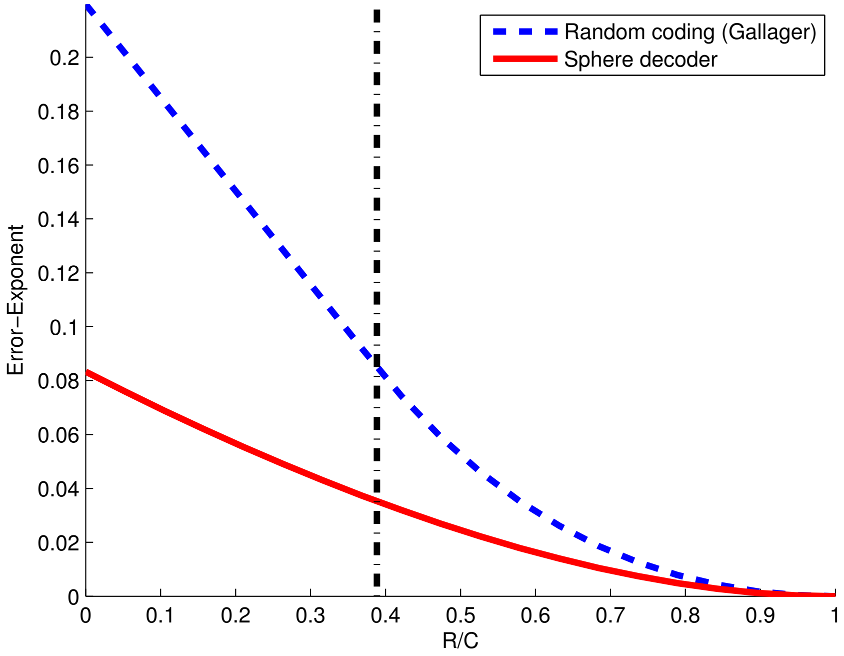


Figure 1: Showing different error exponents for the AWGN channel with SNR of 0dB. The solid curve is the exponent of (43a). The dashed curve is the random-coding error exponent, which is optimal above the critical rate of the channel (shown by the dash-dot vertical line).