19.1 Energy per bit

Consider the additive Gaussian noise channel:

$$Y_i = X_i + Z_i, \quad Z_i \sim \mathcal{N}(0, \frac{N_0}{2}). \tag{19.1}$$

In the last lecture, we analyzed the maximum number of information bits ($M^*(n, \epsilon, P)$) that can be pumped through for given n time use of the channel under the energy constraint P. Today we shall study the counterpart of it: without any time constraint, in order to send k information bits, what is the minimum energy needed? ($E^*(k, \epsilon)$)

Definition 19.1 ($(E, 2^k, \epsilon)$ code)**.** For a channel $W \to X^\infty \to Y^\infty \to \hat{W}$, where $Y^\infty = X^\infty + Z^\infty$, a $(E, 2^k, \epsilon)$ code is a pair of encoder-decoder:

$$\begin{aligned} f : [2^k] \to \mathbb{R}^\infty, &\quad g : \mathbb{R}^\infty \to [2^k], \\ \text{such that 1).} &\quad \forall m, \|f(m)\|_2^2 \le E, \\ \text{2).} &\quad P[g(f(W) + Z^\infty) \ne W] \le \epsilon. \end{aligned}$$

Definition 19.2 (Fundamental limit)**.**

$$E^*(k, \epsilon) = \min\{E : \exists (E, 2^k, \epsilon) \text{ code}\}$$

Note: Operational meaning of $\lim_{\epsilon \to 0} E^*(k, \epsilon)$: it suggests the smallest battery one needs in order to send k bits without any time constraints, below that level reliable communication is impossible.

Theorem 19.1 ($(E_b/N_0)_{min} = -1.6dB$)**.**

$$\lim_{\epsilon \to 0} \limsup_{k \to \infty} \frac{E^*(k, \epsilon)}{k} = \frac{N_0}{\log_2 e}, \quad \frac{1}{\log_2 e} = -1.6dB \tag{19.2}$$

Proof.

1. ("$\geq$" converse part)

$$
\begin{aligned}
-h(\epsilon)+\bar{\epsilon}k &\leq d((1-\epsilon)\|\frac{1}{M}) \qquad \text{(Fano)} \\
&\leq I(W;\hat{W}) \qquad \text{(data processing for divergence)} \\
&\leq I(X^\infty;Y^\infty) \qquad \text{(data processing for M.I.)} \\
&\leq \sum_{i=1}^{\infty} I(X_i;Y_i) \qquad (\lim_{n\to\infty} I(X^n;U)=I(X^\infty;U)) \\
&\leq \sum_{i=1}^{\infty} \frac{1}{2}\log(1+\frac{\mathbb{E}X_i^2}{N_0/2}) \qquad \text{(Gaussian)} \\
&\leq \frac{\log e}{2}\sum_{i=1}^{\infty}\frac{\mathbb{E}X_i^2}{N_0/2} \qquad \text{(linearization)} \\
&\leq \frac{E}{N_0}\log e \\
\Rightarrow \frac{E^*(k,\epsilon)}{k} &\geq \frac{N_0}{\log e}(\bar{\epsilon}-\frac{h(\epsilon)}{k}).
\end{aligned}
$$

2. ("$\leq$" achievability part)
   Notice that a $(n, 2^k, \epsilon, P)$ code for AWGN channel is also a $(nP, 2^k, \epsilon)$ code for the energy problem without time constraint. Therefore,

$$
\log_2 M^*_{max}(n,\epsilon,P) \geq k \Rightarrow E^*(k,\epsilon) \leq nP.
$$

   $\forall P$, take $k_n = \lfloor \log M^*_{max}(n,\epsilon,P)\rfloor$, we have $\frac{E^*(k_n,\epsilon)}{k_n} \leq \frac{nP}{k_n}$, $\forall n$, and take the limit:

$$
\begin{aligned}
\limsup_{n\to\infty}\frac{E^*(k_n,\epsilon)}{k_n} &\leq \limsup_{n\to\infty}\frac{nP}{\log M^*_{max}(n,\epsilon,P)} \\
&= \frac{P}{\liminf_{n\to\infty}\frac{1}{n}\log M^*_{max}(n,\epsilon,P)} \\
&= \frac{P}{\frac{1}{2}\log(1+\frac{P}{N_0/2})}
\end{aligned}
$$

   Choose P for the lowest upper bound:

$$
\begin{aligned}
\limsup_{n\to\infty}\frac{E^*(k_n,\epsilon)}{k_n} &\leq \inf_{P\geq 0}\frac{P}{\frac{1}{2}\log(1+\frac{P}{N_0/2})} \\
&= \lim_{P\to 0}\frac{P}{\frac{1}{2}\log(1+\frac{P}{N_0/2})} \\
&= \frac{N_0}{\log_2 e}
\end{aligned}
$$

□

Note: [Remark] In order to send information reliably at $E_b/N_0 = -1.6dB$, infinitely many time slots are needed, and the information rate (spectral efficiency) is thus 0. In order to have non-zero spectral efficiency, one necessarily has to step back from -1.6 dB.

Note: [PPM code] The following code, pulse-position modulation (PPM), is very efficient in terms of E_b/N_0.

$$\text{PPM encoder: } \forall m, f(m) = (0, 0, \dots, \underbrace{\sqrt{E}}_{m\text{-th location}}, \dots) \tag{19.3}$$

It is not hard to derive an upper bound on the probability of error that this code achieves [PPV11, Theorem 2]:

$$\epsilon \le \mathbb{E}\left[\min\left\{MQ\left(\sqrt{\frac{2E}{N_0}} + Z\right), 1\right\}\right], \qquad Z \sim \mathcal{N}(0,1).$$

In fact, the code can be further slightly optimized by subtracting the common center of gravity $(2^{-k}\sqrt{E}, \dots, 2^{-k}\sqrt{E} \dots)$ and rescaling each codeword to satisfy the power constraint. The resulting constellation (simplex code) is conjectured to be non-asymptotic optimum in terms of E_b/N_0 for small ϵ ("simplex conjecture").

19.2 What is N_0?

In the above discussion, we have assumed $Z_i \sim \mathcal{N}(0, N_0/2)$, but how do we determine N_0?

In reality the signals are continuous time (CT) process, the continuous time AWGN channel for the RF signals is modeled as:

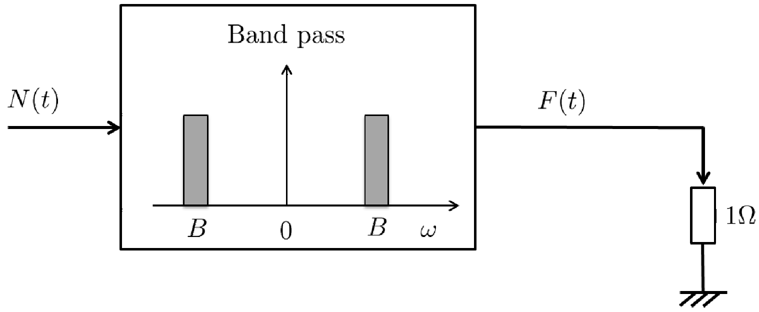
$$Y(t) = X(t) + N(t) \tag{19.4}$$

where noise $N(t)$ (added at the receiver antenna) is a real stationary ergodic process and is assumed to be "white Gaussian noise" with single-sided PSD N_0. Figure 19.1 at the end illustrates the communication architecture. In the following discussion, we shall find the equivalent discrete time (DT) AWGN model for the continuous time (CT) AWGN model in (19.4), and identify the relationship between N_0 in the DT model and $N(t)$ in the CT model.

- Goal: communication in $f_c \pm B/2$ band.
  (the (possibly complex) baseband signal lies in $[-W, +W]$, where $W = B/2$)
- observations:
  1. Any signal band limited to $f_c \pm B/2$ can be produced by this architecture
  2. At the step of C/D conversion, the LPF followed by sampling at B samples/sec is sufficient statistics for estimating $X(t), X_B(t)$, as well as $\{X_i\}$.

First of all, what is $N(t)$ in (19.4)?

Engineers' definition of $N(t)$



Testing whether a process $N(t)$ is "white noise"

Estimate the average power dissipation at the resistor:

$$\lim_{T\to\infty}\frac{1}{T}\int_{t=0}^{T}F_t^2dt \overset{\text{ergodic}}{=} \mathbb{E}[F^2] \overset{(*)}{=} N_0B$$

If for some constant N_0, (*) holds for any narrow band with center frequency f_c and bandwidth B, then $N(t)$ is called a "white noise" with one-sided PSD N_0.

Typically, white noise comes from thermal noise at the receiver antenna. Thus:

$$N_0 \approx k\mathbf{T} \tag{19.5}$$

where $k = 1.38\times10^{-23}$ is the Boltzmann constant, and $\mathbf{T}$ is the absolute temperature. The unit of N_0 is $(Watt/Hz = J)$.

An intuitive explanation to (19.5) is as follows: the thermal energy carried by each microscopic degree of freedom (dof) is approximately $\frac{k\mathbf{T}}{2}$; for bandwidth B and duration T, there are in total $2BT$ dof; by "white noise" definition we have the total energy of the noise to be:

$$N_0BT = \frac{k\mathbf{T}}{2}2BT \;\Rightarrow\; N_0 = k\mathbf{T}.$$

Mathematicians' definition of $N(t)$

Denote the set of all real finite energy signals $f(t)$ by $\mathcal{L}_2(\mathbb{R})$, it is a vector space with the inner product of two signals $f(t), g(t)$ defined by

$$<f,g>= \int_{t=-\infty}^{\infty} f(t)g(t)dt.$$

Definition 19.3 (White noise)**.** $N(t)$ is a white noise with two-sided PSD being constant $N_0/2$ if $\forall f,g\in\mathcal{L}_2(\mathbb{R})$ such that $\int_{-\infty}^{\infty}f^2(t)dt = \int_{-\infty}^{\infty}g^2(t)dt = 1$, we have that

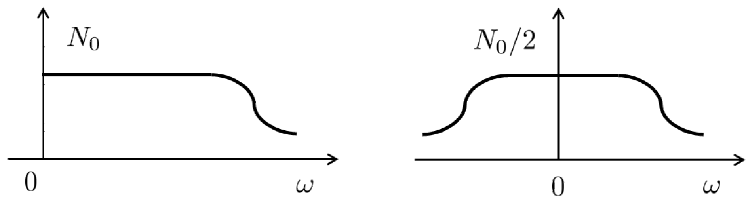
1. $$<f,N>\triangleq \int_{-\infty}^{\infty} f(t)N(t)dt \sim \mathcal{N}(0,\frac{N_0}{2}). \tag{19.6}$$

2. The joint distribution of $(<f,N>,<g,N>)$ is jointly Gaussian with covariance equal to inner product $<f,g>$.

Note: By this definition, $N(t)$ is not a stochastic process, rather it is a collection of linear mappings that map any $f\in\mathcal{L}_2(\mathbb{R})$ to a Gaussian random variable.

Note: Informally, we write:

$$N(t) \text{ is white noise with one-sided PSD } N_0 (or \text{ two-sided PSD } N_0/2) \Longleftrightarrow \mathbb{E}[N(t)N(s)] = \frac{N_0}{2}\delta(t-s) \tag{19.7}$$



Engineers' white noise   Mathematicians' white noise

Note: The concept of one-sided PSD arises when $N(t)$ is necessarily real, since in that case power spectrum density is symmetric around 0, and thus to get the noise power in band $[a,b]$ one can get

$$\text{noise power} = \int_a^b F_{\text{one-sided}}(f)df = \int_a^b + \int_{-b}^{-a} F_{\text{two-sided}}(f)df\,,$$

where $F_{\text{one-sided}}(f) = 2F_{\text{two-sided}}(f)$. In theory of stochastic processes it is uncommon to talk about one-sided PSD, but in engineering it is.

Verify the equivalence between CT /DT models

First, consider the relation between RF signals and baseband signals.

$$X(t) = Re(X_B(t)\sqrt{2}e^{j\omega_c t}),$$
$$Y_B(t) = \sqrt{2}LPF_2(Y(t)e^{j\omega_c t}),$$

where $\omega_c = 2\pi f_c$. The LPF_2 with high cutoff frequency $\sim \frac{3}{4}f_c$ serves to kill the high frequency component after demodulation, and the amplifier of magnitude $\sqrt{2}$ serves to preserve the total energy of the signal, so that in the absence of noise we have that $Y_B(t) = X_B(t)$. Therefore,

$$Y_B(t) = X_B(t) + \widetilde{N}(t) \sim \mathbb{C}$$

where $\widetilde{N}(t)$ is a complex Gaussian white noise and

$$\mathbb{E}[\widetilde{N}(t)\widetilde{N}(s)^*] = N_0\delta(t-s).$$

Notice that after demodulation, the PSD of the noise is $N_0/2$ with $N_0/4$ in the real part and $N_0/4$ in the imaginary part, and after the $\sqrt{2}$ amplifier the PSD of the noise is restored to $N_0/2$ in both real and imaginary part.

Next, consider the equivalent discrete time signals.

$$X_B(t) = \sum_{i=-\infty}^{\infty} X_i sinc_B(t - \frac{i}{B})$$
$$Y_i = \int_{t=-\infty}^{\infty} Y_B(t) sinc_B(t - \frac{i}{B})dt$$
$$Y_i = X_i + Z_i$$

where the additive noise Z_i is given by:

$$Z_i = \int_{t=-\infty}^{\infty} \widetilde{N}(t) sinc_B(t - \frac{i}{B}) dt \sim i.i.d\ \mathbb{C}\mathcal{N}(0, N_0). \qquad \text{(by (19.6))}$$

if we focus on the real part of all signals, it is consistent with the real AWGN channel model in (19.1).

Finally, the energy of the signal is preserved:

$$\sum_{i=-\infty}^{\infty} |X_i|^2 = \|X_B(t)\|_2^2 = \|X(t)\|_2^2.$$

Note: [Punchline]

$$\text{CT AWGN (band limited)} \Longleftrightarrow \text{DT } \mathbb{C}\text{-AWGN}$$
$$\text{two-sided PSD } \frac{N_0}{2} \Longleftrightarrow Z_i \sim \mathbb{C}\mathcal{N}(0, N_0)$$
$$\text{energy=} \int X(t)^2 dt \Longleftrightarrow \text{energy=} \sum |X_i|^2$$

19.3 Capacity of the continuous-time band-limited AWGN channel

Theorem 19.2. *Let $M^*_{CT}(T, \epsilon, P)$ the maximum number of waveforms that can be sent through the channel*

$$Y(t) = X(t) + N(t), \qquad \mathbb{E}\, N(t)N(s) = \frac{N_0}{2}\delta(t-s)$$

such that:

1. *in the duration $[0, T]$;*
2. *band limited to $[f_c - \frac{B}{2}, f_c + \frac{B}{2}]$ for some large carrier frequency*
3. *input energy constrained to $\int_{t=0}^{T} x^2(t) \le TP$;*
4. *error probability $P[\hat{W} \ne W] \le \epsilon$.*

Then

$$\lim_{\epsilon\to 0} \liminf_{n\to\infty} \frac{1}{T} \log M^*_{CT}(T, \epsilon, P) = B \log(1 + \frac{P}{N_0 B}), \qquad (19.8)$$

Proof. Consider the DT equivalent $\mathbb{C}$-AWGN channel of this CT model, we have that

$$\frac{1}{T} \log M^*_{CT}(T, \epsilon, P) = \frac{1}{T} \log M^*_{\mathbb{C}-\text{AWGN}}(BT, \epsilon, P/B)$$

This is because:

- in time T we get to choose BT complex samples
- The power constraint in the DT model changed because for blocklength BT we have

  $$\sum_{i=1}^{BT} |X_i|^2 = \|X(t)\|_2^2 \le PT,$$

  thus per-letter power constraint is $\frac{P}{B}$.

Calculate the rate of the equivalent DT AWGN channel and we are done. □

Note the above "theorem" is not rigorous, since conditions 1 and 2 are mutually exclusive: any time limited non-trivial signal cannot be band limited. Rigorously, one should relax 2 by constraining the signal to have a vanishing out-of-band energy as $T \to \infty$. Rigorous approach to this question lead to the theory of prolate spheroidal functions.

19.4 Capacity of the continuous-time band-unlimited AWGN channel

In the limit of large bandwidth B the capacity formula (19.8) yields

$$C_{B=\infty}(P) = \lim_{B\to\infty} B\log(1+\frac{P}{N_0 B}) = \frac{P}{N_0}\log e\,.$$

It turns out that this result is easy to prove rigorously.

Theorem 19.3. *Let $M^*(T,\epsilon,P)$ the maximum number of waveforms that can be sent through the channel*

$$Y(t) = X(t) + N(t)\,,\qquad \mathbb{E}\,N(t)N(s) = \frac{N_0}{2}\delta(t-s)$$

such that each waveform $x(t)$

1. *is non-zero only on $[0,T]$;*
2. *input energy constrained to $\int_{t=0}^{T} x^2(t) \le TP$;*
3. *error probability $P[\hat{W} \ne W] \le \epsilon$.*

Then

$$\lim_{\epsilon\to0}\liminf_{T\to\infty}\frac{1}{T}\log M^*(T,\epsilon,P) = \frac{P}{N_0}\log e \tag{19.9}$$

Proof. Note that the space of all square-integrable functions on $[0,T]$, denoted $L_2[0,T]$ has countable basis (e.g. sinusoids). Thus, by changing to that basis we may assume that equivalent channel model

$$\tilde{Y}_j = \tilde{X}_j + \tilde{Z}_j\,,\qquad \tilde{Z}_j \sim \mathcal{N}(0,\frac{N_0}{2})\,,$$

and energy constraint (dependent upon duration T):

$$\sum_{j=1}^{\infty}\tilde{X}_j^2 \le PT\,.$$

But then the problem is equivalent to energy-per-bit one and hence

$$\log_2 M^*(T,\epsilon,P) = k \iff E^*(k,\epsilon) = PT\,.$$

Thus,

$$\lim_{\epsilon\to0}\liminf_{n\to\infty}\frac{1}{T}\log_2 M^*(T,\epsilon,P) = \frac{P}{\lim_{\epsilon\to0}\limsup_{k\to\infty}\frac{E^*(k,\epsilon)}{k}} = \frac{P}{N_0}\log_2 e\,,$$

where the last step is by Theorem 19.1. □

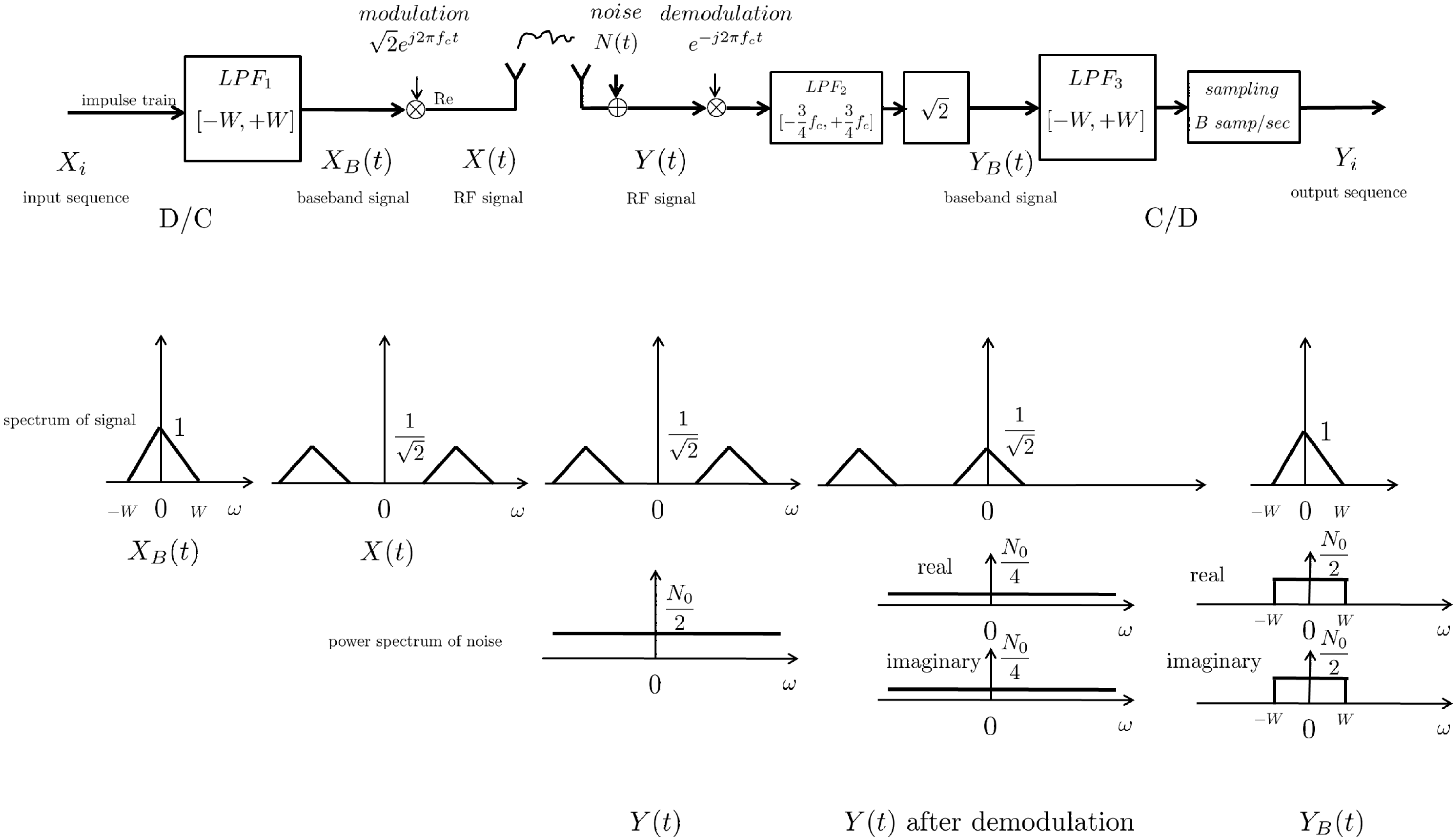


Figure 19.1: DT / CT AWGN model

19.5 Capacity per unit cost

Generalizing the energy-per-bit setting of Theorem 19.1 we get the problem of *capacity per unit cost*:

1. Given a random transformation $P_{Y^\infty|X^\infty}$ and cost function $\mathsf{c} : \mathcal{X} \to \mathbb{R}_+$, we let

$$M^*(E,\epsilon) = \max\{M : (E,M,\epsilon)\text{-code}\}\,,$$

   where (E,M,ϵ)-code is defined as a map $[M] \to \mathcal{X}^\infty$ with every codeword x^∞ satisfying

$$\sum_{t=1}^{\infty} \mathsf{c}(x_t) \le E\,. \tag{19.10}$$

2. Capacity per unit cost is defined as

$$C_{puc} \triangleq \lim_{\epsilon\to 0} \liminf_{E\to\infty} \frac{1}{E} \log M^*(E,\epsilon)\,.$$

3. Let $C(P)$ be the capacity-cost function of the channel (in the usual sense of capacity, as defined in (17.1). Assuming $P_0 = 0$ and $C(0) = 0$ it is not hard to show that:

$$C_{puc} = \sup_P \frac{C(P)}{P} = \lim_{P\to 0} \frac{C(P)}{P} = \left.\frac{d}{dP}\right|_{P=0} C(P)\,.$$

4. The surprising discovery of Verdú is that one can avoid computing $C(P)$ and derive the C_{puc} directly. This is a significant help, as for many practical channels $C(P)$ is unknown. Additionally, this gives a yet another fundamental meaning to KL-divergence.

Theorem 19.4. *For a stationary memoryless channel $P_{Y^\infty|X^\infty} = \prod P_{Y|X}$ with $P_0 = \mathsf{c}(x_0) = 0$ (i.e. there is a symbol of zero cost), we have*

$$C_{puc} = \sup_{x\ne x_0} \frac{D(P_{Y|X=x}\|P_{Y|X=x_0})}{\mathsf{c}(x)}\,.$$

In particular, $C_{puc} = \infty$ if there exists $x_1 \ne x_0$ with $\mathsf{c}(x_1) = 0$.

Proof. Let

$$C_V = \sup_{x\ne x_0} \frac{D(P_{Y|X=x}\|P_{Y|X=x_0})}{\mathsf{c}(x)}\,.$$

Converse: Consider a (E,M,ϵ) code $W \to X^\infty \to Y^\infty \to \hat{W}$. Introduce an auxiliary distribution $Q_{W,X^\infty,Y^\infty,\hat{W}}$, where a channel is a useless one

$$Q_{Y^\infty|X^\infty} = Q_{Y^\infty} \triangleq P^\infty_{Y|X=x_0}\,.$$

That is, the overall factorization is

$$Q_{W,X^\infty,Y^\infty,\hat{W}} = P_W P_{X^\infty|W} Q_{Y^\infty} P_{\hat{W}|Y^\infty}\,.$$

Then, as usual we have from the data-processing for divergence

$$
\begin{aligned}
(1-\epsilon)\log M + h(\epsilon) &\le d(1-\epsilon\|\frac{1}{M}) && (19.11)\\
&\le D(P_{W,X^\infty,Y^\infty,\hat W}\|Q_{W,X^\infty,Y^\infty,\hat W}) && (19.12)\\
&= D(P_{Y^\infty|X^\infty}\|Q_{Y^\infty}|P_{X^\infty}) && (19.13)\\
&= \mathbb{E}\left[\sum_{t=1}^{\infty} d(X_t)\right], && (19.14)
\end{aligned}
$$

where we denoted for convenience

$$d(x) \triangleq D(P_{Y|X=x}\|P_{Y|X=x_0}).$$

By the definition of C_V we have

$$d(x) \le \mathsf{c}(x) C_V .$$

Thus, continuing (19.14) we obtain

$$(1-\epsilon)\log M + h(\epsilon) \le C_V\,\mathbb{E}\left[\sum_{t=1}^{\infty} c(X_t)\right] \le C_V \cdot E\,,$$

where the last step is by the cost constraint (19.10). Thus, dividing by E and taking limits we get

$$C_{puc} \le C_V .$$

Achievability: We generalize the PPM code (19.3). For each $x_1 \in \mathcal{X}$ and $n \in \mathbb{Z}_+$ we define the encoder f as follows:

$$
\begin{aligned}
f(1) &= (\underbrace{x_1, x_1, \dots, x_1}_{n\text{-times}}, \underbrace{x_0, \dots, x_0}_{n(M-1)\text{-times}}) && (19.15)\\
f(2) &= (\underbrace{x_0, x_0, \dots, x_0}_{n\text{-times}}, \underbrace{x_1, \dots, x_1}_{n\text{-times}}, \underbrace{x_0, \dots, x_0}_{n(M-2)\text{-times}}) && (19.16)\\
&\cdots && (19.17)\\
f(M) &= (\underbrace{x_0, \dots, x_0}_{n(M-1)\text{-times}}, \underbrace{x_1, x_1, \dots, x_1}_{n\text{-times}}) && (19.18)
\end{aligned}
$$

Now, by Stein's lemma there exists a subset $S \subset \mathcal{Y}^n$ with the property that

$$
\begin{aligned}
&\mathbb{P}[Y^n \in S | X^n = (x_1, \dots, x_1)] \ge 1 - \epsilon_1 && (19.19)\\
&\mathbb{P}[Y^n \in S | X^n = (x_0, \dots, x_0)] \le \exp\{-nD(P_{Y|X=x_1}\|P_{Y|X=x_0}) + o(n)\}. && (19.20)
\end{aligned}
$$

Therefore, we propose the following (suboptimal!) decoder:

$$
\begin{aligned}
Y^n \in S \quad &\Longrightarrow \quad \hat W = 1 && (19.21)\\
Y_{n+1}^{2n} \in S \quad &\Longrightarrow \quad \hat W = 2 && (19.22)\\
&\cdots && (19.23)
\end{aligned}
$$

From the union bound we find that the overall probability of error is bounded by

$$\epsilon \le \epsilon_1 + M\exp\{-nD(P_{Y|X=x_1}\|P_{Y|X=x_0}) + o(n)\}.$$

At the same time the total cost of each codeword is given by $n\mathsf{c}(x_1)$. Thus, taking $n \to \infty$ and after straightforward manipulations, we conclude that

$$C_{puc} \ge \frac{D(P_{Y|X=x_1} \| P_{Y|X=x_0})}{\mathsf{c}(x_1)}.$$

This holds for any symbol $x_1 \in \mathcal{X}$, and so we are free to take supremum over x_1 to obtain $C_{puc} \ge C_V$, as required. □

19.5.1 Energy-per-bit for AWGN channel subject to fading

Consider a stationary memoryless Gaussian channel with fading H_j (unknown at the receiver). Namely,

$$Y_j = H_j X_j + Z_j, \qquad H_j \sim \mathcal{N}(0,1) \perp\!\!\!\perp Z_j \sim \mathcal{N}(0, \frac{N_0}{2}).$$

The cost function is the usual quadratic one $\mathsf{c}(x) = x^2$. As we discussed previously, cf. (17.8), the capacity-cost function $C(P)$ is unknown in closed form, but is known to behave drastically different from the case of non-fading AWGN (i.e. when $H_j = 1$). So here previous theorem comes handy, as we cannot just compute $C'(0)$. Let us perform a simple computation required, cf. (1.16):

$$C_{puc} = \sup_{x \neq 0} \frac{D(\mathcal{N}(0, x^2 + \frac{N_0}{2}) \| \mathcal{N}(0, \frac{N_0}{2}))}{x^2} \tag{19.24}$$

$$= \frac{1}{N_0} \sup_{x \neq 0} \left(\log e - \frac{\log(1 + \frac{2x^2}{N_0})}{\frac{2x^2}{N_0}} \right) \tag{19.25}$$

$$= \frac{\log e}{N_0} \tag{19.26}$$

Comparing with Theorem 19.1 we discover that surprisingly, the capacity-per-unit-cost is unaffected by the presence of fading. In other words, the random multiplicative noise which is so detrimental at high SNR, appears to be much more benign at low SNR (recall that $C_{puc} = C'(0)$). There is one important difference, however. It should be noted that the supremization over x in (19.25) is solved at $x = \infty$. Following the proof of the converse bound, we conclude that any code hoping to achieve optimal C_{puc} must satisfy a strange constraint:

$$\sum_t x_t^2 1\{|x_t| \ge A\} \approx \sum_t x_t^2 \qquad \forall A > 0$$

I.e. the total energy expended by each codeword must be almost entirely concentrated in very large spikes. Such a coding method is called "flash signalling". Thus, we can see that unlike non-fading AWGN (for which due to rotational symmetry all codewords can be made "mellow"), the only hope of achieving full C_{puc} in the presence of fading is by signalling in huge bursts of energy.

This effect manifests itself in the speed of convergence to C_{puc} with increasing constellation sizes. Namely, the energy-per-bit $\frac{E^*(k,\epsilon)}{k}$ behaves as

$$\frac{E^*(k,\epsilon)}{k} = (-1.59\ dB) + \sqrt{\frac{\text{const}}{k}} Q^{-1}(\epsilon) \qquad \text{(AWGN)} \tag{19.27}$$

$$\frac{E^*(k,\epsilon)}{k} = (-1.59\ dB) + \sqrt[3]{\frac{\log k}{k} (Q^{-1}(\epsilon))^2} \qquad \text{(fading)} \tag{19.28}$$

Fig. 19.2 shows numerical details.

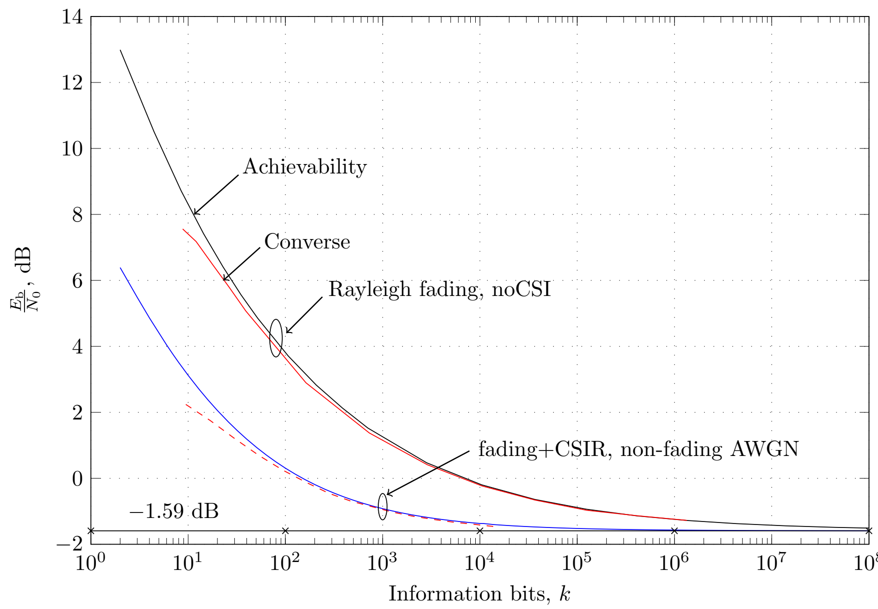


Figure 19.2: Comparing the energy-per-bit required to send a packet of k-bits for different channel models (curves represent upper and lower bounds on the unknown optimal value $\frac{E^*(k,\epsilon)}{k}$). As a comparison: to get to $-1.5\ dB$ one has to code over $6 \cdot 10^4$ data bits when the channel is non-fading AWGN or fading AWGN with H_j known perfectly at the receiver. For fading AWGN without knowledge of H_j (noCSI), one has to code over at least $7 \cdot 10^7$ data bits to get to the same $-1.5\ dB$. Plot generated via [Spe15].