EEE 551 Information Theory (Spring 2022)

Chapter 9: Gaussian Channel

A Bit of Digital Communications

- Many channels of interest can be modeled in continuous-time as

$$R(t) = S(t) + Z(t)$$

- $S(t)$ is the transmitted signal, subject to a power constraint

$$\frac{1}{T}\int_0^T S(t)^2\,dt \le P$$

- $R(t)$ is the received signal
- $Z(t)$ is a Gaussian white noise process, with autocorrelation function

$$R_Z(\tau) = \mathbb{E}[Z(t)Z(t-\tau)] = \frac{N_0}{2}\delta(\tau)$$

Binary Phase-Shift Keying (BPSK)

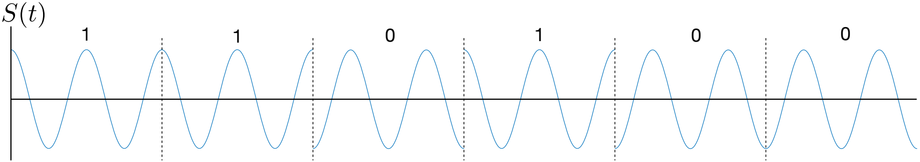
- Let $\phi_1(t)$ be a signal with unit energy: $\int_{-\infty}^{\infty} \phi_1(t)^2 \, dt = 1$

  Typically something like $\phi_1(t) = \begin{cases} \cos(\omega_0 t)/T, & 0 < t < T \\ 0, & \text{otherwise} \end{cases}$
- Given a sequence of bits $(a_1, a_2, \ldots, a_n) \in \{0, 1\}^n$, convert to real number

$$x_i = \sqrt{P}\,(2a_i - 1)$$

  and transmit the modulated signal

$$S(t) = \sum_{i=1}^{n} x_i \phi_1(t - iT)$$



- To recover the transmitted bits, use a matched filter:

$$Y_i = \int_{(i-1)T}^{iT} R(t)\phi_1(t - iT)dt = x_i + Z_i \quad \text{where } Z_i \sim \mathcal{N}\left(0, \frac{N_0}{2}\right)$$

- If we simply threshold Y_i, by taking $Y_i > 0$ as "1" and $Y_i < 0$ as "0", this is a BSC!

More general modulation schemes

- Let $\phi_2(t)$ be a signal with unit-energy that is orthogonal to $\phi_1(t)$:

$$\int_{-\infty}^{\infty} \phi_2(t)^2 \, dt = 1, \qquad \int_{-\infty}^{\infty} \phi_1(t)\phi_2(t) \, dt = 0.$$

For example, $\phi_2(t) = \begin{cases} \sin(\omega_0 t)/T, & 0 < t < T \\ 0, & \text{otherwise} \end{cases}$

- Map a sequence of bits $(a_1, a_2, \ldots, a_k) \in \{0,1\}^k$ to the "in phase" sequence

$$x_{11}, \ldots, x_{1n}$$

and the "quadrature phase" sequence

$$x_{21}, \ldots, x_{2n}$$

then transmit modulated signal

$$S(t) = \sum_{i=1}^{n} \left[x_{1i}\phi_1(t - iT) + x_{2i}\phi_2(t - iT) \right]$$

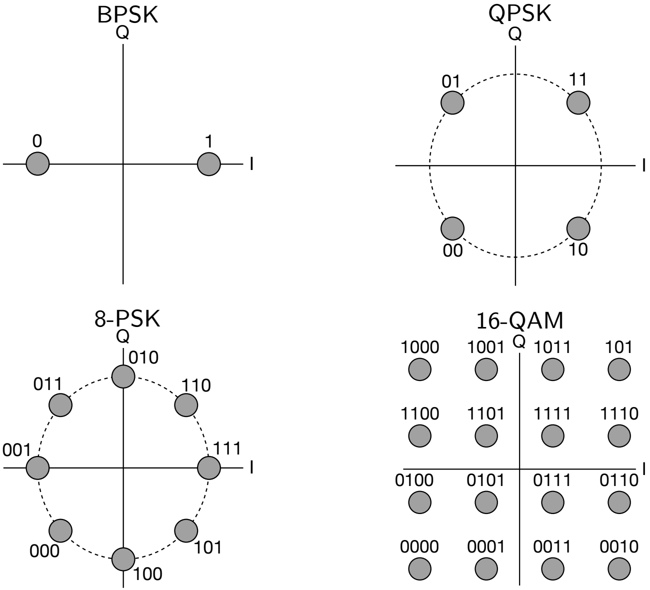
- Recover transmitted values, again using matched filters:

$$Y_{1i} = \int_{(i-1)T}^{iT} R(t)\phi_1(t - iT) \, dt = x_{1i} + Z_{1i}$$

$$Y_{2i} = \int_{(i-1)T}^{iT} R(t)\phi_2(t - iT) \, dt = x_{2i} + Z_{2i}$$

Constellation Diagrams

Constellation diagrams show how to map bits to x_1, x_2



Discrete Time Model for the Gaussian Channel

- We focus only on the **in phase** sequence

$$x_1, x_2, \ldots, x_n$$

The quadrature phase component is orthogonal, so we can consider them separately

- Power constraint is $\frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq P$
- Received signal (after modulation, noise, and matched filter) is

$$Y_i = x_i + Z_i$$

where $Z_i \sim \mathcal{N}(0, N)$

- We allow a generalized modulation strategy; i.e. messages can be mapped to x-sequences in an arbitrary manner

Capacity Definition for the Gaussian Channel

An (M, n) code for the Gaussian channel with power constraint P consists of

- A message set $\{1, 2, \ldots, M\}$
- An encoding function $x^n : \{1, 2, \ldots, M\} \to \mathbb{R}^n$ with codewords $x^n(1), x^n(2), \ldots, x^n(M)$ where for all messages m

$$\sum_{i=1}^{n} x_i(m)^2 \leq nP$$

- A decoding function $g : \mathbb{R}^n \to \{1, 2, \ldots, M\}$

A rate R is **achievable** if there exists a sequence of $(2^{nR}, n)$ codes satisfying the power constraint with maximal probability of error $\lambda^{(n)} \to 0$

The **capacity** C is the supremum of all achievable rates

A guess for the capacity

- Based on the discrete-variable result, we would guess the capacity is

$$\max_{f(x):\mathbb{E}X^2 \le P} I(X;Y)$$

- Here, the power constraint on the sequence X^n is transformed into a single-letter constraint $\mathbb{E}X^2 \le P$
- Expand the mutual information:

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(X+Z|X) \\ &= h(Y) - h(Z|X) \\ &= h(Y) - h(Z) \\ &= h(Y) - \frac{1}{2}\log 2\pi e N \end{aligned}$$

- $\mathbb{E}Y^2 = \mathbb{E}(X+Z)^2 = \mathbb{E}X^2 + \mathbb{E}Z^2 \le P + N$, so

$$h(Y) \le \frac{1}{2}\log 2\pi e(P+N) \quad \text{with equality if } X \sim \mathcal{N}(0,P)$$

- Therefore $\max I(X;Y) = \frac{1}{2}\log 2\pi e(P+N) - \frac{1}{2}\log 2\pi e N = \frac{1}{2}\log\left(1+\frac{P}{N}\right)$

  $P/N =$ signal-to-noise ratio (SNR)

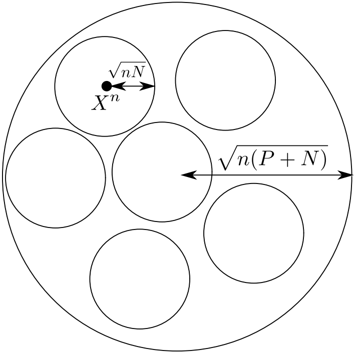
Capacity of the Gaussian Channel

Theorem

For the Gaussian channel, with power P and noise variance N,

$$C = \frac{1}{2}\left(1 + \frac{P}{N}\right)$$

Intuition: sphere packing



- $Y^n = X^n + Z^n$ is likely to be in sphere of radius $\sqrt{n(N+\epsilon)}$ around X^n
- Since X^n has power at most nP, Y^n has power (roughly) at most $n(P+N)$

$$\implies \quad \|Y^n\| \leq \sqrt{n(P+N)}$$

- Volume of n-dimensional sphere of radius r is $C_n r^n$
- Number of non-overlapping spheres that can be packed is at most

$$\frac{C_n \left[\sqrt{n(P+N)}\right]^n}{C_n \left[\sqrt{nN}\right]^n} = \left(\frac{P+N}{N}\right)^{n/2} = 2^{\frac{n}{2}\log(1+P/N)}$$

Achievability proof

Assume $R < \frac{1}{2}\log(1 + \frac{P}{N})$. We show there exists a sequence of codes with $\lambda^{(n)} \to 0$

Random codebook generation

For each $m \in \{1, 2, \ldots, 2^{nR}\}$, generate $X^n(m) \overset{\text{iid}}{\sim} \mathcal{N}(0, P - \epsilon)$

Encoding

Given m, if $\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i(m)^2 \le P$, then send $X^n(m)$, otherwise send 0^n

Decoding

Given Y^n, select the smallest m such that $(X^n(m), Y^n) \in A_\epsilon^{(n)}$.
If none, declare an error

Probability of error analysis

- Let $\bar{P}_e$ be the probability of error averaged over the random choice of codebook
- Let B_m be the probability of error given $W = m$, so $\bar{P}_e = \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} B_m$
- Define events:

$$\mathcal{E}_{0m} = \left\{ \frac{1}{n} \sum_{i=1}^{n} X_i(m)^2 > P \right\}$$

$$\mathcal{E}_{1m} = \left\{ (X^n(m), Y^n) \in A_\epsilon^{(n)} \right\}$$

- Given message m, an error occurs only if $\mathcal{E}_{0m}$, $\mathcal{E}_{1m}^c$, or $\mathcal{E}_{1m'}$ for $m' \neq m$

$$B_m \leq \Pr\{\mathcal{E}_{0m}\} + \Pr\{\mathcal{E}_{1m}^c\} + \sum_{m' \neq m} \Pr\{\mathcal{E}_{1m'}\}$$

By the law of large numbers, $\frac{1}{n}\sum_{i=1}^{n} X_i(m)^2 \to P - \epsilon$, so $\Pr\{\mathcal{E}_{0m}\} \to 0$

By joint AEP, $\Pr\{\mathcal{E}_{1m}^c\} \to 0$

By joint AEP, $\Pr\{\mathcal{E}_{1m'}\} \leq 2^{-n(I(X;Y)-3\epsilon)}$

- For sufficiently large n,

$$B_m \leq 2\epsilon + 2^{nR} 2^{-n(I(X;Y)-3\epsilon)}$$

- Recall $X \sim \mathcal{N}(0, P-\epsilon)$, so $I(X;Y)$ is arbitrarily close to $\frac{1}{2}\log(1+\frac{P}{N})$ for sufficiently small ϵ
- Since $R < \frac{1}{2}\log(1+\frac{P}{N})$ the third term vanishes with n
- Thus there exists at least one code with small avg. probability of error
- Delete the worst half of codewords to get code with small max. probability of error

Converse Proof

Assume there exists a sequence of $(2^{nR}, n)$ codes with $\sum_{i=1}^{n} x_i(m)^2 \leq nP$ and avg. probability of error $P_e^{(n)} \to 0$. We want to prove $R \leq \frac{1}{2}\log(1+\frac{P}{N})$

$$
\begin{aligned}
nR = H(W) &= I(W;Y^n) + H(W|Y^n) \\
&\leq I(W;Y^n) + n\epsilon_n && \text{Fano's inequality} \\
&\leq I(X^n;Y^n) + n\epsilon_n \\
&\leq \sum_{i=1}^{n} I(X_i;Y_i) + n\epsilon_n && \text{usual argument} \\
&\leq \sum_{i=1}^{n} \frac{1}{2}\log\left(1+\frac{P_i}{N}\right) + n\epsilon_n
\end{aligned}
$$

where $P_i = \mathbb{E}X_i^2 = \frac{1}{2^{nR}} \sum_{m=1}^{2^{nR}} x_i(m)^2$

$$\frac{1}{n}\sum_{i=1}^{n} P_i = \frac{1}{2^{nR}}\sum_{m=1}^{2^{nR}} \frac{1}{n}\sum_{i=1}^{n} x_i(m)^2 \leq \frac{1}{2^{nR}}\sum_{m} P = P$$

Therefore

$$\begin{aligned}
R &\leq \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\log\left(1+\frac{P_i}{N}\right)+\epsilon_n \\
&\leq \frac{1}{2}\log\left(1+\frac{1}{n}\sum_{i=1}^{n}\frac{P_i}{N}\right)+\epsilon_n \qquad \text{concavity of log} \\
&\leq \frac{1}{2}\log\left(1+\frac{P}{N}\right)+\epsilon_n
\end{aligned}$$

Channel coding with cost constraint

- The power constraint is a special case of a **cost constraint** in channel coding
- Consider a memoryless channel $(\mathcal{X}, p(y|x), \mathcal{Y})$ (not necessarily discrete)
- A cost function is given by $b : \mathcal{X} \to \mathbb{R}$
- An encoding function

  $$x^n : \{1, 2, \dots, M\} \to \mathcal{X}^n$$

  satisfies cost constraint B if

  $$\frac{1}{n} \sum_{i=1}^{n} b\big(x_i(m)\big) \leq B \quad \text{for all } m$$

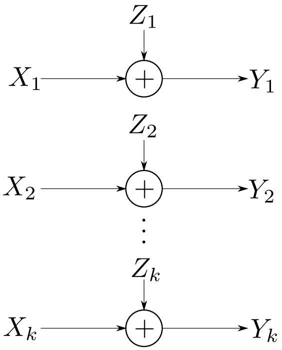
- Define the **capacity-cost** function $C(B)$ as the supremum of achievable rates with encoders satisfying the cost constraint B
- **Example:** Binary-input channel where 1 is more difficult to send than 0

Theorem

The capacity-cost function is

$$C(B) = \max_{p(x):\mathbb{E}[b(X)] \leq B} I(X;Y).$$

Parallel Gaussian channels



- $Z_j \sim \mathcal{N}(0, N_j)$ for $j = 1, \dots, k$ (independent)
- Encoder chooses $(X_{1i}, X_{2i}, \dots, X_{ki})$, decoder receives $(Y_{1i}, Y_{2i}, \dots, Y_{ki})$ at times $i = 1, \dots, n$
- Overall power constraint

$$\sum_{i=1}^{n} \sum_{j=1}^{k} X_{ji}^2 \leq nP$$

This is another cost-constrained channel, so the capacity is

$$C = \max_{f(x_1, \dots, x_k) : \mathbb{E}\left[\sum_{j=1}^{k} X_j^2\right] \leq P} I(X_1, \dots, X_k; Y_1, \dots, Y_k)$$

- Assuming $\mathbb{E}\left[\sum_{j=1}^{k} X_j^2\right] \leq P$,

$$\begin{aligned}
I(X_1, \ldots, X_k; Y_1, \ldots, Y_k) &= h(Y_1, \ldots, Y_k) - h(Y_1, \ldots, Y_k | X_1, \ldots, X_k) \\
&= h(Y_1, \ldots, Y_k) - h(Z_1, \ldots, Z_k) \\
&= h(Y_1, \ldots, Y_k) - \sum_{j=1}^{k} h(Z_j) \\
&\leq \sum_{j=1}^{k} [h(Y_j) - h(Z_j)] \\
&\leq \sum_{j=1}^{k} \frac{1}{2} \log\left(1 + \frac{P_j}{N_j}\right)
\end{aligned}$$

  where $P_j = \mathbb{E}X_j^2$, and $\sum_{j=1}^{k} P_j \leq P$
- Equality above achieved if $(X_1, \ldots, X_k)$ are independent and $X_j \sim \mathcal{N}(0, P_j)$
- Therefore

$$C = \max_{P_1, \ldots, P_k : \sum_{j=1}^{k} P_j \leq P} \sum_{j=1}^{k} \frac{1}{2} \log\left(1 + \frac{P_j}{N_j}\right)$$

Need to solve the constrained convex optimization problem

$$\begin{aligned} \text{maximize} \quad & \sum_{j=1}^{k} \frac{1}{2} \log\left(1 + \frac{P_j}{N_j}\right) \\ \text{subject to} \quad & \sum_{j=1}^{k} P_j = P \\ & P_j \geq 0 \text{ for } j = 1, \ldots, k \end{aligned}$$

Karush-Kuhn-Tucker (KKT) conditions

Consider the generic optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

If $x^\star$ is an optimal point, then there exist $\lambda^\star \in \mathbb{R}^m$ and $\nu^\star \in \mathbb{R}^p$ where

$$\begin{aligned} \nabla\left[f(x^\star) + \sum_{i=1}^m \lambda_i^\star g_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star)\right] &= 0 && \text{(Lagrangian)} \\ g_i(x^\star) \leq 0, \quad & i = 1, \ldots, m \\ h_i(x^\star) = 0, \quad & i = 1, \ldots, p \\ \lambda_i^\star \geq 0, \quad & i = 1, \ldots, m \\ \lambda_i^\star g_i(x^\star) = 0, \quad & i = 1, \ldots, m && \text{(complementary slackness)} \end{aligned}$$

If the problem is convex, then the above conditions are necessary and sufficient (under mild regularity conditions)

Writing our problem in the standard form:

$$
\begin{array}{ll}
\text{minimize} & -\displaystyle\sum_{j=1}^{k} \ln\left(1+\frac{P_j}{N_j}\right) \\
\text{subject to} & -P_j \le 0, \quad j=1,\dots,k \\
& \displaystyle\sum_{j=1}^{k} P_j - P = 0
\end{array}
$$

The KKT conditions are

$$
\begin{aligned}
\nabla\Bigg[-\sum_{j=1}^{k} \ln\left(1+\frac{P_j^\star}{N_j}\right) - \sum_{j=1}^{k}\lambda_j^\star P_j^\star + \nu^\star\left(\sum_{j=1}^{k} P_j^\star - P\right)\Bigg] &= 0, \\
-P_j^\star \le 0, \quad j &= 1,\dots,k, \\
\sum_{j=1}^{k} P_j^\star - P = 0,& \\
\lambda_j^\star \ge 0, \quad j &= 1,\dots,k, \\
\lambda_j^\star P_j^\star = 0, \quad j &= 1,\dots,k
\end{aligned}
$$

Differentiating the Lagrangian with respect to $P_j^\star$:

$$
-\frac{1/N_j}{1+P_j^\star/N_j} - \lambda_j^\star + \nu^\star = 0 \qquad \Longrightarrow \qquad P_j^\star = \frac{1}{\nu^\star - \lambda_j^\star} - N_j
$$

$$P_j^\star = \frac{1}{\nu^\star - \lambda_j^\star} - N_j$$

- If $P_j^\star > 0$, then by complementary slackness $\lambda_j^\star = 0$, so $P_j^\star = 1/\nu^\star - N_j$
- Define $\alpha^\star = 1/\nu^\star$
- Thus either $P_j^\star = 0$ or $P_j^\star = \alpha^\star - N_j$
- Since $\lambda_j^\star \geq 0$, $P_j^\star \geq \alpha^\star - N_j$, meaning that if $\alpha^\star - N_j > 0$, then $P_j^\star = \alpha^\star - N_j$
- Therefore $P_j^\star = (\alpha^\star - N_j)^+$, where

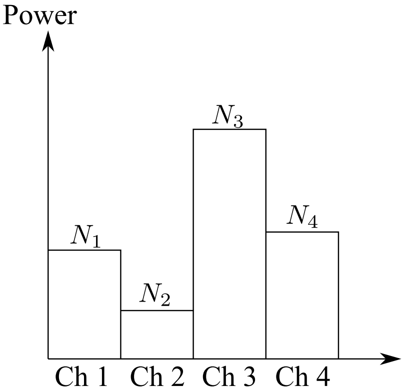
$$(x)^+ = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

and $\alpha^\star$ is selected so that

$$\sum_{j=1}^{k} (\alpha^\star - N_j)^+ = P$$

Water Filling

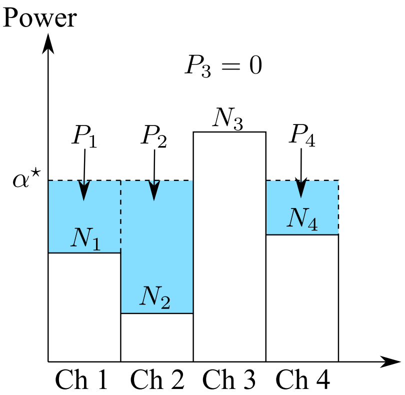
$P_j = (\alpha^\star - N_j)^+$



To maximize rate, allocate lower power (or no power) to noisier channels

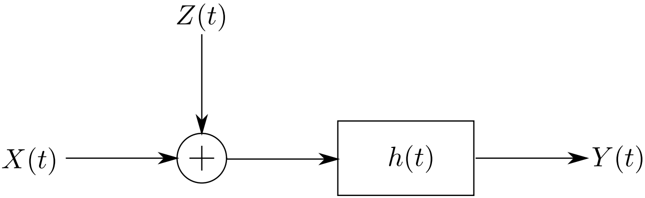
Water Filling

$P_j = (\alpha^\star - N_j)^+$



To maximize rate, allocate lower power (or no power) to noisier channels

Bandlimited Gaussian channels



- $Z(t)$ is white Gaussian noise process with power spectral density $\frac{N_0}{2}$ watts/hertz (i.e. PSD $= \frac{N_0}{2}\delta(t)$)
- $h(t)$ is ideal lowpass filter with bandwidth W
- An encoder is given by a function

$$x : \{1, \ldots, M\} \times [0, T] \to \mathbb{R}$$

  where $x(m, t)$ for $t \in [0, T]$ is the "codeword" for message m
- Power constraint

$$\frac{1}{T} \int_0^T x(m, t)^2 \, dt \leq P$$

- Capacity is the supremum of achievable rates in bits per second for arbitrarily large T

Outline of capacity derivation

- Assume $X(t)$ is bandlimited to W, so by Nyquist theorem $2W$ samples per second are required to describe $X(t)$
- There is a mapping between bandlimited signals $X(t)$ and sequences $X_1, \ldots, X_n$ where $n = 2WT$
- Power constraint $\int_0^T X(t)^2 dt \leq PT$ is equivalent to

$$\sum_{i=1}^{n} X_i^2 \leq PT \quad \Longleftrightarrow \quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 \leq \frac{PT}{n} = \frac{PT}{2WT} = \frac{P}{2W}$$

- Similar mapping between $Y(t)$ and $Y_1, \ldots, Y_n$, where

$$Y_i = X_i + Z_i \qquad Z_i \overset{\text{iid}}{\sim} \mathcal{N}(0, N_0/2)$$

- By discrete-time results, the number of bits that can be sent is roughly

$$\frac{n}{2}\log\left(1 + \frac{P/(2W)}{N_0/2}\right) = WT\log\left(1 + \frac{P}{WN_0}\right)$$

- Channel capacity in bits per second is

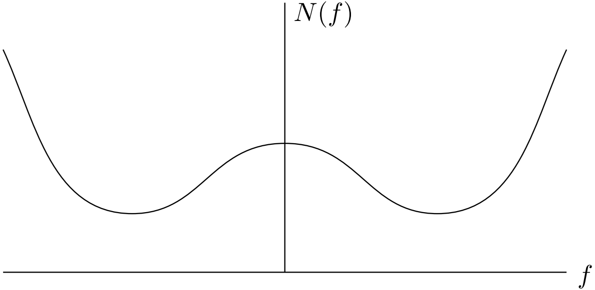
$$C = W\log\left(1 + \frac{P}{WN_0}\right) = W\log\left(1 + \frac{P}{N}\right)$$

  where $N = WN_0$ is the noise power
- The capacity in "bits per second per Hertz" is $\log\left(1 + \frac{P}{N}\right)$

$$C = W \log \frac{P+N}{N}$$

Channels with Colored Gaussian Noise

$Y(t) = X(t) + Z(t)$,
where $Z(t)$ is a Gaussian process with power spectral density $N(f)$

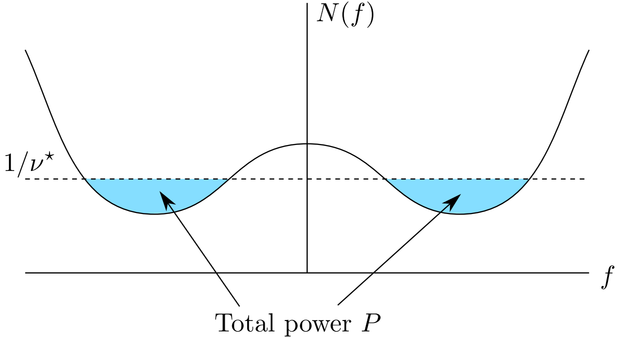


Again, a continuous-time power constraint

$$\frac{1}{T}\int_0^T x(t)^2\,dt \le P$$

Question: How to allocate power across frequency range?

Continuous-Time Water Filling



$$C = \int_{-\infty}^{\infty} \frac{1}{2} \log\left(1 + \frac{\left(1/\nu^\star - N(f)\right)^+}{N(f)}\right) df$$

where water level $1/\nu^\star$ is chosen so that

$$\int_{-\infty}^{\infty} \left(1/\nu^\star - N(f)\right)^+ df = P$$