Fundamentals of Digital Communication

by Upamanyu Madhow

Cambridge University Press, 2008

Lecture Outline for Chapter 3

Introduction

- We know now that information is conveyed by sending one of a selected set of signals
- Must decide based on received signal which signal was sent
  - This provides estimates for the transmitted information
  - Errors can occur due to noise and channel distortions
- In this chapter: framework for receiver design under idealized conditions
  - Focus on the effect of noise: Additive White Gaussian Noise (AWGN) channel
  - Ignore channel distortions
  - Assume ideal carrier and timing synchronization (*coherent* reception)
- **Despite idealizations, the material in this chapter is the core foundation of communication theory**
- Broad goals for this chapter
  - AWGN model
  - Optimal receiver design based on detection theory
  - Performance analysis

Outline

- Gaussian random variables, vectors and processes
- AWGN model
- Detection theory basics
- Signal space concepts
- Optimal reception in AWGN
- Performance analysis of optimal receiver
- Link budget analysis

Gaussian Basics

The ubiquity of Gaussian random variables

- Gaussian random variables arise often because of the central limit theorem (CLT)
  - Sum of i.i.d. random variables with finite variance is approximately Gaussian
  - CLT holds in much greater generality (dependencies should be weak, no one random variable should be dominant)
  - CLT kicks in quickly (e.g., 6-10 terms in sum)
- Gaussian noise arises from random movements of many charge carriers
- Gaussian impulse responses on wireless channels arise due to addition of many multipath components
- Gaussian approximation for sum of residual intersymbol interference after equalization

Gaussian Density

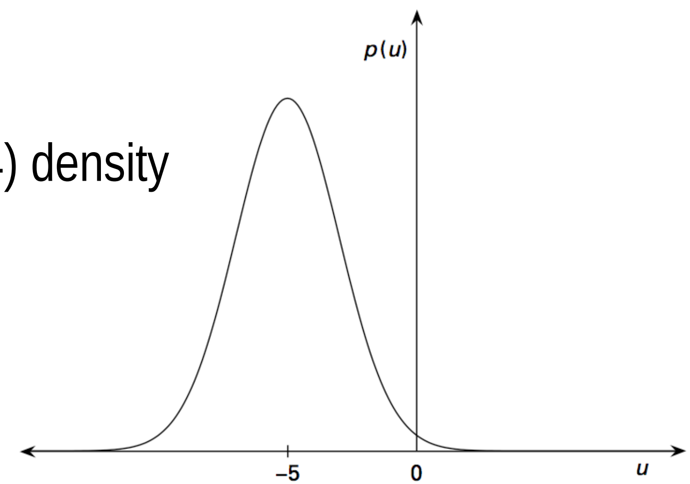
Characterized by 2 parameters, mean and variance
We denote that *X* has a Gaussian distribution with mean *m* and variance v^2 as follows:

$$X \sim N(m, v^2)$$

Density: $$p(x) = \frac{1}{\sqrt{2\pi v^2}} \exp\left(-\frac{(x-m)^2}{2v^2}\right)$$

The famous bell shape around the mean:

N(-5,4) density



Standard Gaussian Random Variable

Standard Gaussian random variable has mean zero and variance one

$$X \sim N(0, 1)$$

Gaussian random variables stay Gaussian under affine transformation

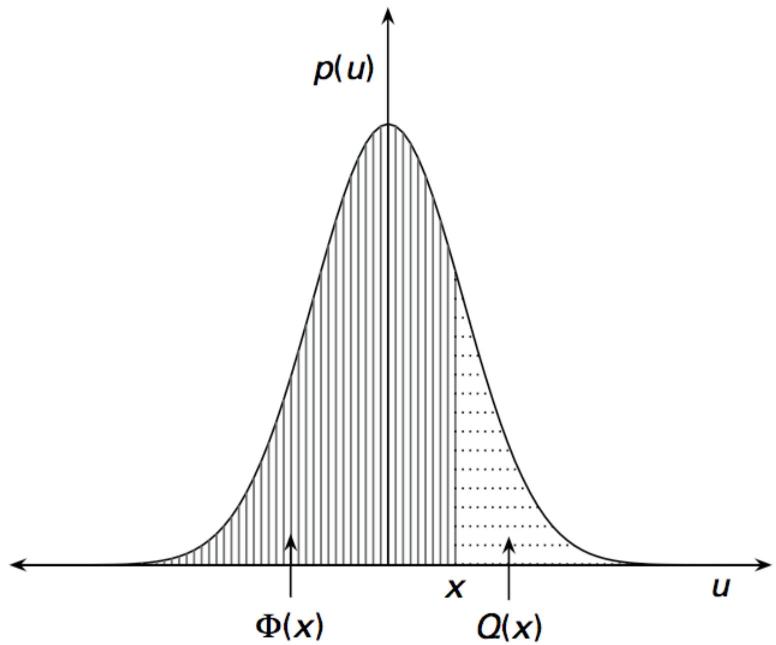
Thus, any Gaussian random variable can be converted to a standard Gaussian using an affine normalization:

$$X \sim N(m, v^2) \longrightarrow (X - m)/v \sim N(0, 1)$$

Can use this transformation to express probabilities for arbitrary Gaussian random variables compactly in terms of standard Gaussian probabilities

Q and Phi Functions

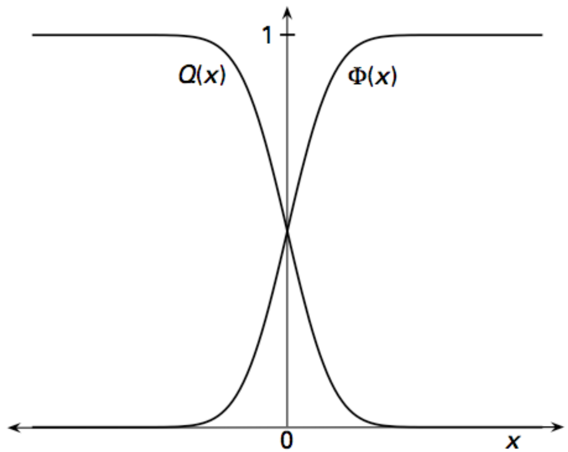
Phi function: CDF of standard Gaussian
Q function: Complementary CDF of standard Gaussian



$$\Phi(x) = P[N(0,1) \le x] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)\,\mathrm{d}t,$$

$$Q(x) = P[N(0,1) > x] = \int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)\,\mathrm{d}t.$$

Q and Phi function properties



By the symmetry of the *N(0,1)* density, we have: $Q(-x) = \Phi(x)$

From their definitions, we have: $Q(x) = 1 - \Phi(x)$

Combining, we get: $Q(-x) = 1 - Q(x)$

Bottomline: we can express Q and Phi functions with any real-valued arguments in terms of the Q function with positive arguments alone. Simplifies computation, enables use of bounds and approximations for Q functions with positive arguments.

Example: Expressing Gaussian probabilities in terms of the Q function

Example:

$$X \sim N(-3,4) \longrightarrow (X-m)/v = (X+3)/\sqrt{4} = (X+3)/2 \sim N(0,1)$$

We can therefore express probabilities involving *X* in terms of the *Q* function:

$$P[X > 5] = P\left[\frac{X+3}{2} > \frac{5+3}{2} = 4\right] = Q(4),$$

$$P[X < -1] = P\left[\frac{X+3}{2} < \frac{-1+3}{2} = 1\right] = \Phi(1) = 1 - Q(1)$$

$$P[1 < X < 4] = P\left[2 = \frac{1+3}{2} < \frac{X+3}{2} < \frac{4+3}{2} = 3.5\right] = \Phi(3.5) - \Phi(2)$$

$$= Q(2) - Q(3.5).$$

Example (contd.)

$$P[X^2 + X > 2] = P[X^2 + X - 2 > 0] = P[(X+2)(X-1) > 0].$$

We can break the event involving a quadratic into events involving intervals
The event occurs if and only if X < -2 or X > 1, so its probability can be expressed in terms of the Q function:

$$P[X^2 + X > 2] = P[X > 1] + P[X < -2] = Q\left(\frac{1+3}{2}\right) + \Phi\left(\frac{-2+3}{2}\right)$$

$$= Q(2) + \Phi\left(\frac{1}{2}\right) = Q(2) + 1 - Q\left(\frac{1}{2}\right).$$

Even if the event is too complicated to be expressed as a disjoint union of intervals, normalizations such as the one to standard Gaussian are useful in keeping our thinking clear on the roles of various parameters in system performance. For example, we shall see that SNR and geometry of signal constellation govern performance over the AWGN channel.

Q function has exponentially decaying tails

Asymptotically tight bounds for large arguments

$$\left(1-\frac{1}{x^2}\right)\frac{e^{-x^2/2}}{x\sqrt{2\pi}} \le Q(x) \le \frac{e^{-x^2/2}}{x\sqrt{2\pi}}, \qquad x \ge 0. \qquad (3.4)$$

Very Important Conclusion: **The asymptotic behavior of the Q function is**

$$\boxed{Q(x) \doteq e^{-x^2/2}, \qquad x \to \infty,}$$ **which is shorthand for** $$\lim_{x\to\infty} \frac{\log Q(x)}{-x^2/2} = 1$$

The rapid decay of the Q function implies, for example, that: $Q(1) + Q(4) \approx Q(1)$

Design implications: We will use this to identify dominant events causing errors

Useful bound for analysis (and works for small arguments):

$$Q(x) \le \frac{1}{2}e^{-x^2/2}, \qquad x \ge 0. \qquad (3.5)$$

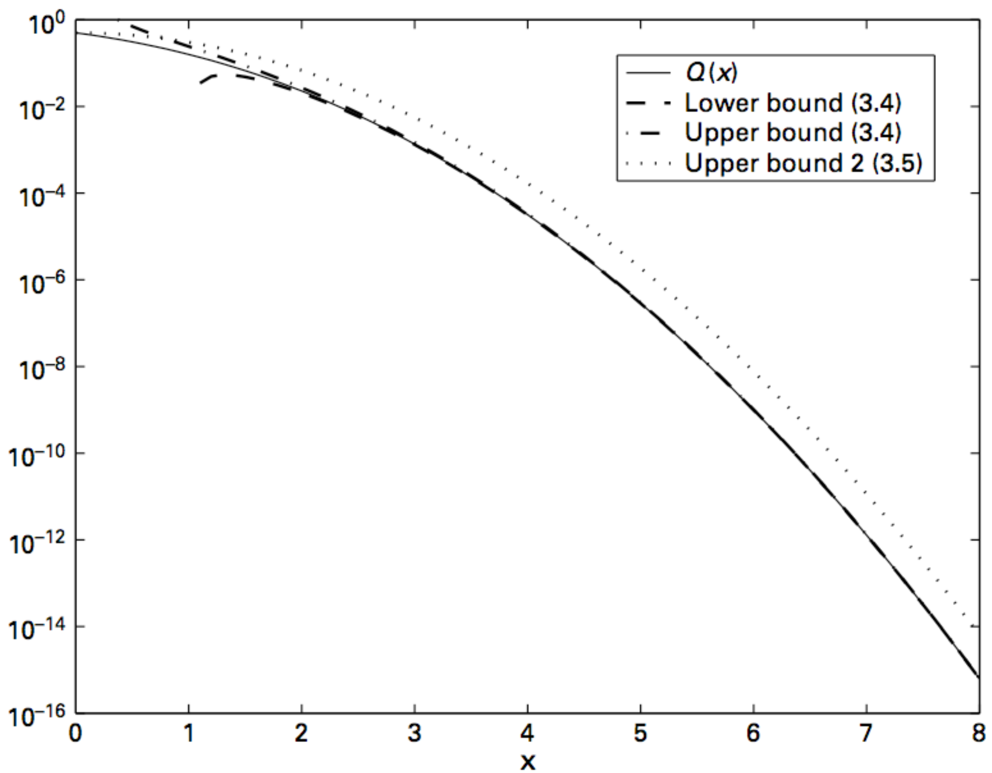
Plots of Q function and its bounds

$$\left(1-\frac{1}{x^2}\right)\frac{e^{-x^2/2}}{x\sqrt{2\pi}} \leq Q(x) \leq \frac{e^{-x^2/2}}{x\sqrt{2\pi}}\,, \qquad x \geq 0. \qquad (3.4)$$

Asymptotically tight

$$Q(x) \leq \frac{1}{2}e^{-x^2/2}\,, \qquad x \geq 0. \qquad (3.5)$$

Useful for analysis and for small arguments



Note the rapid decay: y-axis has log scale

Jointly Gaussian random variables (or Gaussian random vectors)

- Multiple random variables defined on a common probability space are also called random vectors
  - "same probability space" means we can talk about joint distributions
- A random vector is Gaussian (or the random variables concerned are *jointly* Gaussian) if any linear combination is a Gaussian random variable
- These arise naturally when we manipulate Gaussian noise
  - Correlation of Gaussian noise with multiple "templates"
  - Multiple samples of filtered Gaussian noise
- Joint distribution characterized by mean vector and covariance matrix
  - Analogous to mean and variance for scalar Gaussian

Mean vector and covariance matrix

Let's first review these for arbitrary random vectors

$$\mathbf{X} = (X_1, \ldots, X_m)^T$$

m x 1 random vector

Mean Vector (m x 1)

$$\mathbf{m}_X = \mathbb{E}[\mathbf{X}] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_m])^T.$$

Covariance Matrix (m x m)

$$\begin{aligned} \mathbf{C}_X(i,j) = \operatorname{cov}(X_i, X_j) &= \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] \\ &= \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]. \end{aligned}$$

(i,j)th entry

$$\mathbf{C}_X = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T] = \mathbb{E}[\mathbf{X}\mathbf{X}^T] - \mathbb{E}[\mathbf{X}](\mathbb{E}[\mathbf{X}])^T$$

Compact representation

Properties of Covariance

Covariance unaffected when we add constants

$$\mathrm{cov}(X+a, Y+b) = \mathrm{cov}(X, Y)$$

Adding constants changes the mean but not the covariance.
So we can always consider zero mean versions of random variables when computing covariance**.**
Common scenario: Mean is due to signal, covariance is due to noise. So we can often ignore the signal when computing covariance.

Covariance is a bilinear function (i.e., multiplicative constants pull out)

$$\begin{aligned}\mathrm{cov}(a_1X_1 + a_2X_2, a_3X_3 + a_4X_4) &= a_1a_3\mathrm{cov}(X_1, X_3) + a_1a_4\mathrm{cov}(X_1, X_4)\\ &\quad + a_2a_3\mathrm{cov}(X_2, X_3) + a_2a_4\mathrm{cov}(X_2, X_4)\end{aligned}$$

Quantities related to covariance

Variance of a random variable is its covariance with itself

$$\mathrm{var}(X) = \mathrm{cov}(X, X)$$

Correlation coefficient is the normalized covariance

$$\rho(X_1, X_2) = \frac{\mathrm{cov}(X_1, X_2)}{\sqrt{\mathrm{var}(X_1)\mathrm{var}(X_2)}}$$

$|\rho(X_1, X_2)| \leq 1$, with equality if and only if $X_2 = aX_1 + b$

(using Cauchy-Schwartz for random variables)

Mean and covariance evolve separately under affine transformations

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$$

Mean of Y depends only on the mean of X

$$\mathbf{m}_Y = \mathbb{E}[\mathbf{Y}] = \mathbb{E}[\mathbf{A}\mathbf{X} + \mathbf{b}] = \mathbf{A}\mathbb{E}[\mathbf{X}] + \mathbf{b} = \mathbf{A}\mathbf{m}_{\mathbf{X}} + \mathbf{b}$$

Covariance of Y depends only on the covariance of X (and does not depend on the additive constant b)

$$\mathbf{C}_Y = \mathbb{E}[(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])(\mathbf{Y} - \mathbb{E}[\mathbf{Y}])^T] = \mathbb{E}[\mathbf{A}(\mathbf{X} - \mathbf{m}_{\mathbf{X}})(\mathbf{X} - \mathbf{m}_{\mathbf{X}})^T \mathbf{A}^T] = \mathbf{A}\mathbf{C}_{\mathbf{X}}\mathbf{A}^T$$

Back to Gaussian random vectors

$$\mathbf{X} = (X_1, \ldots, X_m)^T$$

is a Gaussian random vector if *any* linear combination

$$a_1 X_1 + \cdots + a_m X_m$$ is a Gaussian random variable

A Gaussian random vector is completely characterized bv its mean vector and covariance matrix. Notation: $\mathbf{X} \sim N(\mathbf{m}, \mathbf{C})$

Why? Consider the characteristic function of X (which specifies its distribution)

$$\Phi_{\mathbf{X}}(u) = E[e^{j(u_1 X_1 + \ldots + u_n X_n)}] = E[e^{j\mathbf{u}^T\mathbf{X}}]$$

But the linear combination $\mathbf{u}^T\mathbf{X} \sim N(\mathbf{u}^T\mathbf{m}_{\mathbf{X}}, \mathbf{u}^T\mathbf{C}_{\mathbf{X}}\mathbf{u})$ is a Gaussian random variable whose distribution depends only on the mean vector and covariance matrix of X

Thus, the characteristic function, and hence the distribution, of X depends only on its mean vector and covariance matrix.

Joint Gaussian Density

Exists if and only if covariance matrix is invertible. If so, is given by:

$$p(x_1, \ldots, x_m) = p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^m |\mathbf{C}|}} \exp\left(-\frac{1}{2} (\mathbf{x} - \mathbf{m})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m}) \right)$$

How would we compute expectation of f(*X*), where *X* is a Gaussian random vector?

We usually avoid integrating over multiple dimensions. Instead use Monte Carlo simulations.

--start with samples of independent N(0,1) random variables

--transform to random vector with desired joint Gaussian stats (how?)

--evaluate function

--average over runs

Often we deal with a linear combination (e.g., sample at output of a filter), which are simply scalar Gaussian, so we do not need multidimensional integration.

Independence and Uncorrelatedness

Two random variables are uncorrelated if their covariance is zero.

Independent random variables are uncorrelated

$$\mathrm{cov}(X_1, X_2) = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = 0$$

Uncorrelated random variables are not necessarily independent

Think of examples!

Uncorrelated, jointly Gaussian, random variables are independent

Diagonal covariance matrix means inverse is also diagonal,
and joint density decomposes into product of marginals.

Joint Gaussianity preserved under affine transformations

$$\mathbf{X} \sim N(\mathbf{m}, \mathbf{C}) \longrightarrow \mathbf{Y} = \mathbf{AX} + \mathbf{b} \sim N(\mathbf{Am} + \mathbf{b}, \mathbf{A}^T\mathbf{CA})$$

Why?

Mean vector and covariance matrix are computed as for arbitrary random vectors

Any linear combination of components of Y can be expressed as linear combinations of components of X, and hence is a Gaussian random variable. Thus, Y inherits the Gaussianity of X.

In-class exercise: X_1 and X_2 are jointly Gaussian with $E[X_1] = -3$, $E[X_2] = 5$, $\mathrm{cov}(X_1,X_2) = -1$, $\mathrm{var}(X_1) = 1$, $\mathrm{var}(X_2) = 9$. Let $Y = 2X_1 + X_2$.

- Find $P[Y < 3]$ in terms of the Q function with positive arguments.
- Let $Z = X_1 + a X_2$ .Find the constant a such that Z is independent of Y..

In-class exercise

X_1 and X_2 are jointly Gaussian with $E[X_1] = -3$, $E[X_2] = 5$, $cov(X_1,X_2) = -1$, $var(X_1) = 1$, $var(X_2) = 9$. Let $Y = 2X_1 + X_2$.

- Find $P[Y < 3]$ in terms of the Q function with positive arguments.
- Let $Z = X_1 + a X_2$ .Find the constant a such that Z is independent of Y.

Answer, part (a):

Y is Gaussian with mean

$E[Y] = 2E[X_1] + E[X_2] = -1$.

$Var(Y)=cov(2X_1+ X_2 ,2X_1+ X_2)=4\,cov(X_1,X_1) + 4\,cov(X_1,X_2) + cov(X_2,X_2)$

$=4(1)+4(-1)+9=9$.

We therefore obtain

$$P[Y < 3] = \Phi\left(\frac{3-(-1)}{\sqrt{9}}\right) = \Phi(4/3) = 1 - Q(4/3)$$

In-class exercise (contd.)

X_1 and X_2 are jointly Gaussian with $E[X_1] = -3$, $E[X_2] = 5$, $\text{cov}(X_1,X_2) = -1$, $\text{var}(X_1) = 1$, $\text{var}(X_2) = 9$. Let $Y = 2X_1 + X_2$.

- Find $P[Y < 3]$ in terms of the Q function with positive arguments.
- Let $Z = X_1 + a X_2$. Find the constant a such that Z is independent of Y.

Answer, part (b):

Y and Z are jointly Gaussian, since they are linear combinations of jointly Gaussian random variables. Thus, they are independent if they are uncorrelated. Computing the covariance,

$$\begin{aligned}\text{cov}(Y,Z) &= \text{cov}(2X_1 + X_2, X_1 + aX_2)\\ &= 2\,\text{cov}(X_1,X_1) + (2a+1)\text{cov}(X_1,X_2) + a\,\text{cov}(X_2,X_2)\\ &= 2(1) + (2a+1)(-1) + 9a = 7a+1\end{aligned}$$

we see that it vanishes for a=-(1/7).

Gaussian random process

Random process: collection of random variables on a common probability space.
(simple generalization of random vectors: instead of the number of random variables being finite, we can have countably or uncountably many of them.)

Gaussian random process: any linear combination of samples is a Gaussian random variable.

$X = \{X(t), t \in T\}$ is a Gaussian random process

$\longleftrightarrow$ $a_1 X(t_1) + \cdots + a_n X(t_n)$ is a Gaussian random variable

for ***any*** choice of number of samples, sampling times and combining coefficients

$\longleftrightarrow$ $X(t_1), \ldots, X(t_n)$ are jointly Gaussian

for ***any*** choice of number of samples, sampling times and combining coefficients

Characterizing a Gaussian random process

Statistics of Gaussian random process completely specified by mean function and autocorrelation/autocovariance function

Why? Need to be able to specify the statistics of any collection of samples.
Since these are jointly Gaussian, only need their means and covariances.

WSS Gaussian random processes are stationary.

Why?
Gaussian random process characterized by second order stats.
If second order stats are shift-invariant, then we cannot distinguish statistically between shifted versions of the random process.

White Gaussian Noise

Real-valued WGN: zero mean, WSS, Gaussian random process with

$$S_n(f) = N_0/2 = \sigma^2 \leftrightarrow R_n(\tau) = (N_0/2)\,\delta(\tau) = \sigma^2 \delta(\tau)$$

Two-sided PSD

(need to integrate over both positive and negative frequencies to get the power)

One-sided PSD: N_0 (need to integrate only over physical band of positive frequencies)

Complex-valued WGN: Real and imaginary components are i.i.d. real valued WGN

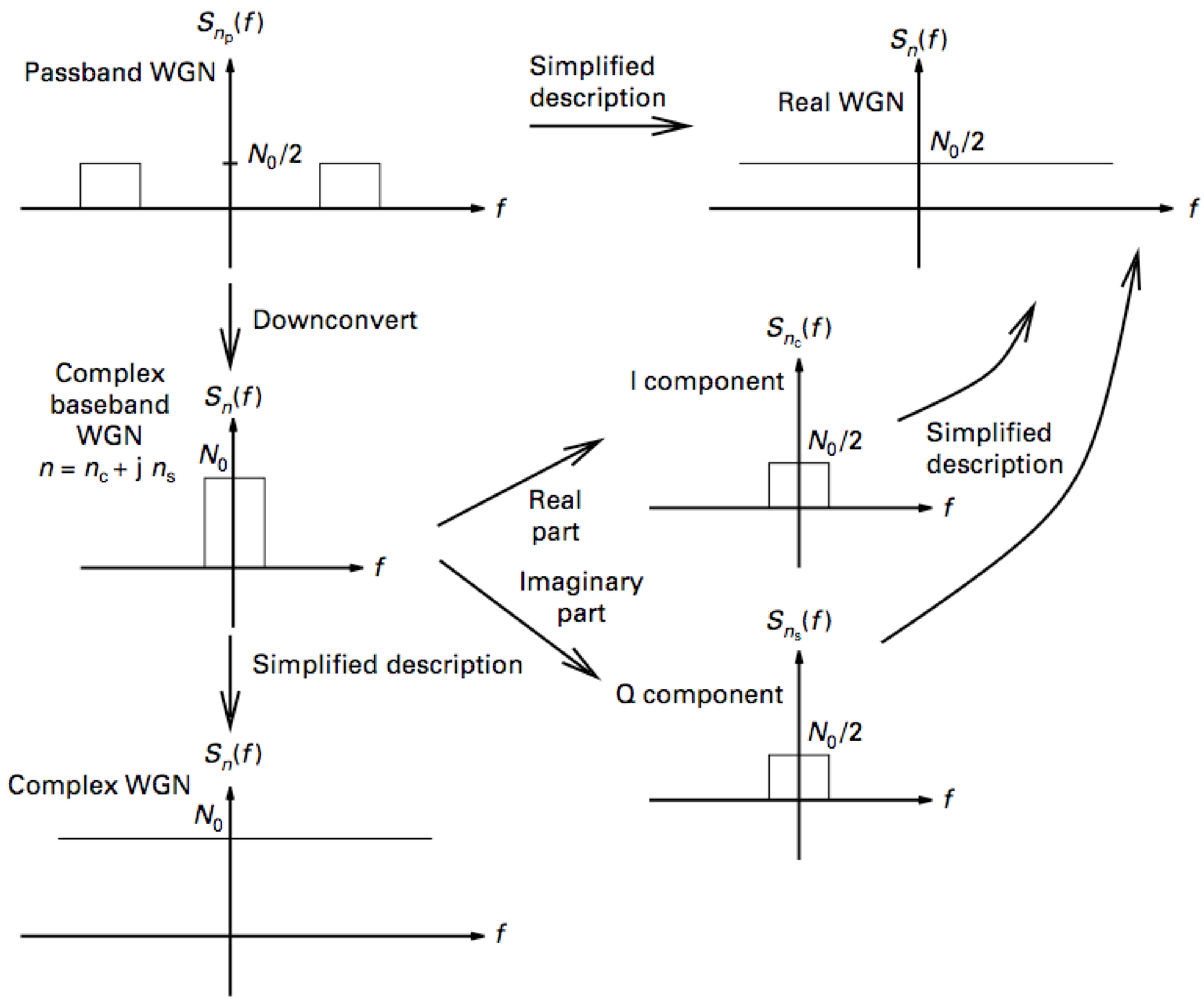
Why we use a physically unrealizable noise model

- WGN has infinite power
  - Not physically realizable
- Actual receiver noise is bandlimited and has finite power
- **OK and convenient** to assume WGN at receiver ***input***
  - Receiver noise PSD is relatively flat over typical receiver bandwidths
  - Receiver always performs some form of bandlimiting (e.g., filtering, correlation), at the output of which we have finite power
  - ***Output*** noise statistics with WGN as input and bandlimited noise at input are identical
- Why is WGN more convenient?
  - Impulsive autocorrelation function makes computation of output second order stats much easier

Modeling using WGN

- Physical baseband system: corrupted by real-valued WGN
  - Replace bandlimited noise by infinite-power WGN at input to receiver
- Physical passband system: complex envelope of passband receiver noise modeled as complex-valued WGN
  - Replace bandlimited noise by infinite-power WGN at input to receiver

Modeling using WGN: the big picture



How much is N_0?

Ideal receiver at "room temperature" $N_0 = kT_0$

$k = 1.38 \times 10^{-23}$ joule/kelvin Boltzmann's constant

T_0 "room temperature" (usually set to **290K**)

Fudge this using the receiver Noise Figure

$N_0 = kT10^{F/10}$ for noise figure of *F* dB

$$P_n = N_0 B = kT_0 10^{F/10} B = (1.38 \times 10^{-23})(290)(10^{6/10})(20 \times 10^6)$$

$$= 3.2 \times 10^{-13} \text{ watt} = 3.2 \times 10^{-10} \text{ milliwatts (mW)}.$$

Noise Power Computation

- Comm theory can work with signal-to-noise *ratios*
- But we do need absolute numbers when figuring out the "link budget": need to figure out required signal power based on the actual value of noise power

Example: *B* = 20 MHz bandwidth, receiver noise figure of 6 dB

Noise power

$$P_n = N_0 B = kT_0 10^{F/10} B = (1.38 \times 10^{-23})(290)(10^{6/10})(20 \times 10^6)$$

$$= 3.2 \times 10^{-13} \text{ watt} = 3.2 \times 10^{-10} \text{ milliwatts (mW)}.$$

Noise power in dBm

$$P_{n,\mathrm{dBm}} = 10 \log_{10} P_n(\mathrm{mW}) = -95\,\mathrm{dBm}$$

AWGN Channel and Hypothesis Testing

One of M signals sent: $s_1(t), \ldots, s_M(t)$

Receiver has to decide between one of M *hypotheses* based on the received signal, which is modeled as:

$$H_i : y(t) = s_i(t) + n(t), \qquad i = 1, \ldots, M,$$

where

$$n(t) \text{ is WGN with PSD } \sigma^2 = N_0/2$$

Consider real-valued signals for now. Will infer demodulator design for complex baseband from that for passband.

Need to learn some detection theory first, before we can solve this hypothesis testing problem.

Detection Theory Basics

Hypothesis Testing Framework

Want to determine which of M possible hypotheses best explains an observation?

Three ingredients:

Hypotheses $H_1, \ldots, H_M$:

Observation Y **takes values in observation space** Γ

(assume finite-dimensional--good enough for our purpose)

Statistical relationship between hypotheses and observation expressed through the conditional densities of the observation given each hypothesis

Conditional densities $p(y|i),\ i = 1, \ldots, M$

Fourth ingredient needed for *Bayesian* hypothesis testing

Prior probabilities $\pi(i) = P[H_i], i = 1, \ldots, M$

$$\left(\sum_{i=1}^{M} \pi(i) = 1\right)$$

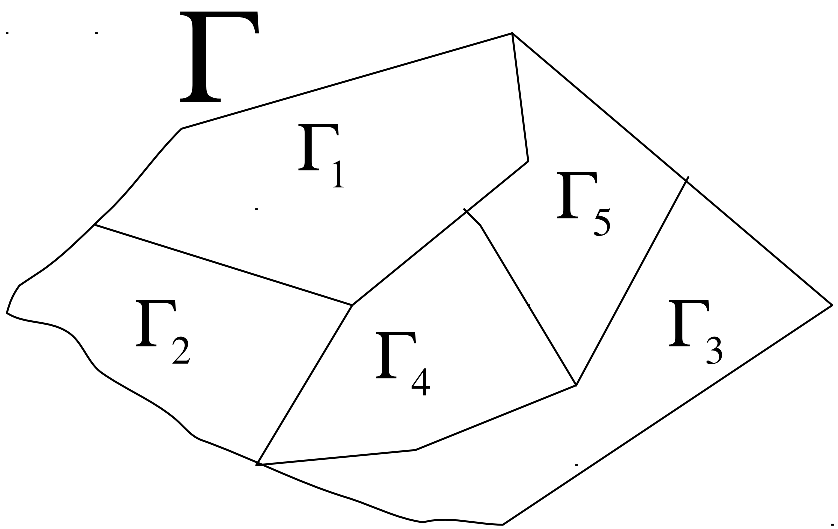
Decision Rule

A decision rule is a mapping from the observation space to the set of hypotheses

$$\delta : \Gamma \to \{1, \ldots, M\}$$

Can also view it as a partition of the observation space into M disjoint regions:

$$\Gamma_i = \{y \in \Gamma : \delta(y) = i\}$$



Basic Gaussian Example

$$H_0 : Y \sim N(0, v^2)$$

$$H_1 : \; Y \sim N(m, v^2)$$

"0 or 1 sent"

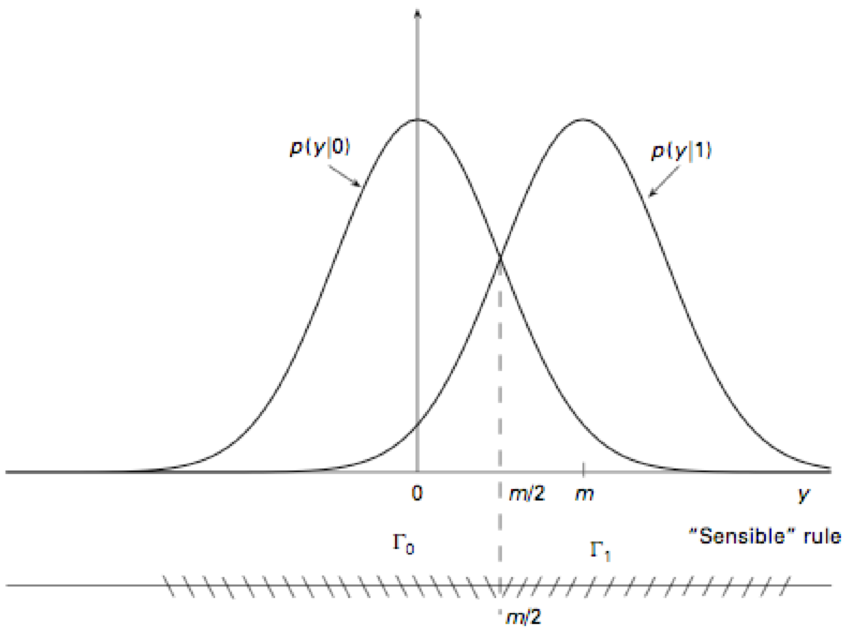
Conditional densities

$$p(y|0) = \frac{\exp\left(-\frac{y^2}{2v^2}\right)}{\sqrt{2\pi v^2}}$$

Could model noisy sample at output of an equalizer

$$p(y|1) = \frac{\exp\left(-\frac{(y-m)^2}{2v^2}\right)}{\sqrt{2\pi v^2}}$$

Basic Gaussian Example (contd.)



"Sensible" rule: split the difference

Would this rule make sense if we know for sure that 0 was sent?

What if we know beforehand that 0 was sent with probability 0.75?

What if the noise is not Gaussian or additive?

Need a systematic framework for deriving good decision rules.

First step: define the performance metrics of interest

Performance Metrics for Evaluating Decision Rules

Conditional error probabilities

$$P_{e|i} = P[\text{say } H_j \text{ for some } j \neq i | H_i \text{ is true}] \quad = \sum_{j \neq i} P[Y \in \Gamma_j | H_i]$$

$$= 1 - P[Y \in \Gamma_i | H_i]$$

Average error probability

$$P_e = \sum_{i=1}^{M} \pi(i) P_{e|i}$$

Error probs for "sensible" rule in basic Gaussian example:

Conditional error probs

$$P_{e|0} = P\left[Y > \frac{m}{2} | H_0\right] = Q\left(\frac{m}{2v}\right) \quad (Y \sim N(0, v^2) \text{ under } H_0)$$

$$P_{e|1} = P\left[Y \leq \frac{m}{2} | H_1\right] = \Phi\left(\frac{\frac{m}{2} - m}{v}\right) = Q\left(\frac{m}{2v}\right) \quad (Y \sim N(m, v^2) \text{ under } H_1)$$

Average error prob $\quad P_e = Q\left(\frac{m}{2v}\right) \quad$ regardless of prior probs

Maximum Likelihood (ML) Rule

Choose the hypothesis that maximizes the conditional probability of the observation:

$$\delta_{ML}(y) = \arg \max_{1 \le i \le M} p(y|i) = \arg \max_{1 \le i \le M} \log p(y|i)$$

Check: "Sensible" rule for the basic Gaussian example is the ML rule

ML rule seems like a good idea. Is there anything optimal about it?

Minimizes error probability if all hypotheses are equally likely (we'll see this soon)

Asymptotically optimal when observations can be trusted more and more (e.g., high SNR, large number of samples). See Poor's text, *Introduction to Signal Detection and Estimation*.

Minimum Probability of Error (MPE) Rule
which turns out to be the
Maximum A Posteriori Probability (MAP) Rule

Minimize average probability of error (assume prior probabilities of the hypotheses are known)

Let's derive it. Convenient to consider maximizing prob of correct decision.

For any given decision rule $\delta \Leftrightarrow$ Decision regions $\{\Gamma_i\}$

$$P_{c|i} = \int_{\_} p(y|i)\mathrm{d}y, \qquad i = 1, \ldots, M$$ Conditional prob of correct decision

$$P_c = \sum_{i=1}^{M} \pi(i) P_{c|i} = \sum_{i=1}^{M} \pi(i) \int_{\Gamma_i} p(y|i)\mathrm{d}y$$ Average prob of correct decision

Consider any potential observation: $y \in \Gamma$

If we put it in the *i*-th decision region ($y \in \Gamma_i$), then our "reward" (contribution to the prob of correct decision) is $\pi(i)p(y|i)$

MPE rule: choose *i* to maximize this contribution

MPE Rule (contd.)

We have derived the MPE rule to be as follows:

$$\delta_{\text{MPE}}(y) = \arg \max_{1 \le i \le M} \pi(i) p(y|i) = \arg \max_{1 \le i \le M} \log \pi(i) + \log p(y|i)$$

1) MPE rule is equivalent to the Maximum A Posteriori Probability (MAP) rule

Posterior probability of hypothesis *i* given the observation

$$P(H_i|y) = \frac{\pi(i) p(y|i)}{p(y)}$$

Can rewrite MPE rule as follows:

$$\delta_{\text{MAP}}(y) = \arg \max_{1 \le i \le M} P(H_i|y)$$

2) MPE rule reduces to ML rule for equal priors $(\pi(i) = 1/M)$

We can drop $\pi(i)$ from the maximization if it does not depend on *i*

Likelihood Ratio Test

For binary hypothesis testing, MPE rule specializes to:

$$\delta_{\text{MPE}}(y) = \begin{cases} 1, & \pi(1)p(y|1) > \pi(0)p(y|0) \\ 0, & \pi(1)p(y|1) < \pi(0)p(y|0) \\ \text{don't care}, & \pi(1)p(y|1) = \pi(0)p(y|0) \end{cases}$$

Can rewrite as a Likelihood Ratio Test (LRT)

$$L(y) = \frac{p(y|1)}{p(y|0)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\pi(0)}{\pi(1)}$$

Often we take log on both sides to get Log LRT (LLRT)

Note: Comparing likelihood ratio to a threshold is a common feature of optimal decision rules resulting from many different criteria--MPE, ML, Neyman-Pearson (radar problem trading off false alarm vs. miss probabilities--not covered here). The threshold changes based on the criterion. The likelihood ratio summarizes all the information relevant to the hypothesis testing problem; that is, it is a **"sufficient statistic**"

Likelihood Ratio for Basic Gaussian Example

$$p(y|0) = \frac{\exp\left(-\frac{y^2}{2v^2}\right)}{\sqrt{2\pi v^2}} \qquad p(y|1) = \frac{\exp\left(-\frac{(y-m)^2}{2v^2}\right)}{\sqrt{2\pi v^2}}$$

$$L(y) = \exp\left(\frac{1}{v^2}\left(my - \frac{m^2}{2}\right)\right)$$

Compare log LR with zero to get "sensible" (ML) rule:

$$\delta(y) = \begin{cases} 1, & y > \frac{m}{2} \\ 0, & y \leq \frac{m}{2} \end{cases} \qquad m > 0$$

Note that the inequalities are reversed when *m<0.*

Irrelevant Statistics

Consider the following hypothesis testing problem:

$$H_1 : Y_1 = m + N_1, \;\; Y_2 = N_2,$$
$$H_0 : Y_1 = N_1, \qquad\quad Y_2 = N_2,$$

When can we throw Y_2 away without performance degradation?

That is, when is Y_2 *irrelevant* to our decision?

Intuition for two scenarios:

If the noises are independent, Y_2 should be irrelevant, since it contains no “signal” contribution.

If the noises are equal (extreme case of highly correlated), Y_2 is very relevant…subtract it out from Y_1 to get perfect detection!

Need a systematic way of recognizing irrelevant statistics…

Irrelevant Statistics (contd.)

Theorem 3.2.2 (Characterizing an irrelevant statistic) *For M-ary hypothesis testing using an observation $Y = (Y_1, Y_2)$, the statistic Y_2 is irrelevant if the conditional distribution of Y_2, given Y_1 and H_i, is independent of i. In terms of densities, we can state the condition for irrelevance as $p(y_2|y_1, i) = p(y_2|y_1)$ for all i.*

BEGIN PROOF

Conditional densities are all that are relevant. Under the given conditions,

$$p(y|i) = p(y_1, y_2|i) = p(y_2|y_1, i)p(y_1|i) = p(y_2|y_1)p(y_1|i)$$

These depend on hypothesis only through the first observation. **END PROOF**

Relation to sufficient statistics: f(Y) is a sufficient statistic if Y is irrelevant for hypothesis testing with (f(Y), Y). That is, once we know f(Y), we have all the information needed for our decision, and no longer need the original observation Y.

Irrelevant Statistics: Example

$$H_1 : Y_1 = m + N_1, \;\; Y_2 = N_2,$$
$$H_0 : Y_1 = N_1, \qquad\quad Y_2 = N_2,$$

N_2 is independent of N_1

Then theorem condition is (more than) satisfied to see *Y_2* is irrelevant:

$$p(y_2|y_1, i) = p(y_2)$$

Why?

$Y_2 = N_2$ is independent of H_i and N_1, and hence of H_i and Y_1

We will apply this argument when deriving optimal receivers for signaling in AWGN.

Signal Space Concepts

M-ary signaling in AWGN

THE MODEL **Received signal = transmitted signal + WGN**

$$H_i : y(t) = s_i(t) + n(t), \qquad i = 1, \ldots, M,$$

$$n(t) \text{ is WGN with PSD } \sigma^2 = N_0/2$$

We consider real-valued signals and noise in our derivations:

applies to real baseband and passband signaling.

Will infer results for complex baseband from those for passband.

KEY RESULTS

1) Can restrict attention to a finite-dimensional "signal space," even though received signal is a continuous-time signal lying in an infinite-dimensional space

2) Optimal receiver matches received signal against each of the *M* possible noiseless transmitted signals, and selects the "best" match.

Performance of optimal reception is scale-invariant

Optimal receiver for

$$H_i : y(t) = s_i(t) + n(t), \qquad i = 1, \ldots, M,$$

should have exactly the same performance as optimal receiver for

$$H_i : \tilde{y}(t) = As_i(t) + An(t), \qquad i = 1, \ldots, M.$$

(if not, can improve performance by scaling and using the model for which the optimal receiver does better, contradicting optimality)

Inference to be drawn later: Performance depends only on *signal to noise ratios* and the *relative geometry* of the signal set.

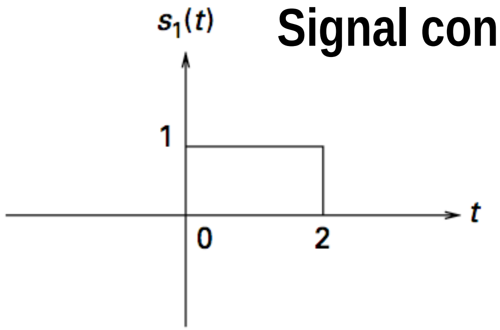
Signal space concepts: Euclidean geometry for finite-dimensional vectors is enough for our purpose

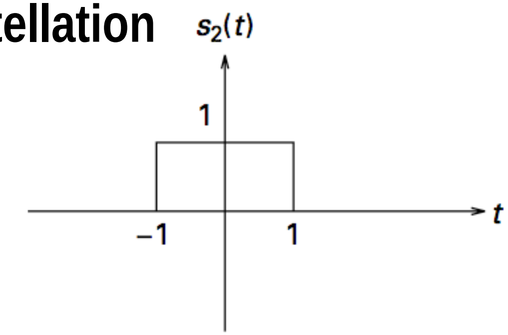
Path to a geometric view: overview

- Signals lie in a finite-dimensional “signal space”
- Noise component outside this finite-dimensional space is irrelevant to deciding on which signal was sent
  - This is because of the geometry of WGN
- We can therefore work with a finite-dimensional received vector to derive optimal receiver
  - Projection of received signal onto signal space

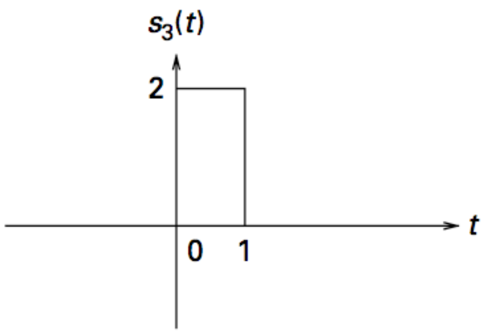
Example: 4-ary signaling in a 3-dimensional signal space

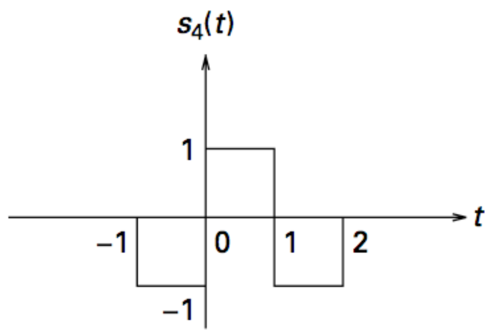
Signal constellation





$$\mathbf{s}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{s}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

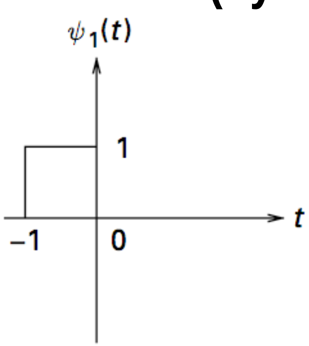


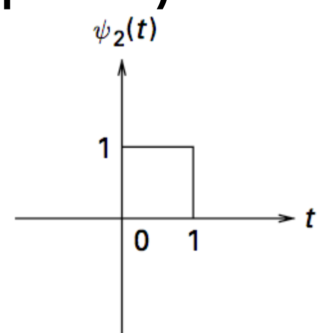


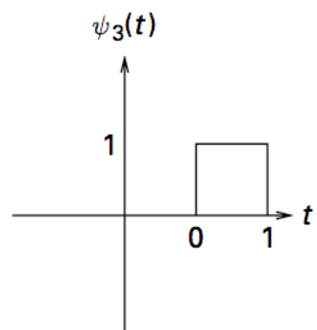
$$\mathbf{s}_3 = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}, \quad \mathbf{s}_4 = \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}$$

Orthonormal basis (by inspection)

Vector representation with respect to basis







Orthonormal basis exists even when inspection does not work
Gramm-Schmidt Orthogonalization

Recursive construction of basis for signal space

Given $\{\psi_1, \ldots \psi_m\}$ orthonormal basis for $\mathcal{S}_k$, the subspace spanned by $s_1, \ldots, s_k$

Update the basis so that it spans $\mathcal{S}_{k+1}$

Step k+1 (we stop at k=M)

Find the component of $s_{k+1}(t)$ orthogonal to $\mathcal{S}_k$

$$\phi_{k+1}(t) = s_{k+1}(t) - \sum_{i=1}^{m} \langle s_{k+1}, \psi_i \rangle \psi_i(t)$$

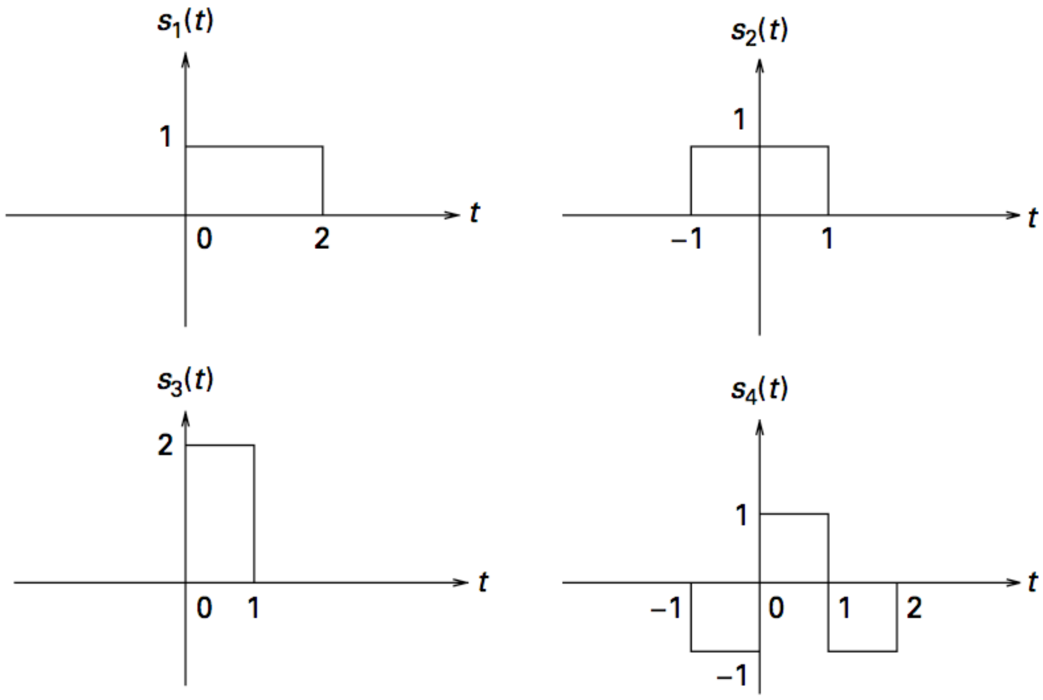
If $\phi_{k+1} \neq 0$, construct a new basis function by normalizing it

$\psi_{m+1}(t) = \phi_{k+1}(t)/||\phi_{k+1}||$ New basis $\{\psi_1, \ldots, \psi_m, \psi_{m+1}\}$

If $\phi_{k+1} = 0$, then $s_{k+1} \in \mathcal{S}_k$ No need for a new basis function.

Step 1 (initialization): $\phi_1 = s_1$

Gramm-Schmidt example

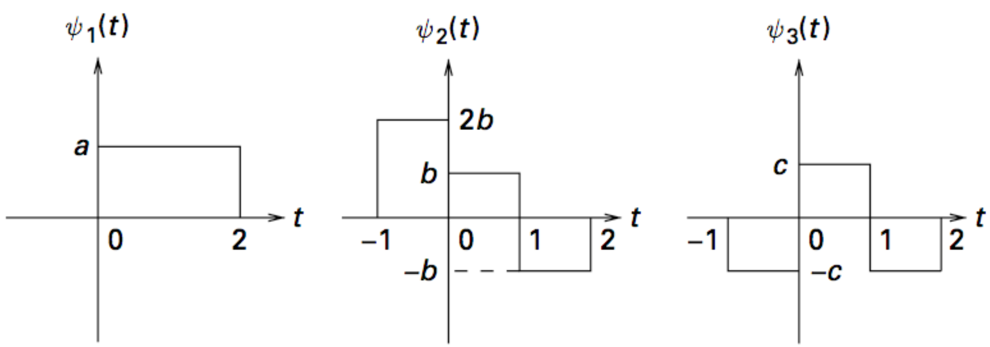


Some observations:

Basis obtained by Gramm-Schmidt depends on the signal ordering

Can be messier than what we get from inspection, but has the virtue of general applicability

Gramm-Schmidt (in order of signal numbering)



Exercise: find *a, b, c*

Signal geometry is preserved

Inner products between signals are preserved when we project onto the signal space

$$\langle s_i, s_j \rangle = \langle \sum_{k=1}^{n} s_i[k]\psi_k, \sum_{l=1}^{n} s_j[l]\psi_l \rangle = \sum_{k=1}^{n}\sum_{l=1}^{n} s_i[k]s_j[l]\langle \psi_k, \psi_l \rangle$$
$$= \sum_{k=1}^{n}\sum_{l=1}^{n} s_i[k]s_j[l]\delta_{kl} = \sum_{k=1}^{n} s_i[k]s_j[k] = \langle \mathbf{s}_i, \mathbf{s}_j \rangle.$$

1) Euclidean geometry depends only on inner products, so relative geometry of signal constellation does not change
2) We shall see that performance depends only on this relative geometry

WGN through correlators

The following proposition is key to understanding the geometry of WGN

Proposition 3.3.1 (WGN through correlators) *Let $u_1(t)$ and $u_2(t)$ denote finite-energy signals, and let $n(t)$ denote WGN with PSD $\sigma^2 = N_0/2$. Then $\langle n, u_1 \rangle$ and $\langle n, u_2 \rangle$ are jointly Gaussian with covariance*

$$\mathrm{cov}(\langle n, u_1 \rangle, \langle n, u_2 \rangle) = \sigma^2 \langle u_1, u_2 \rangle.$$

Covariance computation:

$$\begin{aligned}
\mathrm{cov}(\langle n, u_1 \rangle, \langle n, u_2 \rangle) &= \mathbb{E}[\langle n, u_1 \rangle \langle n, u_2 \rangle] = \mathbb{E}[\int n(t) u_1(t) \mathrm{d}t \int n(s) u_2(s) \mathrm{d}s] \\
&= \int \int u_1(t) u_2(s) \mathbb{E}[n(t) n(s)] \mathrm{d}t \, \mathrm{d}s \\
&= \int \int u_1(t) u_2(s) \sigma^2 \delta(t - s) \mathrm{d}t \, \mathrm{d}s \\
&= \sigma^2 \int u_1(t) u_2(t) \mathrm{d}t = \sigma^2 \langle u_1, u_2 \rangle.
\end{aligned}$$

The Geometry of WGN

From the 'WGN through correlators" result, we infer:

1) Projection of WGN along any "direction," or correlation with a unit energy signal, gives an $N(0, \sigma^2 = N_0/2)$ random variable

2) Projections in orthogonal directions are independent (since jointly Gaussian and uncorrelated)

(which is why we might see in the literature statements such as: the noise is WGN with variance σ^2 per dimension)

Path to a geometric view: overview with a little more detail

1) Signal contribution lies in a finite-dimensional signal space (dimension at most M) spanned by the M possible noiseless received signals

2) Component of received signal orthogonal to signal space is irrelevant to our decision:

Has only contribution from noise. This noise contribution is orthogonal to the signal space, and is independent of noise contributions in the signal space.

Now apply irrelevance criterion: exactly as in earlier example.

3) Can restrict attention to signal and noise contributions in signal space.

These can be represented by finite-dimensional vectors--easy to apply detection theory framework

Now the detailed development…

Decomposition of the received signal

Component of received signal in the signal space can be represented as a finite-dimensional vector

$$y_S(t) = \sum_{j=1}^{n} \langle y, \psi_j \rangle \psi_j(t) = \sum_{j=1}^{n} y[j]\psi_j(t) \iff \mathbf{Y} = (\langle y, \psi_1 \rangle, \ldots, \langle y, \psi_n \rangle)^T$$

Under hypothesis H_i $(i = 1, \ldots, M)$

$$\mathbf{Y} = \mathbf{s}_i + \mathbf{N} \qquad \begin{aligned} \mathbf{s}_i &= (\langle s_i, \psi_1 \rangle, \ldots, \langle s_i, \psi_n \rangle)^T \\ \mathbf{N} &= (\langle n, \psi_1 \rangle, \ldots, \langle n, \psi_n \rangle)^T \end{aligned}$$

Component of received signal orthogonal to signal space

$$y^{\perp}(t) = y(t) - y_S(t) = y(t) - \sum_{j=1}^{n} y_j \psi_j(t)$$

We will show this is irrelevant, so we can work with the finite-dimensional model for the component in the signal space.

Irrelevance of component orthogonal to signal space

Conditioning on hypothesis H_i, $y(t) = s_i(t) + n(t)$

1) Component orthogonal to signal space contains only noise, regardless of which signal was sent:

$$
\begin{aligned}
y^{\perp}(t) &= y(t) - \sum_{j=1}^{n} \langle y, \psi_j \rangle \psi_j(t) = s_i(t) + n(t) - \sum_{j=1}^{n} \langle s_i + n, \psi_j \rangle \psi_j(t) \\
&= n(t) - \sum_{j=1}^{n} \langle n, \psi_j \rangle \psi_j(t) = n^{\perp}(t),
\end{aligned}
$$

2) Noise component orthogonal to signal space is independent of noise component in the signal space.

Jointly Gaussian, hence suffices to show they are uncorrelated

1) and 2) imply irrelevance, using the theorem (similar to scalar example).

Covariance computations for showing irrelevance

We want to show that these covariances are zero:

$$\operatorname{cov}(n^{\perp}(t), N[k]) = \mathbb{E}[n^{\perp}(t)N[k]] = \mathbb{E}[\{n(t) - \sum_{j=1}^{n} N[j]\psi_j(t)\}N[k]]$$

$$= \mathbb{E}[n(t)N[k]] - \sum_{j=1}^{n} \mathbb{E}[N[j]N[k]]\psi_j(t).$$

Now, use

$$\mathbb{E}[n(t)\langle n, \psi_k\rangle] = \mathbb{E}[n(t)\int n(s)\psi_k(s)\mathrm{d}s]$$

$$= \int \mathbb{E}[n(t)n(s)]\psi_k(s)\mathrm{d}s = \int \sigma^2\delta(s-t)\psi_k(s)\mathrm{d}s = \sigma^2\psi_k(t).$$

and

$$\mathbb{E}[N[j]N[k]] = \sigma^2\delta_{jk};$$

to get

$$\operatorname{cov}(n^{\perp}(t), N[j]) = \sigma^2\psi_k(t) - \sigma^2\psi_k(t) = 0.$$

Signal space for passband (2D) signaling

Passband transmitted signal

$$s_{b_c,b_s}(t) = Ab_c p(t)(\sqrt{2}\cos 2\pi f_c t) - Ab_s p(t)(\sqrt{2}\sin 2\pi f_c t)$$

Information encoded in (b_c, b_s)

Two-dimensional signal space with natural choice of basis

$$\psi_c(t) = \alpha\, p(t)\cos 2\pi f_c t, \psi_s(t) = -\alpha\, p(t)\sin 2\pi f_c t$$

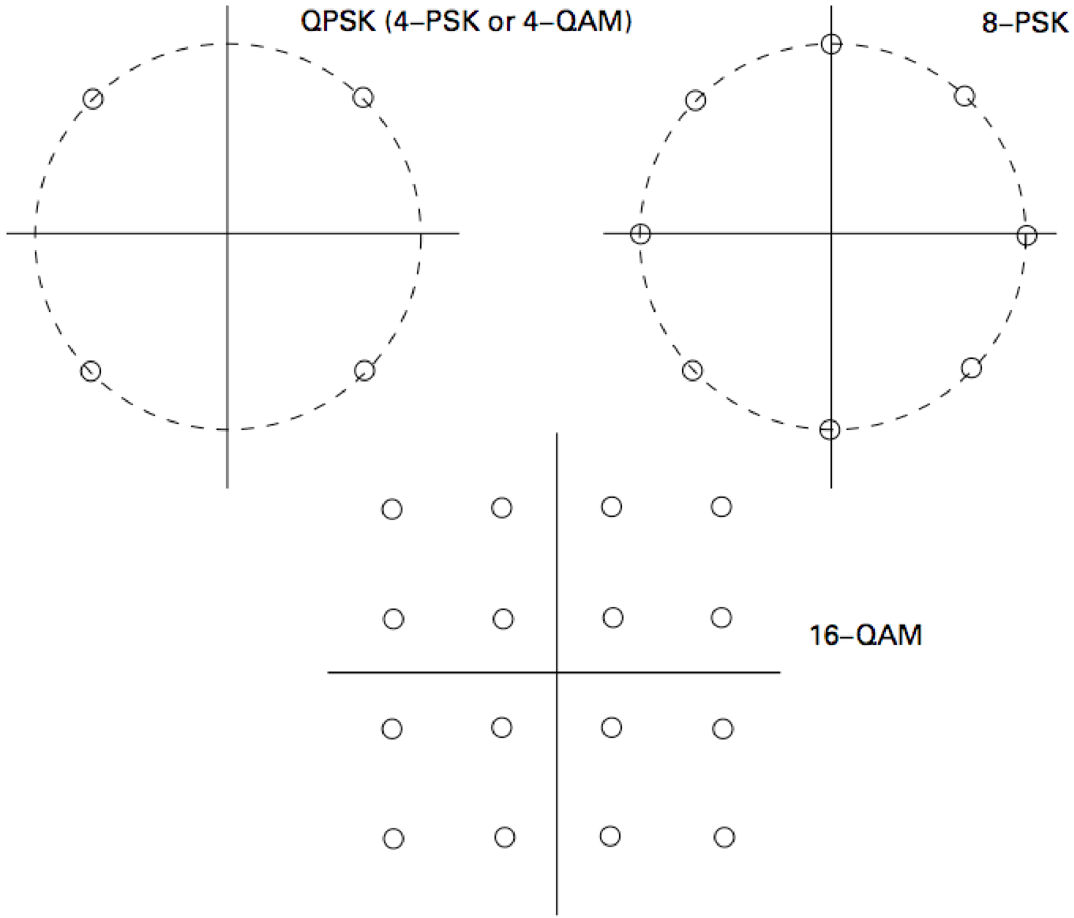
Two-dimensional received signal vector

$$\mathbf{y} = \begin{pmatrix} y_c \\ y_s \end{pmatrix} = \begin{pmatrix} b_c \\ b_s \end{pmatrix} + \begin{pmatrix} N_c \\ N_s \end{pmatrix}$$

$N_c,\ N_s$ are i.i.d. $N(0, \sigma^2)$

(can choose scaling arbitrarily, since what counts is signal-to-noise *ratio*)

Signal space representations for some 2-D constellations



Optimal reception in AWGN

Optimal reception

M-ary signaling over discrete time AWGN channel

$$H_i: \ \mathbf{Y} = \mathbf{s}_i + \mathbf{N} \quad i = 1, \ldots, M,$$
$$\mathbf{N} \sim N(0, \sigma^2 \mathbf{I})$$

Conditional densities

$$p_{\mathbf{Y}|i}(\mathbf{y}|H_i) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{||\mathbf{y}-\mathbf{s}_i||^2}{2\sigma^2}\right)$$

Use log LRT to obtain ML and MPE rules as follows:

ML Rule

$$\delta_{\mathrm{ML}}(\mathbf{y}) = \arg \min_{1 \le i \le M} ||\mathbf{y}-\mathbf{s}_i||^2 = \arg \max_{1 \le i \le M} \langle \mathbf{y}, \mathbf{s}_i \rangle - \frac{||\mathbf{s}_i||^2}{2}$$

"Minimum distance," or "maximum inner product with energy correction term"

MPE Rule

$$\delta_{\mathrm{MPE}}(\mathbf{y}) = \arg \min_{1 \le i \le M} ||\mathbf{y}-\mathbf{s}_i||^2 - 2\sigma^2 \log \pi(i)$$
$$= \arg \max_{1 \le i \le M} \langle \mathbf{y}, \mathbf{s}_i \rangle - \frac{||\mathbf{s}_i||^2}{2} + \sigma^2 \log \pi(i)$$

ML cost function + term accounting for prior probabilities

Optimal reception

M-ary real-valued signaling over continuous time AWGN channel

$$\boxed{H_i : y(t) = s_i(t) + n(t), \quad i = 1, \ldots, M}$$

n(t) WGN, PSD = σ^2

We have shown that there is no loss of optimality in projecting onto signal space.
We can now use the results for discrete time AWGN channel.
Simply replace discrete-time inner products by continuous-time inner products:

Energy term in cost function $||\mathbf{s}_i||^2 = ||s_i||^2$ (had already shown signal inner products are preserved)

Correlation term in cost function

$$\begin{aligned} \langle y, s_i \rangle &= \langle y_{\mathcal{S}} + y^{\perp}, s_i \rangle &&= \langle y_{\mathcal{S}}, s_i \rangle + \langle y^{\perp}, s_i \rangle, \\ &= \langle y_{\mathcal{S}}, s_i \rangle &&= \langle \mathbf{y}, \mathbf{s}_i \rangle, \end{aligned}$$

$$\delta_{\mathrm{ML}}(y) = \arg \max_{1 \le i \le M} \langle y, s_i \rangle - \frac{||s_i||^2}{2}$$

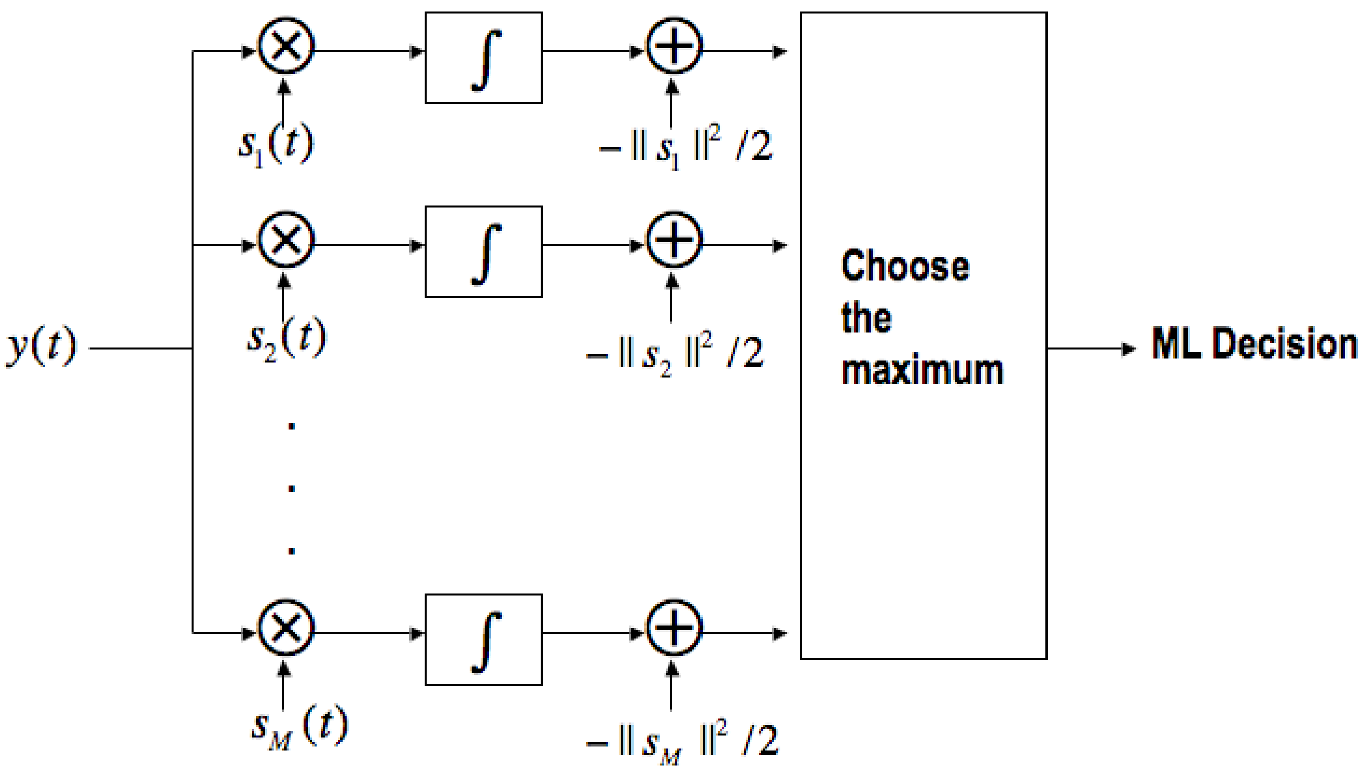
ML Rule

$$\delta_{\mathrm{MPE}}(\mathbf{y}) = \arg \max_{1 \le i \le M} \langle y, s_i \rangle - \frac{||s_i||^2}{2} + \sigma^2 \log \pi(i)$$

MPE Rule

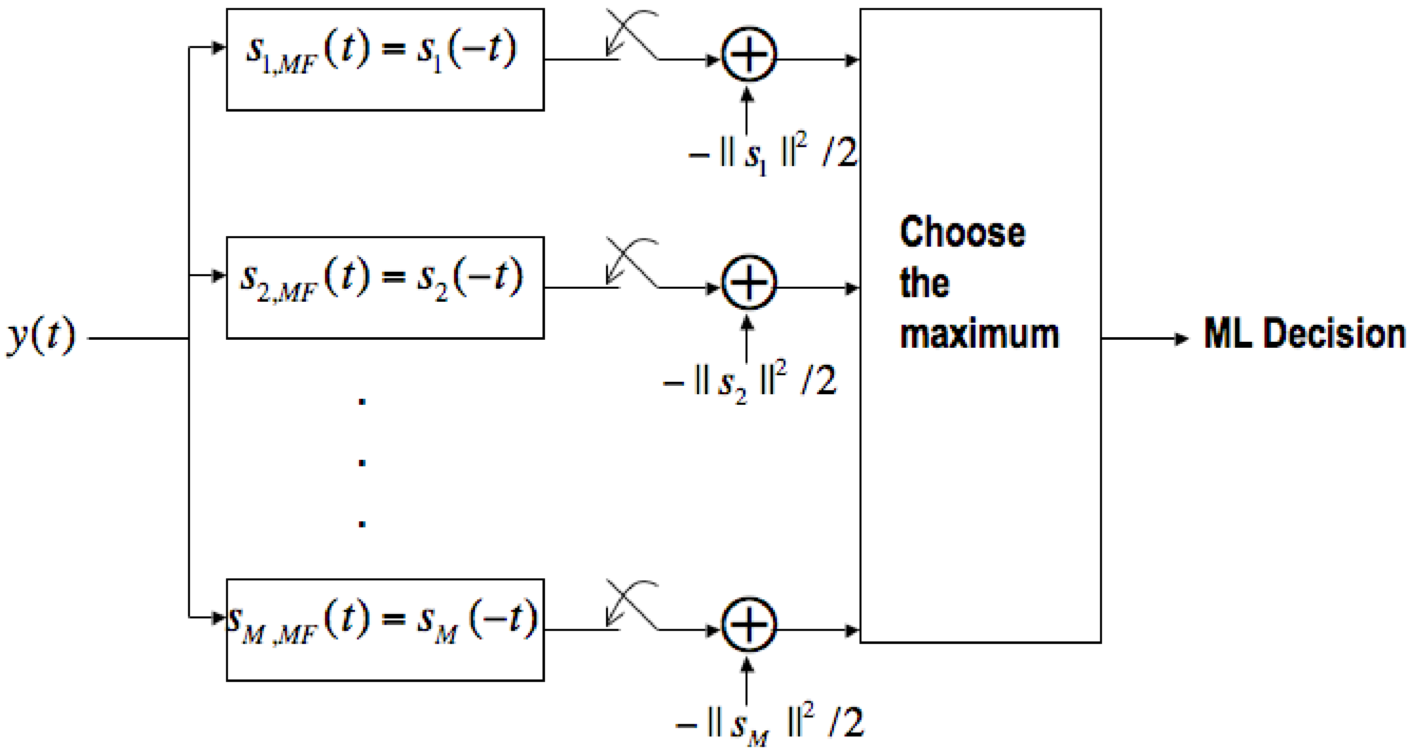
Bank of correlators Implementation

Optimal receiver can be implemented by a bank of correlators



Bank of matched filters implementation

Correlation can be replaced by matched filter sampled at the appropriate time



Implementation in complex baseband

Apply to real-valued passband signals, then write the passband inner products in terms of complex baseband inner products

$$H_i : y_\mathrm{p}(t) = s_{i,\mathrm{p}}(t) + n_\mathrm{p}(t), \quad i = 1, \ldots, M$$ **Passband model**

$$H_i : y(t) = s_i(t) + n(t), \quad i = 1, \ldots, M,$$ **Complex baseband model**

Optimal decision statistics for real-valued passband model are known

$$\langle y_\mathrm{p}, s_{i,\mathrm{p}} \rangle - \frac{||s_{i,\mathrm{p}}||^2}{2} = \mathrm{Re}\left(\langle y, s_i \rangle\right) - \frac{||s_i||^2}{2}$$

Passband inner products can be computed in complex baseband, so we can compute these decision stats in complex baseband

Optimal coherent reception in complex baseband

As before, except correlation is performed in complex baseband:

$$\boxed{\begin{aligned}\delta_{\mathrm{ML}}(y) &= \arg \max_{1 \le i \le M} \mathrm{Re}\left(\langle y, s_i \rangle\right) - \frac{||s_i||^2}{2} \\ \delta_{\mathrm{MPE}}(\mathbf{y}) &= \arg \max_{1 \le i \le M} \mathrm{Re}\left(\langle y, s_i \rangle\right) - \frac{||s_i||^2}{2} + \sigma^2 \log \pi(i)\end{aligned}}$$

No cross-coupling between I and Q for (ideal) coherent receivers

$$\mathrm{Re}(\langle y, s_i \rangle) = \langle y_{\mathrm{c}}, s_{i,\mathrm{c}} \rangle + \langle y_{\mathrm{s}}, s_{i,\mathrm{s}} \rangle \qquad ||s_i||^2 = ||s_{i,\mathrm{c}}||^2 + ||s_{i,\mathrm{s}}||^2$$

Can think of the complex signal as two real-valued signals, and develop signal space ideas in real-valued vector spaces

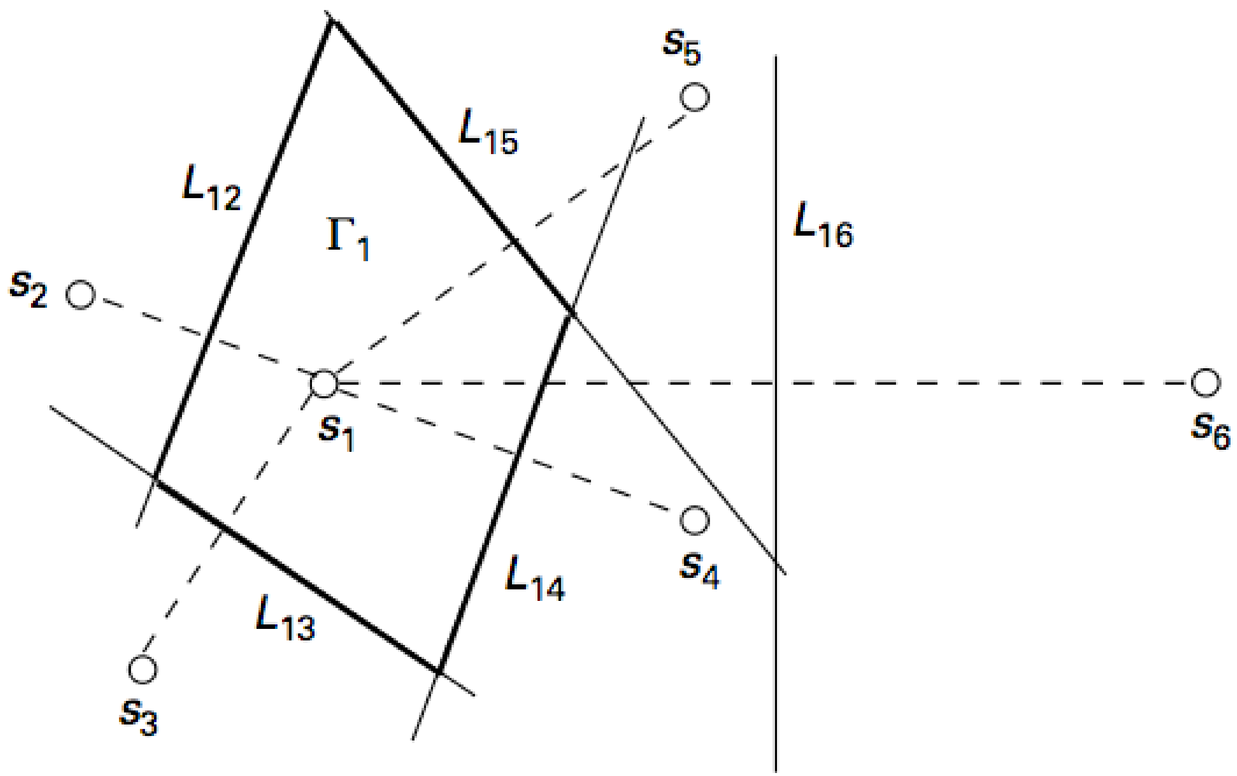
Intrinsically complex-valued operations needed when I and Q components get "mixed up," as is the case for non-ideal carrier sync (Chapter 4: synchronization, noncoherent demodulation)

Geometry of ML Rule

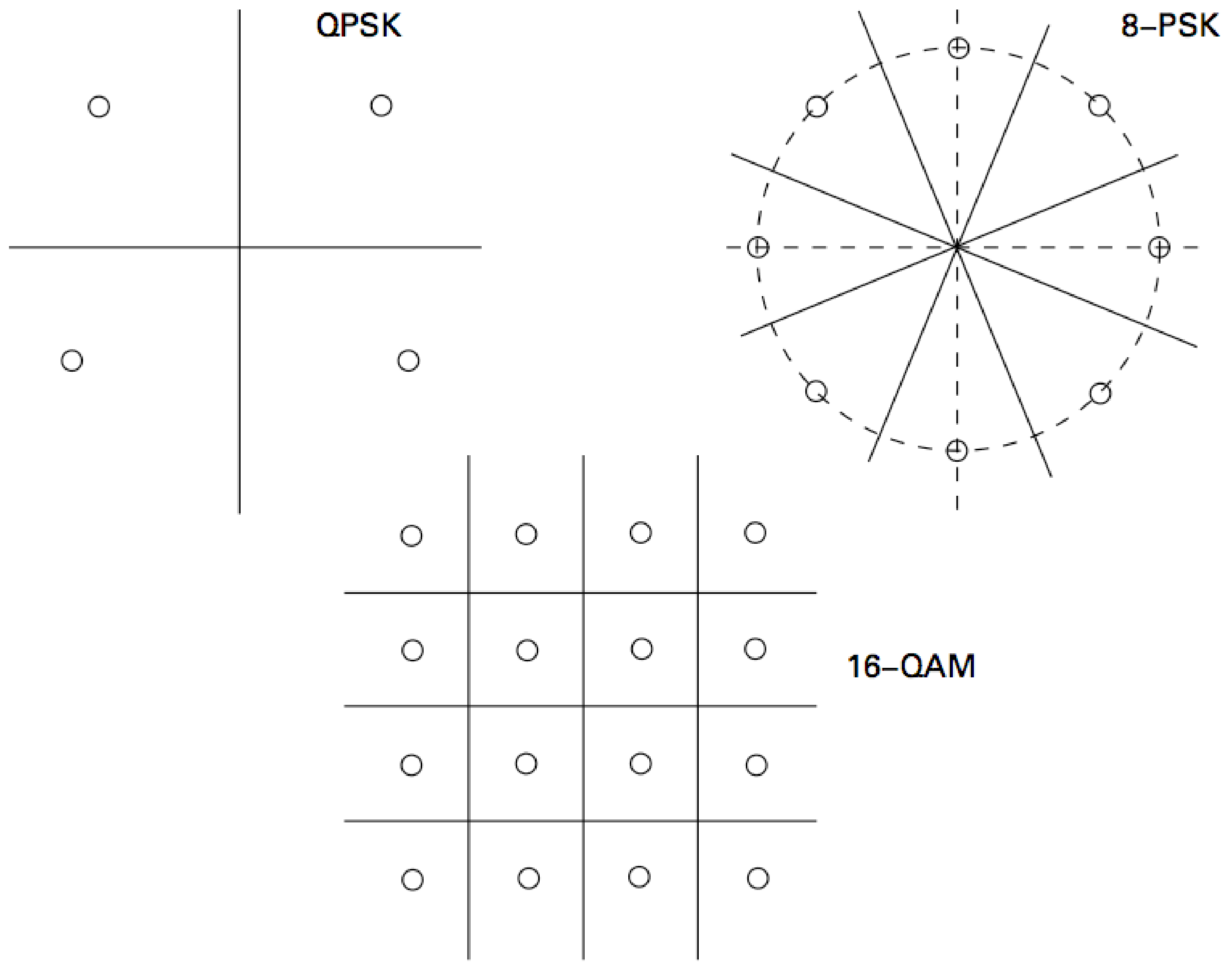
Minimum distance rule

--draw perpendicular bisectors of lines joining signal points

--these pairwise decisions are put together to obtain the overall decision regions



ML decision regions for 2-D constellations



Soft Decisions

We are less confident about a decision when it is close to a decision boundary than when it is well within the decision region

"Soft decisions" convey our confidence in the decision; especially useful when decisions are being processed further by a decoder

Posterior probabilities are the gold standard for soft decisions (they milk all the information available from the received signal)

$$\pi(i|\mathbf{y}) = P[H_i|\mathbf{y}] = \frac{p(\mathbf{y}|i)P[H_i]}{p(\mathbf{y})} = \frac{p(\mathbf{y}|i)P[H_i]}{\sum_{j=1}^{M} p(\mathbf{y}|j)P[H_j]}.$$

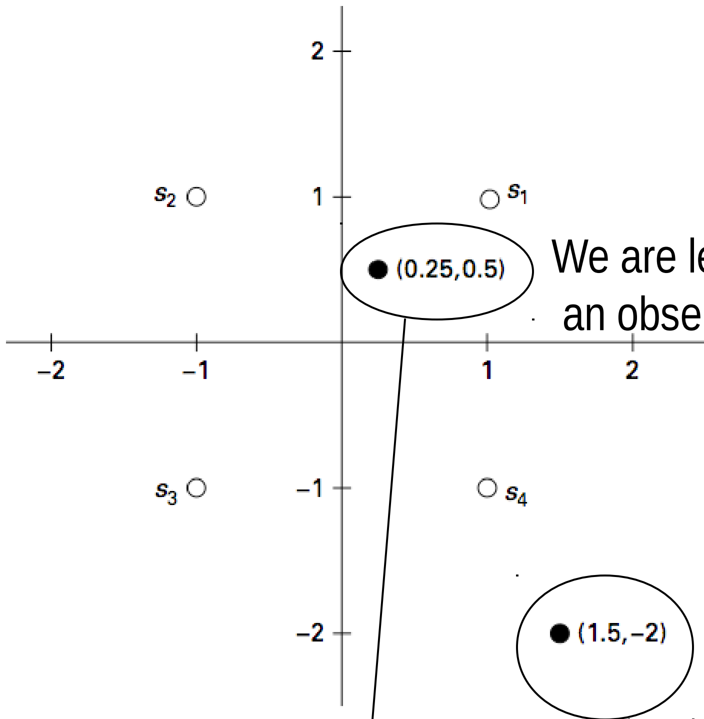
Computation using Bayes' rule

Specializing to AWGN channel:

$$\pi(i|\mathbf{y}) = \frac{\pi(i)\exp\left(-\frac{||\mathbf{y}-\mathbf{s}_i||^2}{2\sigma^2}\right)}{\sum_{j=1}^{M} \pi(j)\exp\left(-\frac{||\mathbf{y}-\mathbf{s}_j||^2}{2\sigma^2}\right)}$$

In practice, we might use logs of the ratios of such probabilities, or devise hand-crafted limited-precision soft decisions

Soft Decisions: Example



We are less confident about the hard decision when we get an observation close to decision boundaries

We are more confident about the hard decision, since it is well within decision region

$\pi(i\|\mathbf{y})$	$\mathbf{y} = (0.25, 0.5)$	$\mathbf{y} = (1.5, -2)$
1	0.455	0.017
2	0.276	0
3	0.102	0.047
4	0.167	0.935

Need to know noise level to compute soft decisions

$$\sigma^2 = 1$$

Dependence of soft decisions on noise level

$\pi(i\|\mathbf{y})$	$\mathbf{y} = (0.25, 0.5)$	$\mathbf{y} = (1.5, -2)$
1	0.455	0.017
2	0.276	0
3	0.102	0.047
4	0.167	0.935

$\sigma^2 = 1$

$\pi(i\|\mathbf{y})$	$\mathbf{y} = (0.25, 0.5)$	$\mathbf{y} = (1.5, -2)$
1	0.299	0.183
2	0.264	0.086
3	0.205	0.235
4	0.233	0.497

$\sigma^2 = 4$

Posterior probs "spread out" at higher noise level to reflect lower confidence in hard decision even when received signal is well within decision region

Performance Analysis of ML Reception

Analysis for binary signaling

We'll now stop making notational distinctions between continuous time signals and vectors, having established their equivalence (for the purpose of optimal reception in AWGN) via signal space concepts.

Start with binary on-off keying

$$H_1 : y(t) = s(t) + n(t),$$
$$H_0 : y(t) = n(t).$$

ML rule: $$\langle y, s \rangle \underset{H_0}{\overset{H_1}{\gtrless}} \frac{||s||^2}{2}$$

Decision statistic: $Z = \langle y, s \rangle$ (correlator output)

Conditional error probabilities:

$$P_{e|1} = P\left[Z < \frac{||s||^2}{2} | H_1\right] \quad P_{e|0} = P\left[Z > \frac{||s||^2}{2} | H_0\right]$$

Need conditional distribution of Z under each hypothesis.
Conditionally Gaussian, so only need means and variances.

Binary OOK (contd.)

Conditional means and variances for $Z = \langle y, s \rangle$

$$\mathbb{E}[Z|H_0] = \mathbb{E}[\langle n, s\rangle] = 0,$$
$$\mathrm{var}(Z|H_0) = \mathrm{cov}(\langle n, s\rangle, \langle n, s\rangle) = \sigma^2 ||s||^2$$

$$H_1 : y(t) = s(t) + n(t),$$
$$H_0 : y(t) = n(t).$$

$$\mathbb{E}[Z|H_1] = \mathbb{E}[\langle s+n, s\rangle] = ||s||^2$$

$$\mathrm{var}(Z|H_1) = \mathrm{cov}(\langle s+n, s\rangle, \langle s+n, s\rangle) = \mathrm{cov}(\langle n, s\rangle, \langle n, s\rangle) = \sigma^2 ||s||^2$$

Reduces to basic Gaussian example with $m = ||s||^2,\ v^2 = \sigma^2 ||s||^2$

$$N(m, v^2)\ vs.\ N(0, v^2) \implies P_{e,ML} = P_{e|0} = P_{e|1} = Q\left(\frac{m}{2v}\right)$$

$$\boxed{P_{\mathrm{e,ML}} = P_{\mathrm{e|1}} = P_{\mathrm{e|0}} = Q\left(\frac{||s||}{2\sigma}\right)}$$

Performance of binary OOK

General binary signaling

$$H_1 : y(t) = s_1(t) + n(t)$$
$$H_0 : y(t) = s_0(t) + n(t)$$

ML Rule

$$\langle y, s_1 \rangle - \frac{||s_1||^2}{2} \mathop{\gtrless}_{H_0}^{H_1} \langle y, s_0 \rangle - \frac{||s_0||^2}{2}$$

Jointly Gaussian conditioned on each hypothesis, need to find means and covariances under each hypothesis to analyze

Simpler to rewrite in terms of a single decision statistic:

$$\langle y, s_1 - s_0 \rangle \mathop{\gtrless}_{H_0}^{H_1} \frac{||s_1||^2}{2} - \frac{||s_0||^2}{2}$$

Conditionally Gaussian under each hypothesis, can be analyzed as for OOK

General binary signaling (contd.)

Even simpler if we transform to OOK:

$$H_1 : y(t) = s_1(t) + n(t)$$
$$H_0 : y(t) = s_0(t) + n(t)$$

$$\xrightarrow{\tilde{y}(t) = y(t) - s_0(t)}$$

Invertible transformation, so optimal receiver must have the same performance in both cases

Binary OOK with $s(t) = s_1(t) - s_0(t)$

Performance of binary signaling in AWGN (with ML reception)

$$\boxed{P_{\mathrm{e,ML}} = P_{\mathrm{e}|1} = P_{\mathrm{e}|0} = Q\left(\frac{||s_1 - s_0||}{2\sigma}\right) = Q\left(\frac{d}{2\sigma}\right)}$$

$$\boxed{d = ||s_1 - s_0||}$$

Depends on distance between signals, normalized by noise standard deviation per dimension

Eb/N0 and power efficiency

Energy per bit, E_b

$$E_b = \frac{1}{2}(||s_0||^2 + ||s_1||^2)$$ (for equal priors)

Scale-invariant measure of power efficiency $$\eta_P = \frac{d^2}{E_b}.$$

(if we scale signals up by a factor of A, numerator and denominator both scale as A^2)

Rewrite performance in terms of Eb/N0 and power efficiency:

$$P_{e,ML} = Q\left(\sqrt{\frac{\eta_P E_b}{2N_0}}\right) = Q\left(\sqrt{\frac{d^2}{E_b}}\sqrt{\frac{E_b}{2N_0}}\right)$$

1) Performance depends on signal-to-noise *ratio Eb/N0* rather than on absolute strengths

2) For given Eb/N0, performance is better for higher $$\eta_P = \frac{d^2}{E_b}.$$

(which is why we call it power efficiency)

Performance computations for common binary signaling schemes

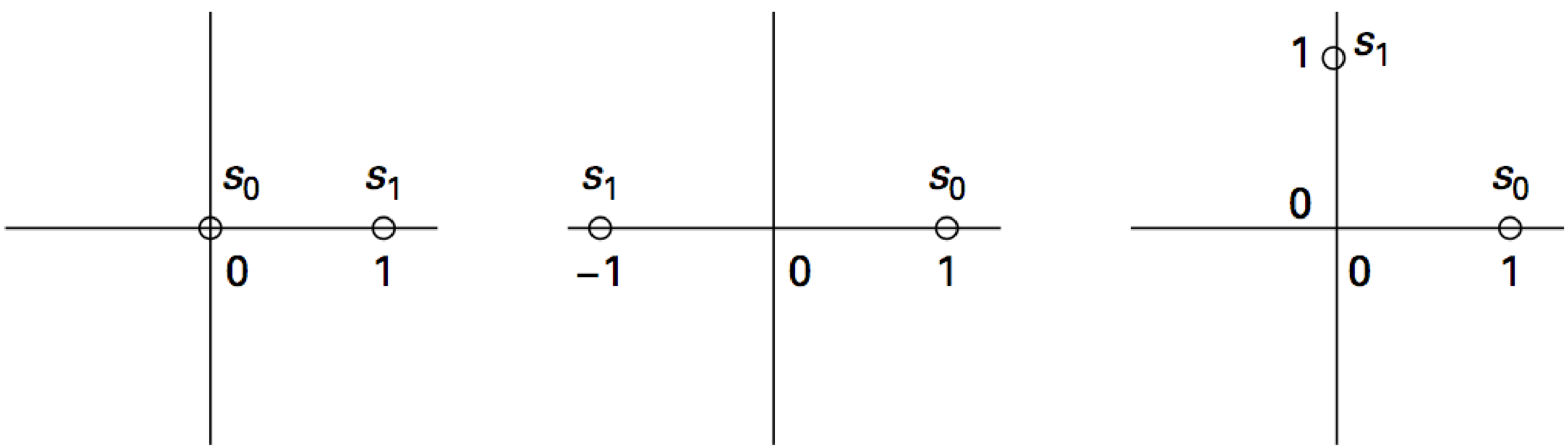
Can work in 2-d signal space

--Inner products, hence energies and distances, preserved in signal space

Can choose any convenient scaling for constellation

--Performance depends on constellation only through the scale-invariant power efficiency parameter

$$d = 1 \text{ and } E_b = 1/2(1^2 + 0^2) = 1/2$$

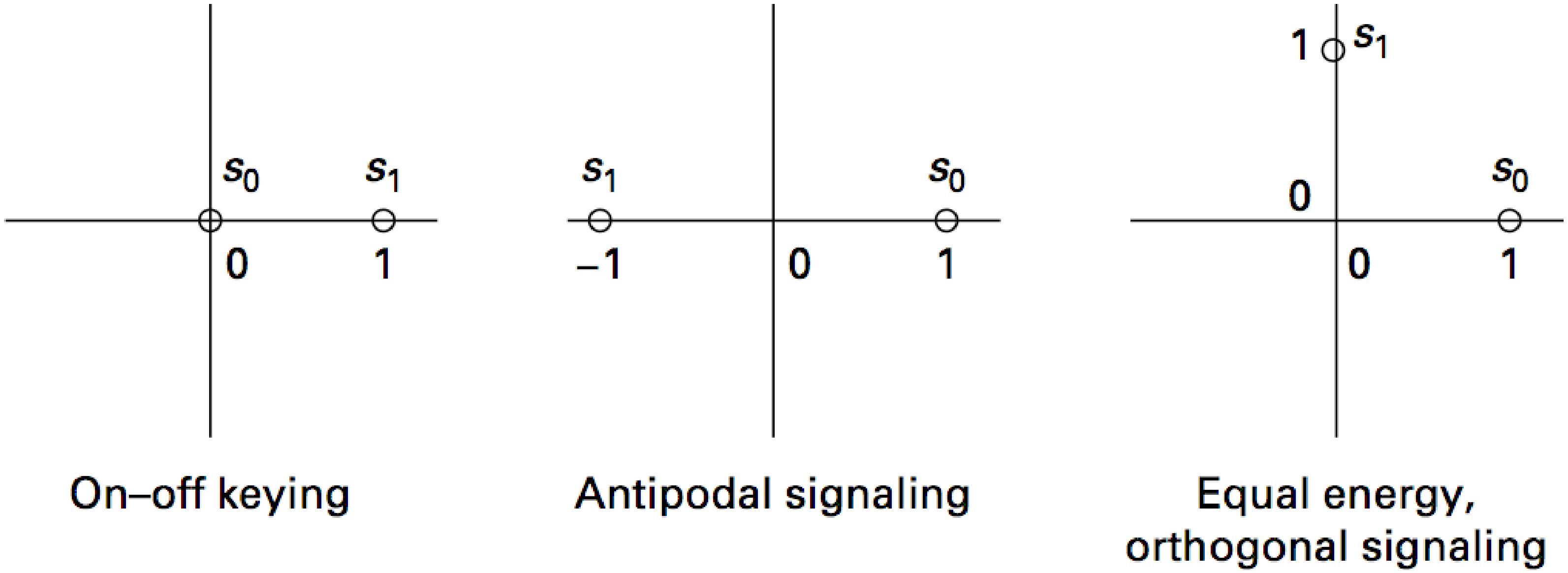


On–off keying

Antipodal signaling

Equal energy, orthogonal signaling

Performance computations (contd.)



OOK $P_{e,ML} = Q(\sqrt{E_b/N_0})$

$d = 1$ and $E_b = 1/2(1^2 + 0^2) = 1/2 \implies \eta_P = d^2/E_b = 2$

Antipodal $P_{e,ML} = Q(\sqrt{2E_b/N_0})$

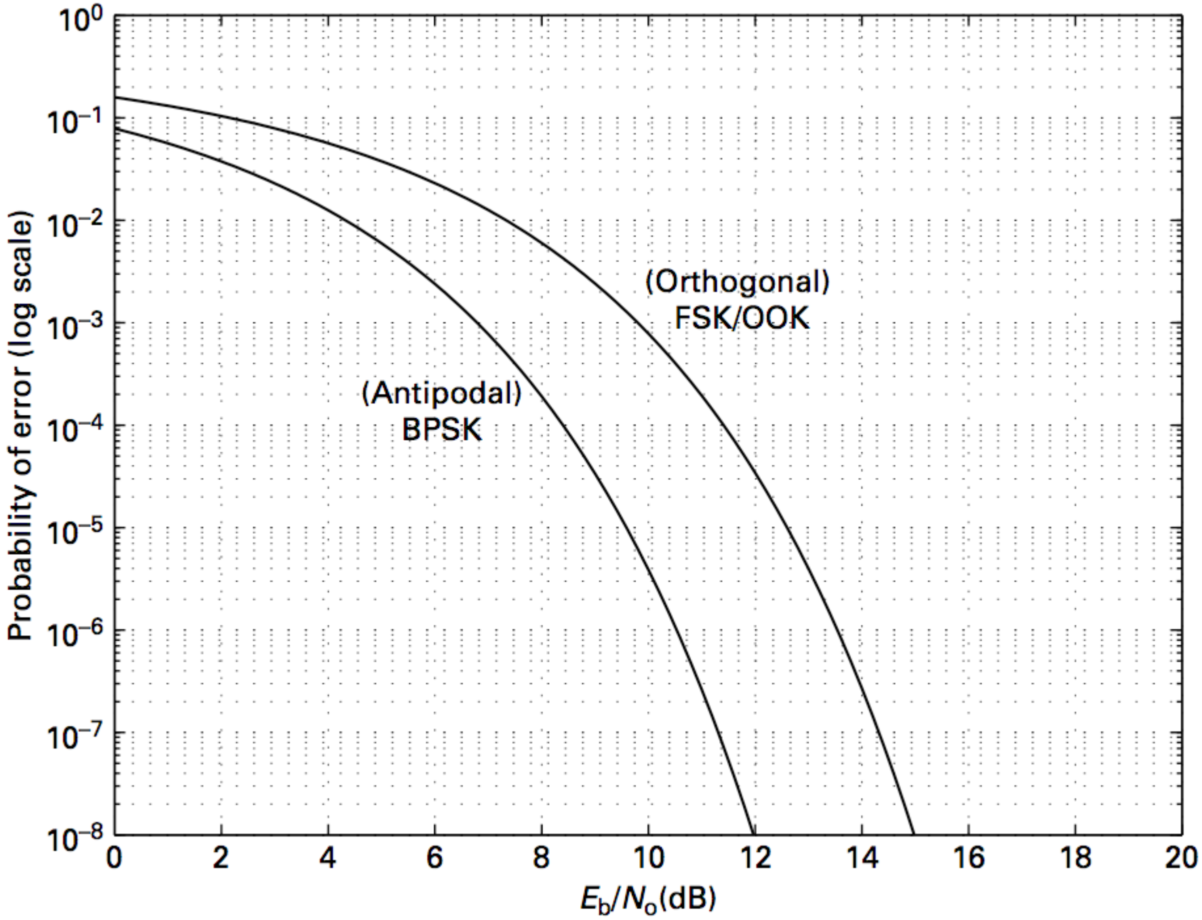
$d = 2$ and $E_b = 1/2(1^2 + (-1)^2) = 1 \implies \eta_P = d^2/E_b = 4$

3 dB better than OOK and orthogonal signaling

Equal energy orthogonal $P_{e,ML} = Q(\sqrt{E_b/N_0})$

$d = \sqrt{2}$ and $E_b = 1 \implies \eta_P = d^2/E_b = 2$

Error probability plots for binary signaling



Rapid decay of error prob with SNR, so y-axis is in log scale
Using dB on x-axis allows us to consider a large range of SNR

M-ary signaling

$$\delta_{\mathrm{ML}}(y) = \arg \max_{1 \le i \le M} Z_i \qquad Z_i = \langle y, s_i \rangle - \frac{1}{2}||s_i||^2.$$

ML rule in terms of correlator outputs

$$\delta_{\mathrm{ML}}(y) = \arg \min_{1 \le i \le M} D_i, \qquad D_i = ||y - s_i||.$$

ML rule in terms of distances

Both forms are going to be useful as we proceed.

ML decision regions

$$\Gamma_i = \{y : \delta_{\mathrm{ML}}(y) = i\} = \{y : Z_i \ge Z_j \text{ for all } j \ne i\} = \{y : D_i \le D_j \text{ for all } j \ne i\}$$

Steps:

1) Performance depends only on the ratio of signal inner products to noise strength, and hence only on Eb/N0 and the scale-invariant geometry of the constellation

2) Exact performance computation is typically difficult; involves integration of multi-dimensional Gaussian densities over decision regions

3) Analysis for binary signaling can be used as building block for bounds and approximations

What the performance depends on

$$P_{e|i} = P[y \notin \Gamma_i | i \text{ sent}] = P[Z_i < Z_j \text{ for some } j \neq i | i \text{ sent}]$$

Need joint distribution of the correlation decision stats conditioned on the hypotheses to evaluate this (never mind that the computation might be difficult)

Conditioning on $y(t) = s_i(t) + n(t)$ **(the *i*th hypothesis)**

$$Z_j = \langle y, s_j \rangle - \frac{1}{2}||s_j||^2 = \langle s_i, s_j \rangle + \langle n, s_j \rangle - \frac{1}{2}||s_j||^2$$

$\{Z_j\}$ **are jointly Gaussian (conditioned on the hypothesis), since *n* is Gaussian.**
Joint distribution of $\{Z_j\}$ **completely characterized by means and covariances.**

$$\mathbb{E}[Z_j] = \langle s_i, s_j \rangle - \frac{1}{2}||s_j||^2$$ **(conditioning suppressed from notation)**

$$\text{cov}(Z_j, Z_k) = \sigma^2 \langle s_k, s_j \rangle$$

Replacing Z_j **by** Z_j / σ^2 **does not change the decisions or the performance.**

But means and covariances now depend only on the normalized inner products:

$$\boxed{\{ \langle s_i, s_j \rangle / \sigma^2, 1 \leq i, j \leq M \}}$$ ← **These completely determine performance**

In other words, only SNR and geometry matter

Energy per symbol $E_s = \frac{1}{M}\sum_{i=1}^{M} ||s_i||^2$

Energy per bit $E_b = \frac{E_s}{\log_2 M}$

Performance depends only on noise-normalized inner products

$$\frac{\langle s_i, s_j \rangle}{\sigma^2} = \frac{\langle s_i, s_j \rangle}{E_b} \frac{2E_b}{N_0}$$

Hence it only depends on

E_b/N_0 and $\{ (s_i, s_j)/E_b \}$

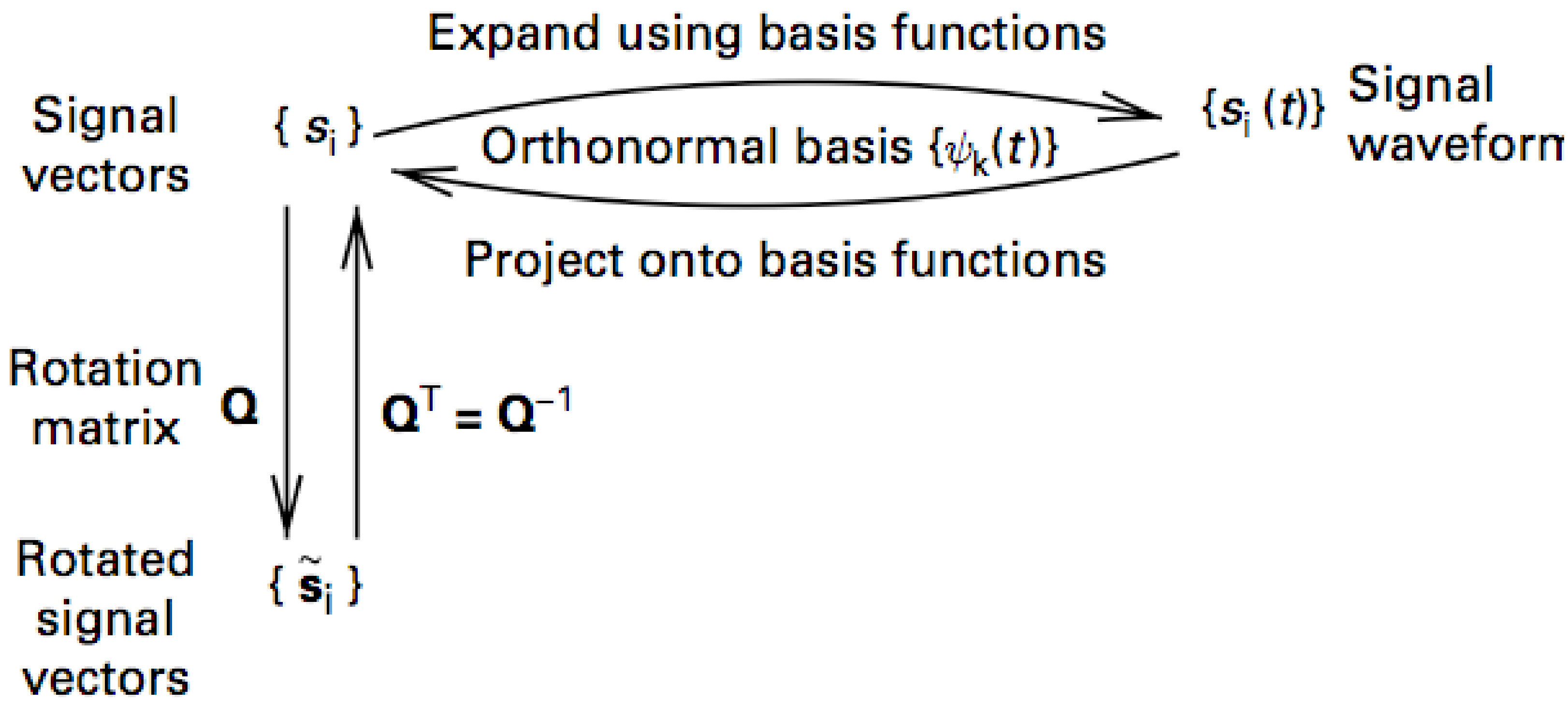
SNR

Scale-invariant, depends only on the constellation geometry

Performance is invariant under rotation

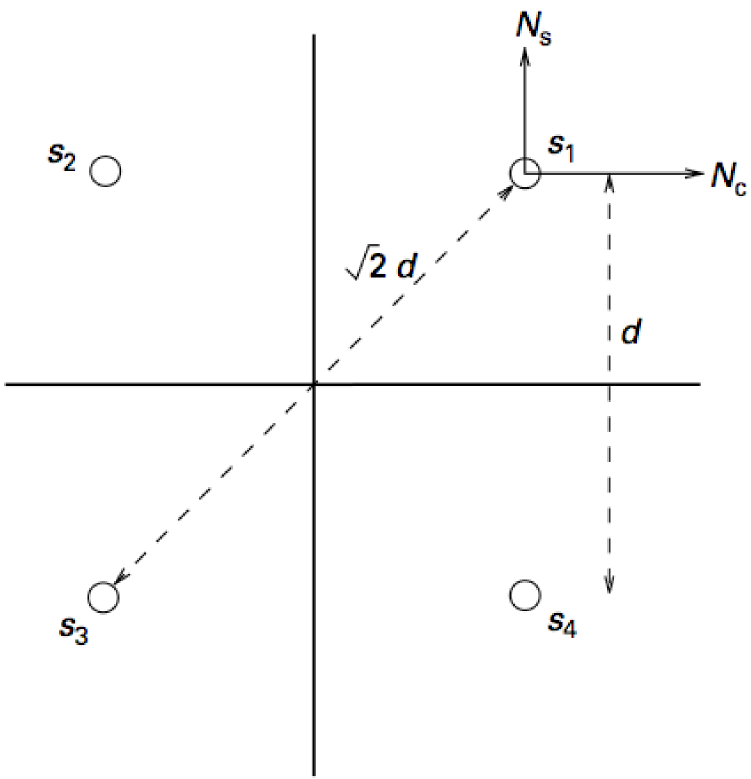
$\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ **Rotation matrix (a change of basis)**

Inner products do not change $\langle \mathbf{Q}\mathbf{s}_i, \mathbf{Q}\mathbf{s}_j \rangle = \mathbf{s}_j^T \mathbf{Q}^T \mathbf{Q}\mathbf{s}_i = \mathbf{s}_j^T \mathbf{s}_i = \langle \mathbf{s}_i, \mathbf{s}_j \rangle$



Performance unchanged under the transformations shown

QPSK: exact analysis



Exact analysis is possible for rectangular QAM.
QPSK or 4-QAM is the simplest example of this.

By symmetry, the conditional error probs are equal (and hence equal to the avg error prob)

Condition on s1 being sent.
Error occurs when noise pushes the received vector out of the first quadrant.

Conditional error probability:

$$P_{e|1} = P\left[N_c + \frac{d}{2} < 0 \text{ or } N_s + \frac{d}{2} < 0\right] = P\left[N_c + \frac{d}{2} < 0\right] + P\left[N_s + \frac{d}{2} < 0\right]$$
$$- P\left[N_c + \frac{d}{2} < 0 \text{ and } N_s + \frac{d}{2} < 0\right].$$

QPSK: exact analysis (contd.)

N_c, N_s are i.i.d. $N(0, \sigma^2)$

$$\Longrightarrow \quad P\left[N_c + \frac{d}{2} < 0\right] = P\left[N_s + \frac{d}{2} < 0\right] = Q\left(\frac{d}{2\sigma}\right)$$

and $$P_{e|1} = 2Q\left(\frac{d}{2\sigma}\right) - \left[Q\left(\frac{d}{2\sigma}\right)\right]^2$$

We now want to express this in terms of SNR $\quad \dfrac{d}{2\sigma} = \sqrt{\dfrac{d^2}{E_b}}\sqrt{\dfrac{E_b}{2N_0}},$

$$E_s = \frac{1}{M}\sum_{i=1}^{M} ||s_i||^2 = ||s_1||^2 = \left(\frac{d}{2}\right)^2 + \left(\frac{d}{2}\right)^2 = \frac{d^2}{2} \qquad E_b = \frac{E_s}{\log_2 M} = \frac{E_s}{\log_2 4} = \frac{d^2}{4}.$$

$$\Longrightarrow \quad d^2 / E_b = 4 \quad \Longrightarrow \quad \boxed{P_e = P_{e|1} = 2Q\left(\sqrt{\frac{2E_b}{N_0}}\right) - Q^2\left(\sqrt{\frac{2E_b}{N_0}}\right)}$$

Union Bound and Variants

QPSK (or in general, rectangular QAM) is the exception rather than the rule: exact performance analysis is not feasible for most constellations.
In practice, we may use bounds and approximations to gain quick insight, and follow up with detailed computer simulations.

Conditional error prob = Prob that some decision stat corresponding to a wrong hypothesis is bigger than the one for the true hypothesis = Prob of a union of events

$$P_{e|i} = P[\cup_{j \neq i} \{Z_i < Z_j\} | i \text{ sent}]$$

Union bound: P(union of events) bounded above by sum of probs of these events

$$P_{e|i} \leq \sum_{j \neq i} P[Z_i < Z_j | i \text{ sent}]$$

Pairwise error probability: error prob for binary hypothesis testing between H_i and H_j

Union Bound (contd.)

Pairwise error probabilities $$P[Z_i < Z_j | i \text{ sent}] = Q\left(\frac{||s_j - s_i||}{2\sigma}\right)$$

Union Bound for conditional error probabilities

$$P_{\mathrm{e}|i} \leq \sum_{j \neq i} Q\left(\frac{||s_j - s_i||}{2\sigma}\right) = \sum_{j \neq i} Q\left(\frac{d_{ij}}{2\sigma}\right)$$

Average over priors to get bound for average error probability

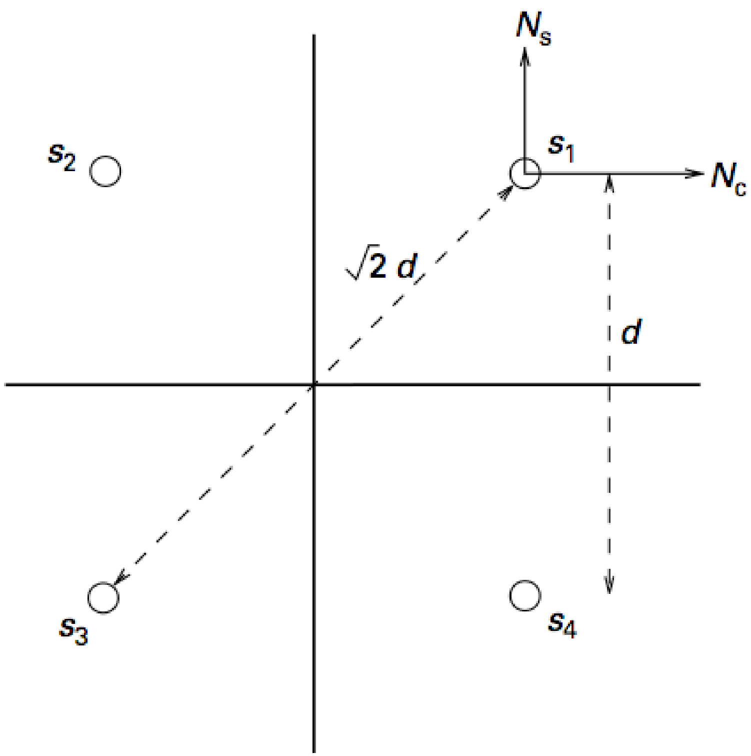
$$P_{\mathrm{e}} = \sum_i \pi(i) P_{\mathrm{e}|i} \leq \sum_i \pi(i) \sum_{j \neq i} Q\left(\frac{||s_j - s_i||}{2\sigma}\right) = \sum_i \pi(i) \sum_{j \neq i} Q\left(\frac{d_{ij}}{2\sigma}\right)$$

In terms of SNR and scale-invariant constellation parameters:

$$\boxed{\begin{aligned} P_{\mathrm{e}|i} &\leq \sum_{j \neq i} Q\left(\sqrt{d_{ij}^2/E_{\mathrm{b}}}\sqrt{E_{\mathrm{b}}/2N_0}\right), \\ P_{\mathrm{e}} = \sum_i \pi(i) P_{\mathrm{e}|i} &\leq \sum_i \pi(i) \sum_{j \neq i} Q\left(\sqrt{d_{ij}^2/E_{\mathrm{b}}}\sqrt{E_{\mathrm{b}}/2N_0}\right). \end{aligned}}$$

Union Bound

Back to QPSK



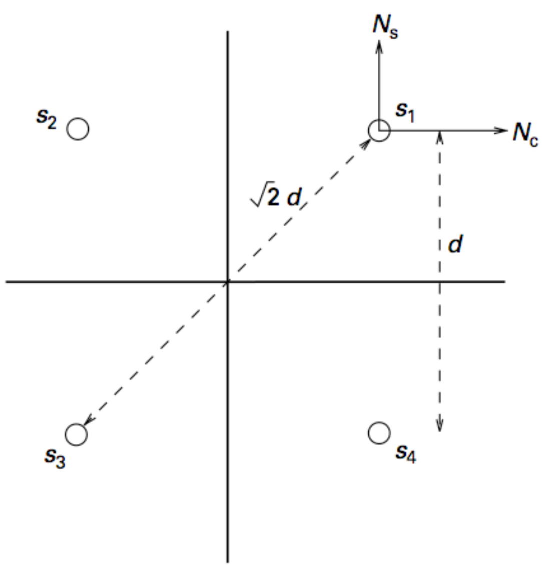
$$P_e = P_{e|1} \le Q\left(\frac{d_{12}}{2\sigma}\right) + Q\left(\frac{d_{13}}{2\sigma}\right) + Q\left(\frac{d_{14}}{2\sigma}\right) = 2Q\left(\frac{d}{2\sigma}\right) + Q\left(\frac{\sqrt{2}d}{2\sigma}\right)$$

$$d^2 / E_b = 4$$

$$\Longrightarrow P_e \le 2Q\left(\sqrt{\frac{2E_b}{N_0}}\right) + Q\left(\sqrt{\frac{4E_b}{N_0}}\right)$$ **QPSK union bound**

Improving the union bound

For large constellations, union bound can be quite loose (esp at low to moderate SNR).
We can sometimes prune it if we exploit the geometry of the constellation.
Let's first try with QPSK. Condition on first signal being sent.



In order to get out of the first quadrant (the correct decision region), hypothesis 1 must lose to either hypothesis 2 or hypothesis 4.
These pairwise decision boundaries correspond to the Q and I axes, respectively.
That is, if an error occurs, then the noise pushes us across either the I or Q axis.
"If A then B" means P(A) <= P(B).
We therefore get an improved bound on the conditional error prob:

$$P_{e|1} \le P[Z_2 > Z_1 \text{ or } Z_4 > Z_1 | s_1 \text{ sent}] \le P[Z_2 > Z_1 | s_1 \text{ sent}] + P[Z_4 > Z_1 | s_1 \text{ sent}] = 2Q\left(\frac{d}{2\sigma}\right).$$

Improving the union bound (contd.)

QPSK example leads to the following "intelligent union bound"

$$P_e = P_{e|1} \le 2Q\left(\sqrt{\frac{2E_b}{N_0}}\right) \quad \textbf{QPSK intelligent union bound.}$$

This generalizes to M-ary signaling:

$$\Gamma_i = \{y : \delta_{ML}(y) = i\} = \{y : Z_i \ge Z_j \ \text{ for all } j \in N_{ML}(i)\}$$

$N_{ML}(i)$ **Indices of neighbors of s_i that characterize its decision region**

If ith signal is sent, then we make an error if and only if we lose to at least one of these neighbors. The conditional error probability is therefore given by:

$$P_{e|i} = P[y \notin \Gamma_i | i \text{ sent}] = P[Z_i < Z_j \text{ for some } j \in N_{ML}(i) | i \text{ sent}]$$

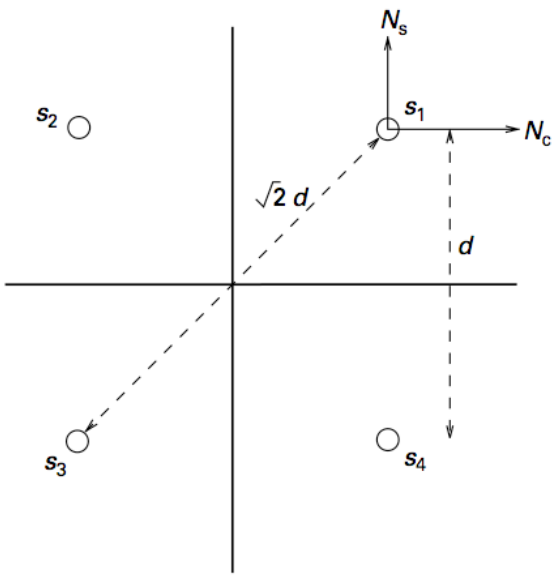
Now apply a union bound: $$P_{e|i} \le \sum_{j \in N_{ML}(i)} Q\left(\frac{||s_j - s_i||}{2\sigma}\right)$$

Intelligent Union Bound

$$P_{e|i} \leq \sum_{j \in N_{ML}(i)} Q\left(\frac{||s_j - s_i||}{2\sigma}\right) = \sum_{j \in N_{ML}(i)} Q\left(\sqrt{\frac{d_{ij}^2}{E_b}}\sqrt{\frac{E_b}{2N_0}}\right)$$

Includes only terms corresponding to neighbors defining the decision boundaries

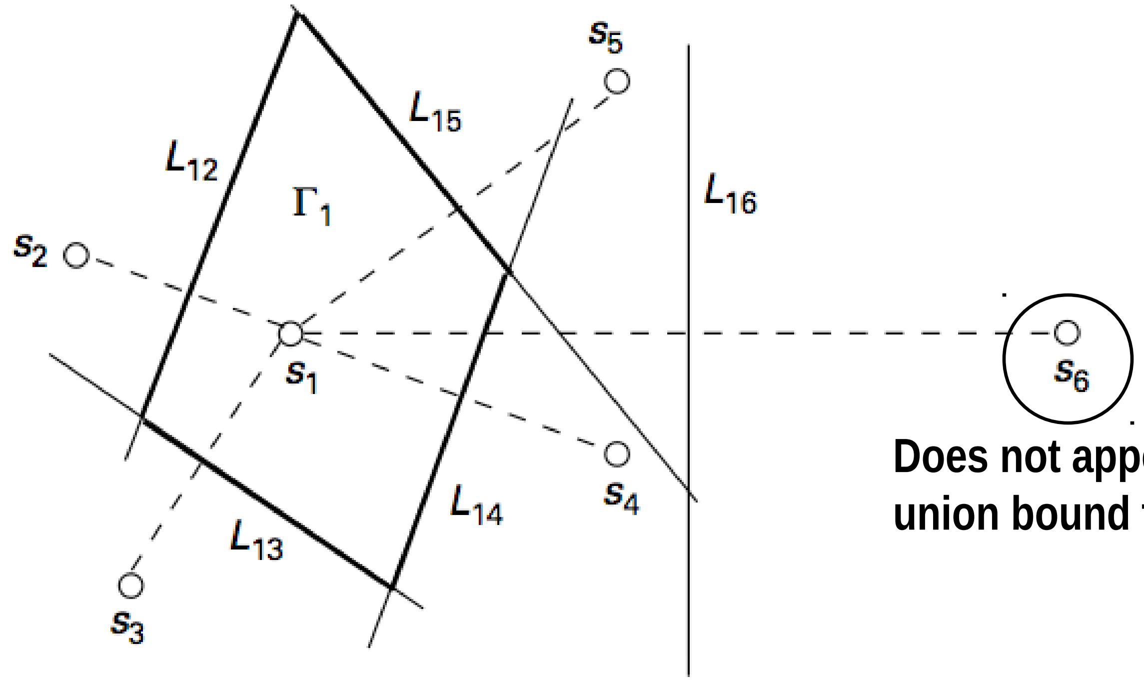
QPSK yet again:



$$N_{ML}(1) = \{2, 4\}$$

$$P_{e|1} \leq 2Q\left(\frac{d}{2\sigma}\right)$$

Intelligent union bound: another example



Does not appear in the intelligent union bound for $P_{e|1}$

$$P_{\mathrm{e}|1} \leq Q\left(\frac{d_{12}}{2\sigma}\right) + Q\left(\frac{d_{13}}{2\sigma}\right) + Q\left(\frac{d_{14}}{2\sigma}\right) + Q\left(\frac{d_{15}}{2\sigma}\right) + Q\left(\frac{d_{16}}{2\sigma}\right)$$

Standard union bound

$$N_{\mathrm{ML}}(1) = \{2, 3, 4, 5\}$$

$$P_{\mathrm{e}|1} \leq Q\left(\frac{d_{12}}{2\sigma}\right) + Q\left(\frac{d_{13}}{2\sigma}\right) + Q\left(\frac{d_{14}}{2\sigma}\right) + Q\left(\frac{d_{15}}{2\sigma}\right)$$

Intelligent union bound

Nearest neighbors approximation

Generic term used to denote approximations based on pruning terms from the union bound. For regular constellations in which each signal point has a number of neighbors at the minimum distance from it, we have

$$P_{e|i} \approx N_{d_{\min}}(i) Q\left(\frac{d_{\min}}{2\sigma}\right)$$

$N_{d_{\min}}(i)$ number of nearest neighbors of s_i

Averaging over *i*,

$$\boxed{P_e \approx \bar{N}_{d_{\min}} Q\left(\frac{d_{\min}}{2\sigma}\right)}$$

$\bar{N}_{d_{\min}}$ average number of nearest neighbors

Good high SNR approximation:

$Q(x) \sim e^{-x^2/2}$ **Decays rapidly in its argument, so terms in union bound corresponding to smallest arguments in the Q function dominate at high SNR**

Nearest neighbors approx (contd.)

Nearest neighbors approximation often coincides with intelligent union bound for common constellations (PSK, QAM) for which ML decision boundaries are governed by the nearest neighbors

Example: QPSK

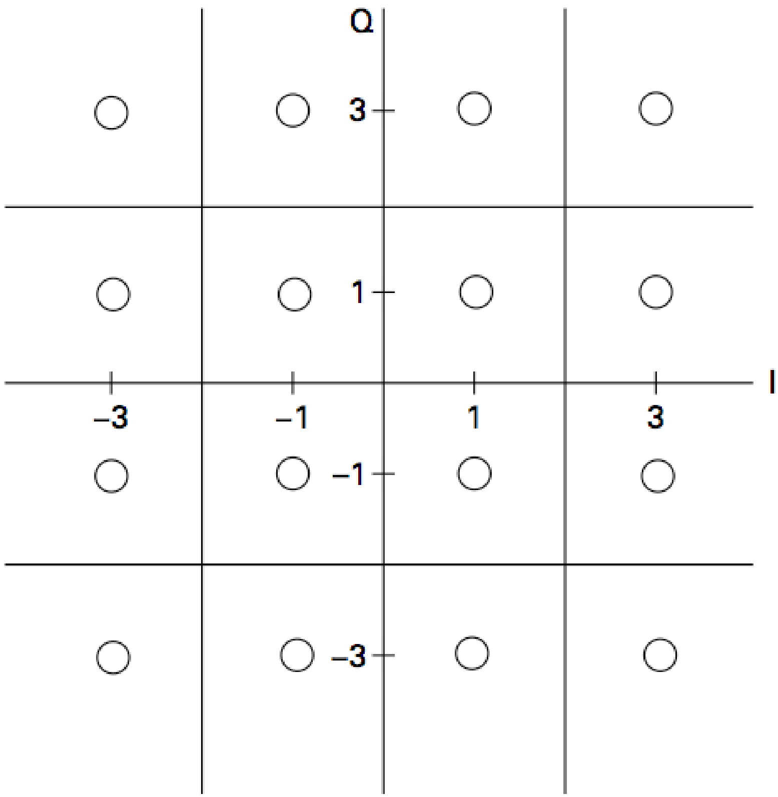
$$N_{d_{\min}}(i) \equiv 2 = \bar{N}_{d_{\min}}$$ and $$\frac{d_{\min}^2}{E_b} = 4$$ $\Longrightarrow$ $$\boxed{P_e \approx 2Q\left(\sqrt{\frac{2E_b}{N_0}}\right)}$$

Power efficiency $$\eta_P = \frac{d_{\min}^2}{E_b}$$

Can rewrite nearest neighbors approx in terms of power efficiency and SNR

$$\boxed{P_e \approx \bar{N}_{d_{\min}} Q\left(\sqrt{\frac{\eta_P E_b}{2N_0}}\right)}$$

16-QAM performance analysis



Nearest nbrs approx coincides with intelligent union bound (ML decision regions determined by nearest nbrs)

$$\boxed{\bar{N}_{d_{\min}} = 3.}$$

(2 nbrs for 4 outermost pts, 4 nbrs for 4 innermost pts, 3 nbrs for remaining 8 pts, gives avg of 3)

$$\boxed{d_{\min} = 2}$$

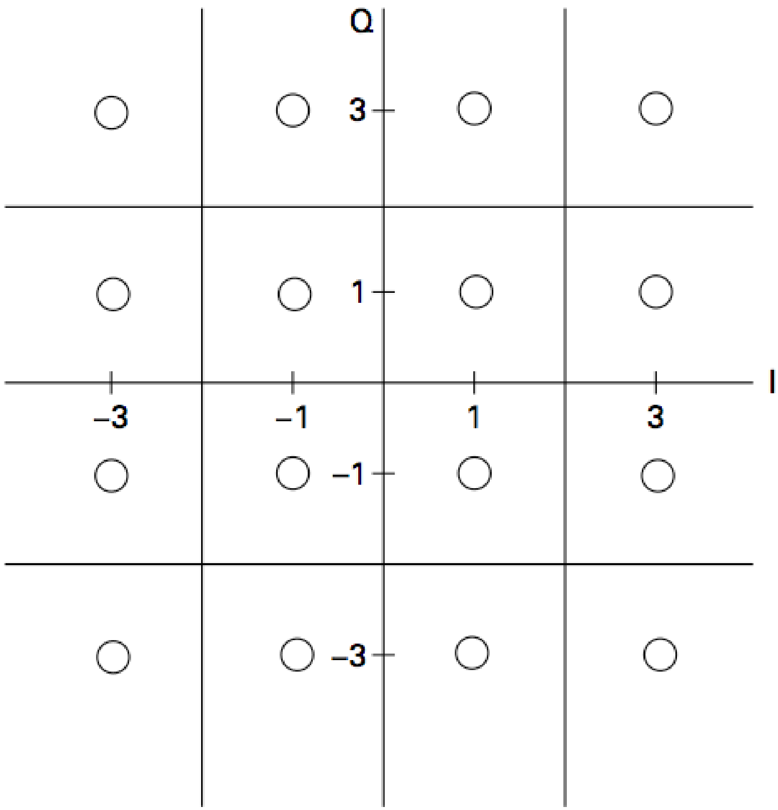
(for the scaling shown)

$$E_s = \text{average energy of I component} + \text{average energy of Q component}$$
$$= 2(\text{average energy of I component})$$

$$\text{average energy of I component} = \frac{1}{2}(1^2 + 3^2) = 5$$

(takes values -3,-1,+1,+3 with equal probability)

16-QAM performance analysis (contd.)



$$E_s = 10$$

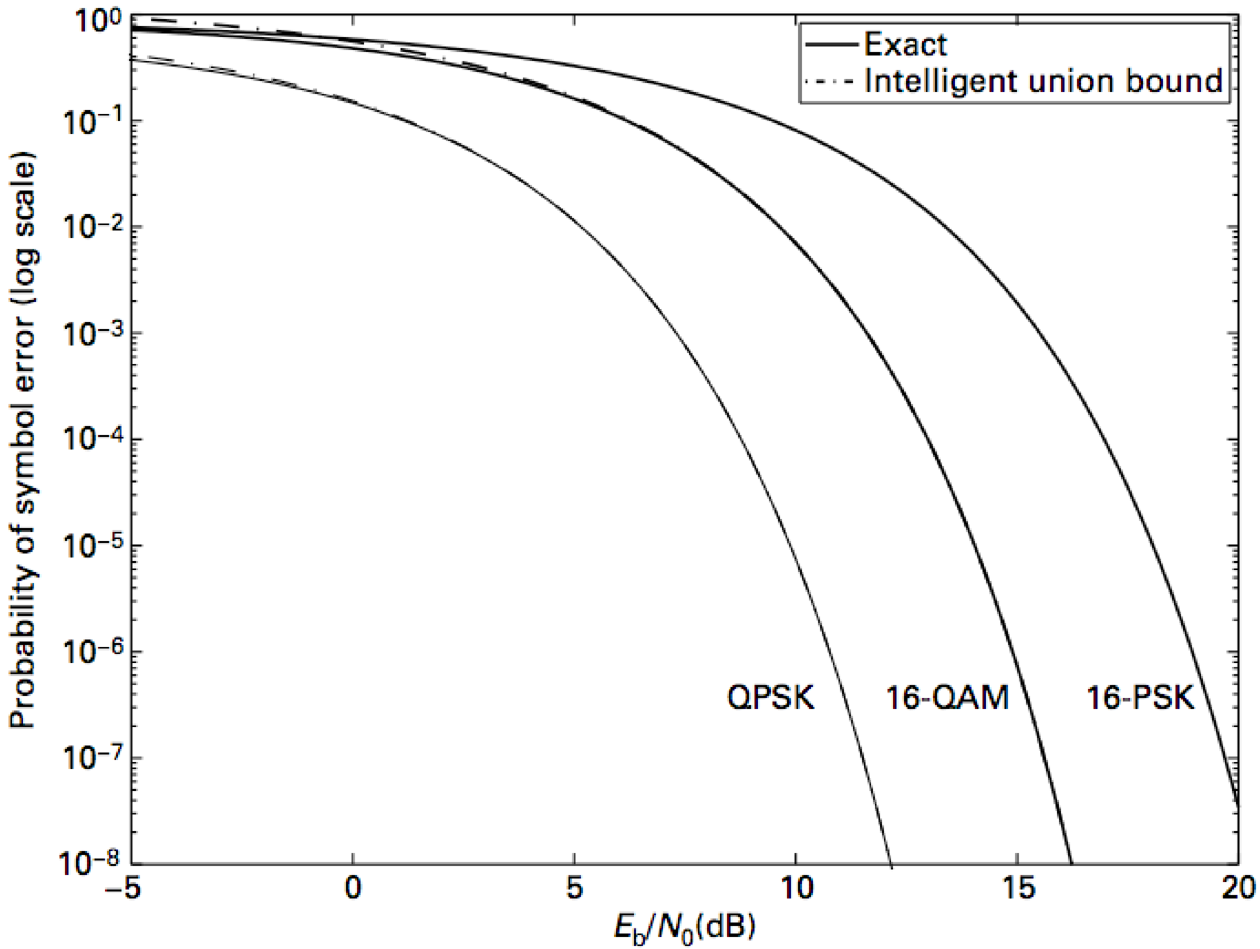
$$E_b = \frac{E_s}{\log_2 M} = \frac{10}{\log_2 16} = \frac{5}{2}$$

$$\boxed{\eta_P = \frac{d_{\min}^2}{E_b} = \frac{2^2}{\frac{5}{2}} = \frac{8}{5}}$$

$$\boxed{P_e(\text{16-QAM}) \approx 3Q\left(\sqrt{\frac{4E_b}{5N_0}}\right)}$$

Compare 16-QAM with QPSK: power-bandwidth tradeoff
Bandwidth efficiency 2x better: 4 bits/2 real dimensions vs. 2 bits/2 real dimensions
Power efficiency 4 dB worse: 1.6 vs. 4

Error probability plots



1) Exact error probability and intelligent union bound are very close
2) Performance gap accurately predicted by power efficiency
(e.g., 4 dB between 16-QAM and QPSK)

M-ary orthogonal signaling

2-d constellations are more and more bandwidth efficient as #pts increases.
M-ary orthogonal signaling is at the other extreme: becomes more and more power-efficient (in fact, information-theoretically optimal in terms of power efficiency as M gets large) and less and less bandwidth efficiency for large M.

Scale signals to unit energy $E_s = 1$,

$$E_b = 1/(\log_2 M) \qquad d_{\min}^2 = 2$$

$$\eta_P = \frac{d_{\min}^2}{E_b} = 2\log_2 M$$

Power efficiency gets better as M increases

$$\eta_B = \frac{\log_2 M}{M}$$

Bandwidth efficiency (bits/dimension) goes to zero as M increases

Can we get reliable performance at arbitrarily small Eb/N0 as M gets large?
Actually, no--there is a threshold that falls out from the error prob analysis.

M-ary orthogonal signaling: error prob

Exact error probability (see Problem 3.25)

$$P_e = (M-1) \int_{-\infty}^{\infty} [\Phi(x)]^{M-2}\ \Phi(x-m) \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\ dx$$

Thresholding behavior (see Problem 3.26)

$$\lim_{M \to \infty} P_e = \begin{cases} 0, & \frac{E_b}{N_0} > \ln 2, \\ 1, & \frac{E_b}{N_0} < \ln 2. \end{cases}$$

Information theory says (Chapter 6): -1.6 dB, or ln 2, is the smallest Eb/N0 at which we can communicate reliably, even if welet the bandwidth efficiency go to zero. M-ary orthogonal signaling as M gets large approaches this limit, hence is asymptotically optimal in terms of power efficiency.

We can show the thresholding phenomenon quickly using the union bound (but are a factor of 2 off on the threshold)…

Union bound for M-ary orthogonal signaling

Union bound $$P_e \le (M-1)Q\left(\sqrt{\frac{E_s}{N_0}}\right) = (M-1)Q\left(\sqrt{\frac{E_b \log_2 M}{N_0}}\right)$$

Letting M gets large gives infinity times zero, or zero by zero, so we use L'Hospital's rule to evaluate the limit

Define $$f_1(M) = Q\left(\sqrt{\frac{E_b}{N_0}\log_2 M}\right) = Q\left(\sqrt{\frac{E_b}{N_0}\frac{\ln M}{\ln 2}}\right), \quad f_2(M) = \frac{1}{M-1}.$$

$$\frac{dQ(x)}{dx} = -\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \implies \frac{df_1(M)}{dM} = \left[\frac{d}{dM}\left(\sqrt{\frac{E_b}{N_0}\frac{\ln M}{\ln 2}}\right)\right]\left[-\frac{1}{\sqrt{2\pi}}e^{-\left(\sqrt{\frac{E_b}{N_0}\frac{\ln M}{\ln 2}}\right)^2/2}\right]$$

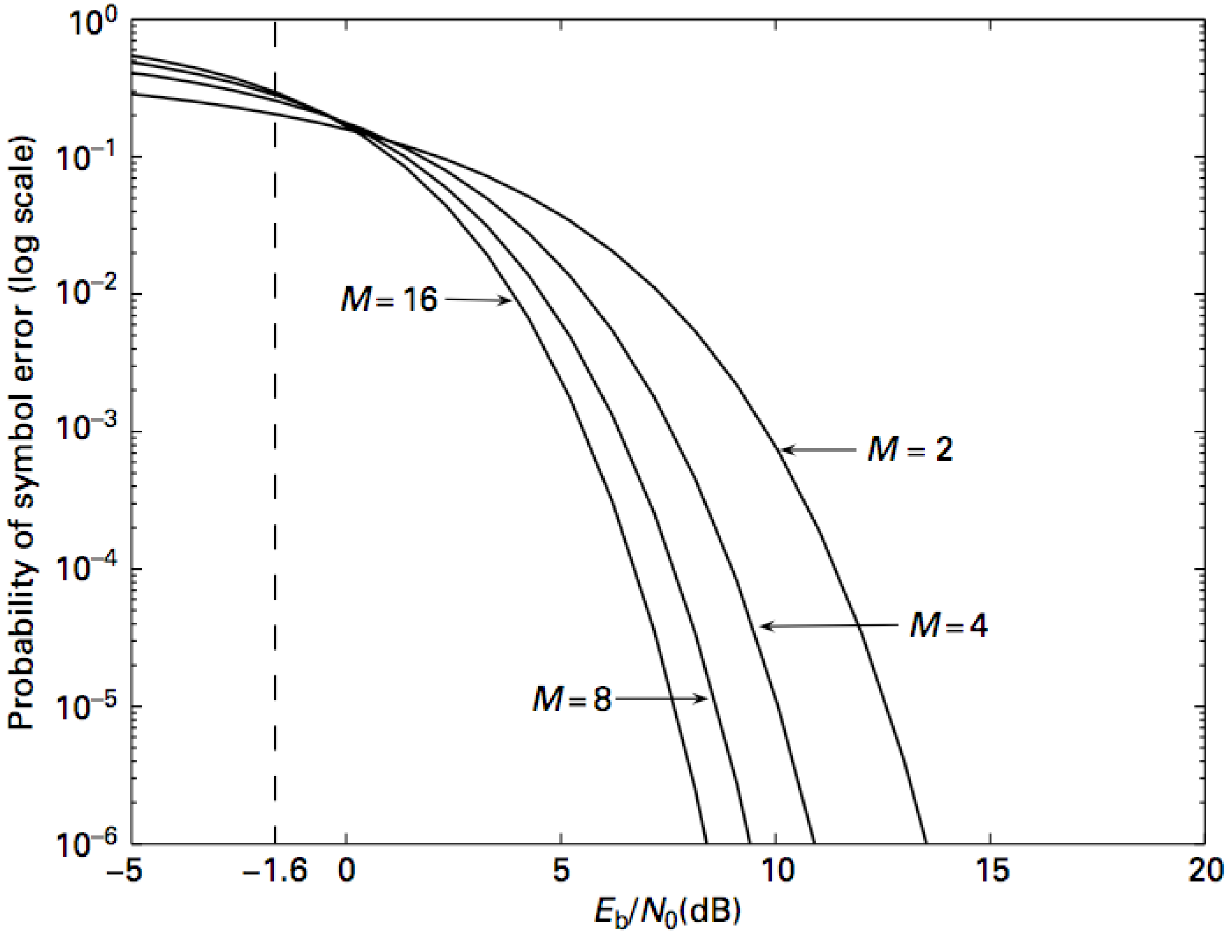
$$\frac{df_2(M)}{dM} = -(M-1)^{-2} \approx -M^{-2}$$

$$\implies \lim_{M\to\infty} P_e \le \lim_{M\to\infty}\frac{\frac{df_1(M)}{dM}}{\frac{df_2(M)}{dM}} = \lim_{M\to\infty} A(\ln M)^{-\frac{1}{2}} M^{1-\frac{E_b/N_0}{2\ln 2}}$$

We conclude that if Eb/N0 > 2 ln 2, then error prob goes to zero as M increases. Shows we can make comm arbitrarily reliable as long as Eb/N0 is above a threshold (the threshold is off by a factor of two).

If Eb/N0 < 2 ln 2, bound blows up, but we can't say what happens to the error prob (which cannot exceed one).

Error prob plots for M-ary orthogonal signaling

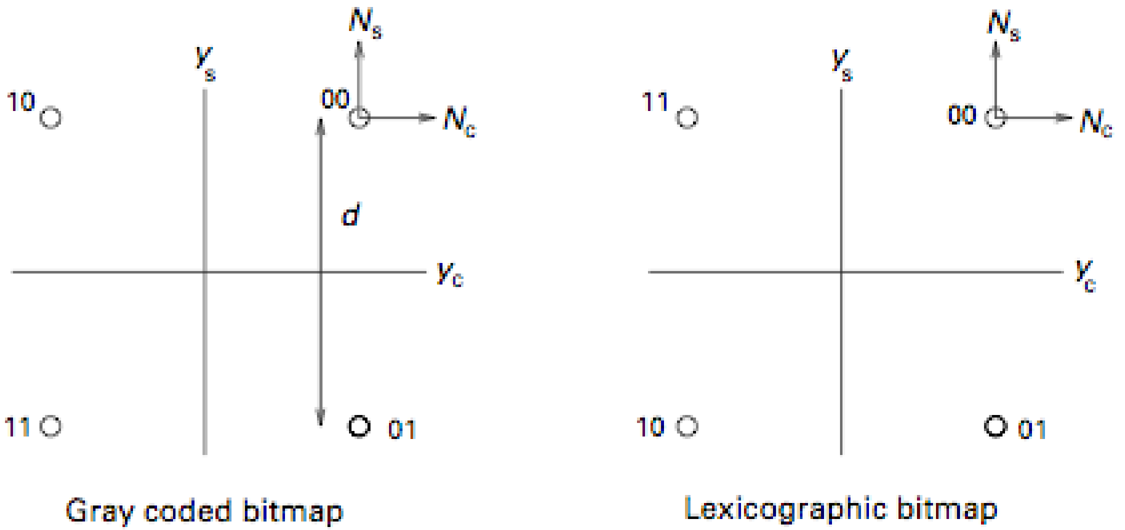


Convergence to -1.6 dB limit is quite slow (9 dB away for M=16 at error prob of 10^{-6}).
Not surprising: need long blocklengths to get to information-theoretic limits

Bit-level demodulation

Symbol error probability depends only on constellation geometry and SNR.
Bit error probability depends, in addition, on the bits-to-symbol mapping.

Example: Two possible bitmaps for QPSK



Bit error probs for Gray code

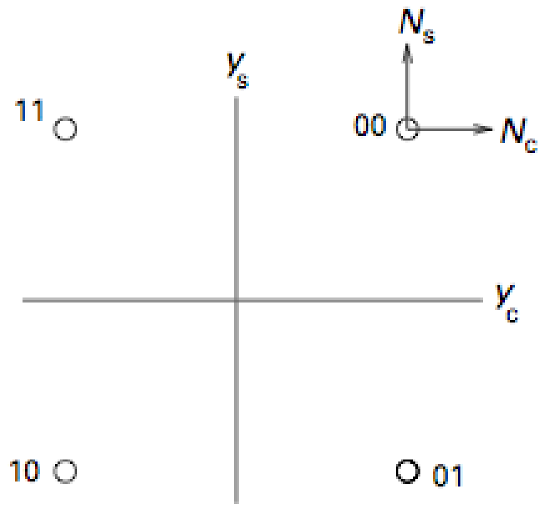
$$p_b = p_1 = p_2 = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$$

Error prob for each bit is the same
(Gray coded QPSK is equivalent to independent BPSK along I and Q channels)

$$P[\hat{b}[1] = 1|00 \text{ sent}] = P[\text{ML decision is 10 or 11}|00 \text{ sent}]$$

$$= P\left[N_c < -\frac{d}{2}\right] = Q\left(\frac{d}{2\sigma}\right) = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$$

Effect of bitmaps



Lexicographic bitmap performs worse than Gray code

$$p_1 = Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$$

Same as for Gray code

Average bit error prob approx 1.5x worse than Gray code

$$P[\hat{b}[2] \neq b[2] | 00 \text{ sent}] = P[\hat{b}[2] = 1 | 00 \text{ sent}]$$
$$= P[\text{ML decision is 01 or 11} | 00 \text{ sent}]$$
$$= P\left[N_c < -\frac{d}{2}, N_s > -\frac{d}{2}\right] + P\left[N_c > -\frac{d}{2}, N_s < -\frac{d}{2}\right].$$

$$p_2 = 2Q\left(\frac{d}{2\sigma}\right)\left[1 - Q\left(\frac{d}{2\sigma}\right)\right] = 2Q\left(\sqrt{\frac{2E_b}{N_0}}\right)\left[1 - Q\left(\sqrt{\frac{2E_b}{N_0}}\right)\right] \approx 2Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$$

2x worse than for Gray code

Gray code

Gray code: Bit assignment for neighboring signal differ by exactly one bit
--does not always exist
--may be hard to find for higher-dimensional signal constellations
--typically easy to do for regular 2-d constellations

While we illustrate its utility using high SNR approximations,
Gray coding is very useful for low SNR heavily coded systems (e.g., Gray coded 2-d constellation followed by a turbo-like code)

Gray coding is by no means the only sensible option
--Depends on the system goals (do we have special bits to protect?) and coding strategy
--We may deliberately map bits to symbols so that different bits get different levels of protection
--Such asymmetric mappings play a role for unequal error protection, trellis coded modulation, multilevel coding

Bit error probability

Exact analysis was possible for QPSK…can be extended to other rectangular QAM constellations
Exact analysis not possible for larger PSK constellations
Bounds are hard to come by; geometry of bit errors not as straightforward as that for symbol errors.
Nearest neighbors approximation is a good approach to obtaining quick estimates for regular constellations.

Distance from its nearest neighbors for the constellation point labeled by b

Conditional error prob for the *i*th bit

$$P(b_i \text{ wrong} | \mathbf{b} \text{ sent}) \approx N_{d_{\min}}(\mathbf{b}, i) Q\left(\frac{d_{\min}(\mathbf{b})}{2\sigma}\right)$$

Number of nearest neighbors of point labeled b whose label differs in position *i*

Remove conditioning $$P(b_i \text{ wrong}) \approx \frac{1}{2^n} \sum_{\mathbf{b}} N_{d_{\min}}(\mathbf{b}, i) Q\left(\frac{d_{\min}(\mathbf{b})}{2\sigma}\right)$$

Average over bits $$P(\text{bit error}) = \frac{1}{n} \sum_{i=1}^{n} P(b_i \text{ wrong})$$

Bit error probability for Gray coding

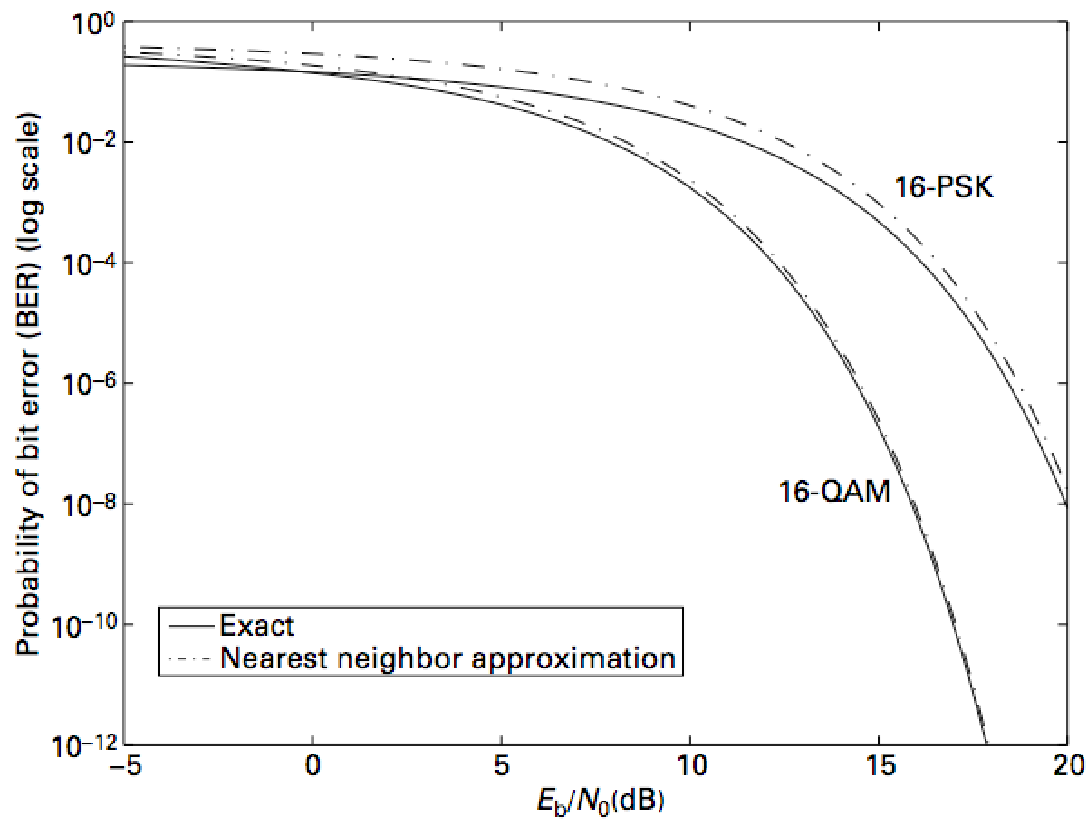
With Gray coding $N_{d_{\min}}(\mathbf{b}, i) \leq 1$

(if there are two neighbors that differ in the *i*th bit from b, then one of them must differ by at least one more bit, which violates the defn of Gray code)

Nearest neighbors approx with Gray coding

$$\boxed{P(\text{bit error}) \approx Q\left(\sqrt{\frac{\eta_P E_b}{2N_0}}\right)}$$

Accurate and slightly pessimistic for 16-QAM and 16-PSK (hence good design tool)



BER for M-ary orthogonal signaling

M-ary orthogonal signaling differs drastically from 2-D constellations

--Gray code out of the question (all M-1 neighbors are equidistant, but only log M bits to encode)

--By symmetry, decoding to each neighbor equally likely given that a symbol error happens.

--M/2 of the neighbors correspond to error in a given bit: each bit partitions the constellation into two equal subsets

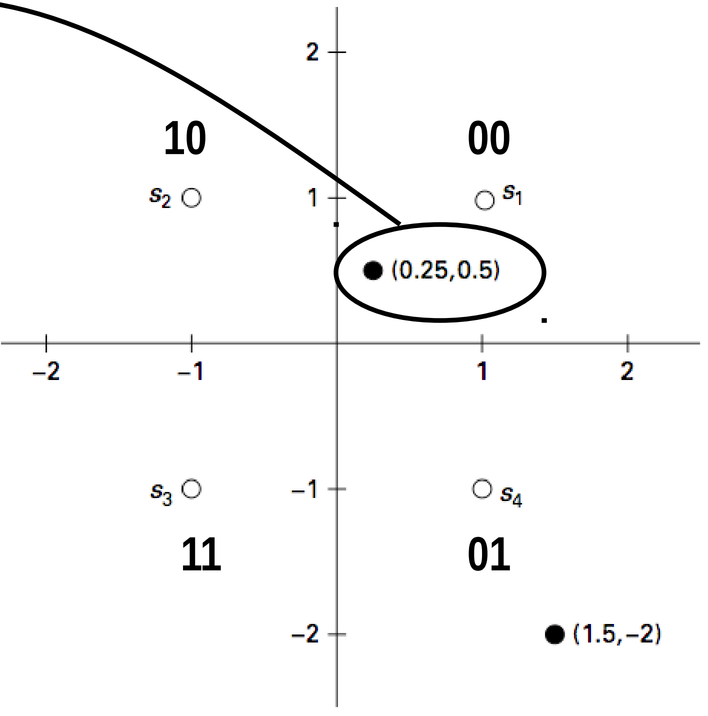
$$\boxed{P(\text{bit error}) = \frac{\frac{M}{2}}{M-1}\, P(\text{symbol error})}$$

Bit-level soft decision computation: example

We discussed earlier how to compute symbol-level soft decisions

$\sigma^2 = 1$

$\pi(i\|\mathbf{y})$	$\mathbf{y} = (0.25, 0.5)$
1	0.455
2	0.276
3	0.102
4	0.167



Bit-level soft decisions can now be computed as follows:

$$P[b[1]=0|y] = P[s_1 \text{ or } s_4 \text{ sent}|y] = \pi_1(y) + \pi_4(y) = 0.455 + 0.167 = 0.622$$

$$P[b[2]=0|y] = P[s_1 \text{ or } s_2 \text{ sent}|y] = \pi_1(y) + \pi_2(y) = 0.455 + 0.276 = 0.731$$

Log likelihood ratio (LLR) computation

Often useful to work in log domain when combining probabilities of independent events, since we can add rather than multiply (such an independence approximation is key to “turbo” or iterative decoding…Chapter 7).

It is convenient to define soft decisions in terms of LLRs:

$$\boxed{\mathrm{LLR}(b) = \log \frac{P[b=0]}{P[b=1]}}$$

In our QPSK example,

$$\mathrm{LLR}(b[1]|\mathbf{y}) = \log \frac{0.622}{1-0.622} = 0.498.$$

For Gray coded QPSK, can compute bit LLRs directly, without going through symbol-level soft decisions, since it is equivalent to independent BPSK along I and Q.

LLRs for BPSK

Bit $b \in \{0, 1\}$ mapped to $(-1)^b \in \{-1, +1\}$

and sent over AWGN channel

$$p(y\,|\,0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-A)^2}{2\sigma^2}\right)$$

$$p(y\,|\,1) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x+A)^2}{2\sigma^2}\right)$$

Model for decision statistic

$$Y = \begin{cases} A + N, & b = 0 \\ -A + N, & b = 1, \end{cases}$$

$$N \sim N(0, \sigma^2)$$

Use Bayes' rule to compute posterior probs

$$P[b|y] = P[b]p(y|b)/p(y)$$

Now compute LLR

$$\mathrm{LLR}(b|y) = \log \frac{P[b=0|y]}{P[b=1|y]} = \log \frac{\pi_0 p(y|0)}{\pi_1 p(y|1)}$$

$$\boxed{\mathrm{LLR}(b|y) = \mathrm{LLR}_{\mathrm{prior}}(b) + \frac{2Ay}{\sigma^2}}$$

$$\mathrm{LLR}_{\mathrm{prior}} = \log \pi_0 / \pi_1$$

LLR contributions due to prior and observation add up: a compelling reason for going to the log domain.

Link Budget Analysis

From performance analysis to physical transceiver specs

- Link budget analysis connects performance analysis to physical transceiver specs
  - Transmit power
  - Antenna directivities at Tx and Rx
  - Quality of receiver circuitry (how much noise do we incur)
  - Desired link range

Taking stock of what we know

- Given bit rate and signal constellation (including any error correction code that is used), we know the required modulation degrees of freedom per unit time (the symbol rate)
  - This gives minimum Nyquist bandwidth B_{min}
  - Now factor in excess bandwidth $B = (1 + a)B_{\mathrm{min}}$
- Given the constellation and desired BER, can find Eb/N0 required, and hence the required SNR
  - SNR=(energy/bit * bits/sec)/(noise PSD*bandwidth)

$$\mathrm{SNR} = E_{\mathrm{b}} R_{\mathrm{b}} / N_0 B \quad \Longrightarrow \quad \mathrm{SNR}_{\mathrm{reqd}} = \left(\frac{E_{\mathrm{b}}}{N_0}\right)_{\mathrm{reqd}} \frac{R_{\mathrm{b}}}{B}$$

- Given receiver noise figure and bandwidth, can find noise power

$$P_{\mathrm{n}} = N_0 B = k T_0 10^{F/10} B$$

We can now specify the receiver sensitivity (min required received power):

$$\boxed{P_{\mathrm{RX}}(\mathrm{min}) = \mathrm{SNR}_{\mathrm{reqd}} P_{\mathrm{n}} = \left(\frac{E_{\mathrm{b}}}{N_0}\right)_{\mathrm{reqd}} \frac{R_{\mathrm{b}}}{B} P_{\mathrm{n}}.}$$

Receiver sensitivity

$$\boxed{P_{\mathrm{RX,dBm}}(\mathrm{min}) = 10 \log_{10} P_{\mathrm{RX}}(\mathrm{min})(\mathrm{mW})}$$

(typically quoted in dBm)

Friis' Formula (free space propagation)

Carrier wavelength

$$P_{\mathrm{RX}} = P_{\mathrm{TX}} G_{\mathrm{TX}} G_{\mathrm{RX}} \frac{\lambda^2}{16\pi^2 R^2}$$

Transmit power

Transmit antenna directivity

Receive antenna directivity

Range

Antenna directivity (gain with respect to isotropic antenna) is often expressed in dB scale

$$G_{\mathrm{dBi}} = 10 \log_{10} G$$

Friis' Formula in dB scale

$$P_{\mathrm{RX,dBm}} = P_{\mathrm{TX,dBm}} + G_{\mathrm{TX,dBi}} + G_{\mathrm{RX,dBi}} + 10 \log_{10} \frac{\lambda^2}{16\pi^2 R^2}$$

Link Budget Equations

More generally, the received power is given by

$$P_{\mathrm{RX,dBm}} = P_{\mathrm{TX,dBm}} + G_{\mathrm{TX,dB}} + G_{\mathrm{RX,dB}} - L_{\mathrm{path,dB}}(R)$$

$$L_{\mathrm{path,dB}}(R) = -10\log_{10}\frac{\lambda^2}{16\pi^2 R^2}$$

for free space propagation

Add on a link margin (to compensate for unforseen or unmodeled performance losses) to get the link budget:

$$\boxed{P_{\mathrm{TX,dBm}} = P_{\mathrm{RX,dBm}}(\mathrm{min}) - G_{\mathrm{TX,dB}} - G_{\mathrm{RX,dB}} + L_{\mathrm{path,dB}}(R) + L_{\mathrm{margin,dB}}}$$

Example: link budget for a WLAN

WLAN link at 5 GHz

20 MHz bandwidth, QPSK, excess bandwidth 33%, receiver noise figure 6 dB

BIT RATE

Symbol rate $R_s = \frac{1}{T} = \frac{B}{1+a} = \frac{20}{1+0.33} = 15 \text{ Msymbol/s}$

Bit rate $R_b = R_s \log_2 M = 15 \text{ Msymbol/s} \times 2 \text{ bit/symbol} = 30 \text{ Mbit/s}$

REQUIRED SNR

10^{-6} BER for QPSK requires Eb/N0 of 10.2 dB (use BER formula $Q\left(\sqrt{2E_b/N_0}\right)$)

)

$$\text{SNR}_{\text{reqd}} = \left(\frac{E_b}{N_0}\right)_{\text{reqd}} \frac{R_b}{B}$$

In dB,

Link budget for WLAN (contd.)

$$P_n = N_0 B = kT_0 10^{F/10} B = (1.38 \times 10^{-23})(290)(10^{6/10})(20 \times 10^6)$$

RECEIVER

Noise power

$$= 3.2 \times 10^{-13} \text{ watt} = 3.2 \times 10^{-10} \text{ milliwatts (mW)}.$$

Receiver sensitivity

$$P_{\mathrm{RX,dBm}}(\min) = P_{\mathrm{n,dBm}} + \mathrm{SNR}_{\mathrm{reqd,dB}} = -95 + 12 = -83 \text{ dBm}$$

RANGE

(transmit power of 100 mW, or 20 dBm, and antenna directivities of 2 dBi at Tx and Rx)

Find the max allowable path loss for the desired receiver sensitivity:

$$L_{\mathrm{path,dB}}(R) = P_{\mathrm{TX,dBm}} - P_{\mathrm{RX,dBm}}(\min) + G_{\mathrm{TX,dBi}} + G_{\mathrm{RX,dBi}} - L_{\mathrm{margin,dB}}$$
$$= 20 - (-83) + 2 + 2 - 20 = \boxed{87 \text{ dB}}$$

Max allowable path loss

Invert path loss formula for free space propagation

$$L_{\mathrm{path,dB}}(R) = -10 \log_{10} \frac{\lambda^2}{16\pi^2 R^2}$$

Range is 107 meters

Of the order of advertised WiFi ranges

Try variants: redo for other constellations; carrier frequency of 60 GHz with antenna directivities of 20 dBi.