CT111 Introduction to Communication Systems
Lecture 14: Modulation

Yash M. Vasavada

Associate Professor, DA-IICT, Gandhinagar

28th March 2018

Overview of Today's Talk

Overview of Today's Talk

Overview of Today's Talk

Overview of Today's Talk

Overview of Today's Talk

Overview of Today's Talk

Overview of Today's Talk

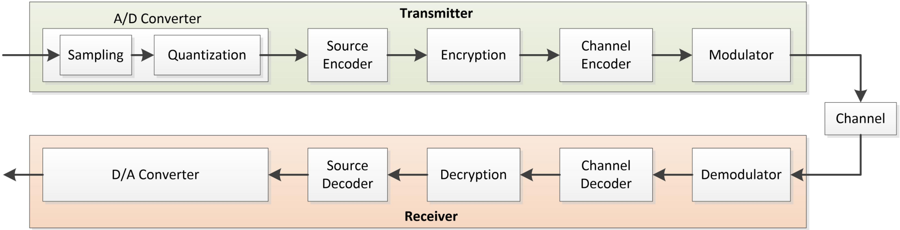
Overview of Today's Talk

Overview of Today's Talk

1. Review of Block Diagram
2. Intro to MOD
3. A Geometric Approach
4. Constellation Diagrams
5. Problem Setup
6. Model of Transmitted Signal
7. Model of Added Noise
8. Model of Received Signal
9. Conclusions

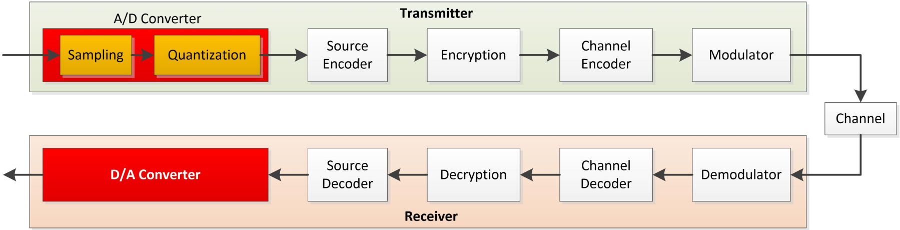
Digital Communication Transceiver
Block Diagram

- We have earlier seen this block diagram model of a digital communication transceiver



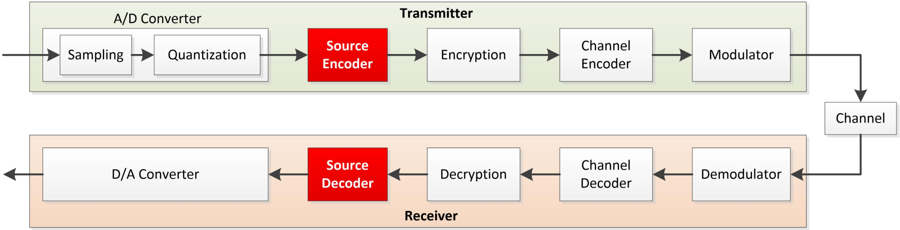
Digital Communication Transceiver
Block Diagram

- We have investigated the process of quantization of an analog information source



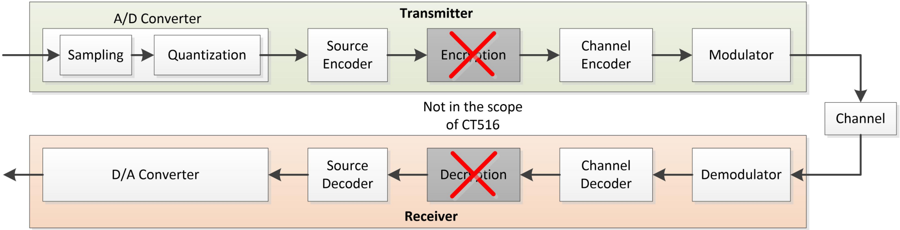
Digital Communication Transceiver
Block Diagram

- We have subsequently looked in detail at the math (Information Theory) and practical algorithms (Huffman Coding and Lempel-Ziv Coding) for source encoding



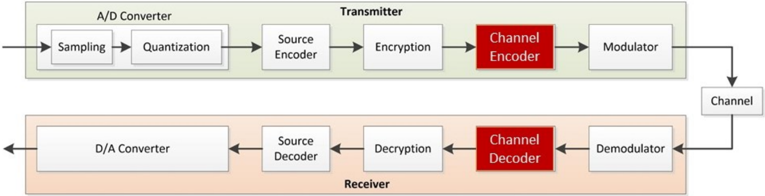
Digital Communication Transceiver
Block Diagram

- We will not be covering the encryption and decryption processes in this class



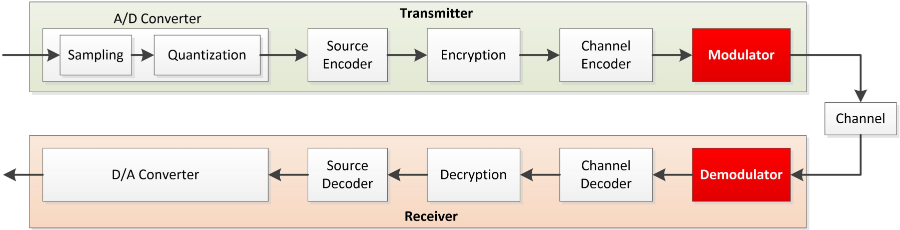
Digital Communication Transceiver
Channel Coding

- We have concluded an introductory study of channel encoding/decoding, also known as FEC (forward error correction) coding



Digital Communication Transceiver
Block Diagram

- We will next be studying the process of modulation and demodulation



Principles of Modulation

- A widely popular scheme for transmission of communication messages uses sinusoidal **carrier** waveforms
  - → *Carrier:* is to be thought of as a truck or goods train.
  - → *Message symbols* are analogous to the actual items that are being transported
- Reasons for the choice of sinusoidal waveforms:
  - → Electromagnetic signals are sinusoidal, and they propagate well in the atmosphere and over the cables
  - → Choice of carrier frequency allows placement of transmitted signal at the desired portion of the spectrum
  - → Choice of carrier frequency can be traded off with antenna size and the path loss that the signal experiences

Principles of Modulation

- A widely popular scheme for transmission of communication messages uses sinusoidal **carrier** waveforms
  - → *Carrier:* is to be thought of as a truck or goods train.
  - → *Message symbols* are analogous to the actual items that are being transported
- Reasons for the choice of sinusoidal waveforms:
  - → Electromagnetic signals are sinusoidal, and they propagate well in the atmosphere and over the cables
  - → Choice of carrier frequency allows placement of transmitted signal at the desired portion of the spectrum
  - → Choice of carrier frequency can be traded off with antenna size and the path loss that the signal experiences

Principles of Modulation

- A widely popular scheme for transmission of communication messages uses sinusoidal **carrier** waveforms
  - → *Carrier:* is to be thought of as a truck or goods train.
  - → *Message symbols* are analogous to the actual items that are being transported
- Reasons for the choice of sinusoidal waveforms:
  - → Electromagnetic signals are sinusoidal, and they propagate well in the atmosphere and over the cables
  - → Choice of carrier frequency allows placement of transmitted signal at the desired portion of the spectrum
  - → Choice of carrier frequency can be traded off with antenna size and the path loss that the signal experiences

Principles of Modulation

- A widely popular scheme for transmission of communication messages uses sinusoidal **carrier** waveforms
  - → *Carrier:* is to be thought of as a truck or goods train.
  - → *Message symbols* are analogous to the actual items that are being transported
- Reasons for the choice of sinusoidal waveforms:
  - → Electromagnetic signals are sinusoidal, and they propagate well in the atmosphere and over the cables
  - → Choice of carrier frequency allows placement of transmitted signal at the desired portion of the spectrum
  - → Choice of carrier frequency can be traded off with antenna size and the path loss that the signal experiences

Digital Modulation
Elementary Methods

Parameters that can be modulated:

- Amplitude: this is called On Off Keying (OOK) or Amplitude Shift Keying (ASK).

$$1 \Rightarrow s(t) = A\cos(2\pi f_c t)$$
$$0 \Rightarrow s(t) = 0$$

- Frequency: this is called Frequency Shift Keying (FSK).

$$1 \Rightarrow s(t) = A\cos(2\pi f_{c,1} t)$$
$$0 \Rightarrow s(t) = A\cos(2\pi f_{c,2} t)$$

- Phase: this is called Phase Shift Keying (PSK).

$$1 \Rightarrow s(t) = A\cos(2\pi f_c t)$$
$$0 \Rightarrow s(t) = A\cos(2\pi f_c t + \pi) = -A\cos(2\pi f_c t)$$

A Visual Way

of Understanding Digital Modulation

- We will examine an elegant and generalized way of visualizing digital modulation methods
- This generalized approach is very powerful and allows us to
  - $\rightarrow$ design the digitally modulated signals that have desirable properties
  - $\rightarrow$ design optimum receivers for these signals
  - $\rightarrow$ analyze the performance of the receiver in presence of noise

Vectors and Vector Spaces

- A **vector** is to be thought of as a line having a length and a direction
- A **vector space** is to be thought of as the "space" that the vector occupies
- **Dimension** N of the vector space is to be thought of as the number of perpendicular axes needed to encompass the vector space, and specify any vector in this vector space

Vectors and Vector Spaces

Several examples:

- One-dimensional vector space is a *line*. Vectors of this $1D$ space are just scalar numbers.
  - → Real numbers can be thought of as $1D$ vectors residing in a $1D$ line.
- A two-dimensional vector space is a *plane*. Vectors of this $2D$ space have length $N = 2$.
  - → Complex numbers can be thought of as $2D$ vectors residing in a $2D$ complex-plane.
- A three-dimensional vector space is the *space* that we see around us. Vectors of this $3D$ space have length $N = 3$.
- Human brain cannot visualize spaces having dimensions $N > 3$.
  - However, a $4D$ vector space is the one in which there is yet another axis $\mathbf{e}_4$ besides the three familiar axes $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ of the $3D$ space.
  - and so on, for $N - D$ vector space

Vectors and Vector Spaces

- We want to be able to specify the locations in a vector space
- Coordinate axes allow us to do precisely that
  - ▷ A formal name of coordinate axes is that they collectively form a set of basis vectors
  - ▷ For $2D$ space, the X and Y coordinate axes are the two basis vectors
- Typical coordinate axes are perpendicular to each other. Also their lengths are unity.
  - ▷ For $2D$ space, the coordinate axes are specified as $\mathbf{e}_1 \stackrel{\text{def}}{=} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = [1, 0]^T$, and $\mathbf{e}_2 \stackrel{\text{def}}{=} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = [0, 1]^T$
    - → Here $[\cdot]^T$ denotes the transposition operation (turns a tall vertical vector into a flat horizontal vector and vice versa).
    - → If the numbers are complex, the terminology is $[\cdot]^H$, and it is called as Hermitial (or complex-conjugate) transpose

Vectors and Vector Spaces

- A collection of coordinate axes, or basis vectors, forms a matrix, and this matrix is known as the coordinate *system*
  - → E.g., $\mathbf{R}_2 = [\mathbf{e}_1, \mathbf{e}_2] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is a coordinate system for $2D$ space
- The size of a matrix is defined by number of its rows and number of its columns. For example, size of
  - ▷ $\mathbf{e}_1$ and $\mathbf{e}_2$ is 2×1
  - ▷ $\mathbf{e}_1^T$ and $\mathbf{e}_2^T$ is 1×2
  - ▷ $\mathbf{R}_2$ is 2×2

Vectors and Vector Spaces

- An N-dimensional identity matrix, which is defined as follows,

$$\mathbf{I}_N = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

  is an example of $\mathbf{R}_N$.
- n^{th} column ($n = 1, 2, \dots, N$) of this matrix, denoted as $\mathbf{e}_n$ is the n^{th} coordinate axis, or n^{th} basis vector
- All N basis vectors have (i) same size of $N \times 1$, (ii) unity length, and (iii) they are perpendicular to each other

Vectors and Vector Spaces

Two Formulations

- There are two formulations of vectors and vector spaces that allow a great amount of flexibility and visualization
  1. Linear superposition of vectors (addition of two vectors)
  2. Projection of one vector onto another (dot product between vectors)
- We will look at these next

Vectors and Vector Spaces
Linear Superposition

- A vector $\mathbf{e}$ can be stretched or shrunk by multiplication with a scalar, say, x. The result is $x\mathbf{e}$
- Two vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ can be combined, or added, to obtain another vector after different amount of stretching or compression, as $x_1\mathbf{e}_1 + x_2\mathbf{e}_2$
- Result of above vectorial addition, combination, or superposition, is another vector that we denote as

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 = [\mathbf{e}_1, \mathbf{e}_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{R}_2\mathbf{x}$$

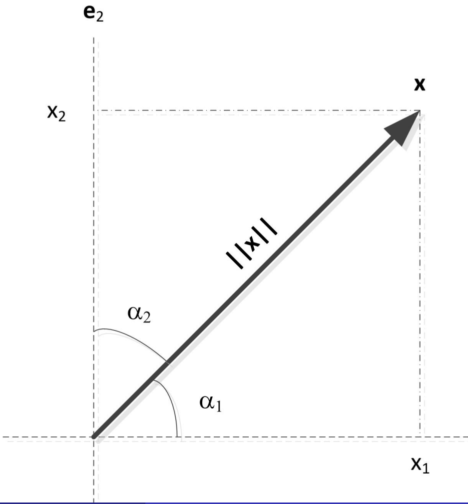
Vectors and Vector Spaces
Linear Superposition

Two terminologies:

1. Span of a basis vector set (coordinate system): is the totality of vector space that can be reached by this set of basis vectors
   - → Any vector in this vector space is expressible as the linear superposition of one, two or more basis vectors in the coordinate system $\mathbf{R}$
   - → E.g., span of $\mathbf{R}_2$ is any vector in a $2D$ plane
2. Linear Independence between vectors
   - A basis vector set, with N different basis vectors, is called linearly independent if no vector in the set can be expressed as linear superposition of remaining $N-1$ basis vectors

Projection of a Vector
on the Coordinate Axes

- Let the length of a vector $\mathbf{x}$ be denoted as $\|\mathbf{x}\|$
- This vector $\mathbf{x}$ is projected on the coordinate axes by means of "dropping perpendiculars" from the tip of $\mathbf{x}$ onto the coordinate axes
- When $N = 2$, the two perpendiculars can be dropped, and these are
  $x_1 = \|\mathbf{x}\| \cos(\alpha_1)$ and
  $x_2 = \|\mathbf{x}\| \cos(\alpha_2)$



Projection of a Vector
on the Coordinate Axes

- A vector $\mathbf{x}$ in N-dimensional $(N - D)$ space will require N different projections on its coordinate system
- Result will be N different scalars, $\{x_n\}_{n=1}^{N}$
  - ▷ These are known as the Cartesian Coordinates of vector $\mathbf{x}$
- Vector $\mathbf{x}$ is specified by a $N \times 1$ vector of its cartesian coordinates, i.e.,

$$\mathbf{x} \stackrel{\text{def}}{=} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = [x_1, x_2, \ldots, x_N]^T$$

Vector
and Its Transpose

- $\mathbf{x} \stackrel{\text{def}}{=} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = [x_1, x_2, \ldots, x_N]^T$ denotes a *column* vector with 1 column and N rows (i.e., of size $N \times 1$)
- $\mathbf{x}^T = [x_1, x_2, \ldots, x_N]$ denotes a *row* vector with 1 row and N column (size $1 \times N$)
  - $\rightarrow$ Think of the row vector as a *horizontal* vector sleeping on a couch, and the column vector as a *vertical* vector standing *tall*
  - $\rightarrow$ $(\cdot)^T$: the transposition operation that turns a $M \times N$ matrix into a $N \times M$ matrix (each row turns into a column and vice versa)

Multiplication of Two Vectors

- Vectors are a lot like the complex numbers
  - → A vector has a length, and a direction, like a complex number does
- There are two differences, however:
  1. A complex number resides in a $2D$ space, whereas the vectors have no such restriction (i.e., N can be greater than 2)
  2. A complex number, like a real number, is commutative for multiplication. It does not matter whether we do the product $x \times y$ or $y \times x$.
  3. For vectors and matrices, $\mathbf{x} \times \mathbf{y} \neq \mathbf{y} \times \mathbf{x}$
     - ▷ In fact, either $\mathbf{x} \times \mathbf{y}$ or $\mathbf{y} \times \mathbf{x}$, or both may not even be defined

Multiplication of Two Vectors

- Vectorial multiplication is defined in multiple ways:
  1. Vector Inner Product (or Dot Product),
  2. Vector Cross Product, and
  3. Vector Outer Product
- In CT-111, we will study only the vector inner product

Vectors and Vector Spaces

Vector Inner Product

- A rule of vector inner or dot product:
  - ▷ Cannot multiply a $1 \times N$ vector with $1 \times N$ vector. The internal dimensions don't match
  - ▷ Can multiply a $1 \times N$ vector with $N \times 1$ vector. The internal dimensions do match.
- Vector (inner or dot) product between a $1 \times N$ vector with $N \times 1$ vector is as follows:

$$\mathbf{x}^T \cdot \mathbf{y} = \mathbf{y}^T \mathbf{x} = x_1 y_1 + x_2 y_2 + \ldots + x_N y_N$$

→ If the vectors have complex-valued coordinates, the dot product requires Hermitian Transpose (transpose and conjugation): $\mathbf{h} \cdot \mathbf{r} = \mathbf{h}^H \mathbf{r}$

Vectors

Two Key Properties

Following are two key properties of the dot products.

1. Dot product of any vector $\mathbf{x}$ with itself is the square of the vector length.
2. Dot product of vectors $\mathbf{x}$ and $\mathbf{y}$ is directly related to the *cosine* of the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$.

We examine each of these two properties next.

Dot Product between Two Vectors

First Key Property: Length of any Vector

- Length of a vector $\mathbf{x}$ is denoted as $\|\mathbf{x}\|$.
- Length of a vector $\mathbf{x} = [x_1, x_2]$ in $2D$ plane is the length of *hypotenuse* of a right angled triangle whose bases have lengths x_1 and x_2.
- The square of length of this hypotenuse is given by Pythagoras theorem:

$$\|\mathbf{x}\|^2 = x_1^2 + x_2^2 = \mathbf{x}^T\mathbf{x}$$

G:/CT111/JP/g1a.jpg

Dot Product between Two Vectors

First Key Property: Length of any Vector

- This can be extended to three dimensional space.

$$\|\mathbf{x}\|^2 = x_1^2 + x_2^2 + x_3^2 = \mathbf{x}^T\mathbf{x}$$

- Therefore, we conclude, generalizing to N dimensional space, that the dot product of any N-dimensional vector $\mathbf{x}$ with itself gives the square of the length of the vector.

$$\mathbf{x}^T\mathbf{x} = x_1^2 + x_2^2 + x_3^2 + \ldots + x_N^2 = \|\mathbf{x}\|^2$$

First Key Property: Length of any Vector

- What does $\|\mathbf{x}\|^2 = x_1^2 + x_2^2 + x_3^2 + \ldots + x_N^2 = \mathbf{x}^T \mathbf{x}$ represent physically?
  - ▷ Approaches $N \times E\left[X^2\right]$ as $N \to \infty$
- Thus, the power of the signal $E\left[X^2\right]$ can be thought of as the normalized length $\|\mathbf{x}\|^2 / N$ of the signal vector

First Key Property: Length of any Vector

- What does $\|\mathbf{x}\|^2 = x_1^2 + x_2^2 + x_3^2 + \ldots + x_N^2 = \mathbf{x}^T\mathbf{x}$ represent physically?
  - Approaches $N \times E\left[X^2\right]$ as $N \to \infty$
- Thus, the power of the signal $E\left[X^2\right]$ can be thought of as the normalized length $\|\mathbf{x}\|^2 / N$ of the signal vector

Dot Product between Two Vectors

Second Key Property: Angle between Two Vectors

- Now, let us consider the dot product of vectors **x** and **y**.
- Length $\|x\|$ is the hypotenuse in the triangle OxQ, and the sine and consine of α are

$$\sin\alpha = \frac{x_2}{\|\mathbf{x}\|}, \quad \cos\alpha = \frac{x_1}{\|\mathbf{x}\|}$$

G:/CT111/JP/g1a.jpg

Dot Product between Two Vectors

Second Key Property: Angle between Two Vectors

- Similarly, for angle β for vector $\mathbf{y}$,

  $\rightarrow$ $\sin\beta = \frac{y_2}{\|\mathbf{y}\|}$, and $\cos\beta = \frac{y_1}{\|\mathbf{y}\|}$.
- Now,

$$
\begin{aligned}
\cos\theta &= \cos(\beta - \alpha) \\
&= \cos\beta\cos\alpha + \sin\beta\sin\alpha \\
&= \frac{x_1 y_1 + x_2 y_2}{\|\mathbf{x}\|\,\|\mathbf{y}\|} \\
&= \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}
\end{aligned}
$$

G:/CT111/JP/g2.jp

Dot Product between Two Vectors

Second Key Property: Angle between Two Vectors

- dot product is proportional to the cosine of the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$:

$$\mathbf{x}^T\mathbf{y} = \cos\theta \times \|\mathbf{x}\| \, \|\mathbf{y}\|$$

- If $\mathbf{x}$ and $\mathbf{y}$ are unit length vectors, dot product equals the cosine of the angle between the vectors.

$$\mathbf{x}^T\mathbf{y} = \cos\theta$$

- If $\mathbf{x}$ and $\mathbf{y}$ point in the same direction, $\theta = 0$ and the dot product is maximized.

$$\mathbf{x}^T\mathbf{y} = \cos(\theta = 0) = 1$$

- If $\mathbf{x}$ and $\mathbf{y}$ are perpendicular, $\theta = 90^o$ and the dot product becomes zero.

$$\mathbf{x}^T\mathbf{y} = \cos(\theta = 90^o) = 0$$

Second Key Property: Alignment between Vectors

- Therefore, whenever we are taking dot products between two vectors (in discrete time) or two signals (in continuous time), think (or visualize) this as if
  - we are measuring how well the two signals are aligned with each other in a multi-dimensional space
- In $\mathbf{h} \cdot \mathbf{r}$, one of the two vectors, let us say $\mathbf{h}$, is sometimes fixed (i.e., it makes up a reference), and $\mathbf{r}$ varies
  - The dot product $\mathbf{h} \cdot \mathbf{r}$ will be high when the variable vector $\mathbf{r}$ happens to be well-aligned to our reference vector $\mathbf{h}$
  - If the dot product turns out to be zero or near zero, think of this as $\mathbf{r}$ being *orthogonal* to $\mathbf{h}$ (i.e., pointing in a direction that is 90^o offset to the direction of $\mathbf{h}$)

Vector Spaces

- An n-dimensional $(n \times 1)$ vector $\mathbf{v} = [v_1, v_2, \ldots, v_n]^T$ has n columns and 1 row and it consists of n scalar components $\{v_i\}, i = 1, \ldots, n$
- Just as the numbers are inhabitants of the real number line, the vectors "live" in a "space" called *vector space*
- What is a mathematical description of the vector space?

Vector Spaces

1. Inner product of two vectors $\mathbf{v}_1 = [v_{11}, v_{12}, \ldots, v_{1n}]$ and $\mathbf{v}_2 = [v_{21}, v_{22}, \ldots, v_{2n}]$ is given as

$$\mathbf{v}_1 \bullet \mathbf{v}_2 = \mathbf{v}_1^T \mathbf{v}_2 = \mathbf{v}_2^T \mathbf{v}_1 = \sum_{i=1}^{n} v_{1i} v_{2i}$$

2. Norm (or length) of a vector $\mathbf{v}$ is given as

$$\|\mathbf{v}\| = \sqrt{\sum_{i=1}^{n} v_i^2} = \sqrt{\mathbf{v}_1 \bullet \mathbf{v}_1}$$

3. A vector $\mathbf{v}$ may be expressed as a linear combination of its *basis* vectors:

$$\mathbf{v} = \sum_{i=1}^{n} c_i \mathbf{e}_i$$

Here, $\mathbf{e}_i$ are the basis vectors, and $c_i = \mathbf{e}_i \bullet \mathbf{v}$. Note that $c_i = v_i$ if $\mathbf{e}_i$ is an all-zero vector except a single 1 at i^{th} location (i.e., if $\mathbf{e}_i$ is i^{th} axis of conventional Cartesian coordinate system).

Vector Spaces

1. Inner product of two vectors $\mathbf{v}_1 = [v_{11}, v_{12}, \ldots, v_{1n}]$ and $\mathbf{v}_2 = [v_{21}, v_{22}, \ldots, v_{2n}]$ is given as

$$\mathbf{v}_1 \bullet \mathbf{v}_2 = \mathbf{v}_1^T \mathbf{v}_2 = \mathbf{v}_2^T \mathbf{v}_1 = \sum_{i=1}^{n} v_{1i} v_{2i}$$

2. Norm (or length) of a vector $\mathbf{v}$ is given as

$$\|\mathbf{v}\| = \sqrt{\sum_{i=1}^{n} v_i^2} = \sqrt{\mathbf{v}_1 \bullet \mathbf{v}_1}$$

3. A vector $\mathbf{v}$ may be expressed as a linear combination of its *basis* vectors:

$$\mathbf{v} = \sum_{i=1}^{n} c_i \mathbf{e}_i$$

Here, $\mathbf{e}_i$ are the basis vectors, and $c_i = \mathbf{e}_i \bullet \mathbf{v}$. Note that $c_i = v_i$ if $\mathbf{e}_i$ is an all-zero vector except a single 1 at i^{th} location (i.e., if $\mathbf{e}_i$ is i^{th} axis of conventional Cartesian coordinate system).

Vector Spaces

1. Inner product of two vectors $\mathbf{v}_1 = [v_{11}, v_{12}, \ldots, v_{1n}]$ and $\mathbf{v}_2 = [v_{21}, v_{22}, \ldots, v_{2n}]$ is given as

$$\mathbf{v}_1 \bullet \mathbf{v}_2 = \mathbf{v}_1^T \mathbf{v}_2 = \mathbf{v}_2^T \mathbf{v}_1 = \sum_{i=1}^{n} v_{1i} v_{2i}$$

2. Norm (or length) of a vector $\mathbf{v}$ is given as

$$\|\mathbf{v}\| = \sqrt{\sum_{i=1}^{n} v_i^2} = \sqrt{\mathbf{v}_1 \bullet \mathbf{v}_1}$$

3. A vector $\mathbf{v}$ may be expressed as a linear combination of its *basis* vectors:

$$\mathbf{v} = \sum_{i=1}^{n} c_i \mathbf{e}_i$$

Here, $\mathbf{e}_i$ are the basis vectors, and $c_i = \mathbf{e}_i \bullet \mathbf{v}$. Note that $c_i = v_i$ if $\mathbf{e}_i$ is an all-zero vector except a single 1 at i^{th} location (i.e., if $\mathbf{e}_i$ is i^{th} axis of conventional Cartesian coordinate system).

Vector Spaces

- Think of the basis vectors as forming an arbitrary coordinate system for describing the vector $\mathbf{v}$
- What constitutes a good set of basis vectors?

Complete Orthonormal Basis

What constitutes a good set of basis vectors?

- The set of basis vectors $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ should be *complete*, i.e., it should *span* the n-dimensional vector space $\mathcal{R}^n$.
  - $\rightarrow$ If you pick any vector $\mathbf{v}$ in $\mathcal{R}^n$, it should be possible to determine values $\{v_i\}$ such that $\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{e}_i$
- Each basis vector should be orthogonal to all the other basis vectors:

$$\mathbf{e}_i \bullet \mathbf{e}_l = 0, \forall i \neq l$$

- Each basis vector should be normalized (i.e., should have a length of unity)

Any set of basis vectors that satisfied these three properties is called a *complete orthonormal basis*.

Complete Orthonormal Basis

What constitutes a good set of basis vectors?

- The set of basis vectors $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ should be *complete*, i.e., it should *span* the n-dimensional vector space $\mathcal{R}^n$.
  - $\rightarrow$ If you pick any vector $\mathbf{v}$ in $\mathcal{R}^n$, it should be possible to determine values $\{v_i\}$ such that $\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{e}_i$
- Each basis vector should be orthogonal to all the other basis vectors:

$$\mathbf{e}_i \bullet \mathbf{e}_l = 0, \forall i \neq l$$

- Each basis vector should be normalized (i.e., should have a length of unity)

Any set of basis vectors that satisfied these three properties is called a *complete orthonormal basis*.

Complete Orthonormal Basis

What constitutes a good set of basis vectors?

- The set of basis vectors $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ should be *complete*, i.e., it should *span* the n-dimensional vector space $\mathcal{R}^n$.
  - → If you pick any vector $\mathbf{v}$ in $\mathcal{R}^n$, it should be possible to determine values $\{v_i\}$ such that $\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{e}_i$
- Each basis vector should be orthogonal to all the other basis vectors:

$$\mathbf{e}_i \bullet \mathbf{e}_l = 0, \forall i \neq l$$

- Each basis vector should be normalized (i.e., should have a length of unity)

Any set of basis vectors that satisfied these three properties is called a *complete orthonormal basis*.

Complete Orthonormal Basis

What constitutes a good set of basis vectors?

- The set of basis vectors $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ should be *complete*, i.e., it should *span* the n-dimensional vector space $\mathcal{R}^n$.
  - → If you pick any vector $\mathbf{v}$ in $\mathcal{R}^n$, it should be possible to determine values $\{v_i\}$ such that $\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{e}_i$
- Each basis vector should be orthogonal to all the other basis vectors:

$$\mathbf{e}_i \bullet \mathbf{e}_l = 0, \forall i \neq l$$

- Each basis vector should be normalized (i.e., should have a length of unity)

Any set of basis vectors that satisfied these three properties is called a *complete orthonormal basis*.

Complete Orthonormal Basis

What constitutes a good set of basis vectors?

- The set of basis vectors $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ should be *complete*, i.e., it should *span* the n-dimensional vector space $\mathcal{R}^n$.
  - → If you pick any vector $\mathbf{v}$ in $\mathcal{R}^n$, it should be possible to determine values $\{v_i\}$ such that $\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{e}_i$
- Each basis vector should be orthogonal to all the other basis vectors:

$$\mathbf{e}_i \bullet \mathbf{e}_l = 0, \forall i \neq l$$

- Each basis vector should be normalized (i.e., should have a length of unity)

Any set of basis vectors that satisfied these three properties is called a *complete orthonormal basis*.

Complete Orthonormal Basis

What constitutes a good set of basis vectors?

- The set of basis vectors $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ should be *complete*, i.e., it should *span* the n-dimensional vector space $\mathcal{R}^n$.
  - $\rightarrow$ If you pick any vector $\mathbf{v}$ in $\mathcal{R}^n$, it should be possible to determine values $\{v_i\}$ such that $\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{e}_i$
- Each basis vector should be orthogonal to all the other basis vectors:

$$\mathbf{e}_i \bullet \mathbf{e}_l = 0, \forall i \neq l$$

- Each basis vector should be normalized (i.e., should have a length of unity)

Any set of basis vectors that satisfied these three properties is called a *complete orthonormal basis*.

Complete Orthonormal Basis

Why?

1. Why to insist on completeness?
2. Why is orthogonality useful?
3. Why is the normalization useful?

Why Complete Orthonormal Basis

1. Why to insist on completeness?
   → To ensure that *any* given vector $\mathbf{v}$ can be described as a linear combination of the basis vectors, i.e., as $\mathbf{v} = \sum_{i=1}^{n} c_i \mathbf{e}_i$
2. Why is orthogonality useful?
   → So that c_i can be easily calculated as the dot product $\mathbf{v} \bullet \mathbf{e}_i$. In absence of orthogonality, the dot product $\mathbf{v} \bullet \mathbf{e}_i$ does not depend only on c_i.
3. Why is the normalization useful?
   → So that c_i can be easily calculated as the dot product $\mathbf{v} \bullet \mathbf{e}_i$. If $\mathbf{e}_i$ is not normalized, this dot product has an additional term due to the length of the vector $\mathbf{e}_i$ itself.

Why Complete Orthonormal Basis

An alternate explanation:

- Matrix equation that captures representation of vector $\mathbf{v}$ as a linear combination of the basis vectors, i.e., as $\mathbf{v} = \sum_{i=1}^{n} c_i \mathbf{e}_i$, is as follows:

$$\mathbf{v} = \mathbf{E}\mathbf{c}$$

→ Here $\mathbf{c} = [c_1, c_2, \ldots, c_n]^T$ is an $n \times 1$ vector, and
→ $\mathbf{E} = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n]$ is an $n \times n$ matrix

Why Complete Orthonormal Basis

An alternate explanation (continued):

- **Completeness** ensures that the matrix $\mathbf{E}$ is full rank, i.e., it is invertible
- **Orthonormality** ensures that the inverse of matrix $\mathbf{E}$ is simply the transpose, i.e., $(\mathbf{E})^{-1} = \mathbf{E}^T$ (or, more generally, for complex-valued $\mathbf{E}$, this is the Hermitian transpose $(\mathbf{E})^{-1} = \mathbf{E}^H$.
  - $\rightarrow$ Thus, it is easy to take any vector $\mathbf{v}$ and compute $\mathbf{c}$ simply as $\mathbf{c} = \mathbf{E}^H\mathbf{v}$
- We will see that there is often a vast number of possible basis vector sets. Completeness and orthonormality allows an easy *transformation* of a given vector from one basis set to another.

Signal Spaces

- Surprisingly, the signals that we have to work with in design of a digital modulation system can be treated in a manner similar to the vectors

Signal Spaces

1. Inner product of two signals $x_1(t)$ and $x_2(t)$ defined over $a \leq t \leq b$ is given as

$$< x_1(t), x_2(t) >= \int_a^b x_1(t) x_2^*(t) dt$$

2. Norm (or length) signal $x(t)$ is given as

$$\|x(t)\| = \sqrt{< x(t), x(t) >} = \left(\int_a^b x(t) x^*(t) dt \right)^{1/2} = \sqrt{E_x}$$

3. A signal $x(t)$ may be expressed as a linear combination of its *basis functions* $\{f_i(t)\}, i = 1, 2, \ldots, n$:

$$x(t) = \sum_{i=1}^{n} x_i f_i(t)$$

Here, $f_i(t)$ are basis functions, and $x_i =< x(t), f_i(t) >$

Signal Spaces

1. Inner product of two signals $x_1(t)$ and $x_2(t)$ defined over $a \leq t \leq b$ is given as

$$< x_1(t), x_2(t) >= \int_a^b x_1(t) x_2^*(t) dt$$

2. Norm (or length) signal $x(t)$ is given as

$$\|x(t)\| = \sqrt{< x(t), x(t) >} = \left(\int_a^b x(t) x^*(t) dt \right)^{1/2} = \sqrt{E_x}$$

3. A signal $x(t)$ may be expressed as a linear combination of its *basis functions* $\{f_i(t)\}$, $i = 1, 2, \ldots, n$:

$$x(t) = \sum_{i=1}^{n} x_i f_i(t)$$

Here, $f_i(t)$ are basis functions, and $x_i =< x(t), f_i(t) >$

Signal Spaces

1. Inner product of two signals $x_1(t)$ and $x_2(t)$ defined over $a \leq t \leq b$ is given as

$$< x_1(t), x_2(t) >= \int_a^b x_1(t) x_2^*(t) dt$$

2. Norm (or length) signal $x(t)$ is given as

$$\|x(t)\| = \sqrt{< x(t), x(t) >} = \left(\int_a^b x(t) x^*(t) dt \right)^{1/2} = \sqrt{E_x}$$

3. A signal $x(t)$ may be expressed as a linear combination of its *basis functions* $\{f_i(t)\}$, $i = 1, 2, \ldots, n$:

$$x(t) = \sum_{i=1}^{n} x_i f_i(t)$$

Here, $f_i(t)$ are basis functions, and $x_i =< x(t), f_i(t) >$

Basis Functions

for a Signal Space

- Suppose we want to design a modulation scheme in which one of M different symbols is transmitted at a time.
- For transmission of m^{th} symbol, we use a signal $s_m(t)$, which belongs to the set $\{s_1(t), s_2(t), \ldots, s_M(t)\}$

Basis Functions
for a Signal Space

- Functions $\{f_1(t), f_2(t), \ldots, f_K(t)\}, (K \leq M)$ form a *complete orthonormal basis* for the signal set $\{s_1(t), s_2(t), \ldots, s_M(t)\}$ if:
  1. Any signal $s_m(t)$ in the signal set $\{s_1(t), s_2(t), \ldots, s_M(t)\}$ can be described by a linear combination of the basis functions
  $$s_m(t) = \sum_{i=1}^{K} s_{m,i} f_i(t), \quad m = 1, 2, \ldots, M$$
  2. Basis functions are orthogonal to each other
  $$< f_i(t), f_j(t) >= \int_a^b f_i(t) f_j^*(t) dt = 0, \forall i \neq j$$
  3. Basis functions are normalized to have unit energy
  $$\|f_i(t)\| = \sqrt{< f_i(t), f_i(t) >} = \left(\int_a^b f_i(t) f_i^*(t) dt \right)^{1/2} = \sqrt{E_{f_i}} = 1, \forall i$$

Basis Functions
for a Signal Space

- Functions $\{f_1(t), f_2(t), \ldots, f_K(t)\}, (K \leq M)$ form a *complete orthonormal basis* for the signal set $\{s_1(t), s_2(t), \ldots, s_M(t)\}$ if:
  1. Any signal $s_m(t)$ in the signal set $\{s_1(t), s_2(t), \ldots, s_M(t)\}$ can be described by a linear combination of the basis functions

$$s_m(t) = \sum_{i=1}^{K} s_{m,i} f_i(t), \quad m = 1, 2, \ldots, M$$

  2. Basis functions are orthogonal to each other

$$< f_i(t), f_j(t) >= \int_a^b f_i(t) f_j^*(t) dt = 0, \forall i \neq j$$

  3. Basis functions are normalized to have unit energy

$$\|f_i(t)\| = \sqrt{< f_i(t), f_i(t) >} = \left(\int_a^b f_i(t) f_i^*(t) dt \right)^{1/2} = \sqrt{E_{f_i}} = 1, \forall i$$

Basis Functions

for a Signal Space

- Functions $\{f_1(t), f_2(t), \ldots, f_K(t)\}, (K \leq M)$ form a *complete orthonormal basis* for the signal set $\{s_1(t), s_2(t), \ldots, s_M(t)\}$ if:

  1. Any signal $s_m(t)$ in the signal set $\{s_1(t), s_2(t), \ldots, s_M(t)\}$ can be described by a linear combination of the basis functions

$$s_m(t) = \sum_{i=1}^{K} s_{m,i} f_i(t), \quad m = 1, 2, \ldots, M$$

  2. Basis functions are orthogonal to each other

$$< f_i(t), f_j(t) >= \int_a^b f_i(t) f_j^*(t) dt = 0, \forall i \neq j$$

  3. Basis functions are normalized to have unit energy

$$\|f_i(t)\| = \sqrt{< f_i(t), f_i(t) >} = \left(\int_a^b f_i(t) f_i^*(t) dt \right)^{1/2} = \sqrt{E_{f_i}} = 1, \forall i$$

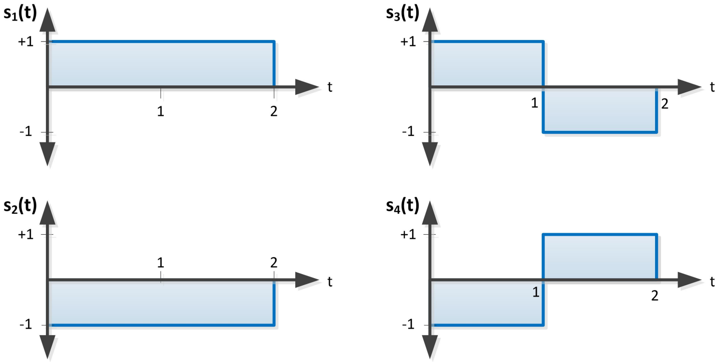
Basis Functions

for a Signal Space

- Functions $\{f_1(t), f_2(t), \ldots, f_K(t)\}, (K \leq M)$ form a *complete orthonormal basis* for the signal set $\{s_1(t), s_2(t), \ldots, s_M(t)\}$ if:
  1. Any signal $s_m(t)$ in the signal set $\{s_1(t), s_2(t), \ldots, s_M(t)\}$ can be described by a linear combination of the basis functions
  $$s_m(t) = \sum_{i=1}^{K} s_{m,i} f_i(t), \quad m = 1, 2, \ldots, M$$
  2. Basis functions are orthogonal to each other
  $$< f_i(t), f_j(t) >= \int_a^b f_i(t) f_j^*(t) dt = 0, \forall i \neq j$$
  3. Basis functions are normalized to have unit energy
  $$\|f_i(t)\| = \sqrt{< f_i(t), f_i(t) >} = \left(\int_a^b f_i(t) f_i^*(t) dt \right)^{1/2} = \sqrt{E_{f_i}} = 1, \forall i$$

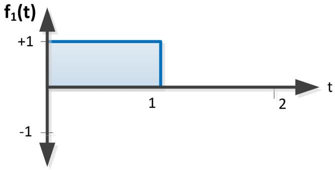
Signal Spaces
An Example

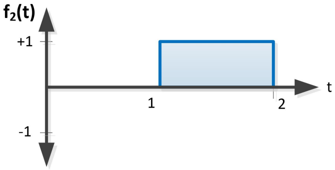
Consider the following signal set.



Signal Spaces
An Example

- We can express each of the four signals of the previous slide in terms of the following basis functions

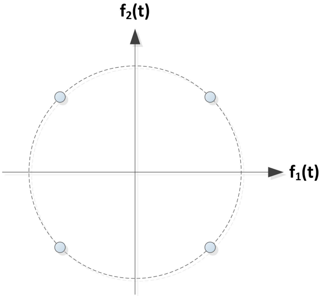




$$s_1(t) = 1 \times f_1(t) + 1 \times f_2(t), \qquad s_2(t) = -1 \times f_1(t) - 1 \times f_2(t)$$
$$s_3(t) = 1 \times f_1(t) - 1 \times f_2(t), \qquad s_4(t) = -1 \times f_1(t) + 1 \times f_2(t)$$

i.e., the basis is *complete*.

Signal Constellation



Signal Spaces
An Example

- Basis is orthogonal since

$$\int_0^2 f_1(t) f_2^*(t)\,dt = 0$$

- Basis is normal since

$$\int_0^2 |f_i(t)|^2 dt = 1, \quad i = 1, 2$$

Signal Spaces
Another Example

- Next consider the functions:

$$f_1(t) = \sqrt{\frac{2}{T}}\cos(2\pi f_c t), \qquad f_2(t) = \sqrt{\frac{2}{T}}\sin(2\pi f_c t)$$

These also form a complete orthonormal basis.

Signal Spaces

Another Example

Note that

- the basis is orthogonal

$$\begin{aligned} < f_1(t), f_2(t) > &= \frac{2}{T}\int_0^T \cos(2\pi f_c t)\sin(2\pi f_c t)dt \\ &= \frac{2}{T}\int_0^T \sin(0) + \sin(4\pi f_c t)dt \\ &= \frac{-1}{4\pi f_c T}\left[\cos(4\pi f_c t)\right]_0^T \\ &\approx 0, \qquad \text{for } f_c T >> 1 \end{aligned}$$

- Similarly prove that the basis is normalized, i.e.,

$$< f_1(t), f_1(t) >=< f_2(t), f_2(t) >= 1$$

Digital Modulation

Elementary Methods

Parameters that can be modulated:

- Amplitude: this is called On Off Keying (OOK) or Amplitude Shift Keying (ASK).

$$1 \Rightarrow s(t) = A\cos(2\pi f_c t)$$
$$0 \Rightarrow s(t) = 0$$

- Frequency: this is called Frequency Shift Keying (FSK).

$$1 \Rightarrow s(t) = A\cos(2\pi f_{c,1} t)$$
$$0 \Rightarrow s(t) = A\cos(2\pi f_{c,2} t)$$

- Phase: this is called Phase Shift Keying (PSK).

$$1 \Rightarrow s(t) = A\cos(2\pi f_c t)$$
$$0 \Rightarrow s(t) = A\cos(2\pi f_c t + \pi) = -A\cos(2\pi f_c t)$$

Binary Phase Shift Keying
BPSK

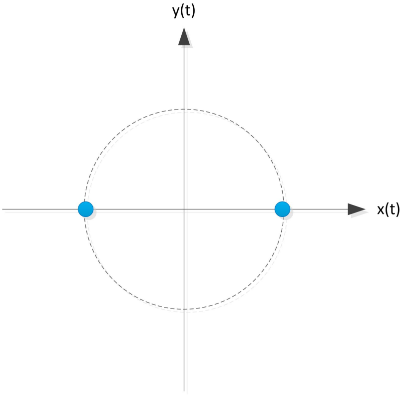
- BPSK signal description:
  - → $K = 1$, one basis function $f_1(t) = \sqrt{\frac{2}{T}} \cos(2\pi f_c t)\,|_0^T$
  - → $M = 2$ symbols
  - → Transmitted symbol waveforms:
    - ▷ $s_1(t) = A\cos(2\pi f_c t)\,|_0^T = \sqrt{P \times T} f_1(t)$, and
    - ▷ $s_2(t) = -A\cos(2\pi f_c t)\,|_0^T = -\sqrt{P \times T} f_1(t)$
  - → A is the transmitted peak voltage, $P = A^2$ is the transmit power, and $E_S = E_b = P \times T$ is the energy per symbol which equals energy per bit
  - Verify above yourself, using the following definitions:
    - ▷ Signal power: $\frac{1}{T}\int_0^T s^2(t)dt$,
    - ▷ Symbol energy: $\int_0^T s^2(t)dt$
- What is the symbol set for the above?

Binary Phase Shift Keying
BPSK

- Symbol set description for the formulation of BPSK on the prior slide:
  - Symbol set: $\mathbf{S} = \left\{+\sqrt{E_b}, -\sqrt{E_b}\right\}$,

Binary Phase Shift Keying
BPSK

Vectorial representation for BPSK:

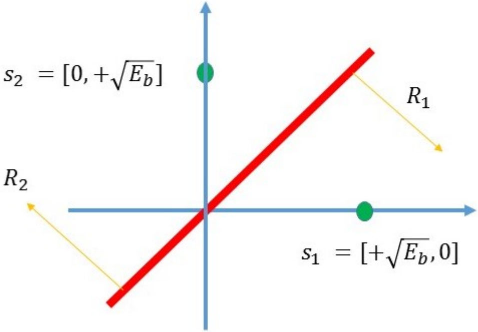


Binary Frequency Shift Keying
BFSK

- Binary Frequency Shift Keying (BFSK) signal description:
  - → $K = 2$, two basis functions:
    - ▷ $f_1(t) = \sqrt{\frac{2}{T}} \cos(2\pi f_1 t)\,|_0^T$
    - ▷ $f_2(t) = \sqrt{\frac{2}{T}} \cos(2\pi f_2 t)\,|_0^T$
  - → $M = 2$ symbols:
    - ▷ $s_1(t) = \sqrt{2P} \cos(2\pi f_1 t)\,|_0^T = \sqrt{P \times T} f_1(t)$, and
    - ▷ $s_2(t) = -\sqrt{2P} \cos(2\pi f_2 t)\,|_0^T = \sqrt{P \times T} f_2(t)$
- What is the symbol set for the above?

Binary Frequency Shift Keying
BFSK

- Symbol set description for the formulation of BFSK on the prior slide:
  - $\mathbf{s}_1 = \left[+\sqrt{E_b}, 0\right]^T$,
  - $\mathbf{s}_2 = \left[0, +\sqrt{E_b}\right]^T$,



Signal Space Representations
Summary

- The signal coming into the modulator block is entirely digital
  - → This is typically just a sequence of bits out of channel encoding
- We have seen that it is not possible to transmit the bits over a wireline or wireless channel
- Job of the modulator is, therefore, to convert the bits into analog signals $\{s_m(t)\}_{m=1}^{M}$
  - → Modulator can be thought of as a mapping function that takes a sequence of $\log_2 M$ bits and maps each such sequence to one of M analog signals $\{s_m(t)\}_{m=1}^{M}$

Signal Space Representations
Summary

- These M analog symbol waveforms are to be thought of as vectors residing in a K dimensional space.
- K can be the same value as M, and it can be smaller than M, but it should never be greater than M
- A common case is $K = 2$, and the basis functions are $f_1(t) = \sqrt{2/T_c}\cos(2\pi f_c t)$ and $\sqrt{2/T_c}\sin(2\pi f_c t)$ functions of the carrier frequency $f_c = 1/T_c$
- In this case, the two dimensional space *spanned* by the basis functions becomes the familiar $2D$ plane of complex numbers
- Furthermore, the *pair* of the basis functions $f_1(t)$ and $f_2(t)$ is to be viewed as defining the complex phasor $f(t) = \exp(j2\pi f_c t)$, and the $2D$ modulation signals $s_1(t), s_2(t), \ldots, s_m(t)$ as being just complex numbers

Signal Space Representations
Summary

- These M analog symbol waveforms are to be thought of as vectors residing in a K dimensional space.
- K can be the same value as M, and it can be smaller than M, but it should never be greater than M
- A common case is $K = 2$, and the basis functions are $f_1(t) = \sqrt{2/T_c}\cos(2\pi f_c t)$ and $\sqrt{2/T_c}\sin(2\pi f_c t)$ functions of the carrier frequency $f_c = 1/T_c$
- In this case, the two dimensional space *spanned* by the basis functions becomes the familiar $2D$ plane of complex numbers
- Furthermore, the *pair* of the basis functions $f_1(t)$ and $f_2(t)$ is to be viewed as defining the complex phasor $f(t) = \exp(j2\pi f_c t)$, and the $2D$ modulation signals $s_1(t), s_2(t), \ldots, s_m(t)$ as being just complex numbers

Signal Space Representations
Summary

- These M analog symbol waveforms are to be thought of as vectors residing in a K dimensional space.
- K can be the same value as M, and it can be smaller than M, but it should never be greater than M
- A common case is $K = 2$, and the basis functions are $f_1(t) = \sqrt{2/T_c} \cos(2\pi f_c t)$ and $\sqrt{2/T_c} \sin(2\pi f_c t)$ functions of the carrier frequency $f_c = 1/T_c$
- In this case, the two dimensional space *spanned* by the basis functions becomes the familiar $2D$ plane of complex numbers
- Furthermore, the *pair* of the basis functions $f_1(t)$ and $f_2(t)$ is to be viewed as defining the complex phasor $f(t) = \exp(j2\pi f_c t)$, and the $2D$ modulation signals $s_1(t), s_2(t), \ldots, s_m(t)$ as being just complex numbers

Signal Space Representations
Summary

- These M analog symbol waveforms are to be thought of as vectors residing in a K dimensional space.
- K can be the same value as M, and it can be smaller than M, but it should never be greater than M
- A common case is $K = 2$, and the basis functions are $f_1(t) = \sqrt{2/T_c}\cos(2\pi f_c t)$ and $\sqrt{2/T_c}\sin(2\pi f_c t)$ functions of the carrier frequency $f_c = 1/T_c$
- In this case, the two dimensional space *spanned* by the basis functions becomes the familiar $2D$ plane of complex numbers
- Furthermore, the *pair* of the basis functions $f_1(t)$ and $f_2(t)$ is to be viewed as defining the complex phasor $f(t) = \exp(j2\pi f_c t)$, and the $2D$ modulation signals $s_1(t), s_2(t), \ldots, s_m(t)$ as being just complex numbers

Signal Space Representations
Summary

- These M analog symbol waveforms are to be thought of as vectors residing in a K dimensional space.
- K can be the same value as M, and it can be smaller than M, but it should never be greater than M
- A common case is $K = 2$, and the basis functions are $f_1(t) = \sqrt{2/T_c}\cos(2\pi f_c t)$ and $\sqrt{2/T_c}\sin(2\pi f_c t)$ functions of the carrier frequency $f_c = 1/T_c$
- In this case, the two dimensional space *spanned* by the basis functions becomes the familiar $2D$ plane of complex numbers
- Furthermore, the *pair* of the basis functions $f_1(t)$ and $f_2(t)$ is to be viewed as defining the complex phasor $f(t) = \exp(j2\pi f_c t)$, and the $2D$ modulation signals $s_1(t), s_2(t), \ldots, s_m(t)$ as being just complex numbers

Digital Modulation

Basis Functions Approach

This results in the following three, completely equivalent, ways of representing the transmitted signal for $K = 2$ case:

1. **Quadrature Notation:**

$$s(t) = x(t)\cos(2\pi f_c t) - y(t)\sin(2\pi f_c t)$$

   Here $x(t)$ and $y(t)$ are real-valued baseband signals. These are called inphase and quadrature components generated using the bits out of channel encoder

2. **Complex Envelope Notation:**

$$s(t) = Re\left[(x(t) + jy(t))\exp\{-j2\pi f_c t\}\right] = Re\left[s_l(t)\exp\{-j2\pi f_c t\}\right]$$

   Here $s_l(t) = x(t) + jy(t)$ is called the *complex envelope* of $s(t)$

3. **Magnitude and Phase Representation:**

$$s(t) = a(t)\cos(2\pi f_c t + \theta(t))$$

   Here $a(t) = \sqrt{x^2(t) + y^2(t)}$ is the *magnitude* of $s(t)$, and $\theta(t) = \tan^{-1}\left(\frac{y(t)}{x(t)}\right)$ is the *phase* of $s(t)$.

Digital Modulation
Basis Functions Approach

This results in the following three, completely equivalent, ways of representing the transmitted signal for $K = 2$ case:

1. **Quadrature Notation:**

$$s(t) = x(t)\cos(2\pi f_c t) - y(t)\sin(2\pi f_c t)$$

   Here $x(t)$ and $y(t)$ are real-valued baseband signals. These are called inphase and quadrature components generated using the bits out of channel encoder

2. **Complex Envelope Notation:**

$$s(t) = Re\left[(x(t) + jy(t))\exp\{-j2\pi f_c t\}\right] = Re\left[s_l(t)\exp\{-j2\pi f_c t\}\right]$$

   Here $s_l(t) = x(t) + jy(t)$ is called the *complex envelope* of $s(t)$

3. **Magnitude and Phase Representation:**

$$s(t) = a(t)\cos(2\pi f_c t + \theta(t))$$

   Here $a(t) = \sqrt{x^2(t) + y^2(t)}$ is the *magnitude* of $s(t)$, and $\theta(t) = \tan^{-1}\left(\frac{y(t)}{x(t)}\right)$ is the *phase* of $s(t)$.

Digital Modulation

Basis Functions Approach

This results in the following three, completely equivalent, ways of representing the transmitted signal for $K = 2$ case:

1. **Quadrature Notation:**

$$s(t) = x(t)\cos(2\pi f_c t) - y(t)\sin(2\pi f_c t)$$

   Here $x(t)$ and $y(t)$ are real-valued baseband signals. These are called inphase and quadrature components generated using the bits out of channel encoder

2. **Complex Envelope Notation:**

$$s(t) = Re\left[(x(t) + jy(t))\exp\{-j2\pi f_c t\}\right] = Re\left[s_l(t)\exp\{-j2\pi f_c t\}\right]$$

   Here $s_l(t) = x(t) + jy(t)$ is called the *complex envelope* of $s(t)$

3. **Magnitude and Phase Representation:**

$$s(t) = a(t)\cos(2\pi f_c t + \theta(t))$$

   Here $a(t) = \sqrt{x^2(t) + y^2(t)}$ is the *magnitude* of $s(t)$, and $\theta(t) = \tan^{-1}\left(\frac{y(t)}{x(t)}\right)$ is the *phase* of $s(t)$.

Digital Modulation

Basis Functions Approach

This results in the following three, completely equivalent, ways of representing the transmitted signal for $K = 2$ case:

1. **Quadrature Notation:**

$$s(t) = x(t)\cos(2\pi f_c t) - y(t)\sin(2\pi f_c t)$$

   Here $x(t)$ and $y(t)$ are real-valued baseband signals. These are called inphase and quadrature components generated using the bits out of channel encoder

2. **Complex Envelope Notation:**

$$s(t) = Re\left[(x(t) + jy(t))\exp\{-j2\pi f_c t\}\right] = Re\left[s_l(t)\exp\{-j2\pi f_c t\}\right]$$

   Here $s_l(t) = x(t) + jy(t)$ is called the *complex envelope* of $s(t)$

3. **Magnitude and Phase Representation:**

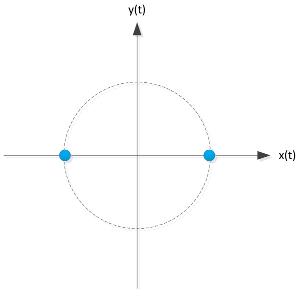
$$s(t) = a(t)\cos(2\pi f_c t + \theta(t))$$

   Here $a(t) = \sqrt{x^2(t) + y^2(t)}$ is the *magnitude* of $s(t)$, and $\theta(t) = \tan^{-1}\left(\frac{y(t)}{x(t)}\right)$ is the *phase* of $s(t)$.

Digital Modulation
Basis Functions Approach

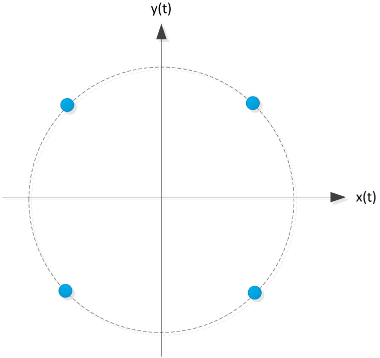
Quadrature method of signal representation allows us to:

- Look at the bandpass signal independent of the carrier frequency
- Set up coordinate system for looking at the common modulation types; this is called signal *constellation* diagram
- **Constellation diagram** for Binary Phase Shift Keying or BPSK: $x(t) \in \{\pm 1\}, y(t) = 0$



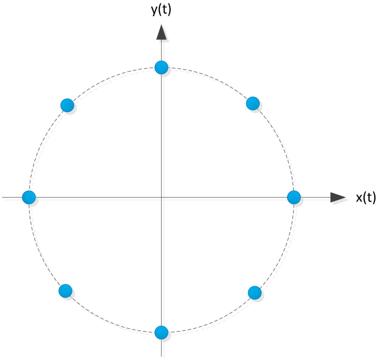
QPSK

→ Constellation diagram for Quadrature Phase Shift Keying or QPSK: $x(t) \in \{\pm 1\}, y(t) \in \{\pm 1\}$



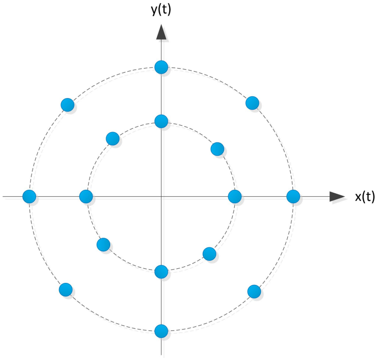
M-ary PSK

→ Constellation diagram for 8-ary Phase Shift Keying or 8-APSK:
$x(t) \in \left\{\pm 1, \pm\sqrt{2}\right\}, y(t) \in \left\{\pm 1, \pm\sqrt{2}\right\}$



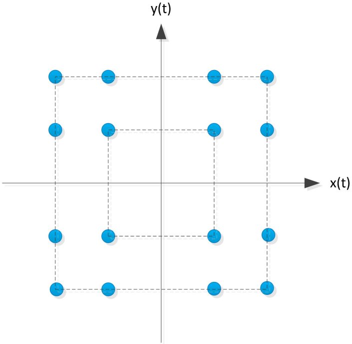
16-APSK

→ Constellation diagram for 16-ary Amplitude Phase Shift Keying or 16-APSK:
$x(t) \in \left\{\pm 1, \pm\sqrt{2}, \pm r, \pm r\sqrt{2}\right\}, y(t) \in \left\{\pm 1, \pm\sqrt{2}, \pm r, \pm r\sqrt{2}\right\}$



QAM

→ Constellation diagram for 16-ary Quadrature Amplitude Modulation or 16-QAM: $x(t) \in \{\pm 1, \pm 3\}$, $y(t) \in \{\pm 1, \pm 3\}$



Summary
of Constellation Diagrams

- Signals $x(t)$ and $y(t)$ make up the coordinate system
  - $\rightarrow$ Recall, $x(t)$ modulates the carrier $\cos(2\pi f_c t)$, and $y(t)$ modulates the quadrature carrier $\sin(2\pi f_c t)$
- Possible values of $[x(t), y(t)]$ are plotted on $2D$ plane with the above coordinate system as different points
- Probability of mistaking one transmitted symbol with another is dependent on the distance between the points
- At the receiver, the decision regarding which symbol was transmitted is made by choosing that point of the constellation diagram that is *closest* to the actual received signal location (recall the landmark point mentioned at the beginning of the lecture)

Summary
of Constellation Diagrams

- Signals $x(t)$ and $y(t)$ make up the coordinate system
  - $\rightarrow$ Recall, $x(t)$ modulates the carrier $\cos(2\pi f_c t)$, and $y(t)$ modulates the quadrature carrier $\sin(2\pi f_c t)$
- Possible values of $[x(t), y(t)]$ are plotted on $2D$ plane with the above coordinate system as different points
- Probability of mistaking one transmitted symbol with another is dependent on the distance between the points
- At the receiver, the decision regarding which symbol was transmitted is made by choosing that point of the constellation diagram that is *closest* to the actual received signal location (recall the landmark point mentioned at the beginning of the lecture)

Summary
of Constellation Diagrams

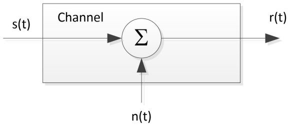
- Signals $x(t)$ and $y(t)$ make up the coordinate system
  - $\rightarrow$ Recall, $x(t)$ modulates the carrier $\cos(2\pi f_c t)$, and $y(t)$ modulates the quadrature carrier $\sin(2\pi f_c t)$
- Possible values of $[x(t), y(t)]$ are plotted on $2D$ plane with the above coordinate system as different points
- Probability of mistaking one transmitted symbol with another is dependent on the distance between the points
- At the receiver, the decision regarding which symbol was transmitted is made by choosing that point of the constellation diagram that is *closest* to the actual received signal location (recall the landmark point mentioned at the beginning of the lecture)

Summary
of Constellation Diagrams

- Signals $x(t)$ and $y(t)$ make up the coordinate system
  - $\rightarrow$ Recall, $x(t)$ modulates the carrier $\cos(2\pi f_c t)$, and $y(t)$ modulates the quadrature carrier $\sin(2\pi f_c t)$
- Possible values of $[x(t), y(t)]$ are plotted on $2D$ plane with the above coordinate system as different points
- Probability of mistaking one transmitted symbol with another is dependent on the distance between the points
- At the receiver, the decision regarding which symbol was transmitted is made by choosing that point of the constellation diagram that is *closest* to the actual received signal location (recall the landmark point mentioned at the beginning of the lecture)

Problem Statement

- We have seen earlier that in the digital communication system, we want to transmit one of M messages.
  - $\rightarrow$ How much information gets conveyed is determined by how much the receiver considers this transmitted symbol to be likely. Those symbols that the receiver considers as highly likely convey less information than those that the receiver views as improbable.
- Symbol gets mapped to a waveform $s_m(t)$. Since there are M symbols, we need a total of M waveforms; i.e., $s_m(t) \in \{s_1(t), s_2(t), \ldots, s_M(t)\}$. These are nonzero over $0 \leq t < T$.
- Let the probabilities of these M symbols be $\{p_1, p_2, \ldots, p_M\}$.
- The received signal is given as $r(t) = s_m(t) + n(t)$, where $n(t)$ is the added noise.



Problem Statement

Goal

- Given $r(t)$, we want to form an estimate $\hat{s}(t)$ of the transmitted signal $s(t) = s_m(t)$
- We want our estimated $\hat{s}(t)$ such that the probability of symbol error, i.e., $P_e = P\left[\hat{s}(t) \neq s_m(t)\right]$, is minimized
- This depends on the probability distribution of the additive noise $n(t)$, which we will take to be Additive White Gaussian Noise or AWGN
  - $\rightarrow$ *Additive* because $n(t)$ adds to $s_m(t)$ (does not get multiplied)
  - $\rightarrow$ *White* because $n(t)$ has autocorrelation function $\phi_{nn}(\tau) = \frac{N_0}{2}\delta(\tau)$
  - $\rightarrow$ *Gaussian* implies that not only $n(t)$ but any linear function of $n(t)$ is also Gaussian distributed

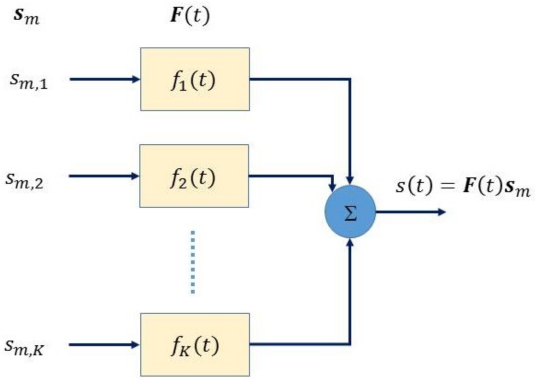
Signal Space Representations

- Transmitted signal $s_m(t)$ is expressed as follows:

$$s_m(t) = \sum_{k=1}^{K} s_{m,k} f_k(t) = \begin{bmatrix} | & | & & | \\ f_1(t) & f_2(t) & \dots & f_K(t) \\ | & | & & | \end{bmatrix} \begin{bmatrix} s_{m,1} \\ s_{m,2} \\ \vdots \\ s_{m,K} \end{bmatrix}$$

$$= \mathbf{F}(t)\mathbf{s}_m$$

Signal Space Representations
Block Diagram



Block diagram description of
$s(t) = \boldsymbol{F}(t)\boldsymbol{s}_m$

Signal Space Representations

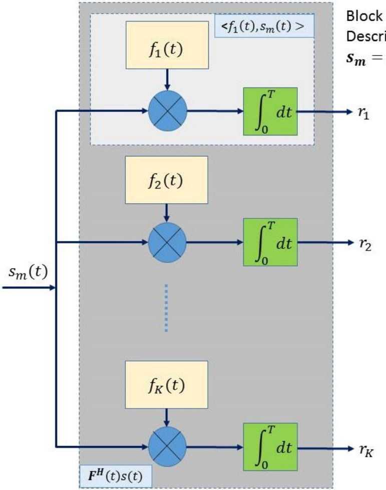
- Here, $\mathbf{F}(t) = \begin{bmatrix} | & | & & | \\ f_1(t) & f_2(t) & \ldots & f_K(t) \\ | & | & & | \end{bmatrix}$ is a "matrix" of continuous-time basis functions $\{f_k(t)\}$. This has a dimension of $\infty \times K$
- Let us also denote $\mathbf{F}^H(t) = \begin{bmatrix} - & f_1(t) & - \\ - & f_2(t) & - \\ & \vdots & \\ - & f_K(t) & - \end{bmatrix}$ is a transposed version of $\mathbf{F}(t)$ with dimensions of $K \times \infty$
- Orthonormality of the basis functions ensures that

$$< \mathbf{F}^H(t), \mathbf{F}(t) >= \mathbf{I}_{K \times K}$$

- Furthermore,

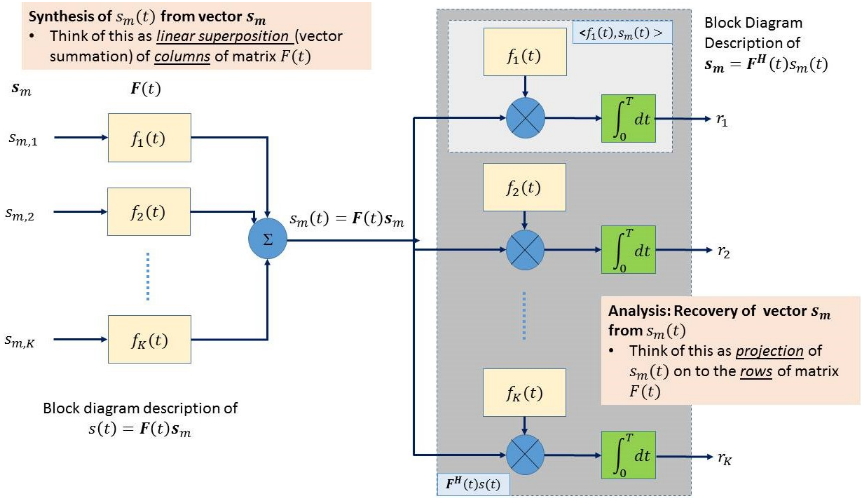
$$< \mathbf{F}^H(t), s_m(t) >= \mathbf{s}_m$$

Signal Space Representations
Block Diagram



Block Diagram Description of $\boldsymbol{s_m} = \boldsymbol{F}^H(t)s_m(t)$

Signal Space Representations
Block Diagram



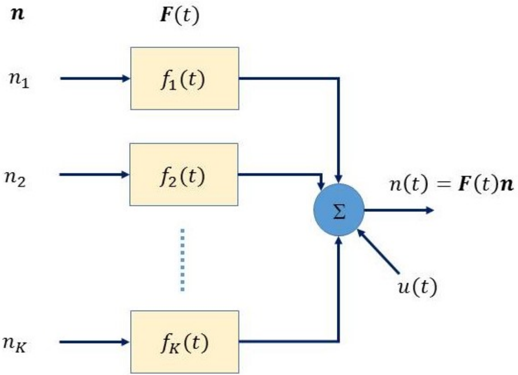
Representation of Noise
in Signal Space

- Consider the AWGN $n(t)$. We can write it as follows:

$$n(t) = \sum_{k=1}^{K} n_k f_k(t) + u(t) = \mathbf{F}(t)\mathbf{n} + u(t)$$

- $K \times 1$ vector $\mathbf{n}$ is given as $\mathbf{n} =< \mathbf{F}^H(t), n(t) >$
- $u(t) = n(t) - \mathbf{F}(t)\mathbf{n}$ is that part of the noise $n(t)$ which is not "spanned" by the basis functions $\{f_k(t)\}$

Representation of Noise
in Signal Space



Block diagram description of
$n(t) = \boldsymbol{F}(t)\boldsymbol{n}$

Noise Has Infinite Dimensions

We Don't Care about Most of Them

We now show that $u(t)$ is orthogonal to any of M signals $\{s_m(t)\}$

$$\begin{aligned}
< s_m^H(t), u(t) > &=< \mathbf{s}_m^H \mathbf{F}^H(t), n(t) - \mathbf{F}(t)\mathbf{n} > \\
&= \mathbf{s}_m^H < \mathbf{F}^H(t), n(t) > - \mathbf{s}_m^H < \mathbf{F}^H(t)\mathbf{F}(t) > \mathbf{n} \\
&= \mathbf{s}_m^H \mathbf{n} - \mathbf{s}_m^H \mathbf{n} \\
&= 0
\end{aligned}$$

This shows that $u(t)$ is orthogonal to any of the signals $\{s_m(t)\}$. As we will show next, the orthogonality implies that $u(t)$ can be ignored in our analysis. It does not affect the operation or design of the receiver.

Statistical Characterization of
Noise Vector $\mathbf{n}$

- Recall that we have earlier taken the noise to be White
- Specifically, $n(t)$ is taken to have an autocorrelation function or ACF
  $\phi_{nn}(\tau) = \frac{N_0}{2}\delta(\tau)$
- How to convert the above ACF to the statistical properties of vector $\mathbf{n}$?
  - → Note that we will be mostly working with vector $\mathbf{n}$ instead of the time-signal $n(t)$
- Let us look at $E\left[\mathbf{n}\mathbf{n}^H\right]$.

Statistical Characterization of
Noise Vector $\mathbf{n}$

$$
\begin{aligned}
E\left[\mathbf{n}\mathbf{n}^H\right] &= E\left[< \mathbf{F}^H(t), n(t) >< n^H(t), \mathbf{F}(t) >\right] \\
&= \mathbf{F}^H(t), E\left[< n(t), n^H(t) >\right] \mathbf{F} \\
&= \mathbf{F}^H(t)_{K\times\infty}, \frac{N_0}{2}\mathbf{I}_{\infty\times\infty}\mathbf{F}(t)_{\infty\times K} \\
&= \frac{N_0}{2}\mathbf{I}_{K\times K}
\end{aligned}
$$

- In summary, the samples of vector $\mathbf{n}$ have
  1. Mean-squared value of $\frac{N_0}{2}$ (this is the same as the variance since the noise samples are taken to have zero mean) and
  2. are uncorrelated

Statistical Characterization of
Noise Vector $\mathbf{n}$

$$
\begin{aligned}
E\left[\mathbf{n}\mathbf{n}^H\right] &= E\left[< \mathbf{F}^H(t), n(t) >< n^H(t), \mathbf{F}(t) >\right] \\
&= \mathbf{F}^H(t), E\left[< n(t), n^H(t) >\right] \mathbf{F} \\
&= \mathbf{F}^H(t)_{K\times\infty}, \frac{N_0}{2}\mathbf{I}_{\infty\times\infty}\mathbf{F}(t)_{\infty\times K} \\
&= \frac{N_0}{2}\mathbf{I}_{K\times K}
\end{aligned}
$$

- In summary, the samples of vector $\mathbf{n}$ have
  1. Mean-squared value of $\frac{N_0}{2}$ (this is the same as the variance since the noise samples are taken to have zero mean) and
  2. are uncorrelated

Statistical Characterization of
Noise Vector $\mathbf{n}$

- Let us look at the joint distribution of vector $\mathbf{n}$

$$
\begin{aligned}
p\left(\mathbf{n}\right) &= p\left(n_1, n_2, \ldots, n_K\right) \\
&= \prod_{k=1}^{K} \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{n_k^2}{N_0}\right) \\
&= \left(\pi N_0\right)^{-K/2} \exp\left(-\sum_{k=1}^{K} \frac{n_k^2}{N_0}\right)
\end{aligned}
$$

Received Signal
in Transmitted Signal's Space

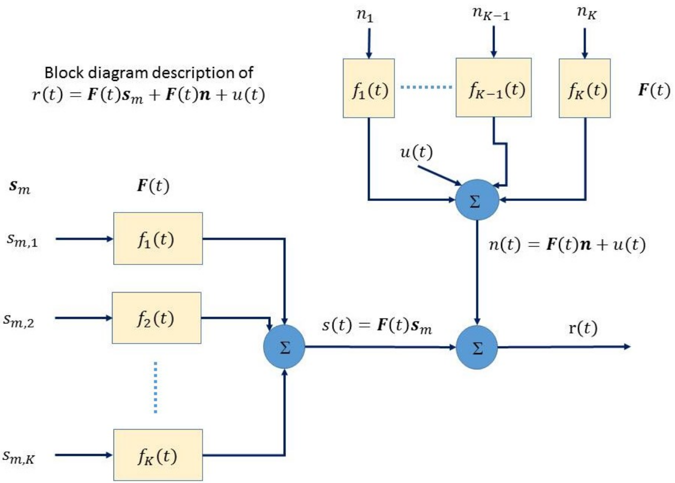
- Received signal $r(t)$ can be expressed as follows:

$$\begin{aligned} r(t) &= s_m(t) + n(t) \\ &= \mathbf{F}(t)\left(\mathbf{s}_m + \mathbf{n}\right) \\ &= \mathbf{F}(t)\mathbf{r} \end{aligned}$$

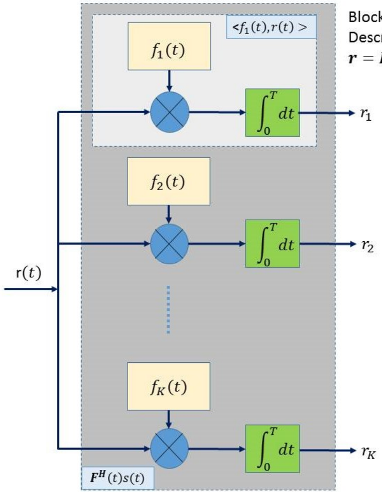
- Alternatively, the continuous-time signal $r(t)$ is equivalent to the following vector:

$$\mathbf{r} = \mathbf{F}^H(t)r(t)$$

Received Signal
in Transmitted Signal's Space

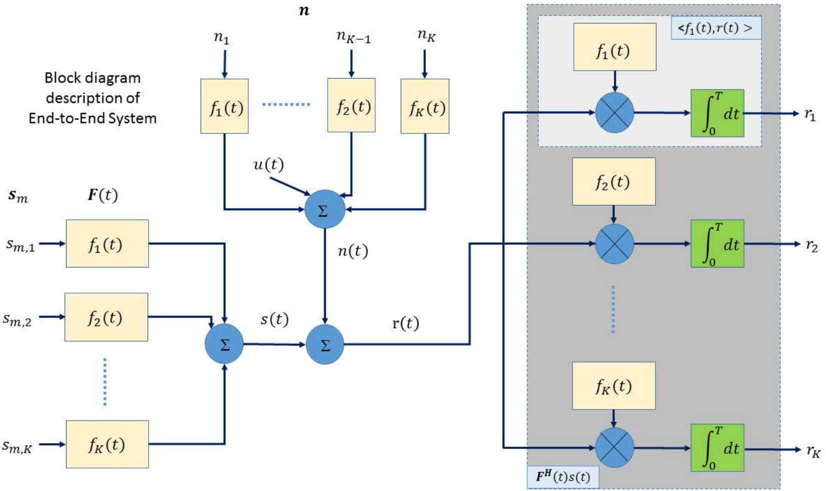


Received Signal
in Transmitted Signal's Space

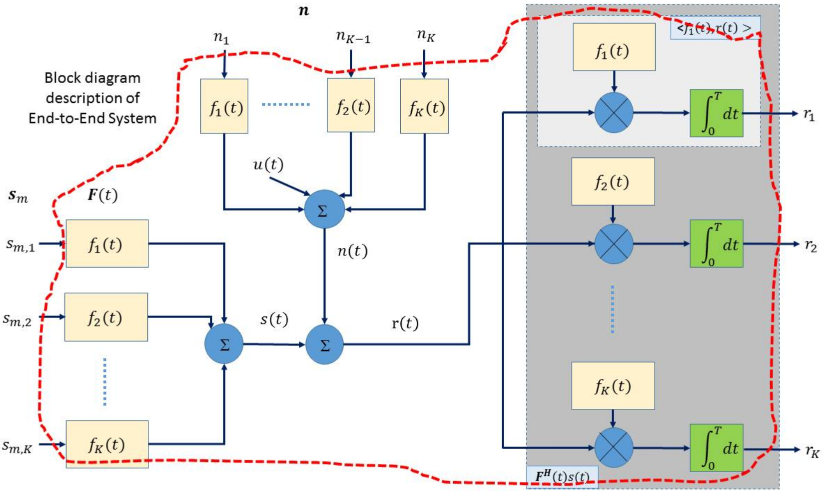


Block Diagram Description of $\boldsymbol{r} = \boldsymbol{F}^{\boldsymbol{H}}(t)s(t)$

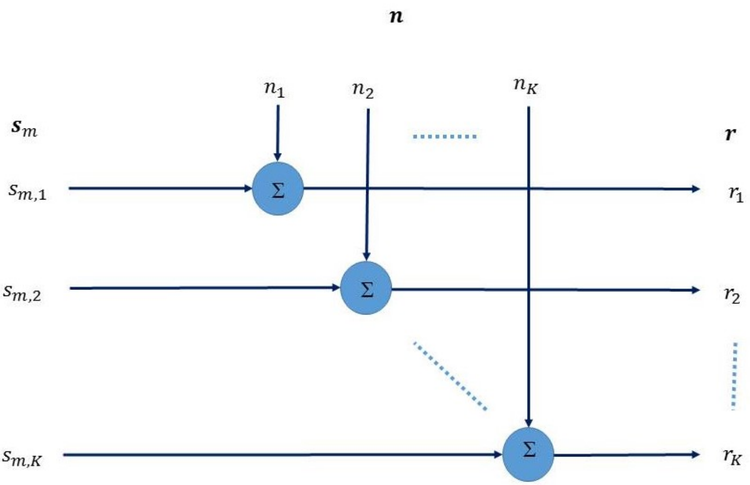
Received Signal
in Transmitted Signal's Space



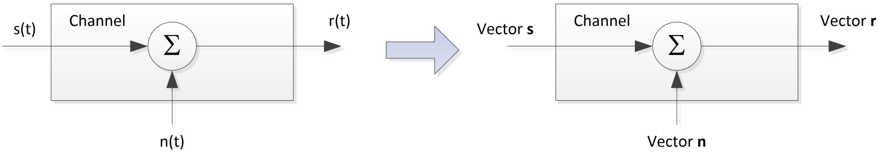
Received Signal
in Transmitted Signal's Space



Received Signal
in Transmitted Signal's Space



Received Signal
in Transmitted Signal's Space



Note: all vectors are K dimensional,
K is the dimension of the transmitted signal subspace

Continuous Time (C-T) Model **Equivalent Discrete Time (D-T) Model**

Received Signal
in Transmitted Signal's Space

- We have earlier derived the joint distribution of vector $\mathbf{n}$

$$p(\mathbf{n}) = (\pi N_0)^{-K/2} \exp\left(-\sum_{k=1}^{K} \frac{n_k^2}{N_0}\right)$$

- From this analysis, we see that $r_k = s_{m,k} + n_k$, i.e., $n_k = r_k - s_{m,k}$
- This shows that the joint distribution of vector $\mathbf{n}$ is the same as the conditional distribution of the received vector $\mathbf{r}$ given the transmitted signal vector $\mathbf{s}_m$.

$$\begin{aligned} p(\mathbf{r}|\mathbf{s}_m) &= (\pi N_0)^{-K/2} \exp\left(-\sum_{k=1}^{K} \frac{(r_k - s_{m,k})^2}{N_0}\right) \\ &= (\pi N_0)^{-K/2} \exp\left(-\frac{\|\mathbf{r} - \mathbf{s}_m\|^2}{N_0}\right) \end{aligned}$$

Signal Space Representations
Summary

- The signal coming into the modulator block is entirely digital
  - → This is typically just a sequence of bits out of channel encoding
- We have seen that it is not possible to transmit the bits over a wireline or wireless channel
- Job of the modulator is, therefore, to convert the bits into analog signals $\{s_m(t)\}_{m=1}^{M}$
  - → Modulator can be thought of as a mapping function that takes a sequence of $\log_2 M$ bits and maps each such sequence to one of M analog signals $\{s_m(t)\}_{m=1}^{M}$

Signal Space Representations

Summary

- While the modulator is a function that maps a sequence of $\log_2 M$ bits into one of M analog signals, the demodulator at the receiver is an inverse function that recovers (ideally the same) sequence of $\log_2 M$ bits given the analog signal that is received
- Thus, the modulator and the demodulator act as interfaces to the actual transmission channel which lives in the analog world. Channel side of this interface is analog and the other side of this interface is digital (until it reaches the point of human interface, at which it often needs to be converted into analog domain)
- This justifies calling this scheme of communication as digital method, although the actual transmission is fully analog

Signal Space Representations

Summary

- Signal space representation provides even further justification (on a more minute level) for considering this as digital communication scheme
- By the virtue of the basis function set $\{f_k(t)\}_{k=1}^{K}$, or equivalently the pseudo-matrix $\mathbf{F}(t) = \begin{bmatrix} | & | & & | \\ f_1(t) & f_2(t) & \dots & f_K(t) \\ | & | & & | \end{bmatrix}$, that are analog-valued, the analog signal $s_m(t)$ that is transmitted over the channel can itself be thought of as essentially digital in nature since $s(t) = \mathbf{F}(t)\mathbf{s}_m$
- Similarly, the received signal $r(t)$ which is analog-valued can also be thought of having essentially a digital nature, since $r(t)$ can be decomposed as vector $\mathbf{r}$ by decomposition $\mathbf{r} = \mathbf{F}^H(t)r(t)$