IK2651 Principles of Communications

Lecture 12
Information Theory

Richard Schatz
rschatz@kth.se
08-7904069
KTH 2011

Outline

Introduction to information theory

The source: definitions of information content, source entropy etc.

The channel: definition of channel matrix, channel capacity etc.

Channel capacity vs E_b/N_0 for different modulation formats

Shannon´s Theorem:

- Minimum SNR for Error-Free Transmission
- Spectral Efficiency vs SNR in Digital and Analog Systems

Introduction

Consider a communication system with limited transmitter power (and/or bandwidth) and a certain noise spectral density. We want to optimize the capacity, i.e., transmit as much error-free information per time unit as possible.

Which modulation format should we select?

Should we increase the number of modulation levels, M?
(increases bitrate but also bit error rate)

Should we use higher symbol rate?
(increases bitrate but also bit error rate, costs bandwidth)

Should we use error correcting coding?
(reduces bit error rate but costs overhead, i.e., extra redundant bits)

What maximum capacity can we obtain?

Information theory is needed to tackle these questions!

Definition of Information

$$I\left(x_j \text{ occurs}\right) = \log_2\left(\frac{1}{p(x_j)}\right) = -\log_2\left(p(x_j)\right)\text{bits}$$

$$p(x_j) = 0.5 \Rightarrow I = 1\,\text{bit}$$

$$p(x_j) = 1 \Rightarrow I = 0\,\text{bit}$$

$$p(x_j) = 0 \Rightarrow I = \infty\,\text{bit}$$

<u>Units</u>

$\log_2$	[bits] or [Sh] (Shannon)
$\log_{10}$	[Hartley]
ln	[nats]

The information content is larger for uncommon events than for common. Compare with news reports.

An outcome of two, equally probably alternatives, have information content 1 bit

Entropy

The entropy of a information source with different possible outcomes (e.g. symbol alternatives) is the average information content of these outcomes

$$H[X] = E[I(x_j)] = \sum_{j=1}^{n} p(x_j)\log_2\left(\frac{1}{p(x_j)}\right) =$$

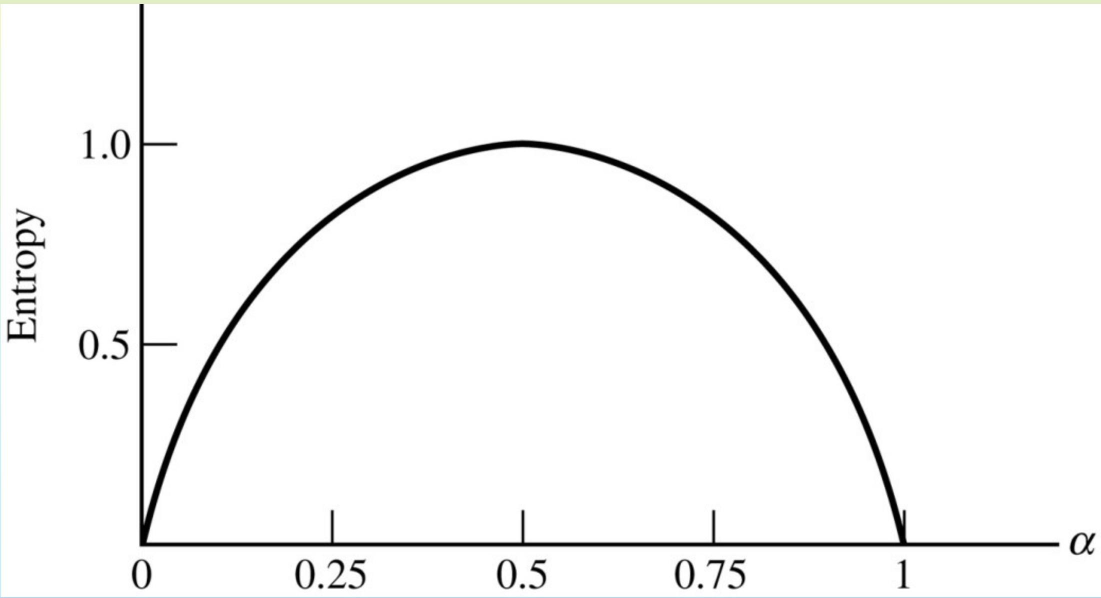
$$= -\sum_{j=1}^{n} p(x_j)\log_2(p(x_j))$$

X: discrete random variable with outcomes $x_1, x_2, \ldots, x_n$

Entropy of a Binary Source

Binary Source: $p(x_1)=\alpha \quad p(x_1)=1-\alpha$

$$H[X] = -\sum_{j=1}^{n} p(x_j)\log_2(p(x_j)) =$$
$$= -\alpha\log_2(\alpha) - (1-\alpha)\log_2(1-\alpha)$$



The entropy of a binary source is maximum 1 bit when both signal alternatives are equally probable

Maximum Entropy of a Source with n Possible Outcomes

X: discrete random variable with outcomes $x_1, x_2, \ldots, x_n$

Entropy is maximum when all outcomes are equally probable

$$p(x_j) = \frac{1}{n} \text{ for } j = 1 \ldots n$$

$$\boxed{H_{\max}[X] = -\sum_{j=1}^{n} p(x_j) \log_2 (p(x_j)) = \log_2(n)}$$

Proof: $\quad p_n = 1 - \sum_{1}^{n-1} p_i$

ln is chosen for simplicity

$$\frac{dH}{dp_k} = \frac{d}{dp_k}\left[-p_k \ln(p_k) - p_n \ln(p_n)\right]$$

$$= -p_k \frac{1}{p_k} - \ln(p_k) - p_n\left(\frac{-1}{p_n}\right) - (-1)\ln(p_n)$$

$$= \ln \frac{p_n}{p_k} = 0 \Rightarrow p_n = p_k$$

$$p_k \text{ arbitrary} \Rightarrow p_i = \frac{1}{n}$$

Laboratory of PHOTONICS and MICROWAVE Eng.

Outline

Introduction to information theory

The source: definitions of information content, source entropy etc.

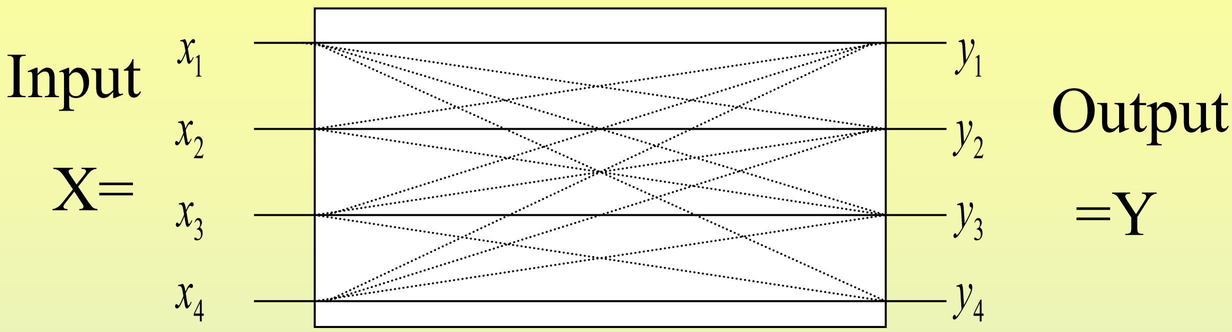
<u>The channel: definition of channel matrix, channel capacity etc.</u>

Channel capacity vs E_b/N_0 for different modulation formats

Shannon´ s Theorem:

- Minimum SNR for Error-Free Transmission
- Spectral Efficiency vs SNR in Digital and Analog Systems

Probability Definitions



$$[P(X)] = [p(x_1) \quad \ldots \quad p(x_n)]$$

Source probability vector
Vector of probabilites that a certain symbol is sent.

$$[P(Y)] = [p(y_1) \quad \ldots \quad p(y_n)]$$

Output probability vector

$$[P(X|Y)] = \begin{bmatrix} p(x_1|y_1) & \cdots & p(x_n|y_1) \\ \vdots & \ddots & \vdots \\ p(x_1|y_n) & \cdots & p(x_n|y_n) \end{bmatrix}$$

Y- **Conditional probability matrix**
Matrix of probabilities that a certain symbol x_i was sent *when* a symbol y_j was received

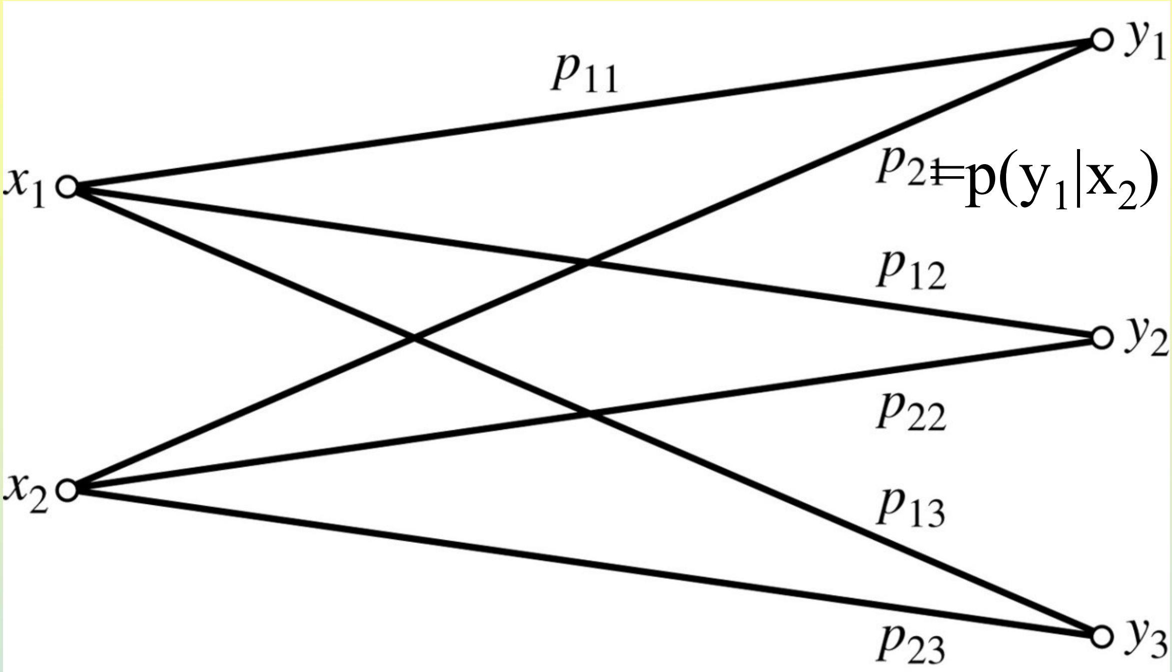
$$[P(X,Y)] = \begin{bmatrix} p(x_1,y_1) & \cdots & p(x_n,y_1) \\ \vdots & \ddots & \vdots \\ p(x_1,y_n) & \cdots & p(x_n,y_n) \end{bmatrix}$$

Joint probability matrix
Matrix of probabilities that a certain symbol x_i was sent *and* a symbol y_j was received

$$p(x_i, y_j) \equiv p(x_i \text{ and } y_j) = p(x_i|y_j)\,p(y_j) = p(y_j|x_i)\,p(x_i)$$

Laboratory of PHOTONICS and MICROWAVE Eng.

Channel Matrix



$$P(X) = [p(x_1), p(x_2)] \qquad P(Y) = [p(y_1), p(y_2), p(y_3)]$$

$$P(Y|X) = \text{channel matrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \end{bmatrix}$$

$$\boxed{P(Y) = P(X)P(Y|X)}$$

Joint Probability Matrix

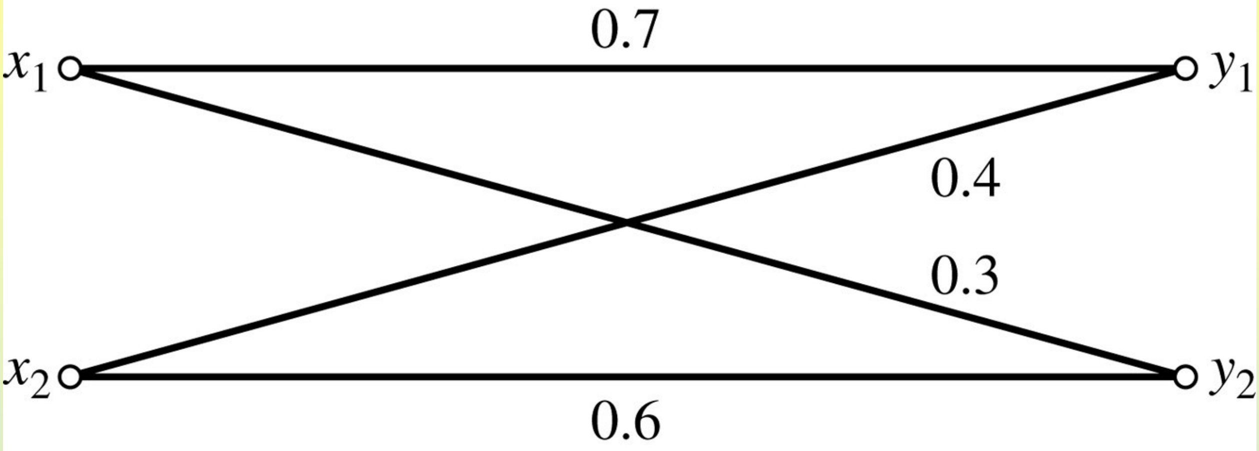
$$P[X,Y] = \begin{bmatrix} p(x_1,y_1) & p(x_1,y_2) & p(x_1,y_3) \\ p(x_2,y_1) & p(x_2,y_2) & p(x_2,y_3) \end{bmatrix} = \left\{ \begin{matrix} \text{joint} \\ \text{probility} \\ \text{matrix} \end{matrix} \right\}$$

$$= \underbrace{P[X]}_{\text{diagonal}} P[Y|X] = \begin{bmatrix} p(x_1) & 0 \\ 0 & p(x_2) \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \end{bmatrix}$$

$$= \begin{bmatrix} p(x_1) & 0 \\ 0 & p(x_2) \end{bmatrix} \begin{bmatrix} p(y_1|x_1) & p(y_2|x_1) & p(y_3|x_1) \\ p(y_1|x_2) & p(y_2|x_2) & p(y_3|x_2) \end{bmatrix}$$

$$= P[X,Y]$$

Example:Binary Channel



$$P(X)=[0.5,0.5] \quad P(Y|X)=\text{channel matrix }=\begin{bmatrix}0.7 & 0.3\\ 0.4 & 0.6\end{bmatrix}$$

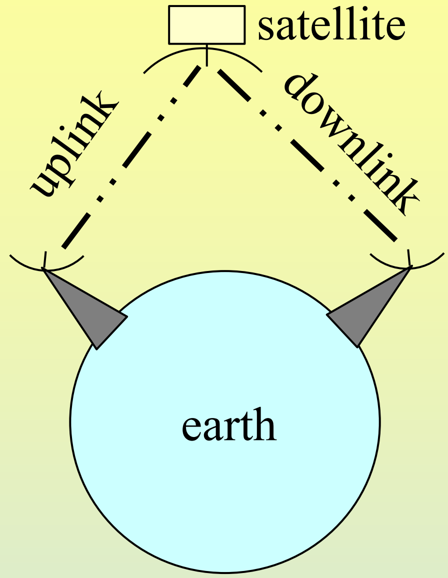
Calculate $P(Y)$ and $P(X,Y)$

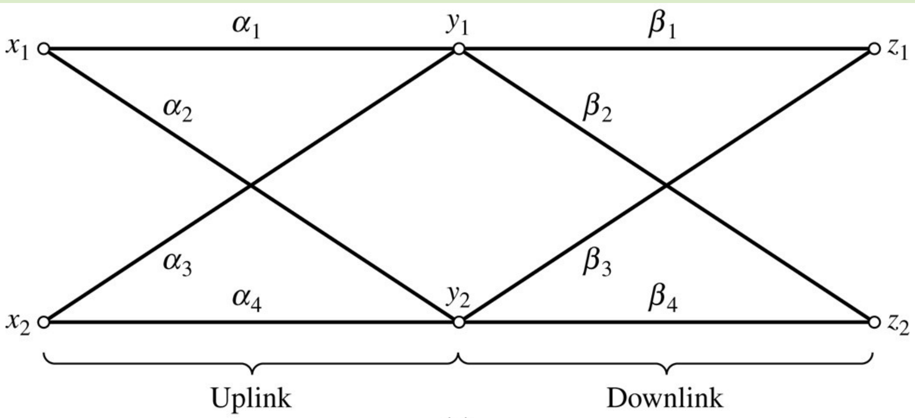
$$P(Y)=P(X)P(Y|X)=[0.5,0.5]\begin{bmatrix}0.7 & 0.3\\ 0.4 & 0.6\end{bmatrix}=[0.55,0.45]$$

$$P(X,Y)=\underset{diagonal}{P(X)}P(Y|X)=\begin{bmatrix}0.5 & 0\\ 0 & 0.5\end{bmatrix}\begin{bmatrix}0.7 & 0.3\\ 0.4 & 0.6\end{bmatrix}=\begin{bmatrix}0.35 & 0.15\\ 0.2 & 0.3\end{bmatrix}$$

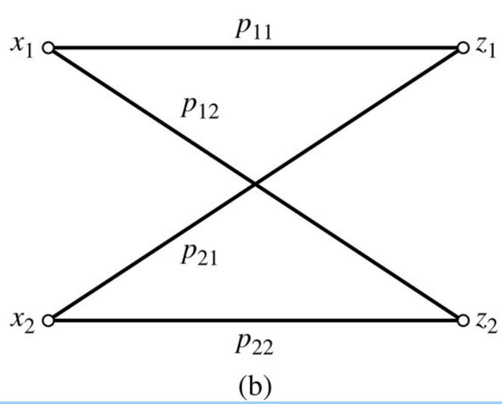
Laboratory of PHOTONICS and MICROWAVE Eng.

Two-Hop Satellite Channel





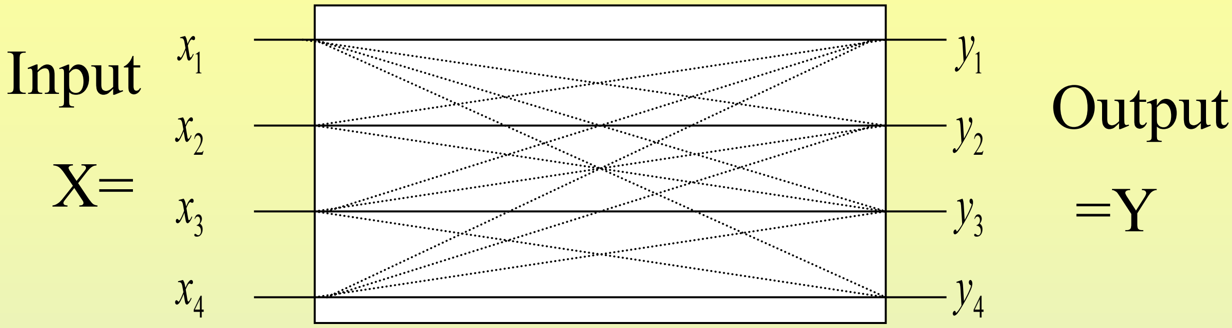
(a)



(b)

$$P(Z|X) = P(Y|X)P(Z|Y) =$$
$$= \begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{bmatrix} \begin{bmatrix} \beta_1 & \beta_2 \\ \beta_3 & \beta_4 \end{bmatrix} =$$
$$= \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix}$$

Important Entropy Definitions



All values are scalars!

$$I(x_i) = -log_2\, p(x_i)$$

Information content
A priori uncertainty of symbol i

$$H(X) = -\sum_i p(x_i)\log_2 p(x_i)$$

Entropy
(Average) uncertainty of input symbol

$$H(X|Y) = -\sum_{i,j} p(x_i, y_j)\log_2 p(x_i|y_j)$$

Y- **Conditional Entropy**
(Average) uncertainty of input symbol when we have received an output symbol

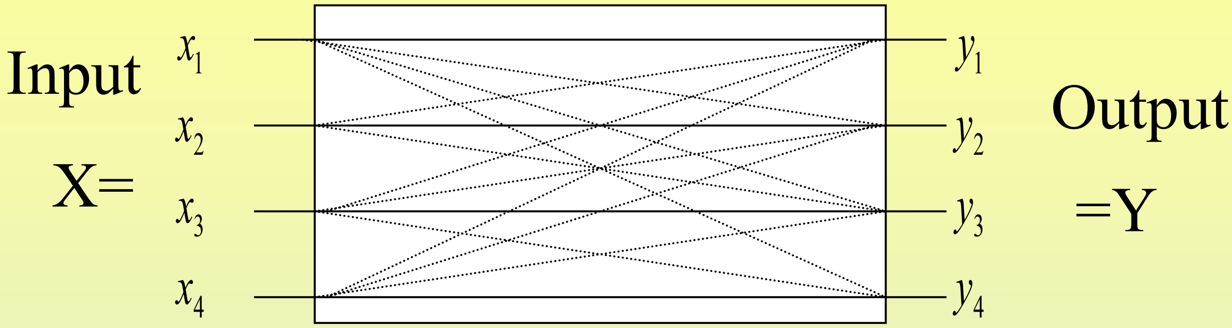
$$C_{bit/symbol} = \mathrm{Max}_X\left[H(X) - H(X|Y)\right]$$

Channel capacity
Decrease of input uncertainty after reception using an optimal input signal

$$C_{bps} = R_s \cdot C_{bit/symbol}$$

R_S is the symbol rate

Additional Entropy Definitions



$$H(Y|X) = -\sum_{i,j} p(x_i, y_j)\log_2 p(y_i|x_j)$$

X- **Conditional entropy**
Uncertainty of output symbol when we know the sent input symbol

$$I(X;Y) = H(X) - H(X|Y)$$
$$I(X;Y) = H(Y) - H(Y|X)$$

Mutual information
(Average) decrease of uncertainty of input symbol when we have received output symbol

$$H(X,Y) = -\sum_{i,j} p(x_i, y_j)\log_2 p(x_i, y_j)$$

Joint entropy
Average uncertainty of communication system as a whole

$$H(X,Y) = H(X|Y) + H(Y) = H(Y|X) + H(X)$$

Outline

Introduction to information theory

The source: definitions of information content, source entropy etc.

The channel: definition of channel matrix, channel capacity etc.

<u>Channel capacity vs E_b/N_0 for different modulation formats</u>

Shannon´s Theorem:

- Minimum SNR for Error-Free Transmission
- Spectral Efficiency vs SNR in Digital and Analog Systems

Channel Capacity

$I(X;Y)$ = mutual information = $H(X) - H(X|Y)$

Average decrease of uncertainty of input symbol when we have received output symbol=average transmitted information

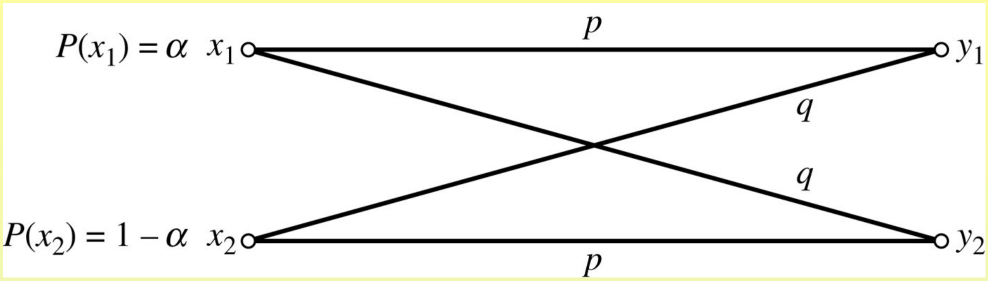
$$I(X;Y) = -\sum_i p(x_i)\log_2 p(x_i) - \left[-\sum_{i,j} p(x_i,y_j)\log_2 p(x_i|y_j)\right] =$$
$$= -\sum_{i,j} p(x_i,y_j)\log_2 p(x_i) - \left[-\sum_{i,j} p(x_i,y_j)\log_2 p(x_i|y_j)\right] =$$
$$= -\sum_{i,j} p(x_i,y_j)\log_2\left[\frac{p(x_i)}{p(x_i|y_j)}\right] = -\sum_{i,j} p(x_i,y_j)\log_2\left[\frac{p(x_i)p(y_j)}{p(x_i,y_j)}\right] =$$
$$= H(Y) - H(Y|X) \geq 0$$

$I(X;Y)$ depends both on the source probabilities, $P(X)$, and the channel matrix, $P(Y|X)$. The channel capacity, C, is the mutual information for the optimum source

$$\boxed{C = \max_{\substack{\text{with respect}\\ \text{to P(X) of}\\ \text{source}}} \left[I(X;Y)\right]}$$

For a noisless channel $I(X;Y)=H(X)$ and hence $C=\log_2(n)$

Capacity for Binary Channel



Find C

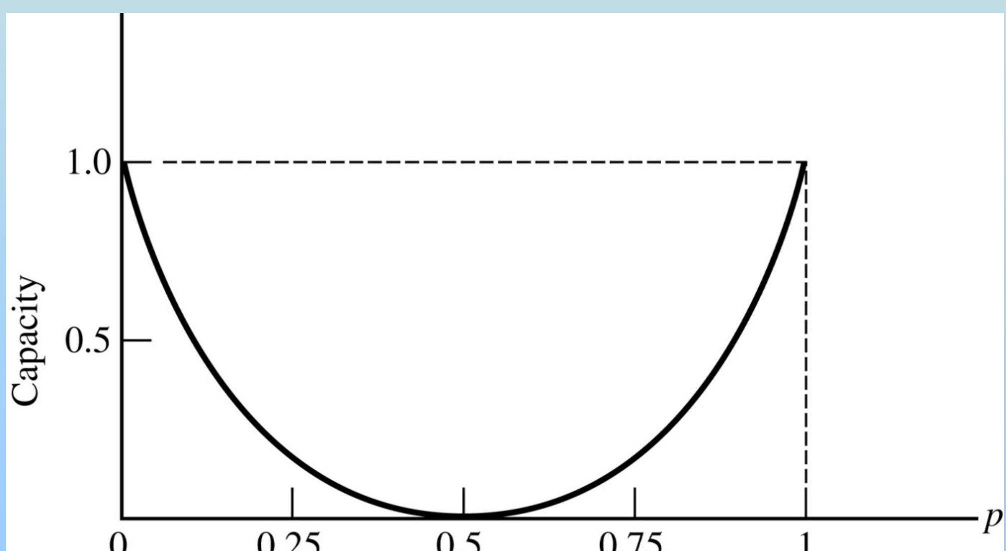
$$I(X;Y) = H(Y) - H(Y \mid X) \qquad H(Y \mid X) = -\sum_{i,j} p(x_i, y_j) \log_2 p(y_i | x_j)$$

$$P(X,Y) = \underbrace{P(X)}_{diagonal} P(Y|X) = \begin{bmatrix} \alpha & 0 \\ 0 & 1-\alpha \end{bmatrix} \begin{bmatrix} p & q \\ q & p \end{bmatrix} = \begin{bmatrix} \alpha p & \alpha q \\ (1-\alpha)q & (1-\alpha)p \end{bmatrix}$$

$$H(Y|X) = -\alpha p \log_2 p - \alpha q \log_2 q - (1-\alpha) q \log_2 q - (1-\alpha) p \log_2 p =$$

$$= -q \log_2 q - p \log_2 p \text{ (independent of source!)}$$

$$\boxed{C = \max[I(X;Y)] = \max(H(Y)) - H(Y \mid X) = 1 + q \log_2 q + p \log_2 p}$$



Laboratory of PHOTONICS and MICROWAVE Eng.

Bit Error Rate vs E_b/N_0

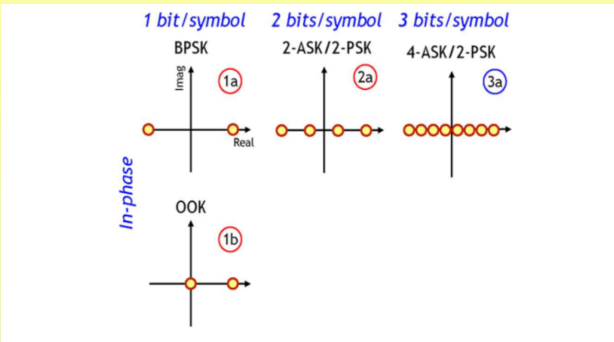


Fig. 8. Examples of constellations using only one quadrature of the field (here the real part). The number of bits/symbol is given by $\log_2(M)$ where M is the total number of symbols. The number $\log_2(M)$ of symbols is used as the first digit of the format label.

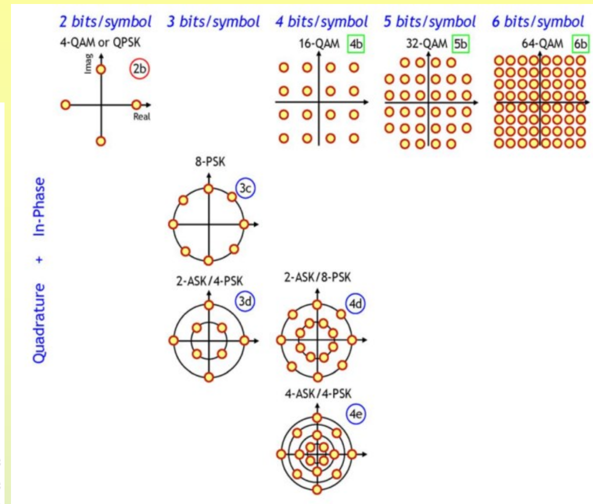


Fig. 9. Examples of constellations that use both quadratures of the field.

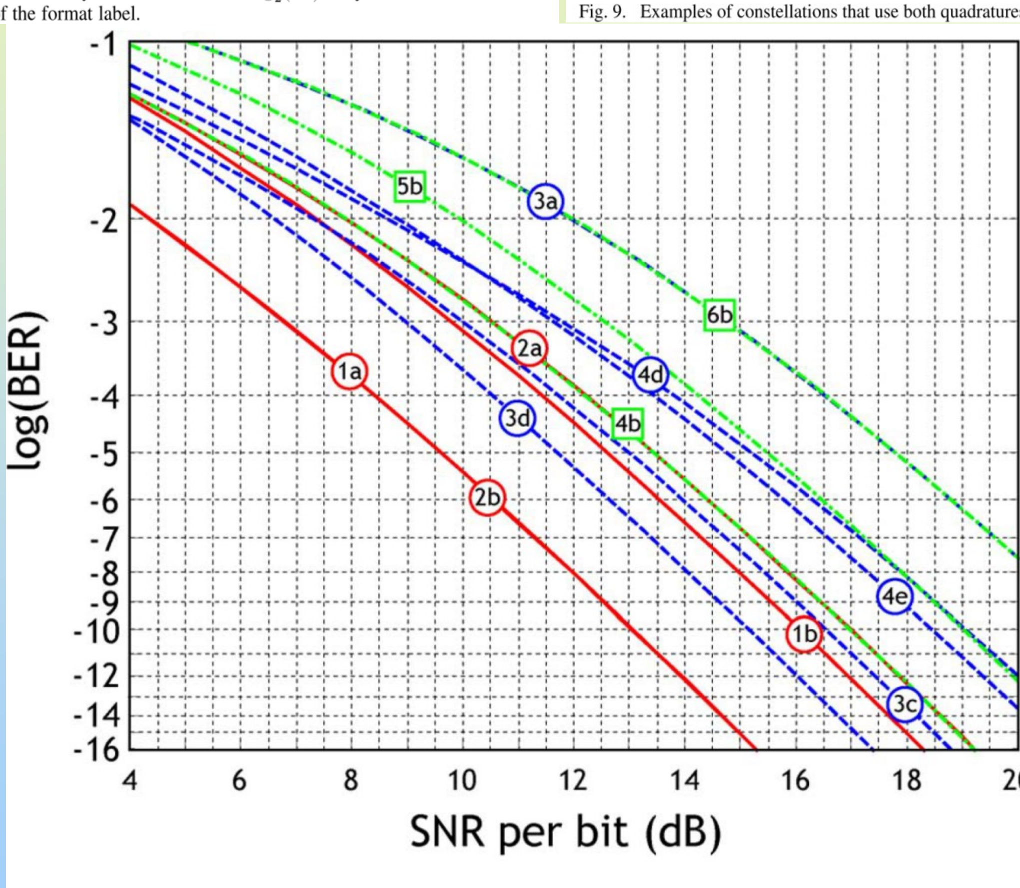


Fig. 11. BER as a function of *SNR per bit* for the modulation formats of Figs. 8 and 9.

R. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity Limits of Optical Fiber Networks," J. Lightwave Technol. **28**, 662-701 (2010)

Laboratory of PHOTONICS and MICROWAVE Eng.

Capacity vs E_b/N_0

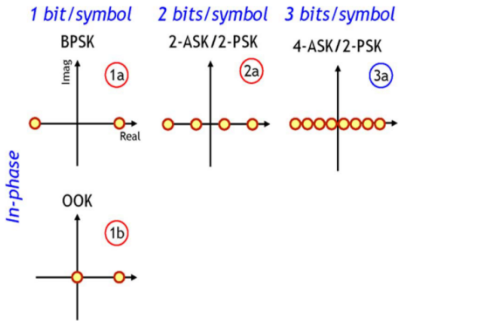


Fig. 8. Examples of constellations using only one quadrature of the field (here the real part). The number of bits/symbol is given by $\log_2(M)$ where M is the total number of symbols. The number $\log_2(M)$ of symbols is used as the first digit of the format label.

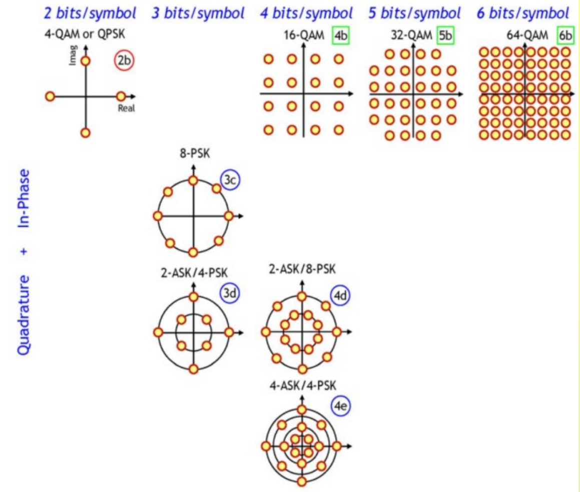


Fig. 9. Examples of constellations that use both quadratures of the field.

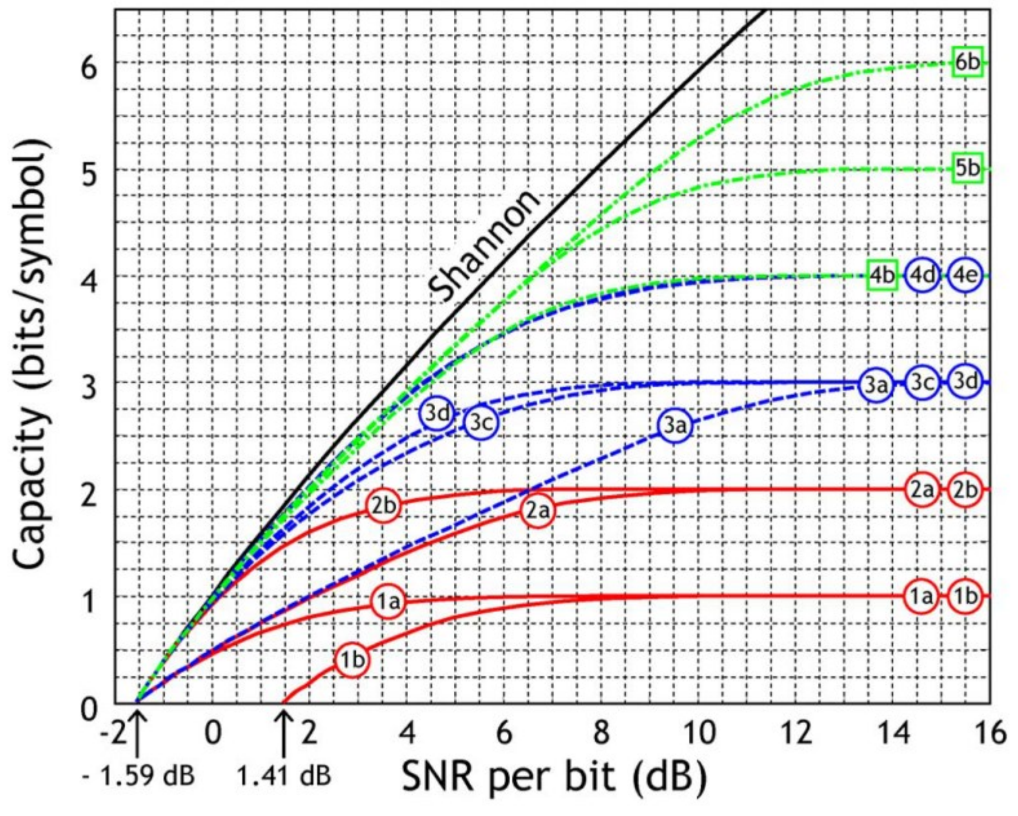


Fig. 13. Capacity as a function of SNR per bit of information for the modulation formats of Figs. 8 and 9, respectively.

R. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity Limits of Optical Fiber Networks," J. Lightwave Technol. **28**, 662-701 (2010)

Laboratory of PHOTONICS and MICROWAVE Eng.

Outline

Introduction to information theory

The source: definitions of information content, source entropy etc.

The channel: definition of channel matrix, channel capacity etc.

Channel capacity vs E_b/N_0 for different modulation formats

Shannon´s Theorem:

•Minimum SNR for Error-Free Transmission

•Spectral Efficiency vs SNR in Digital and Analog Systems

Laboratory of PHOTONICS and MICROWAVE Eng.

Shannon´s Theorem

Shannons fundamental theorem:

Claude Elwood Shannon
USA, 1916-2001

The maximum capacity of a physical link with bandwidth B and AWGN (additive white Gaussian Noise) is

$$\boxed{C_{\max} = B\log_2\left(1+SNR\right)}$$

A discrete memoryless source with information rate R[bits/s] can be transmitted with negligable error through a noisy channel with capacity C_c[bit/s] if $R<C_c$

The theorem does not state *how* this should be achieved (modulation format etc.)

One can compensate a low system bandwidth with increased transmitter power!

Alternatively, if one has a lot of bandwidth, one can decrease the transmitter power. However, in this case there is a limit how far the transmitter power can be decreased.

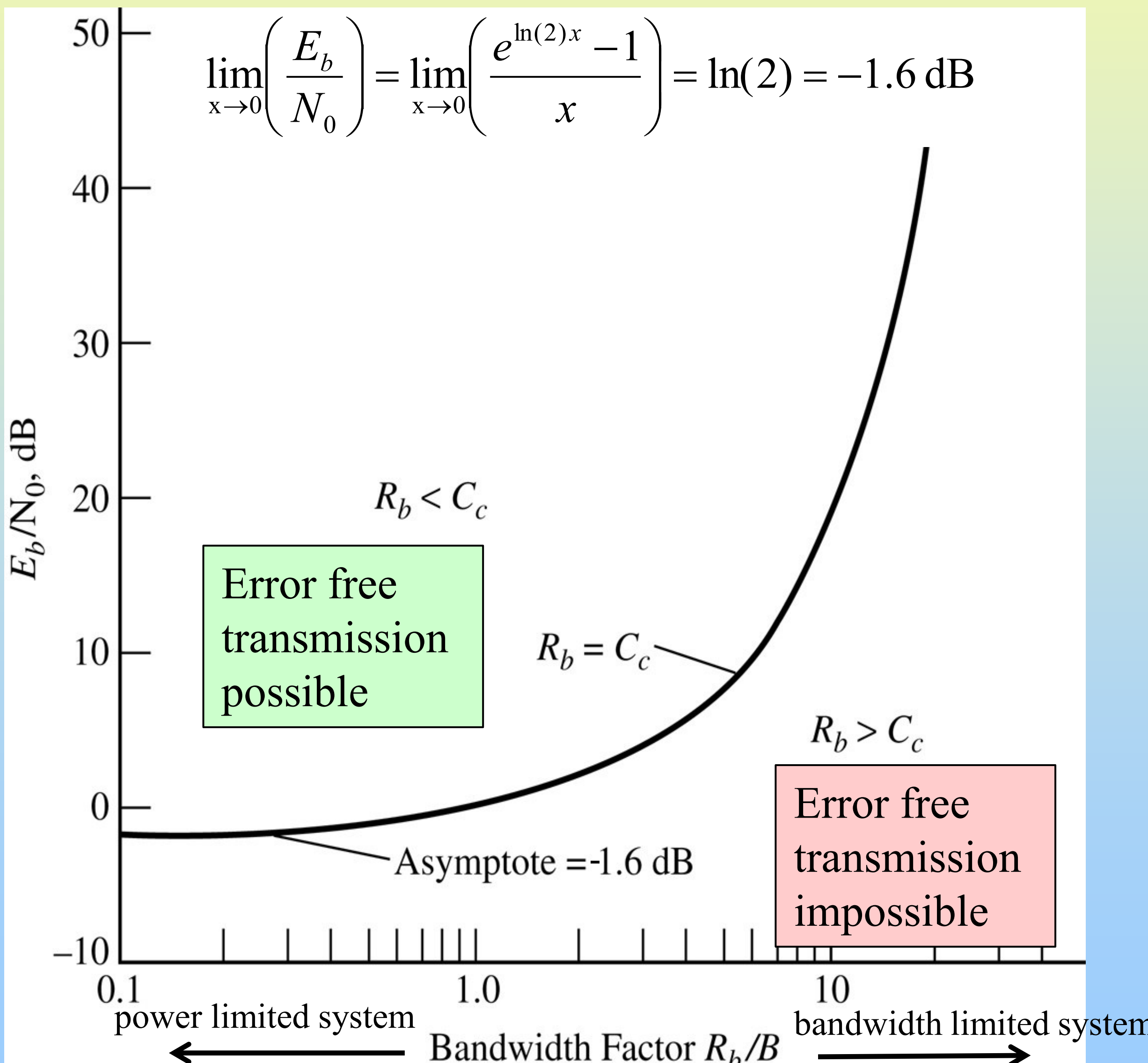
Minimum E_b/N_0 to Transmit Information Error-Free

$$\frac{C_c}{B} = \log_2(1+SNR) = \log_2\left(1+\frac{R_b E_b}{N_0 B}\right) = \left\{\begin{matrix} R_b = C_c : \text{ideal} \\ \text{modulation} \end{matrix}\right\} = \log_2\left(1+\frac{C_c}{B}\frac{E_b}{N_0}\right)$$

$$\frac{E_b}{N_0} = \frac{2^x - 1}{x} \qquad \text{where } x = \frac{C_c}{B} = \frac{R_b}{B}$$

"spectral efficiency"

$$\lim_{x\to 0}\left(\frac{E_b}{N_0}\right) = \lim_{x\to 0}\left(\frac{e^{\ln(2)x}-1}{x}\right) = \ln(2) = -1.6\text{ dB}$$



Capacity vs E_b/N_0

Similar figure as last but with axes exchanged

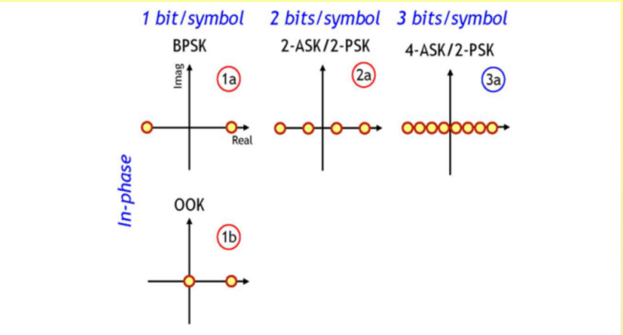


Fig. 8. Examples of constellations using only one quadrature of the field (here the real part). The number of bits/symbol is given by $\log_2(M)$ where M is the total number of symbols. The number $\log_2(M)$ of symbols is used as the first digit of the format label.

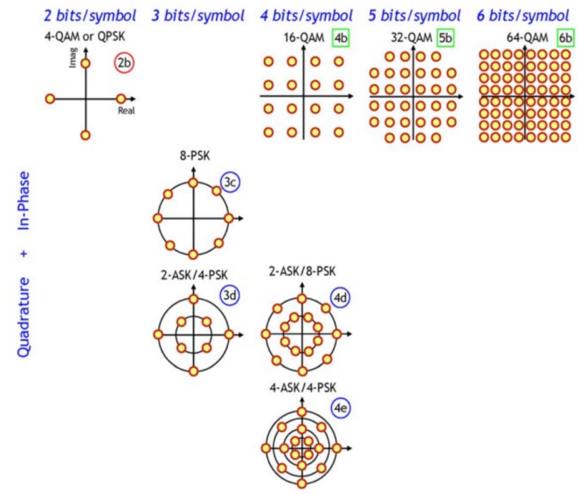


Fig. 9. Examples of constellations that use both quadratures of the field.

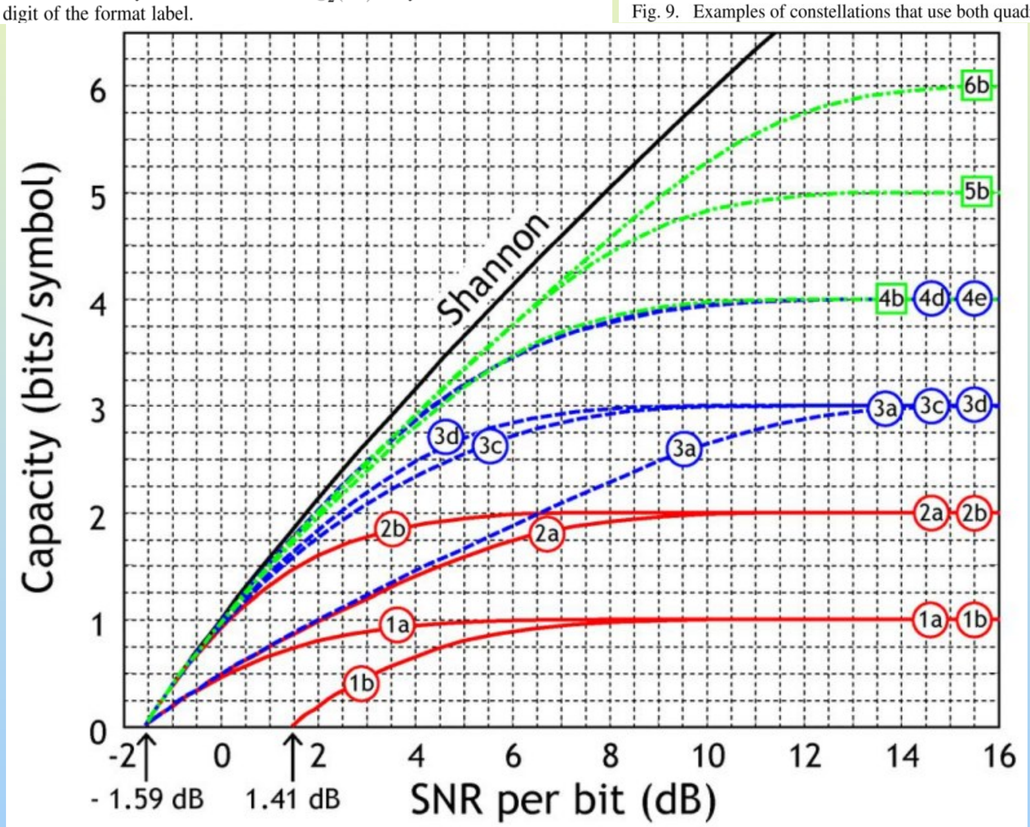
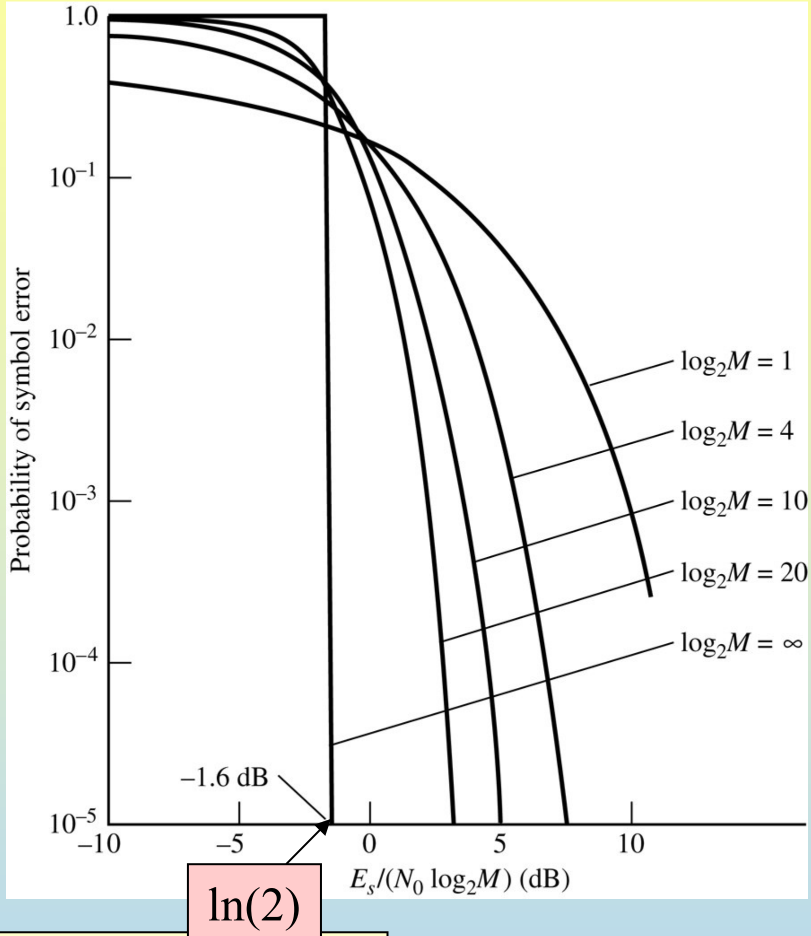


Fig. 13. Capacity as a function of SNR per bit of information for the modulation formats of Figs. 8 and 9, respectively.

R. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity Limits of Optical Fiber Networks," J. Lightwave Technol. **28**, 662-701 (2010)

Symbol Error Probability for M-ary Orthogonal Signals



It is possible to reach the ultimate power limit:

$$\frac{E_b}{N_0} = \ln(2)$$

by increasing the number of orthogonal functions, i.e. OFDM with many subcarriers, but it costs bandwidth

For OFDM the power consumption in DA/AD converters and DSP-units can be several orders of magnitude larger than the transmitted power

Capacity vs E_b/N_0

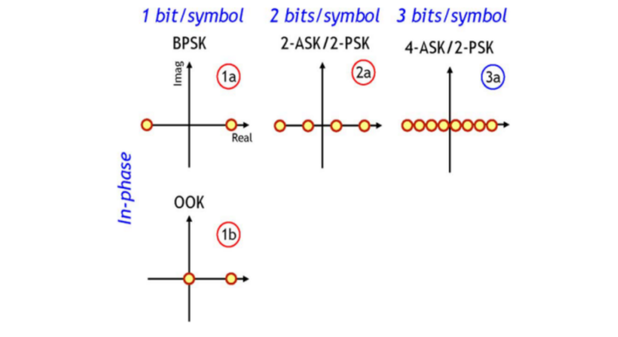


Fig. 8. Examples of constellations using only one quadrature of the field (here the real part). The number of bits/symbol is given by $\log_2(M)$ where M is the total number of symbols. The number $\log_2(M)$ of symbols is used as the first digit of the format label.

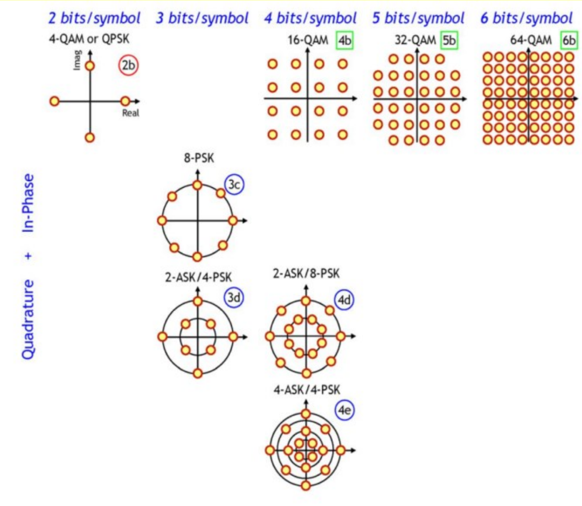


Fig. 9. Examples of constellations that use both quadratures of the field.

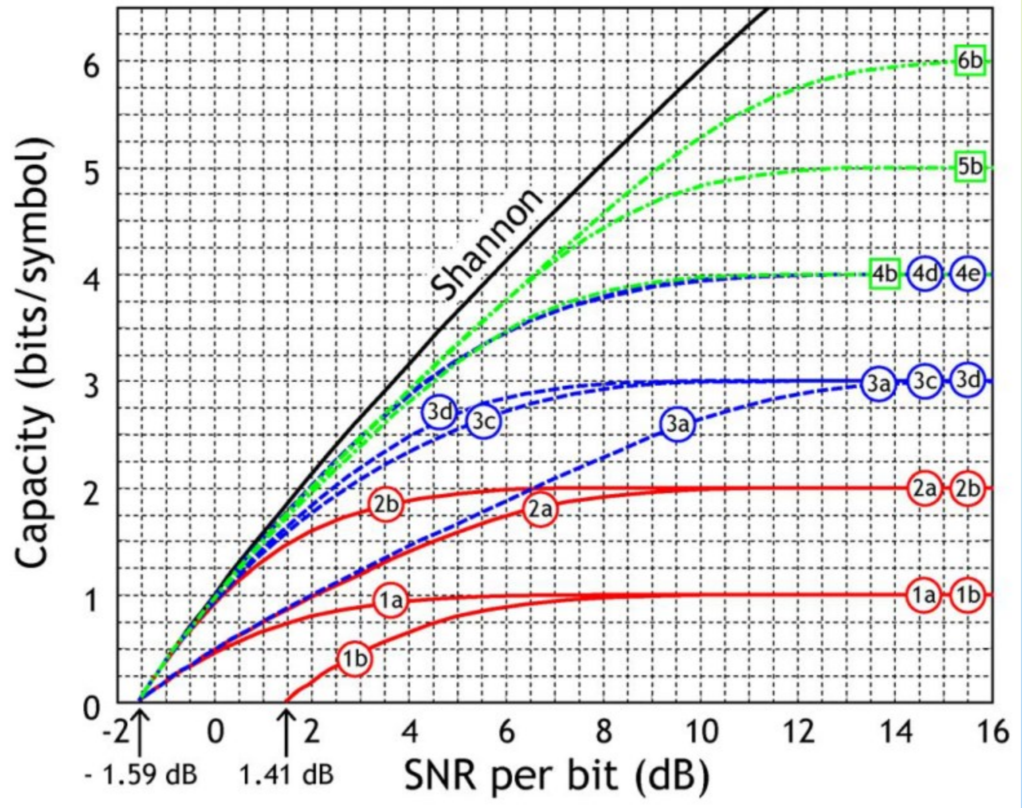
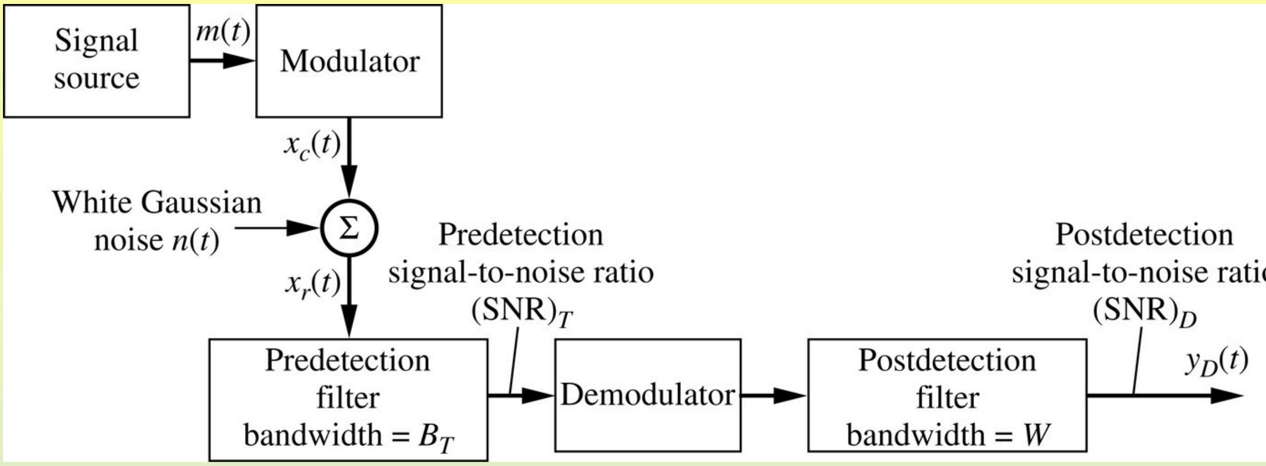


Fig. 13. Capacity as a function of SNR per bit of information for the modulation formats of Figs. 8 and 9, respectively.

R. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity Limits of Optical Fiber Networks," J. Lightwave Technol. **28**, 662-701 (2010)

Laboratory of PHOTONICS and MICROWAVE Eng.

SNR vs Spectral Efficiency in Analog Systems



Predetection capacity

$$\boxed{C_T = B_T \log_2\left(1 + SNR_T\right)}$$

Postdetection capacity

$$\boxed{C_D = W \log_2\left(1 + SNR_D\right)}$$

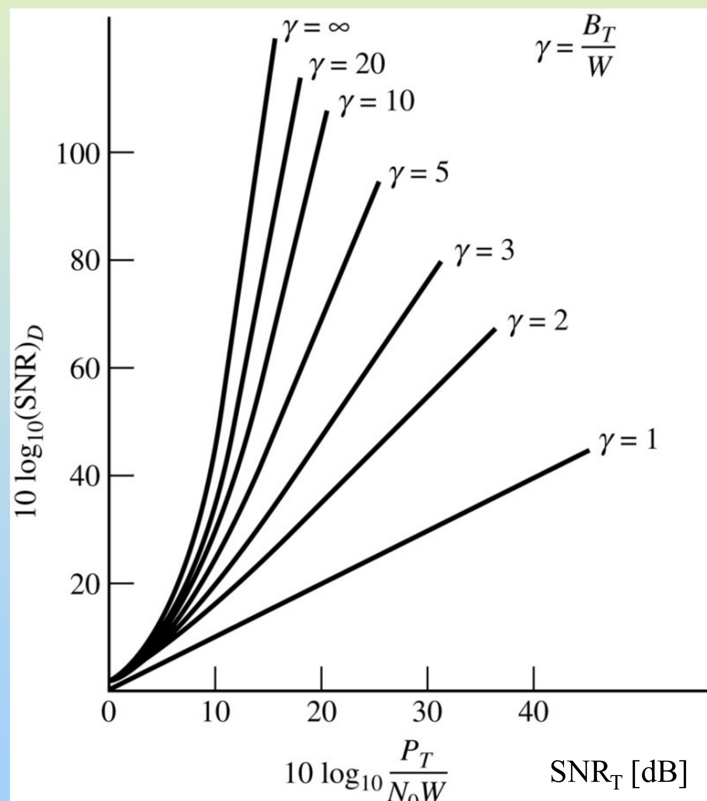
Optimal modulation $C_D = C_T$

⇨

$$SNR_D = \left[1 + SNR_T\right]^{B_T/W} - 1 =$$

$$= \left[1 + \frac{1}{\gamma}\left(\frac{P_T}{N_0 W}\right)\right]^{\gamma} - 1$$

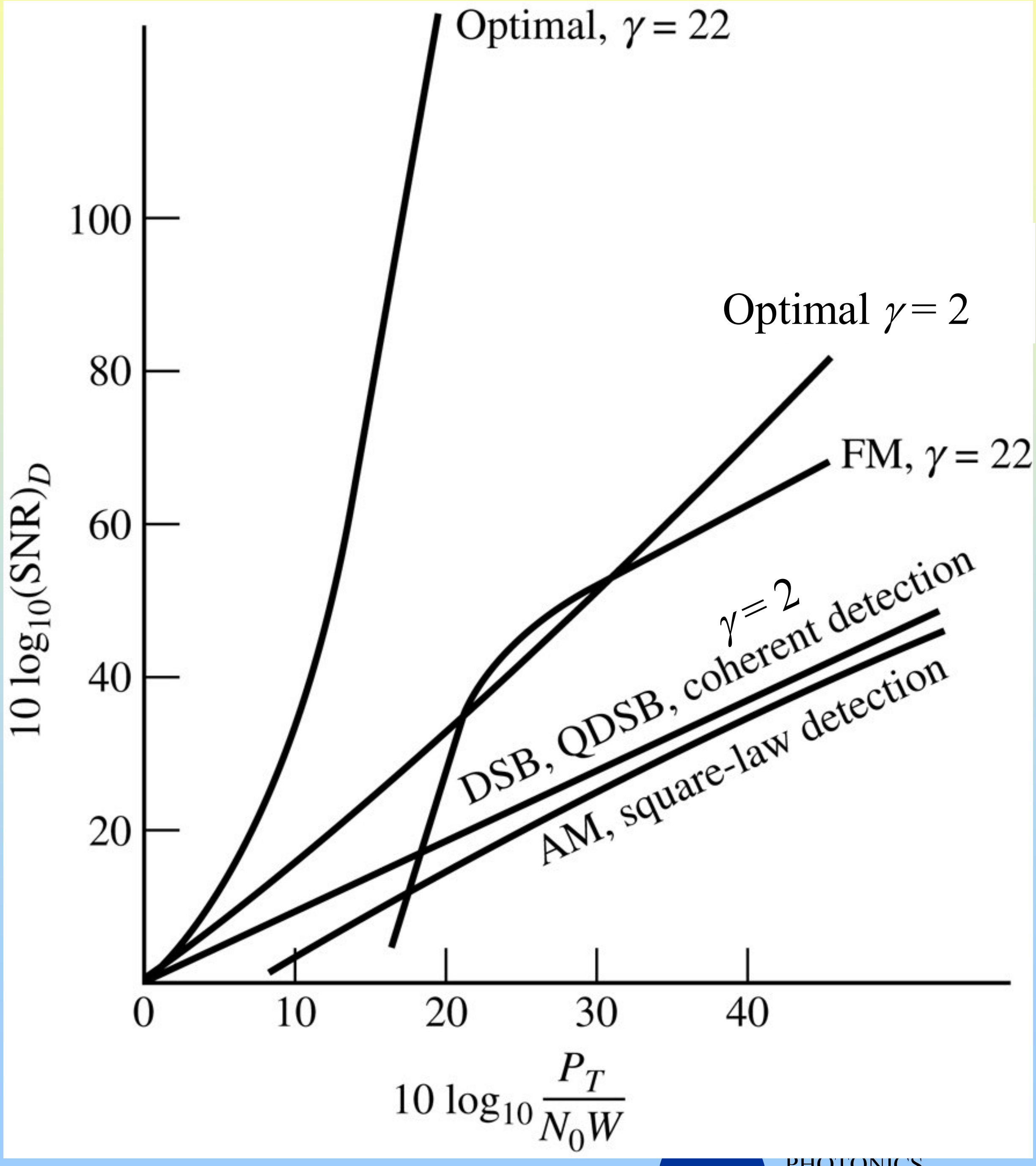
with $\gamma = \dfrac{B_T}{W}$ bandwidth expansion factor



Improved SNR costs bandwidth (or power)!

Laboratory of PHOTONICS and MICROWAVE Eng.

SNR Performance Comparison of Analog Systems



PHOTONICS and MICROWAVE Eng.