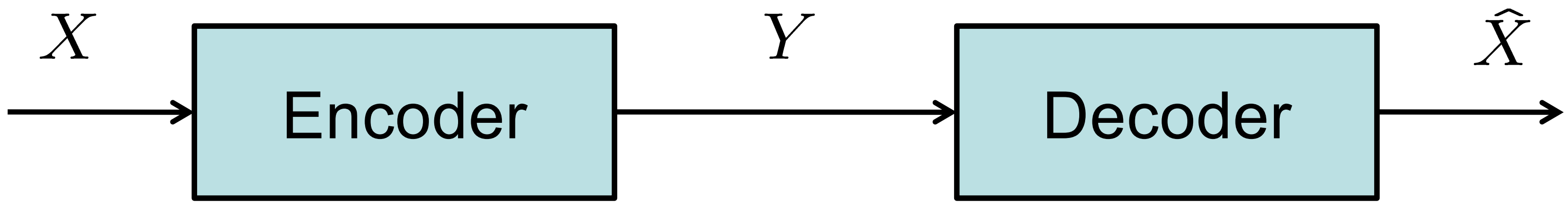
Rate Distortion Theory

- Distortion measure
- Rate distortion function
- Shannon’s noisy source coding theorem
- Convexity of rate distortion function
- Shannon lower bound
- Rate distortion function for Gaussian source

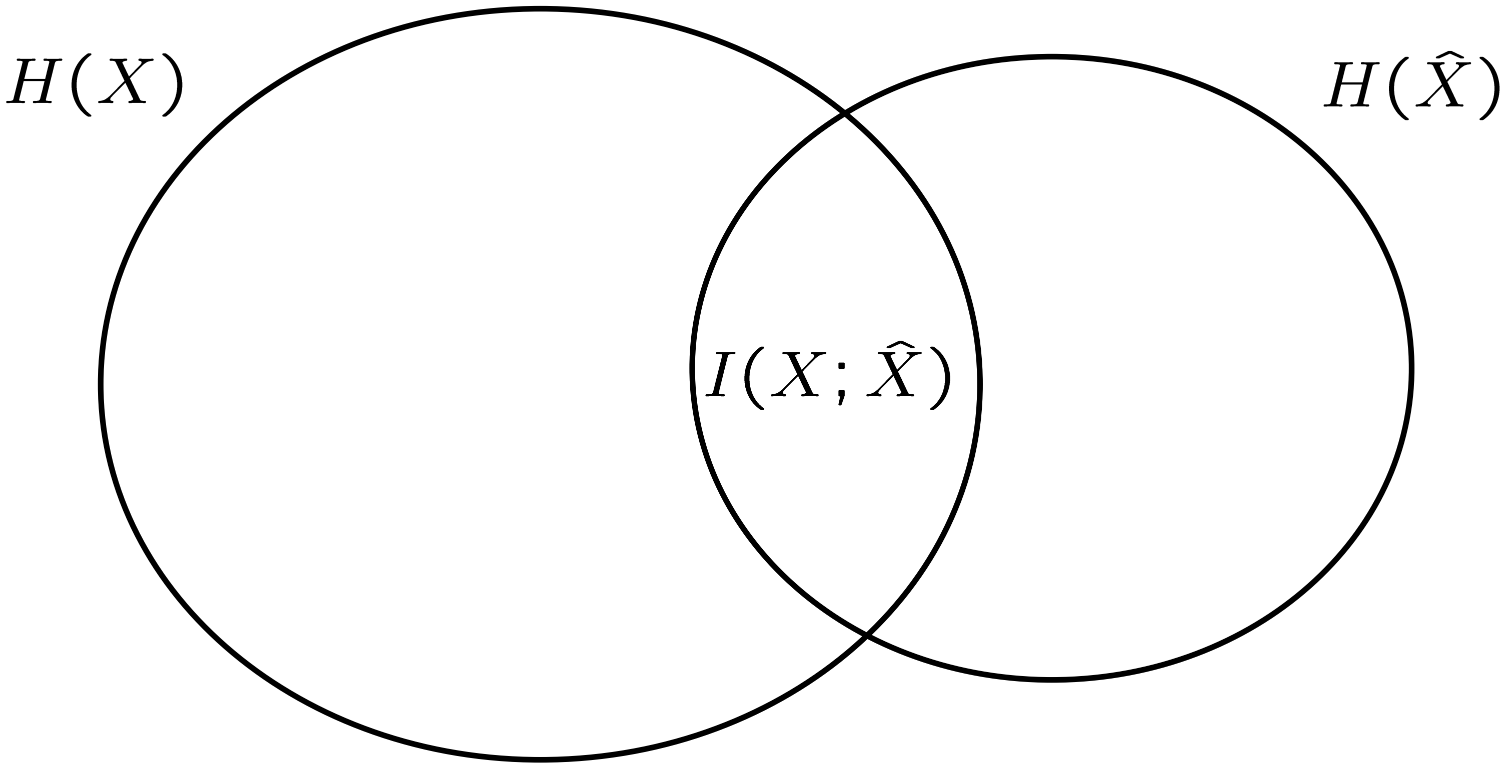
Encoding and Decoding

- Encoding and decoding as data processing



- Lossy coding: $X \approx \hat{X}$
- Data processing inequality: $I(X;\hat{X}) \leq I(X;Y)$

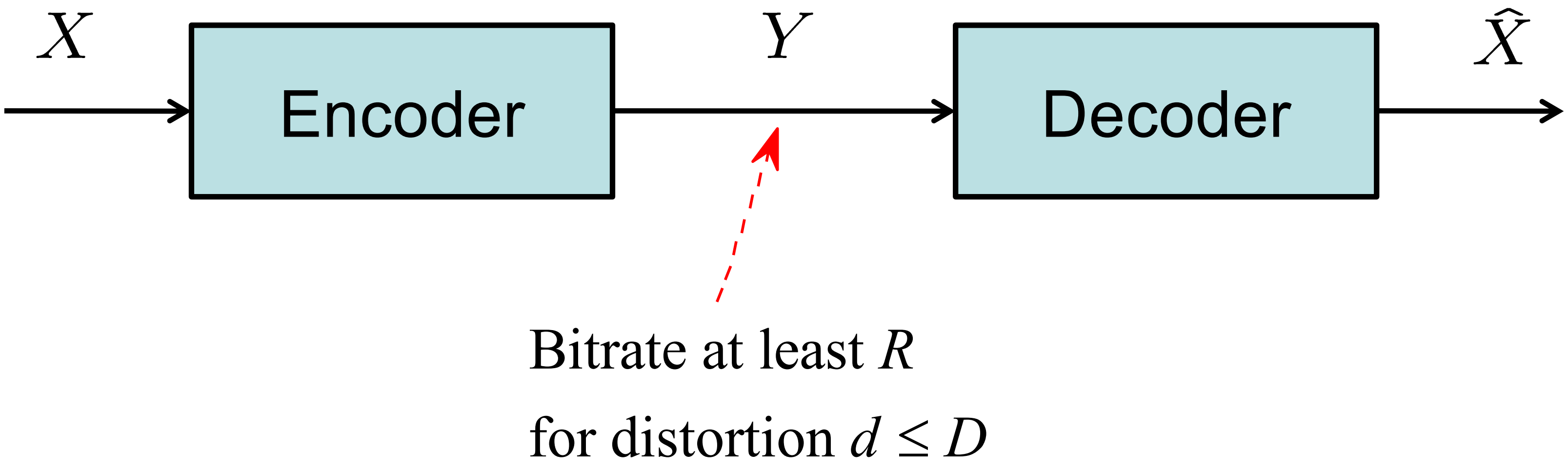
Encoding and Decoding



$$I(X;\hat{X}) \leq I(X;Y) \leq H(Y)$$

Rate Distortion Theory

- Rate distortion theory calculates the minimum transmission bitrate R for a required reconstruction quality



- Results of rate distortion theory are obtained without consideration of a specific coding method

Distortion Measure

- Symbol (signal, image . . .) x sent, $\hat{x}$ received
- Single-letter distortion measure:

$$\begin{aligned} d(x,\widehat{x}) &\geq 0 \\ d(x,\widehat{x}) &= 0 \quad \text{for} \quad x=\widehat{x} \end{aligned}$$

- Average distortion:

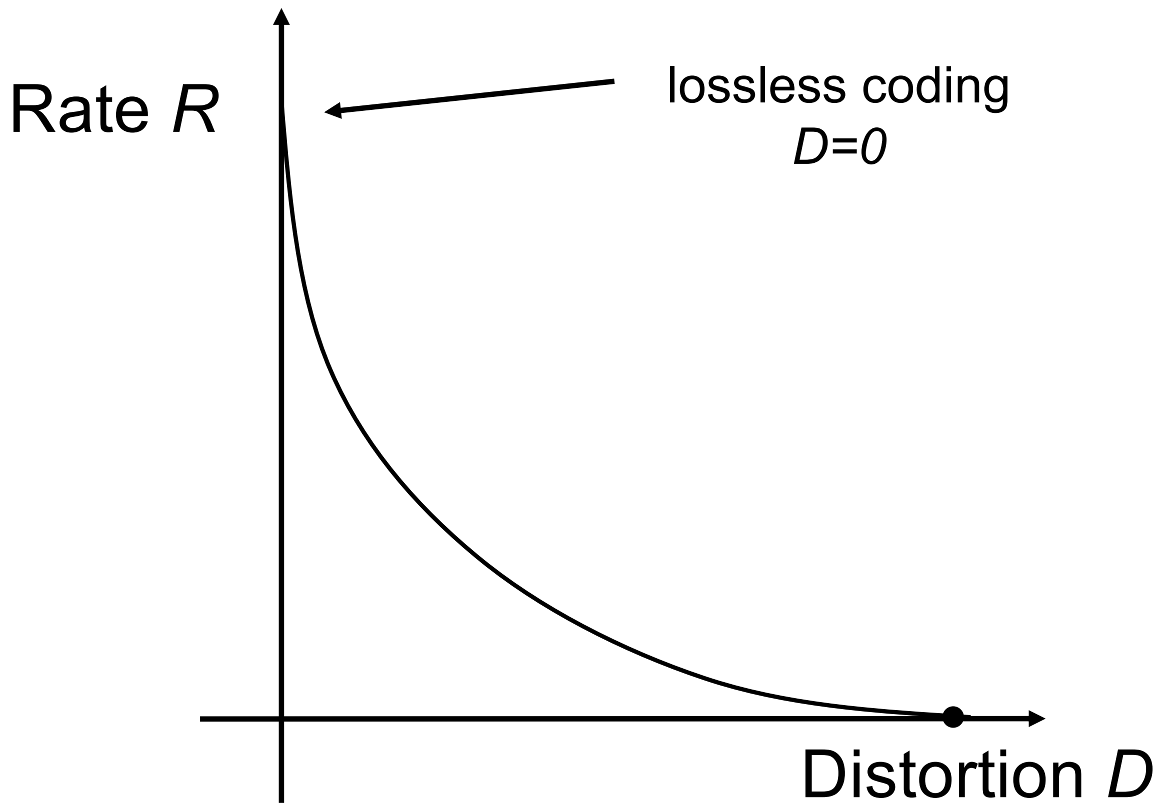
$$E\{d(X,\widehat{X})\} = \sum_{x}\sum_{\widehat{x}} f_{X,\widehat{X}}(x,\widehat{x})d(x,\widehat{x})$$

- Distortion criterion: $E\{d(X,\widehat{X})\} \leq D$

Maximum permissible average distortion

Rate Distortion Function

- Lower the bit-rate *R* by allowing some acceptable distortion *D* of the signal



Rate Distortion Function

- Definition:

$$R(D) = \inf_{f_{\hat{X}|X}} I(X;\hat{X}) \quad \text{s.t.} \quad E\{d(X,\hat{X})\} \leq D$$

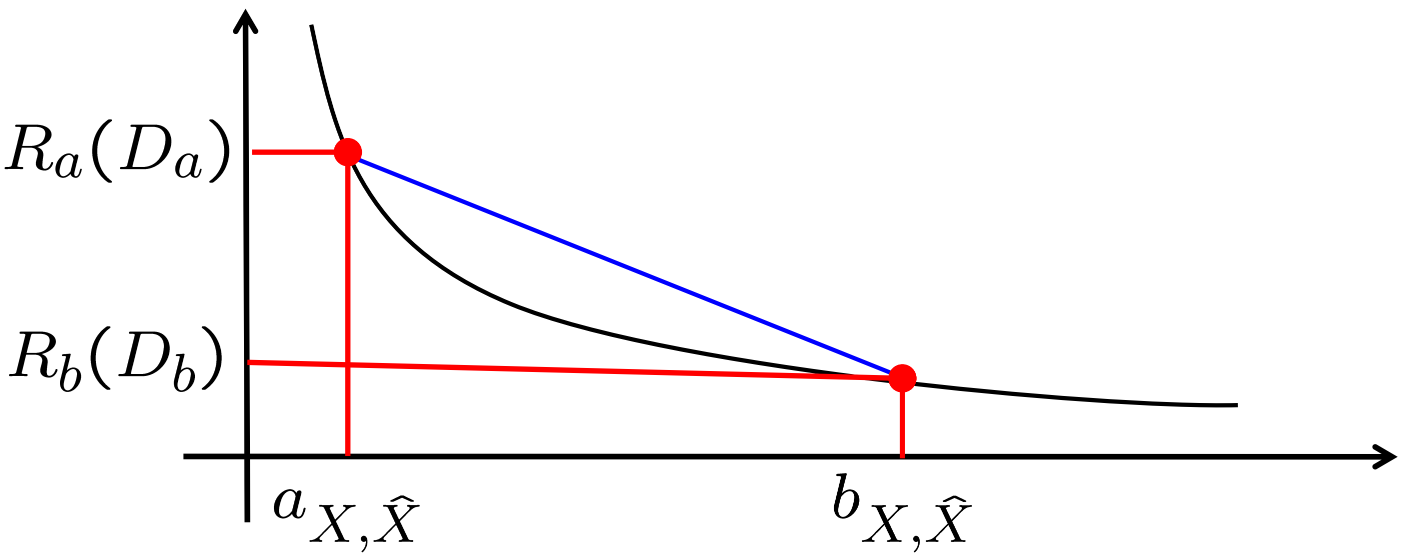
- **Shannon's Noisy Source Coding Theorem:**

  For a given maximum average distortion *D*, the rate distortion function *R(D)* is the (achievable) lower bound for the transmission bit-rate.

- *R(D)* is continuous, monotonically decreasing for *R>0* and convex
- Equivalently use distortion-rate function *D(R)*

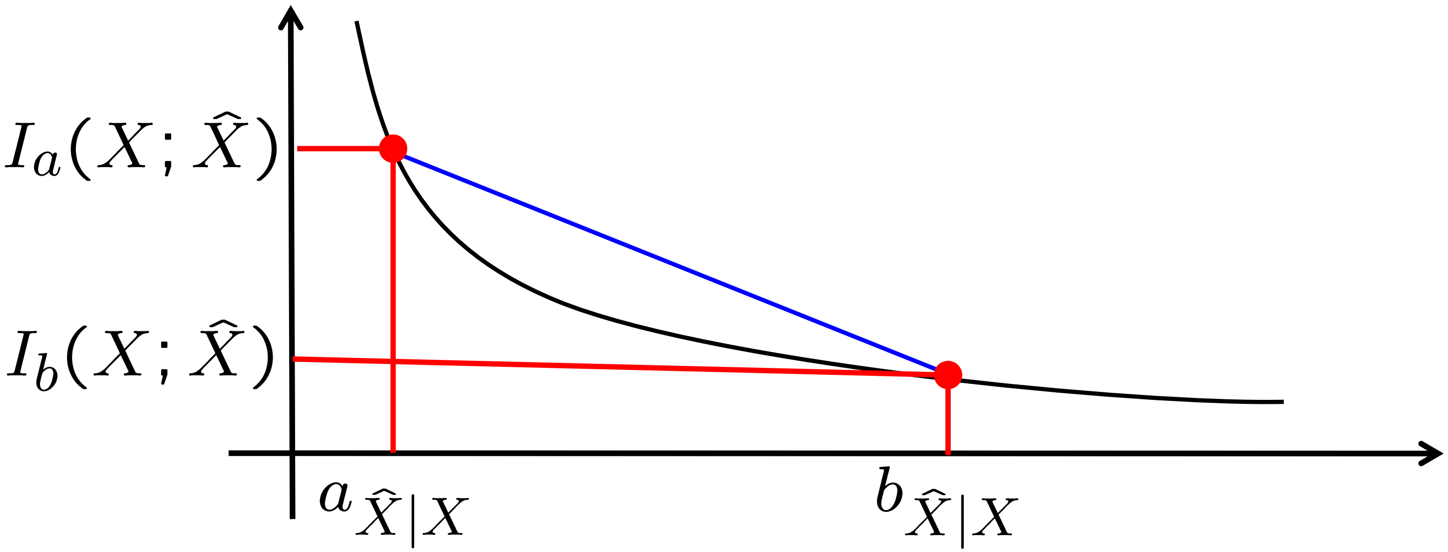
Convexity of the Rate Distortion Function

- Consider two rate distortion pairs (R_a, D_a) and (R_b, D_b) which lie on the rate distortion curve.
- Let the joint distributions $a_{X,\hat{X}}$ and $b_{X,\hat{X}}$ achieve these pairs.
- Consider the distribution: $f_{X,\hat{X}} = \lambda a_{X,\hat{X}} + [1-\lambda] b_{X,\hat{X}}$



Remark: Convexity of Mutual Information

- Mutual information is a convex function of the conditional distribution
- Consider the distribution: $f_{\hat{X}|X} = \lambda a_{\hat{X}|X} + [1-\lambda] b_{\hat{X}|X}$



Remark: Convexity of Mutual Information

- For the conditional distribution $f_{\hat{X}|X}$ we have

$$I_f(X;\hat{X}) = \sum f_X f_{\hat{X}} \frac{f_{\hat{X}|X}}{f_{\hat{X}}} \log_2 \frac{f_{\hat{X}|X}}{f_{\hat{X}}}$$

$$= \sum f_X f_{\hat{X}} \left[\lambda \frac{a}{f_{\hat{X}}} + [1-\lambda]\frac{b}{f_{\hat{X}}}\right] \log_2 \left[\lambda \frac{a}{f_{\hat{X}}} + [1-\lambda]\frac{b}{f_{\hat{X}}}\right]$$

$$\leq \sum f_X f_{\hat{X}} \left[\lambda \frac{a}{f_{\hat{X}}} \log_2 \frac{a}{f_{\hat{X}}} + [1-\lambda]\frac{b}{f_{\hat{X}}} \log_2 \frac{b}{f_{\hat{X}}}\right]$$

$$= \lambda I_a(X;\hat{X}) + [1-\lambda] I_b(X;\hat{X})$$

- Note that $g(t) = t \log_2 t$ is convex, i.e.,

$$g(\lambda t_1 + [1-\lambda]t_2) \leq \lambda g(t_1) + [1-\lambda] g(t_2)$$

Convexity of the Rate Distortion Function

- For the joint distribution $f_{X,\hat{X}}$ we have

$$R_f(D_f) = \inf_{f_{\hat{X}|X}} I_f(X;\hat{X}) \quad \text{s.t.} \quad E_f\{d(X,\hat{X})\} \leq D$$

- Distribution of the source is given: f_X
- Mutual information is a convex function of $f_{\hat{X}|X}$

$$\begin{aligned} R_f(D_f) &= \inf I_f(X;\hat{X}) \\ &\leq \inf \left\{\lambda I_a(X;\hat{X}) + [1-\lambda] I_b(X;\hat{X})\right\} \\ &= \lambda \inf I_a(X;\hat{X}) + [1-\lambda] \inf I_b(X;\hat{X}) \\ &= \lambda R_a(D_a) + [1-\lambda] R_b(D_b) \end{aligned}$$

Characterization of the Rate Distortion Function

- Solve unconstrained problem $\inf_{f_{\hat{X}|X}} J(X;\hat{X})$ with $\lambda > 0$ and

$$
\begin{aligned}
J(X;\hat{X}) &= E\left\{\ln \frac{f_{\hat{X}|X}(\hat{X}|X)}{f_{\hat{X}}(\hat{X})} + \lambda[d(X,\hat{X}) - D]\right\} \\
&= E\left\{-\ln\left[\frac{f_{\hat{X}}(\hat{X})}{f_{\hat{X}|X}(\hat{X}|X)} e^{-\lambda[d(X,\hat{X})-D]}\right]\right\} \\
&\geq -\ln E\left\{\frac{f_{\hat{X}}(\hat{X})}{f_{\hat{X}|X}(\hat{X}|X)} e^{-\lambda[d(X,\hat{X})-D]}\right\} \\
&= -\ln E\{g(\hat{X}|X)\}
\end{aligned}
$$

Characterization of the Rate Distortion Function

- Jensen: Lower bound is tight iff $g(\hat{x}|x)$ is constant in $\hat{x}$
- We obtain the conditional PDF

$$f_{\hat{X}|X}(\hat{x}|x) = \frac{f_{\hat{X}}(\hat{x})e^{-\lambda d(x,\hat{x})}}{\mu(x)}$$

- Normalizing the conditional PDF

$$f_{\hat{X}|X}(\hat{x}|x) = \frac{f_{\hat{X}}(\hat{x})e^{-\lambda d(x,\hat{x})}}{\int f_{\hat{X}}(\xi)e^{-\lambda d(x,\xi)}d\xi}$$

- Basis for Blahut algorithm, which is an iterative method to find a numerical solution for the rate distortion function

Shannon Lower Bound

- It can be shown that $h(X-\hat{X} \mid \hat{X}) = h(X \mid \hat{X})$
- Thus

$$R(D) = \inf_{d \le D} \{h(X) - h(X \mid \hat{X})\}$$

$$= h(X) - \sup_{d \le D} \{h(X \mid \hat{X})\}$$

$$= h(X) - \sup_{d \le D} \{h(X - \hat{X} \mid \hat{X})\}$$

- Ideally, the source coder would introduce errors $x - \hat{x}$ that are statistically independent from the reconstructed signal $\hat{x}$ (not always possible!).
- Shannon lower bound:

$$\boxed{R(D) \ge h(X) - \sup_{d \le D} h(X - \hat{X})}$$

Shannon Lower Bound

- Mean squared error distortion measure: Gaussian PDF possesses largest entropy for given variance

$$\boxed{\begin{aligned} R(D) &\geq h(X) - \sup_{d \leq D} h(X - \hat{X}) \\ &= h\left(X\right) - \tfrac{1}{2}\log_2 2\pi e D \end{aligned}}$$

- Distortion reduction by 6 dB requires 1 bit/sample

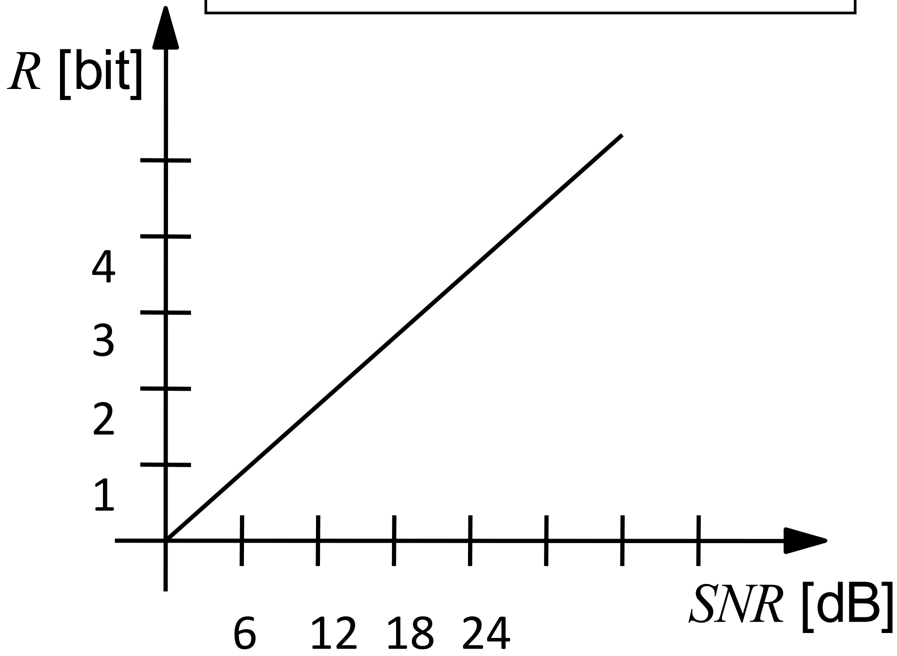
$R(D)$ Function for a Memoryless Gaussian Source and MSE Distortion

- Gaussian source, variance σ^2
- Mean squared error

$$E\{(X-\hat{X})^2\} \le D$$

- Rule of thumb: 6 dB $\cong$ 1 bit
- *R(D)* for non-Gaussian sources with the same variance σ^2 is always <u>below</u> this Gaussian *R(D)* curve.

$$\boxed{R(D) = \frac{1}{2}\log_2\left(\frac{\sigma^2}{D}\right)}$$



$$SNR = 10\log_{10}\left(\frac{\sigma^2}{D}\right) \text{ [dB]}$$

R(D) for Gaussian Source with Memory

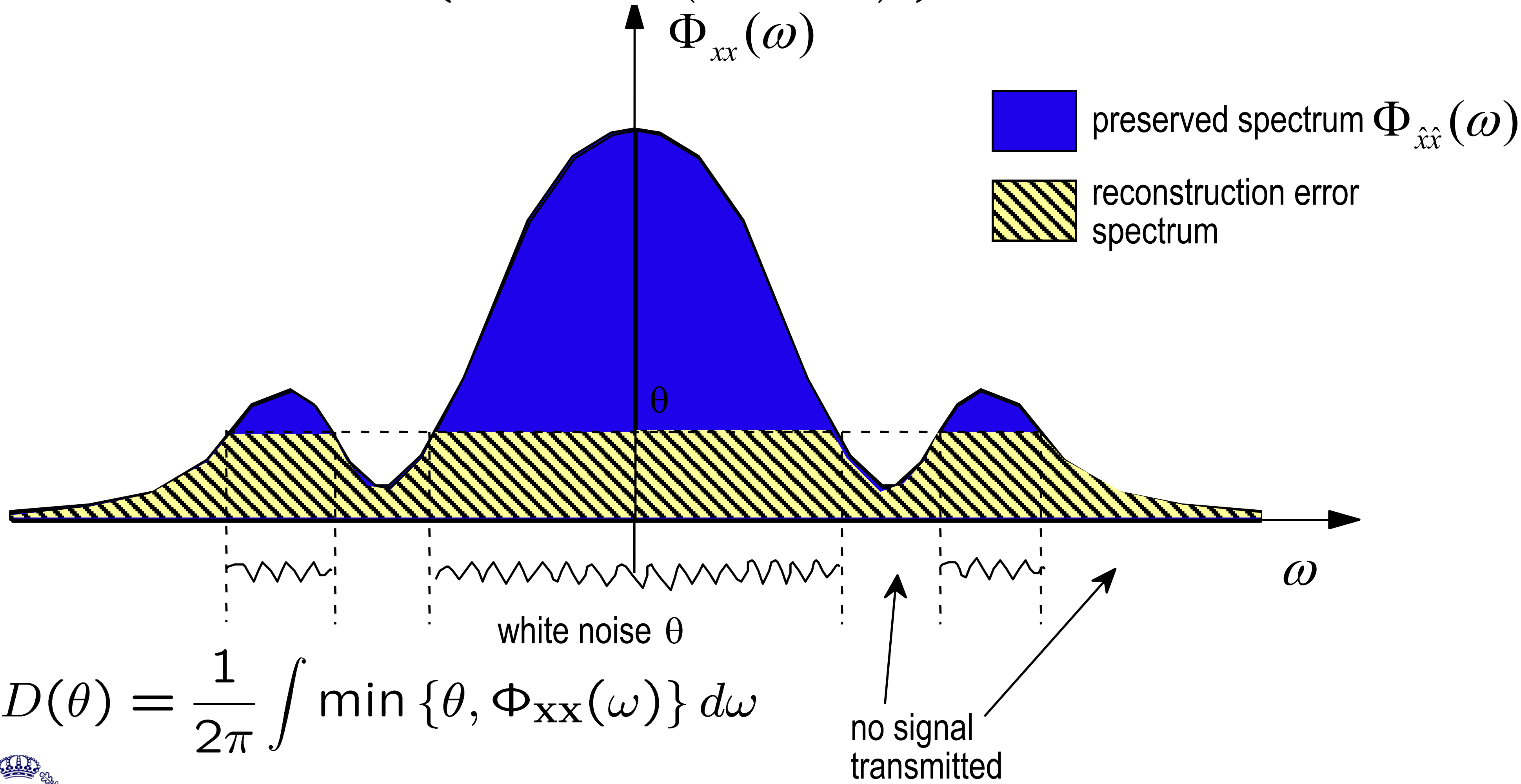
- Bandlimited, jointly Gaussian source with power spectral density $\Phi_{xx}(\omega)$
- Mean squared error distortion $E\{(X - \hat{X})^2\} \leq D$
- Parametric formulation of the $R(D)$ function

$$\boxed{\begin{aligned} D(\theta) &= \frac{1}{2\pi}\int_{\omega} \min\left\{\theta, \Phi_{xx}(\omega)\right\} d\omega \\ R(\theta) &= \frac{1}{2\pi}\int_{\omega} \max\left\{0, \frac{1}{2}\log\frac{\Phi_{xx}(\omega)}{\theta}\right\} d\omega \end{aligned}}$$

- $R(D)$ for non-Gaussian sources with the same power spectral density is always lower.

R(D) for Gaussian Source with Memory

$$R(\theta) = \frac{1}{2\pi} \int \max\left\{0, \frac{1}{2} \log_2 \left(\frac{\Phi_{\mathbf{xx}}(\omega)}{\theta} \right) \right\} d\omega$$



$$D(\theta) = \frac{1}{2\pi} \int \min\left\{\theta, \Phi_{\mathbf{xx}}(\omega)\right\} d\omega$$