Fundamentals of Information Theory

Rate Distortion Theory

Yayu Gao
School of Electronic Information and Communications
Huazhong University of Science and Technology
Email: yayugao@hust.edu.cn

Outline

- Do all information sources need error-free coding?
- System model for rate-distortion coding
- How to evaluate distortion?——distortion function
- Optimization problem for rate-distortion coding
- Rate distortion function
- Shannon’s third theorem: Rate-distortion source coding theorem
- Distortion rate function
- Practical insights

本节学习目标

1. 理解有损压缩的动机与意义
2. 说出率失真信源编码的建模过程
3. 说出失真函数的定义
4. 说出平均失真的定义
5. 说出率失真函数的定义及意义
6. 写出香农第三定理及其意义
7. 理解率失真理论与信道容量的联系

重难点：

- 率失真信源编码的建模
- 率失真函数的定义
- 香农第三定理

01

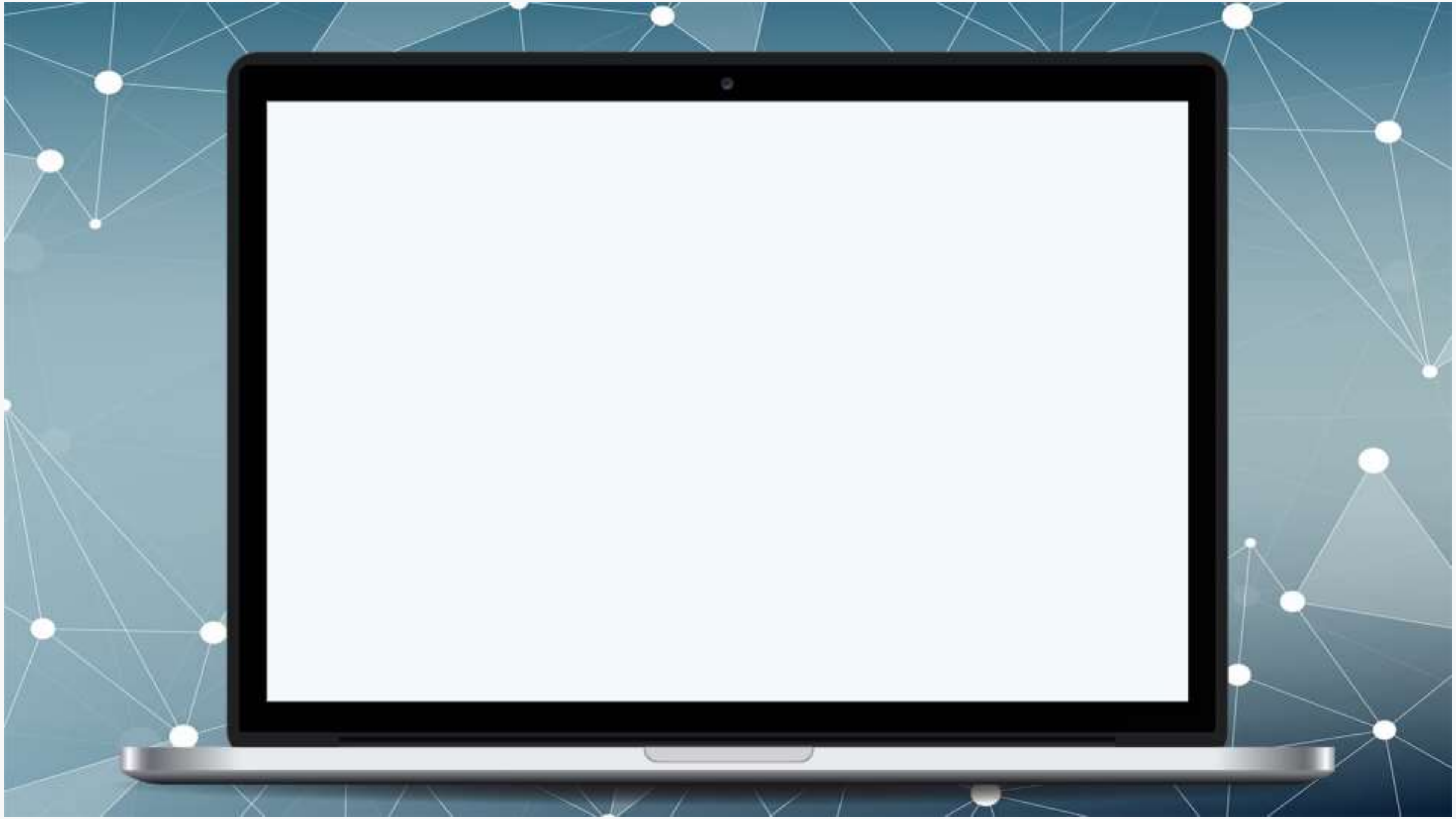
Do all information sources need error-free coding?

Revisiting

- Source coding
  - Eliminate redundancy to compress data and improve efficiency.
  - Represent the source efficiently and **without error**.
- Channel coding
  - Increase redundancy to combat transmission errors.
  - Transmit information reliably over channels **without error**
- **Preserve entropy**
  - To guarantee reliable and error-free transmission.

Do all information sources need error-free coding?

Do all information sources need error-free coding?



Error-free coding are not always needed.



We do not need to reconstruct all the information in continuous sources

Lossy source coding

Fundamental question of rate-distortion theory

Tradeoff between Rate and Distortion

Compression rate

VS

Compression quality

We are all imperfect. But how well can we do?

- Given a requirement on distortion, how small can the source be compressed?
  - What is the minimum description rate?

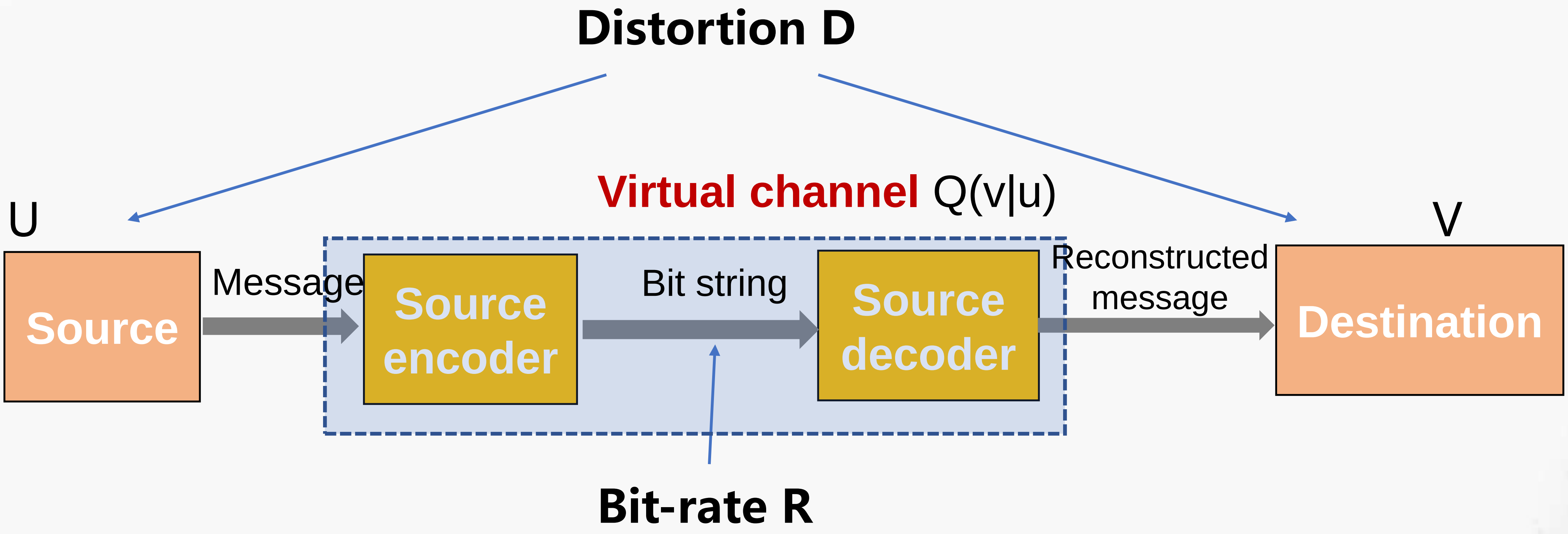
We are all imperfect. But how well can we do?

- Given a specific transmission bit-rate, how high-definition video I can watch?
  - What is the minimum distortion?

02

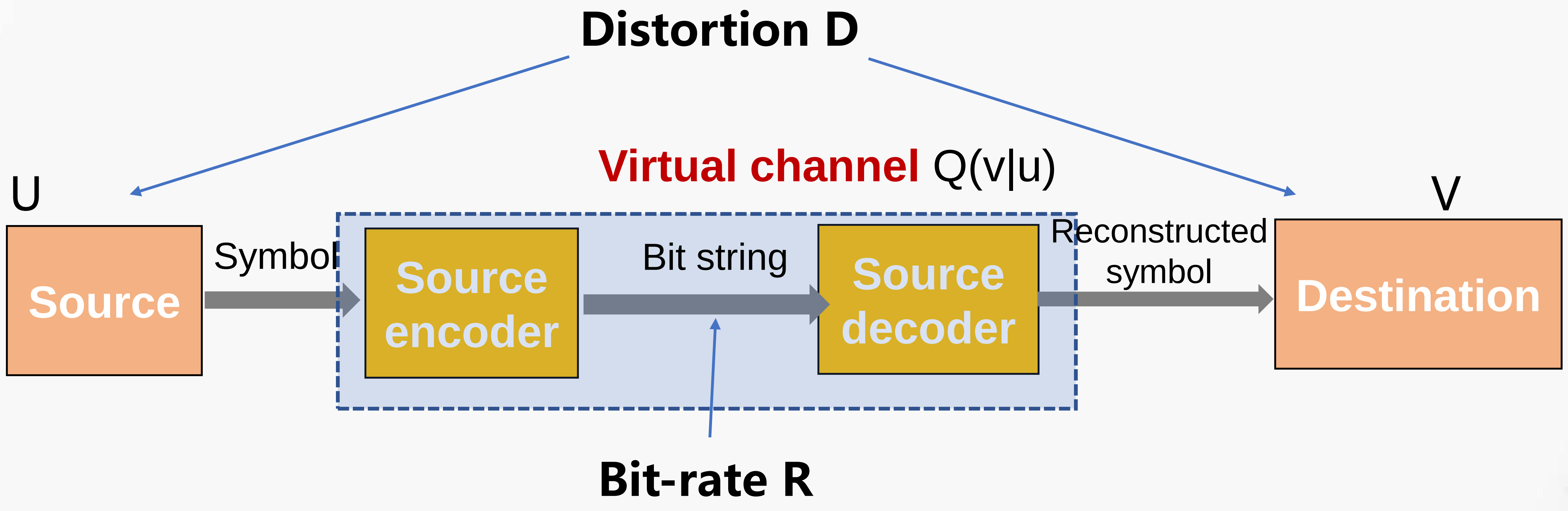
System model for rate-distortion coding

Rate-distortion source coding: system model



- Objective: Establish functional relationship between source U, destination V, distortion D and information rate R.
- Idea: **Consider the process of rate distortion encoding and decoding as a virtual channel**.

Rate-distortion source coding: system model



conditional probability distribution

Source symbols

$$U = (u_0, u_1, \ldots, u_{M-1})$$

$$\boxed{Q = \{Q(v|u), u \in \mathcal{U}, v \in \mathcal{V})\}}$$

Reconstruction symbols

$$V = (v_0, v_1, \ldots, v_{N-1})$$

Source symbols and reconstruction symbols

- Source symbols are given by the random sequence $\{U_k\}$. Each U_k assumes values in the discrete set $U = (u_0, u_1, \ldots, u_{M-1})$.
  For simplicity, let us assume U_k to be independent and identically distributed (i.i.d.) with the distribution $P(u)$, $u \in \mathcal{U}$.
  Example:
  - For a binary source: $U = (0, 1)$.
  - For a picture: $U = (0, 1, \ldots, 255)$.
- Reconstruction symbols are given by the random sequence $\{V_k\}$ with distribution $P(v)$, $v \in \mathcal{V}$.
  Each V_k assumes values in the discrete set $V = (v_0, v_1, \ldots, v_{N-1})$.
- The sets $\mathcal{U}$ and $\mathcal{V}$ are usually the same.

Coder/Decoder

- The statistical description of the coder/decoder defines the mapping from the source symbols to the reconstruction symbols, via

$$Q = \{Q(v|u), u \in \mathcal{U}, v \in \mathcal{V})\}.$$

- Q is the conditional probability distribution over the letters of the reconstruction alphabet $\mathcal{V}$ given a letter of the source alphabet $\mathcal{U}$.
- The transmission system is described via the joint p.d.f.: $P(u, v)$.

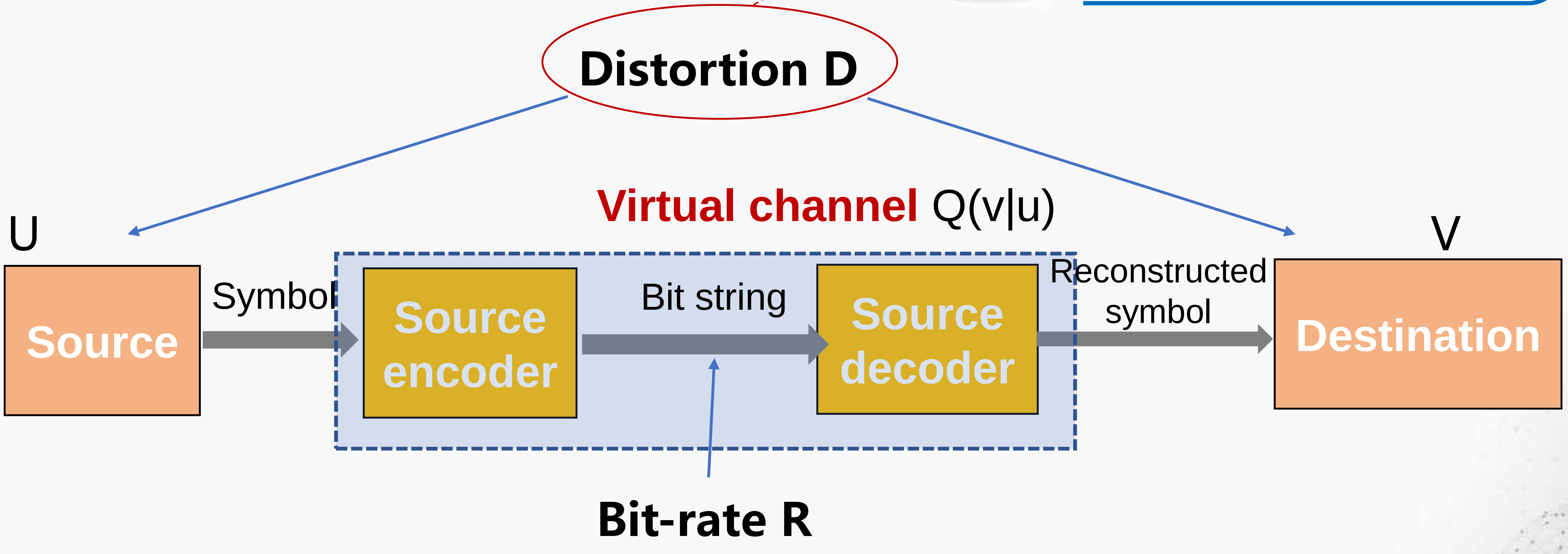
$$P(u, v) = P(u) \cdot Q(v|u) \text{ (Bayes' rule)}$$

03

How to characterize distortion?
——distortion function

A natural question...

How to characterize the distortion?



How to characterize distortion?——distortion function

- **Definition**: the distortion between the input symbol u by the output symbol v is measured by a **non-negative** cost function $d(u,v)$.

$$d(u,v) = \begin{cases} 0, & u = v \\ a, & u \neq v \end{cases}$$

- A mapping from the set of source-reconstruction alphabet pairs into the set of nonnegative real numbers.
  - For discrete alphabets, distortion function can be described with **distortion matrix**.

$$d : \mathcal{U} \times \mathcal{V} \rightarrow [0, \infty)$$

- **Physical meaning**: the cost of representing u by v

Some common distortion functions

- Hamming distortion

$$d(u,v) = \begin{cases} 0, & \text{for } u = v \\ 1, & \text{for } u \neq v \end{cases}$$

- Hamming distortion matrix

$$\mathcal{D} = \begin{bmatrix} 0 & 1 & 1 & \dots & 1 \\ 1 & 0 & 1 & \dots & 1 \\ \dots & & & & \dots \\ 1 & 1 & 1 & \dots & 0 \end{bmatrix}$$

- Squared-error distortion

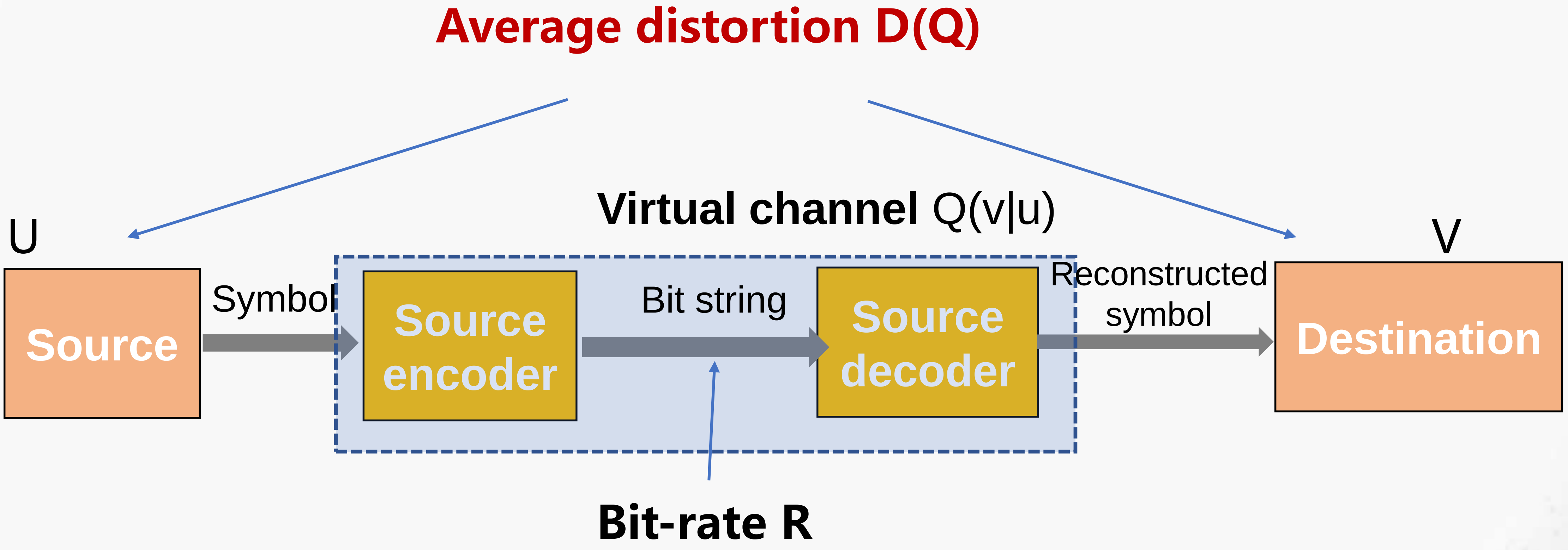
$$d(u,v) = |u - v|^2$$

- Given $U = \{0,1,2\},\ V = \{0,1,2\}$
- Distortion matrix

$$\mathcal{D} = \begin{bmatrix} 0 & 1 & 4 \\ 1 & 0 & 1 \\ 4 & 1 & 0 \end{bmatrix}$$

- Widely adopted for continuous alphabets.

How to measure the overall distortion?



- Average distortion: statistical average of the distortion function

$$D(Q) = E[d(u, v)]$$

Average distortion: consider the source distribution

- Average distortion: statistical average of the distortion function

$$D(Q) = E[d(u, v)]$$

$$D(Q) = \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} P(u, v) \cdot d(u, v)$$
$$= \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} P(u) \cdot Q(v|u) \cdot d(u, v)$$

- Information source: $P(u)$
- Coder/decoder: $Q(v|u)$
- Distortion function: $d(u, v)$

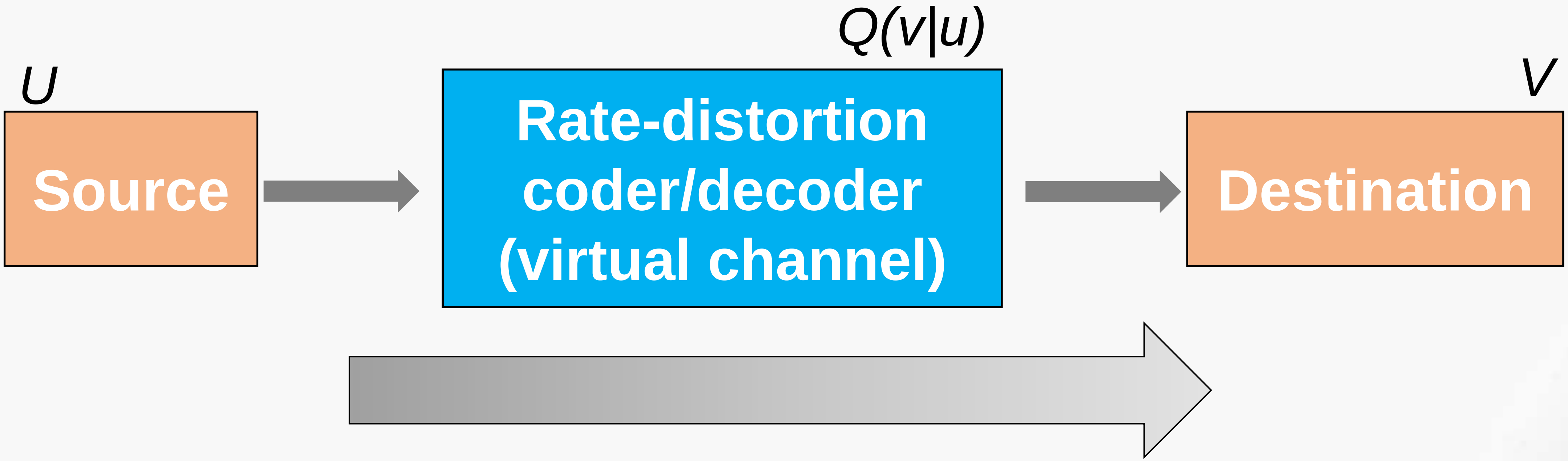
- Given source distribution and the transition probability distribution, the **average measure of distortion** over the channel.

04

Optimization problem for rate-distortion coding

Question

After compression, how much information of the source are kept in the reconstruction symbols?



Mutual information $I(U;V)$

Information rate

- The Shannon average mutual information is expressed via entropy.

$$I(U;V) = H(U) - H(U|V),$$

where

  - *H(U)*: Source entropy
  - *H(U|V)*: Equivocation (conditional entropy).

- **Equivocation**:
  - The conditional entropy (uncertainty) about the source U given the reconstruction V.
  - A measure for the amount of **missing** [quantized] information in the received signal V.

- I(U; V) denotes **the amount of average information of the source U that contains in the reconstruction one V**.

Optimization problem: objective function

Q: How well can we compress the source?

Less kept information, better compression.

Minimize mutual information

$$\min \ I(U;V)$$

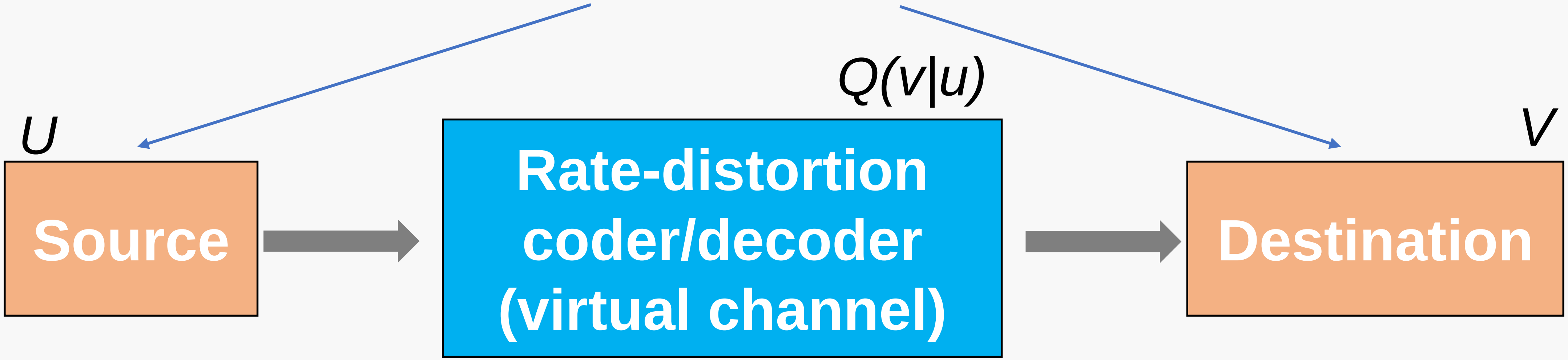
Let V be a constant, then

$$I(U;V) = 0$$

Any problem?

Optimization problem: constraint on average distortion

Average distortion D(Q) $D(Q) = E[d(u,v)]$



- Fidelity criteria: constraint on average distortion

$$D(Q) \leq D^*.$$

- D^* : maximum average distortion

05

Rate Distortion Function

Rate distortion function: definition

- Definition: For a source *U* and distortion function *d(u, v)*, given the maximum allowable distortion *D**, the **minimum information rate *R*(*D**)**

$$R(D^*) = \min_{\boxed{Q: D(Q) \leq D^*}} \{I(U;V)\}$$

$$\boxed{D(Q) = \sum_{u \in \mathcal{U}} \sum_{v \in \mathcal{V}} P(u) \cdot \boxed{Q(v|u)} \cdot d(u,v)}$$

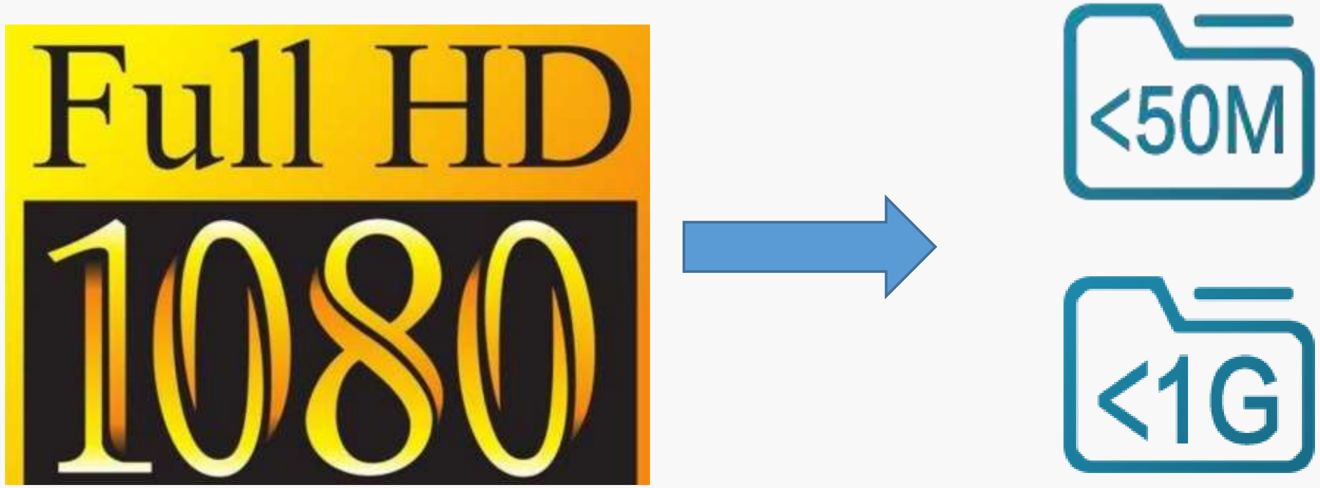
coder/decoder

- The minimization is conducted for **all possible mappings Q** that satisfy the average distortion constraint.

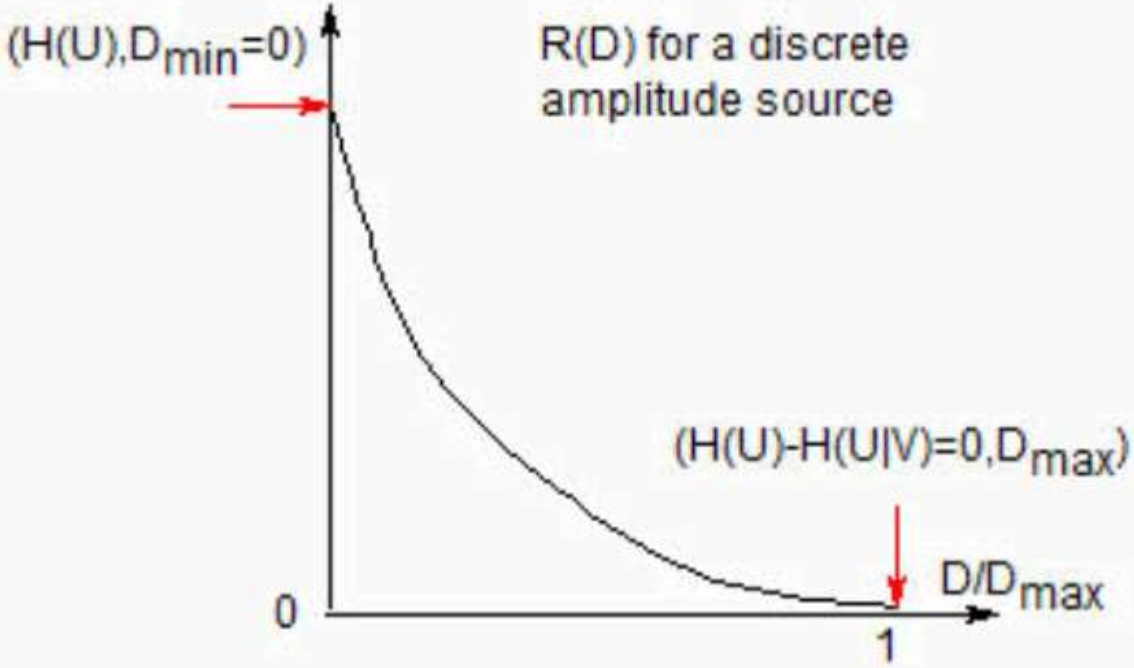
Rate distortion function: Physical meaning

$$R(D^*) = \min_{Q:D(Q)\leq D^*} \{I(U;V)\}$$

- Data compression limit for lossy source coding
- Given a requirement on distortion, how small can the source be compressed? What is the minimum description rate?

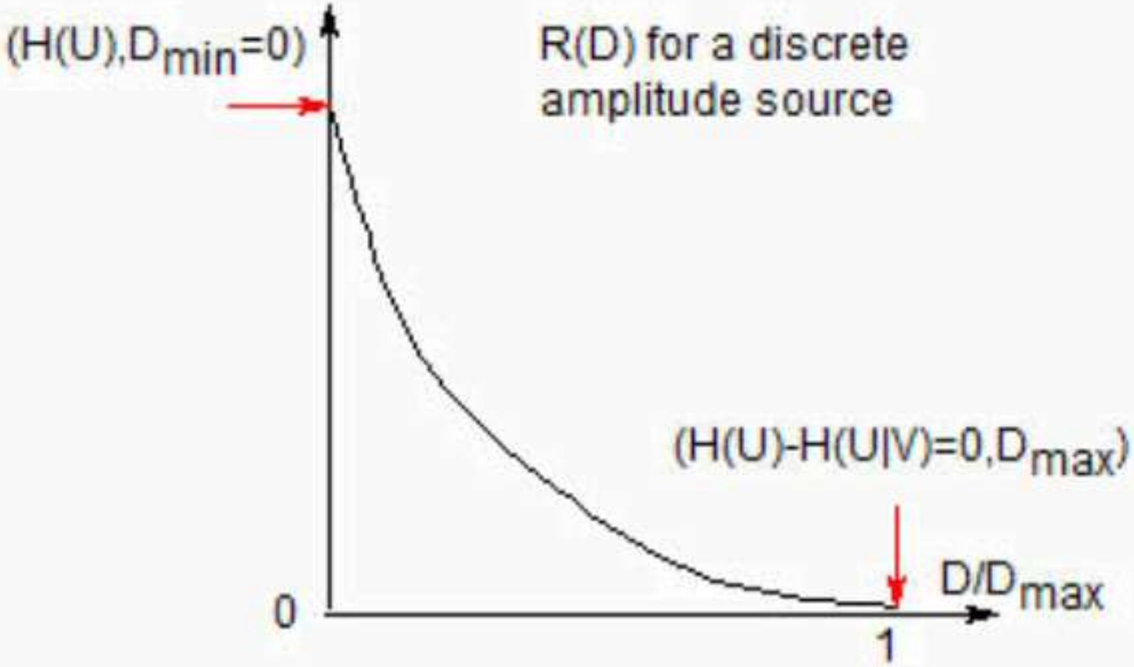


Rate distortion function: properties



- $R(D)$ is well defined for $D \in (D_{\min}, D_{\max})$.
- For discrete amplitude sources, $D_{\min} = 0$.
- $R(D) = H(U)$, if $D = D_{\min} = 0$. (not always true.)
- $R(D) = 0$, if $D \geq D_{\max}$.
- $H(U) > R(D) > 0$, if $0 < D < D_{\max}$.

Rate distortion function: properties



- $R(D)$ is always non-negative.

$$0 \leq I(U;V) \leq H(U)$$

- $R(D)$ is decreasing in the range $(D_{\min}, D_{\max})$.
- $R(D)$ is strictly convex upward in the range $(D_{\min}, D_{\max})$.
- The slope of $R(D)$ is continuous in the range $(D_{\min}, D_{\max})$.

Rate distortion function: discrete source

$$\boxed{R(D^*) = \min_{Q:D(Q)\leq D*} \{I(U;V)\}}$$

- For discrete sources, calculating $R(D^*)$ is to find the local minimum mutual information problem under some constraint conditions.
- Given $p(u)$ and $d(u,v)$, find the minimum $I(U;V)$ under the constraint condition $D(Q) \leq D$. The typical solution applies the Lagrange multiplier method.

Rate distortion function: continuous source

$$\boxed{R(D^*) = \min_{Q:D(Q)\leq D^*} \{I(U;V)\}}$$

$R(D^*)$ for memoryless Gaussian sources.

- Gaussian source, variance σ^2.
- Mean squared error (MSE) $D = E\{(u-v)^2\}$

$$R(D^*) = \begin{cases} \frac{1}{2}\log\frac{\sigma^2}{D^*}, & 0 \leq D^* \leq \sigma^2 \\ 0, & D^* > \sigma^2 \end{cases}$$

$$SNR = 10 \cdot \log_{10}\frac{\sigma^2}{D}$$

- Rule of thumb: 6dB $\sim$ 1 bit

- The $R(D^*)$ for non-Gaussian sources with the same variance σ^2 is always below the Gaussian $R(D^*)$ curve.

Rate distortion function: Memoryless Gaussian source

$$R(D) = \min_{f(V|U): E(U-V)^2 \le D} I(U;V).$$

$$
\begin{aligned}
I(U;V) &= h(U) - h(U|V) && (h(X+a) = h(X)) \\
&= h(U) - h(U-V|V) && (r.v. U \text{ is Gaussian.}) \\
&\ge \frac{1}{2}\log[(2\pi e)\sigma^2] - h(U-V) && (\text{Conditioning reduce entropy.}) \\
&\ge \frac{1}{2}\log[(2\pi e)\sigma^2] - h\left(\mathcal{N}\left(0, E(U-V)^2\right)\right) && (\text{Gaussian maximum entropy.}) \\
&= \frac{1}{2}\log[(2\pi e)\sigma^2] - \frac{1}{2}\log[(2\pi e)E(U-V)^2] \\
&\ge \frac{1}{2}\log[(2\pi e)\sigma^2] - \frac{1}{2}\log[(2\pi e)D^*] && (\text{Distortion fidelity criteria.}) \\
&= \frac{1}{2}\log\frac{\sigma^2}{D^*}
\end{aligned}
$$

Rate distortion function: Example

Let $d(x, \hat{x})$ be a distortion function. We have a source $X \sim p(x)$. Let $R(D)$ be the associated rate distortion function.

(a) Find $\widetilde{R}(D)$ in terms of $R(D)$, where $\widetilde{R}(D)$ is the rate distortion function associated with the distortion $\widetilde{d}(x, \hat{x}) = d(x, \hat{x}) + a$ for some constant $a > 0$. (They are not equal)

(b) Now suppose that $d(x, \hat{x}) \geq 0$ for all $x, \hat{x}$ and define a new distortion function $d^*(x, \hat{x}) = bd(x, \hat{x})$, where b is some number ≥ 0. Find the associated rate distortion function $R^*(D)$ in terms of $R(D)$.

(c) Let $X \sim N(0, \sigma^2)$ and $d(x, \hat{x}) = 5(x - \hat{x})^2 + 3$. What is $R(D)$?

Rate distortion function: Example

(a)

$$\widetilde{R}(D) = \inf_{p(\hat{x}|x):E\left(\widetilde{d}(x,\hat{x})\right)\leq D} I(X;\hat{X})$$

$$= \inf_{p(\hat{x}|x):E(d(x,\hat{x}))+a\leq D} I(X;\hat{X})$$

$$= \inf_{p(\hat{x}|x):E(d(x,\hat{x}))\leq D-a} I(X;\hat{X})$$

$$= R(D-a)$$

Rate distortion function: Example

(b) If $b > 0$,

$$
\begin{aligned}
R^*(D) &= \inf_{p(\hat{x}|x):E(d^*(x,\hat{x}))\le D} I(X;\hat{X}) \\
&= \inf_{p(\hat{x}|x):E(bd(x,\hat{x}))\le D} I(X;\hat{X}) \\
&= \inf_{p(\hat{x}|x):E(d(x,\hat{x}))\le \frac{D}{b}} I(X;\hat{X}) \\
&= R\left(\frac{D}{b}\right).
\end{aligned}
$$

If $b = 0$, then $d^* = 0$ and $R^*(D) = 0$.

Rate distortion function: Example

(c) Let $R_{se}(D)$ be the rate distortion function associate with the distortion $d_{se}(x, \hat{x}) = (x - \hat{x})^2$: Then from parts (a) and (b) we have

$$R(D) = R_{se}\left(\frac{D-3}{5}\right).$$

We know that

$$R_{se}(D) = \begin{cases} \frac{1}{2}\log\frac{\sigma^2}{D}, & 0 \le D \le \sigma^2 \\ 0, & D > \sigma^2 \end{cases}.$$

Therefore, we have

$$R(D) = \begin{cases} \frac{1}{2}\log\frac{5\sigma^2}{D-3}, & 3 \le D \le 5\sigma^2 + 3 \\ 0, & D > 5\sigma^2 + 3 \end{cases}.$$

06

Shannon’s third theorem:

Rate-distortion source coding theorem

Rate-distortion source coding theorem

$R \geq R(D) \Rightarrow$ There exists a coding method C, which satisfies $D(C) \leq D + \varepsilon$ for any given positive D and any minimum ε.

$R < R(D) \Rightarrow$ For any coding method C, $D(C) > D$.

- Known as Shannon's third theorem
- Limits of data compression
  - Zero-error source compression (1^{st} theorem): **H(S)**
  - Distortion source compression (3^{rd} theorem): **R(D)**
  - Given D, normally R(D)<H(S).

07

Distortion Rate Function

Motivation: Another perspective

- Given a requirement of distortion D from the source, what is the minimum transmission bit-rate R?
  - Rate distortion function R(D)

- Given a specific transmission bit-rate R, what is the minimum distortion D?
  - **Distortion rate function D(R)**

- The calculation of the rate distortion function R(D) and the distortion rate function D(R) are called as dual-problems.

Distortion rate function: definition

- **Definition**: For a source U and distortion function d(u, v), given the maximum average rate R*, the **minimum average distortion D(R*)**

$$D(R^*) = \min_{Q:I(U;V)\leq R^*} \{d(Q)\}$$

- We can set R* to the capacity C of the transmission channel and determine the minimum distortion for this ideal communication system.

Distortion rate function: physical meaning

- Given a specific transmission bit-rate, how high-definition video I can watch?
  - What is the minimum distortion?

Distortion rate function: continuous source

$$\boxed{D(R^*) = \min_{Q:I(U;V)\leq R^*} \{d(Q)\}}$$

$D(R^*)$ for memoryless Gaussian sources.

- Gaussian source, variance σ^2.
- Mean squared error (MSE) $D = E\left\{(u-v)^2\right\}$

$$R(D^*) = \begin{cases} \frac{1}{2}\log\frac{\sigma^2}{D^*}, & 0 \leq D^* \leq \sigma^2 \\ 0, & D^* > \sigma^2 \end{cases} \qquad D(R^*) = \sigma^2 \cdot 2^{-2R^*}, R \geq 0.$$

08

Practical Insights

Rate distortion function vs. Channel capacity

$$R(D^*) = \min_{Q:D(Q)\leq D^*} \{I(U;V)\}$$

$$C = \max_{p(x)} \{I(X;Y)\}$$

- **Rate distortion theory**
  - **minimize** mutual information
  - **Source** is given
  - Search all possible **channels (coder/decoder design)** for the optimal solution
  - **Efficiency** for compression
  - **Decrease redundancy**
  - **Source coding**

- **Channel capacity**
  - **maximize** mutual information
  - **Channel** is given
  - Search all possible **input distributions** for the optimal solution
  - **Reliability** for communication
  - **Increase redundancy**
  - **Channel coding**

Source coding vs. Channel coding

- **Source coding**
  - Core problem: **efficiency**
  - Efficiency: having an average code length that is as small as possible
  - Example: to use shorter code for the English letters which appear frequently, so as to reduce the average code length

- **Channel Coding**
  - Core problem: **reliability**
  - Reliability: to cope with the errors in the transmission
  - Example: to send the same sequence multiple times, so as to recover from the errors in channel

本节学习目标

1. 理解有损压缩的动机与意义
2. 说出率失真信源编码的建模过程
3. 说出失真函数的定义
4. 说出平均失真的定义
5. 说出率失真函数的定义及意义
6. 写出香农第三定理及其意义
7. 理解率失真理论与信道容量的联系

重难点：

- 率失真信源编码的建模
- 率失真函数的定义
- 香农第三定理

Thank you!

My Homepage

Yayu Gao

School of Electronic Information and Communications

Huazhong University of Science and Technology

Email: yayugao@hust.edu.cn