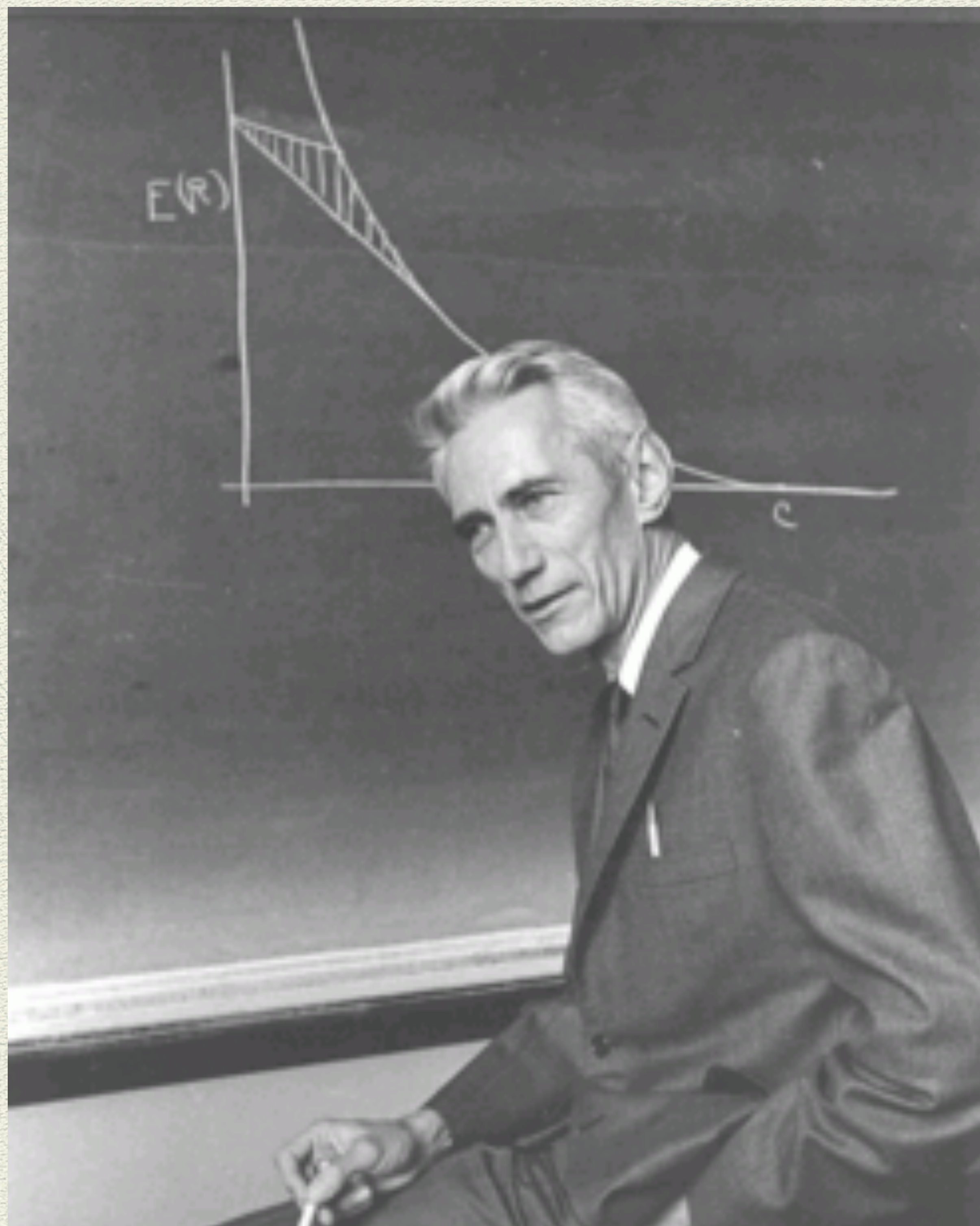
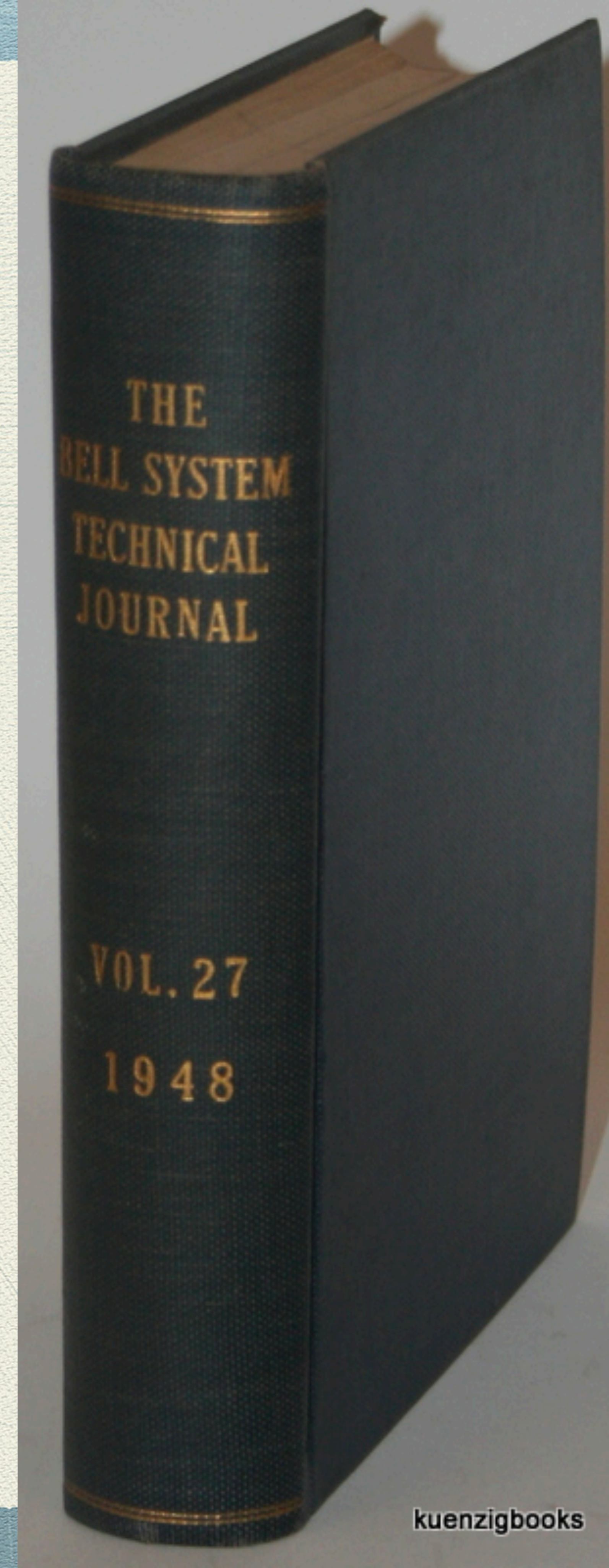
Introduction to Quantum Information-I

Vahid Karimipour,
Sharif University of Technology



Claude Shannon (1916-2001)



VOLUME XXVII JULY, 1948 NO. 3

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50¢ per copy *$1.50 per Year*

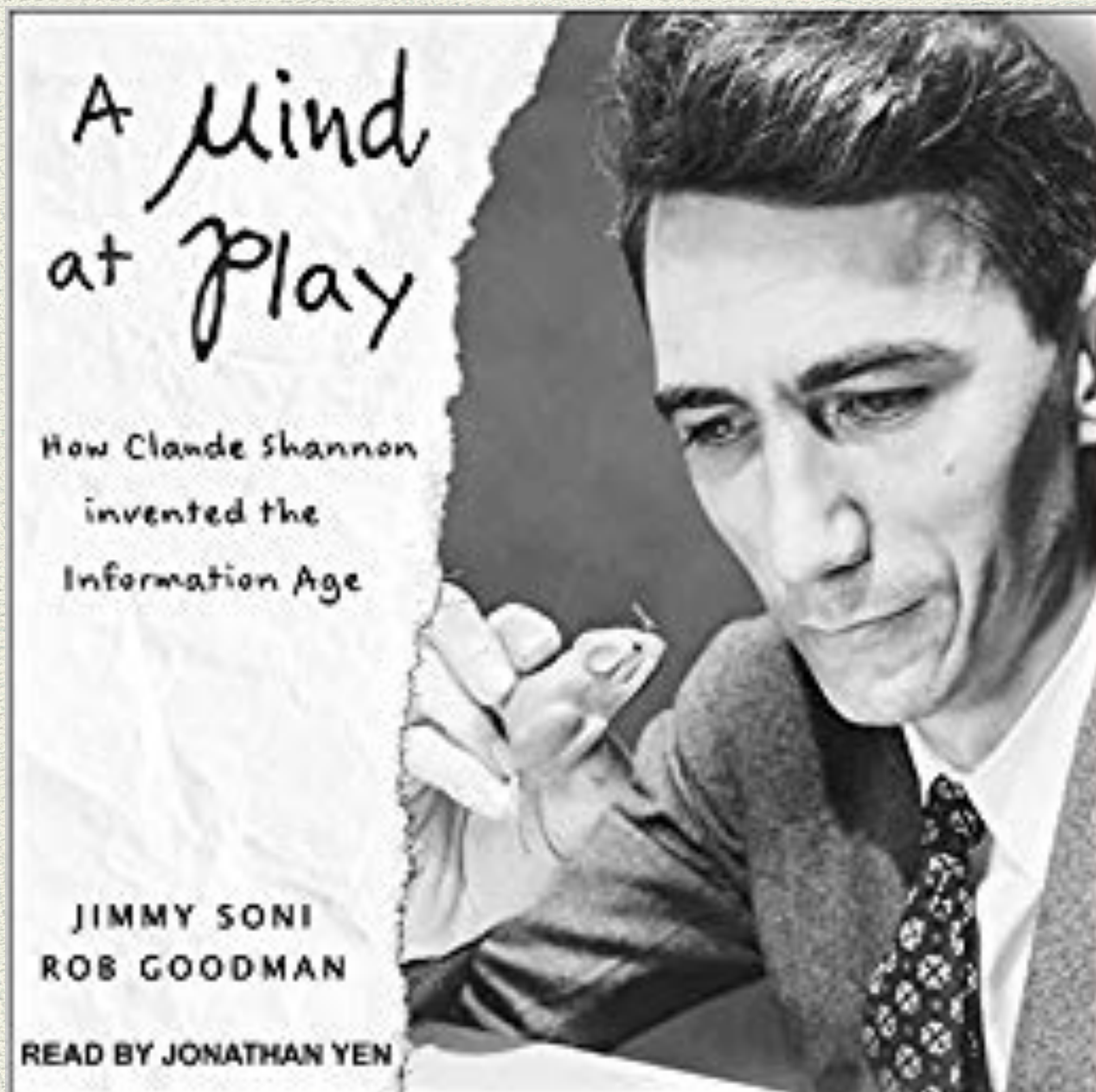
VOLUME XXVII OCTOBER, 1948 NO. 4

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50¢ per copy *$1.50 per Year*



Claude Shannon (1916-2001)

The Bell System Technical Journal

Vol. XXVII July, 1948 No. 3

A Mathematical Theory of Communication

By C. E. SHANNON

Introduction

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

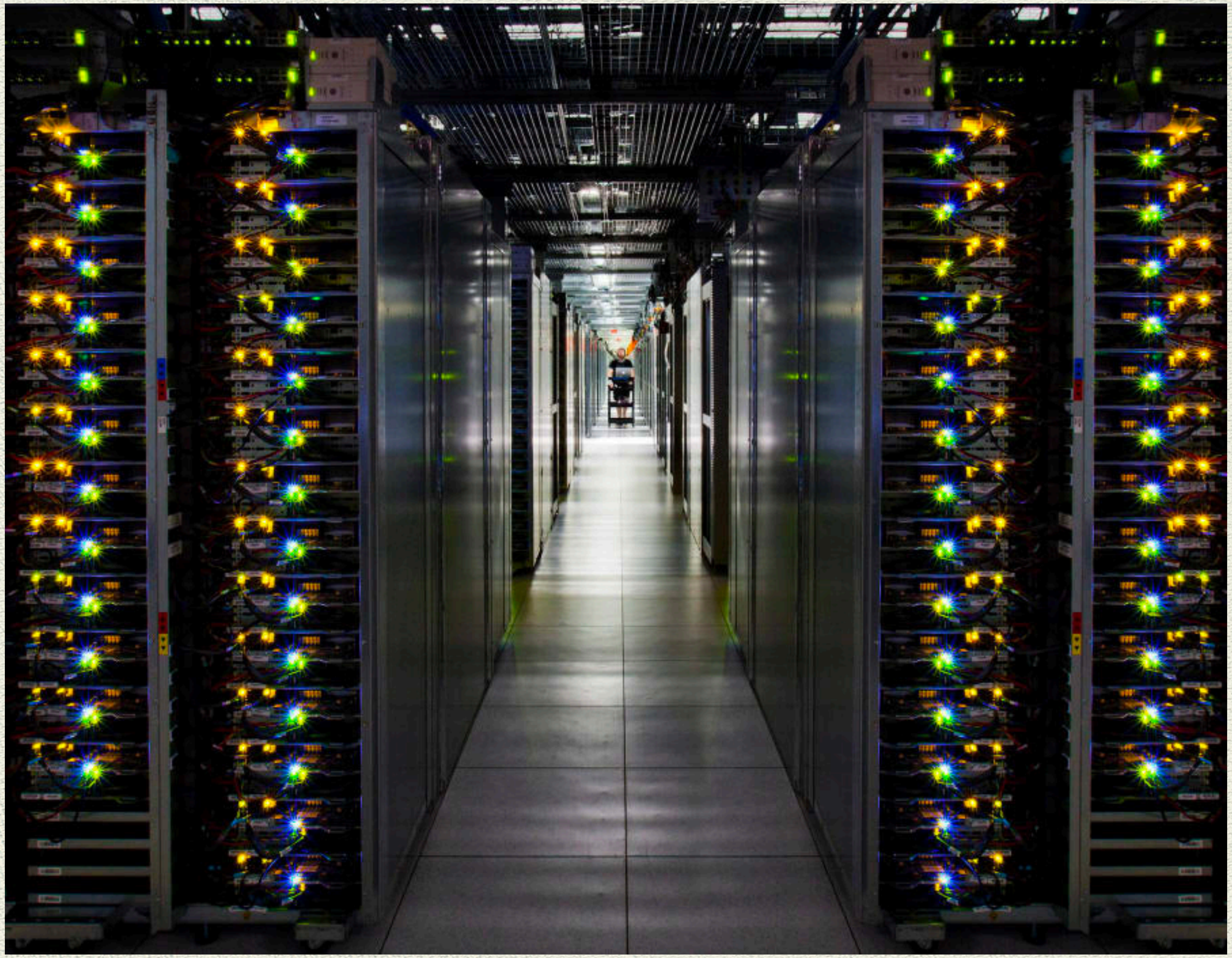
If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic function. Although this definition must be generalized considerably when we consider the influence of the statistics of the message and when we have a continuous range of messages, we will in all cases use an essentially logarithmic measure.

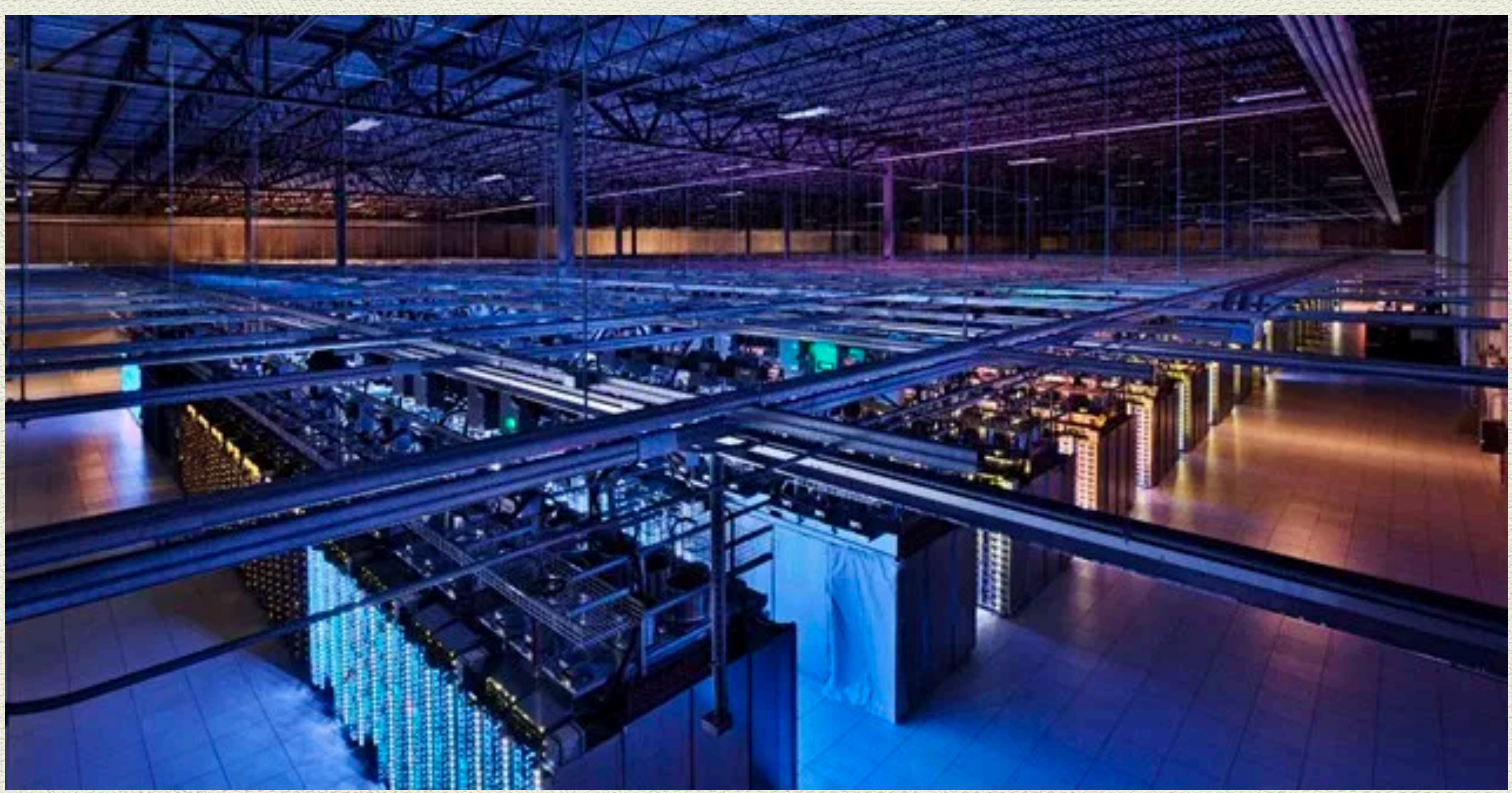
The logarithmic measure is more convenient for various reasons:

1. It is practically more useful. Parameters of engineering importance

[1] Nyquist, H., "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal*, April 1924, p. 324; "Certain Topics in Telegraph Transmission Theory," *A. I. E. E. Trans.*, v. 47, April 1928, p. 617.
[2] Hartley, R. V. L., "Transmission of Information," *Bell System Technical Journal*, July 1928, p. 535.

Storage of Information





Google

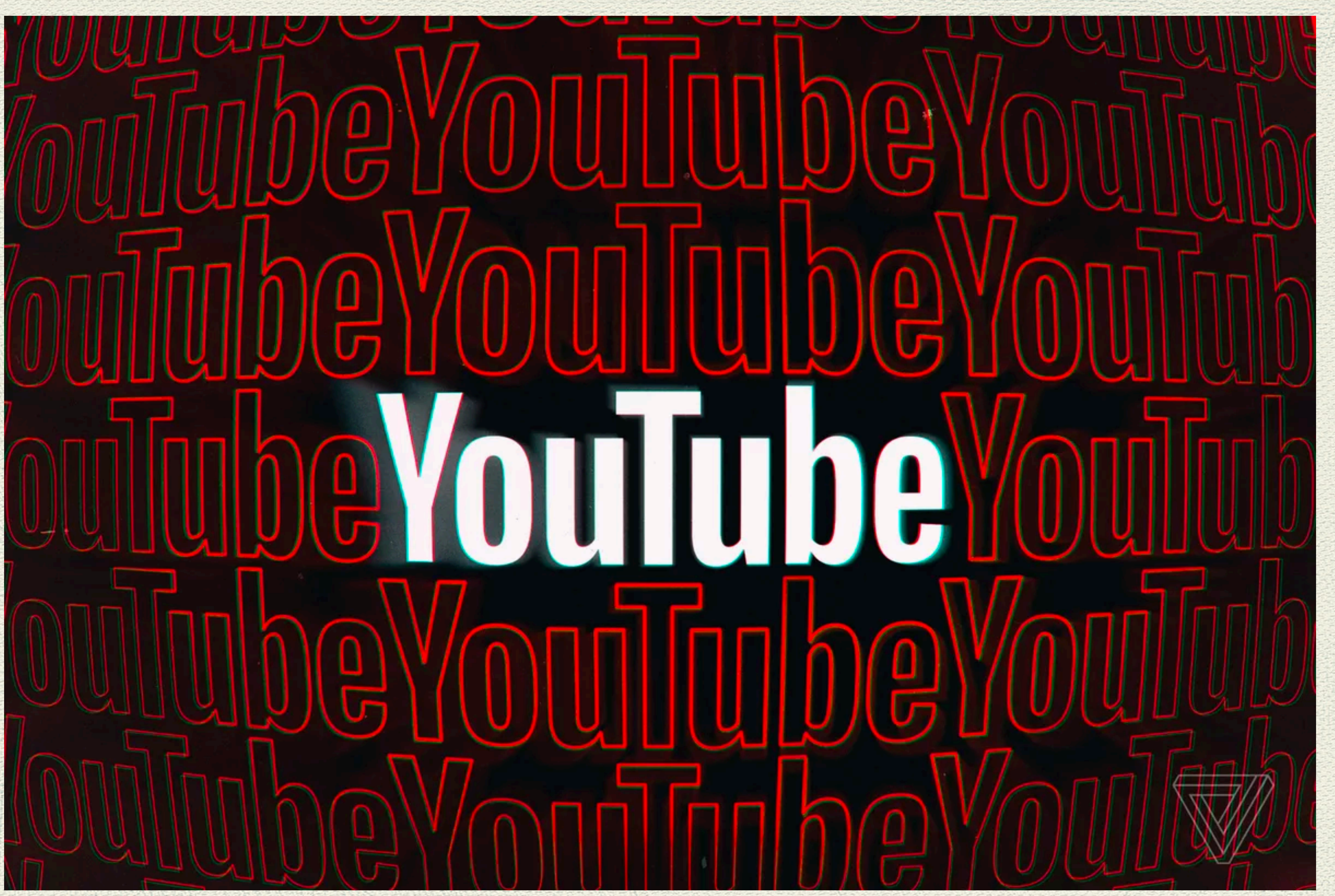
Google: 40,000 Searches /sec

3.5 Billion Searches / day

20 Petabyte/day=20 million GigaBytes/day

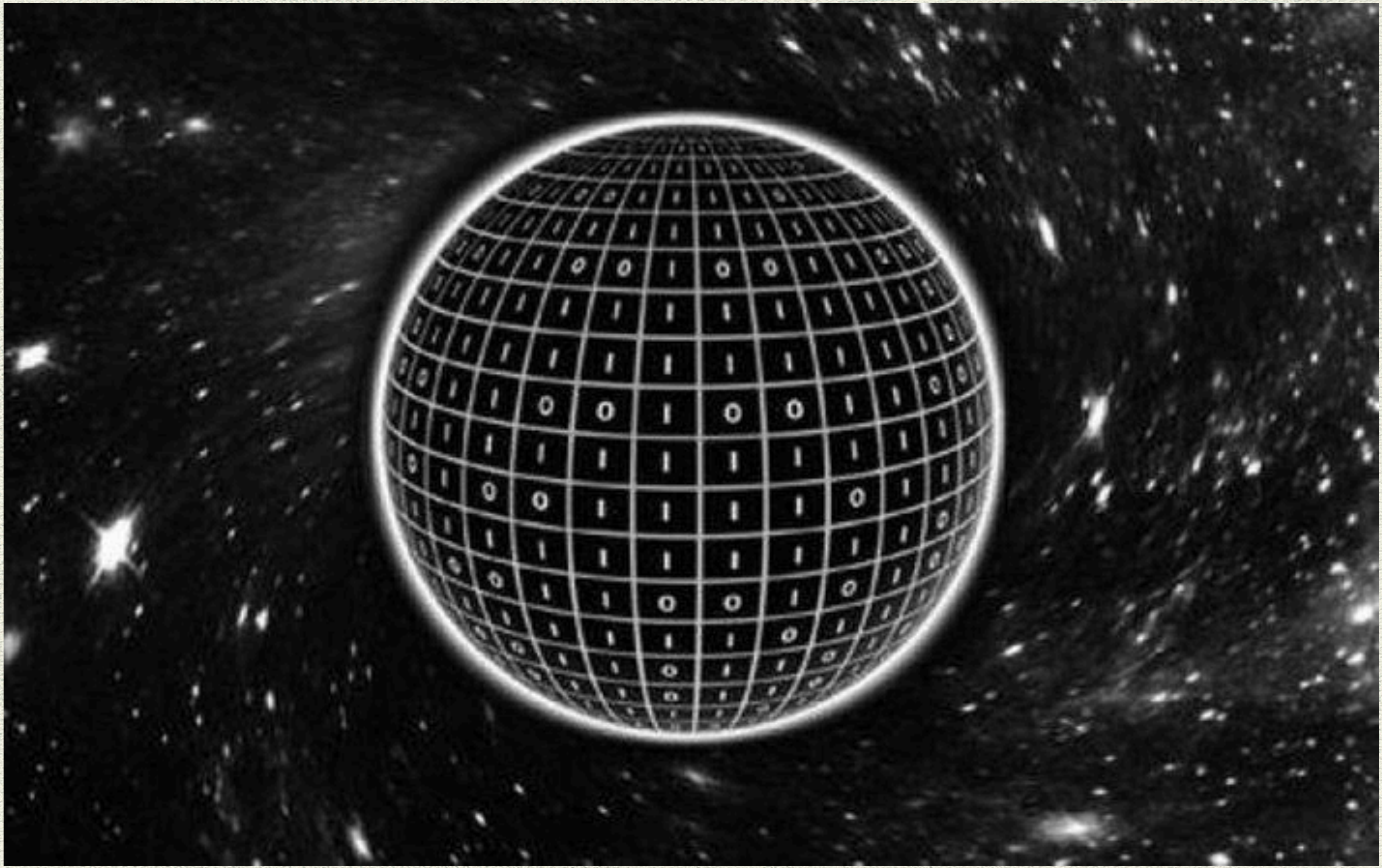
The current capacity of Google

~1000 Exabytes = ~1000 billion Gigabyte



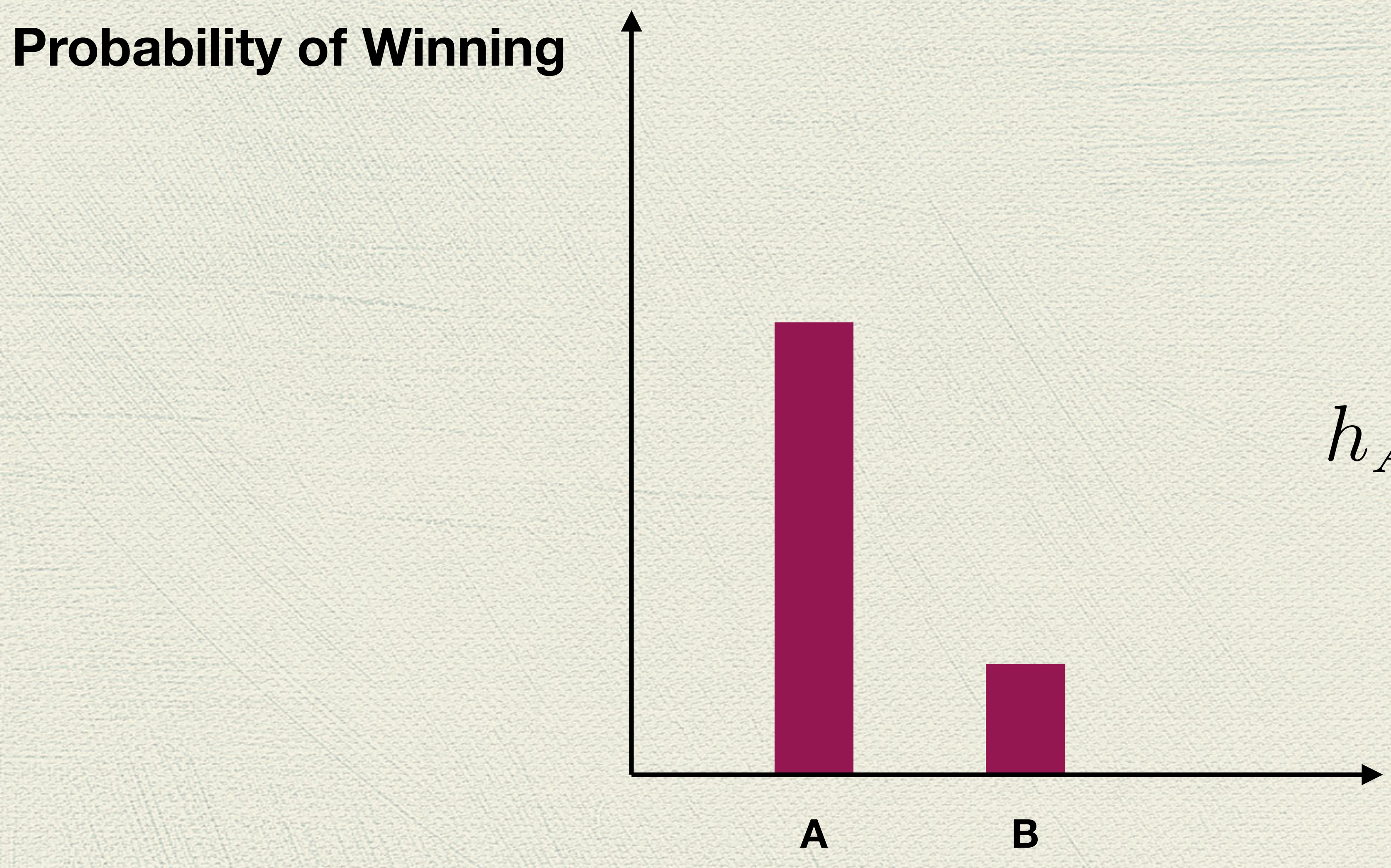
Youtube: 4 Million hours of new clips each day

Black hole and information

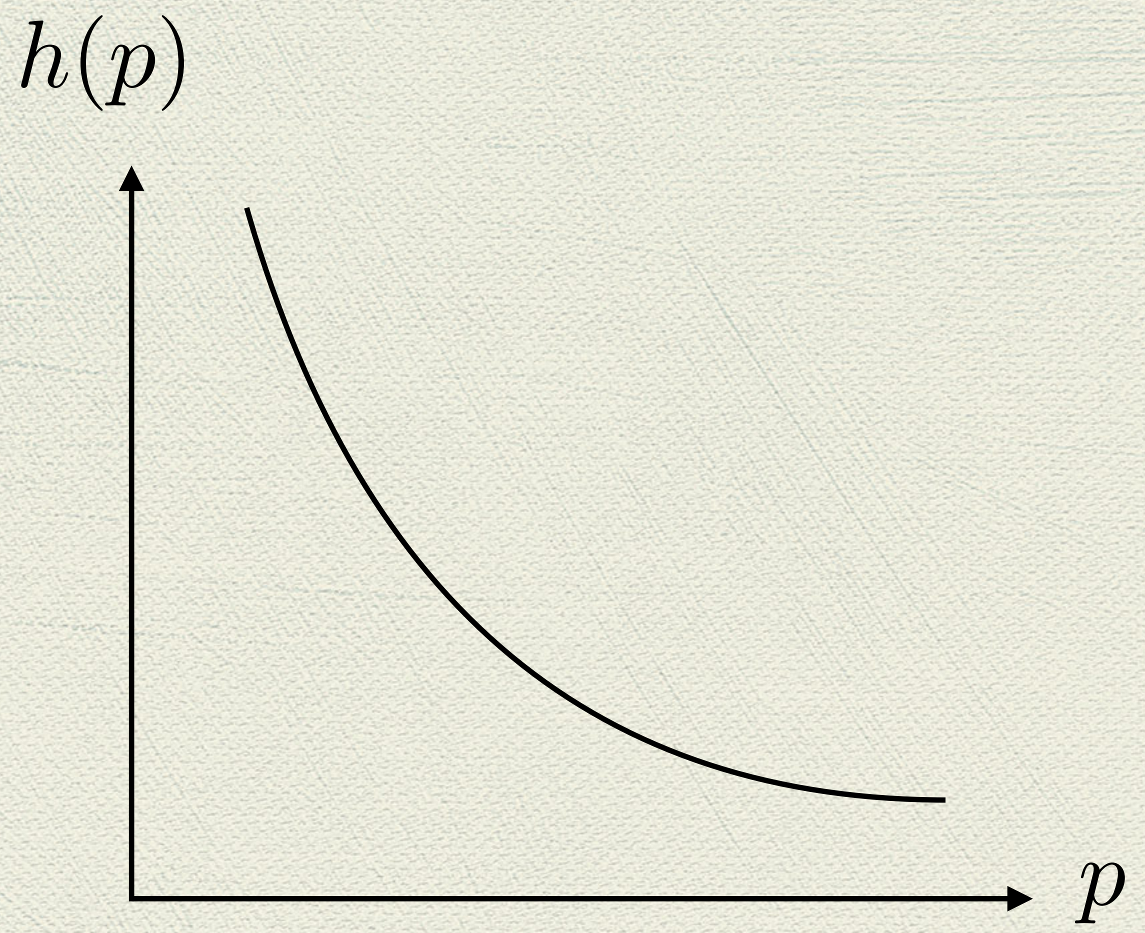


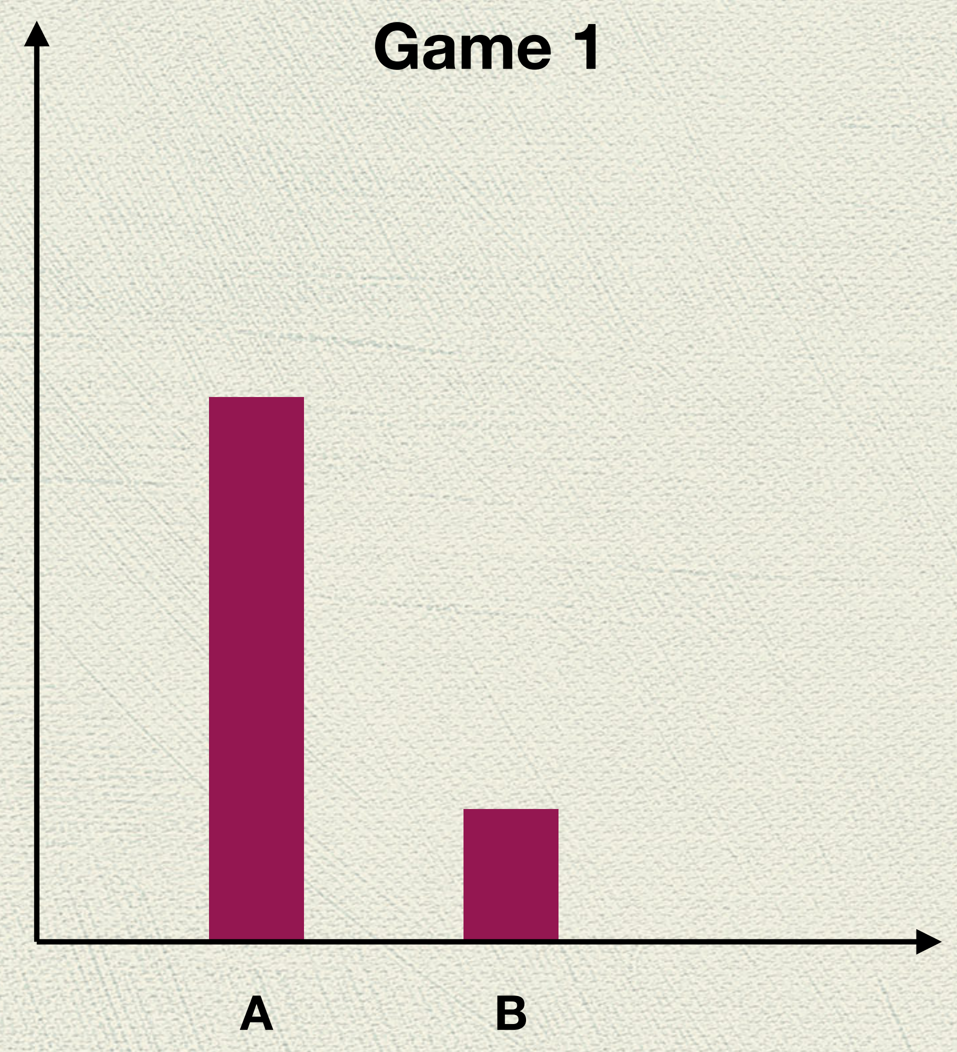
Information, entropy and Surprize

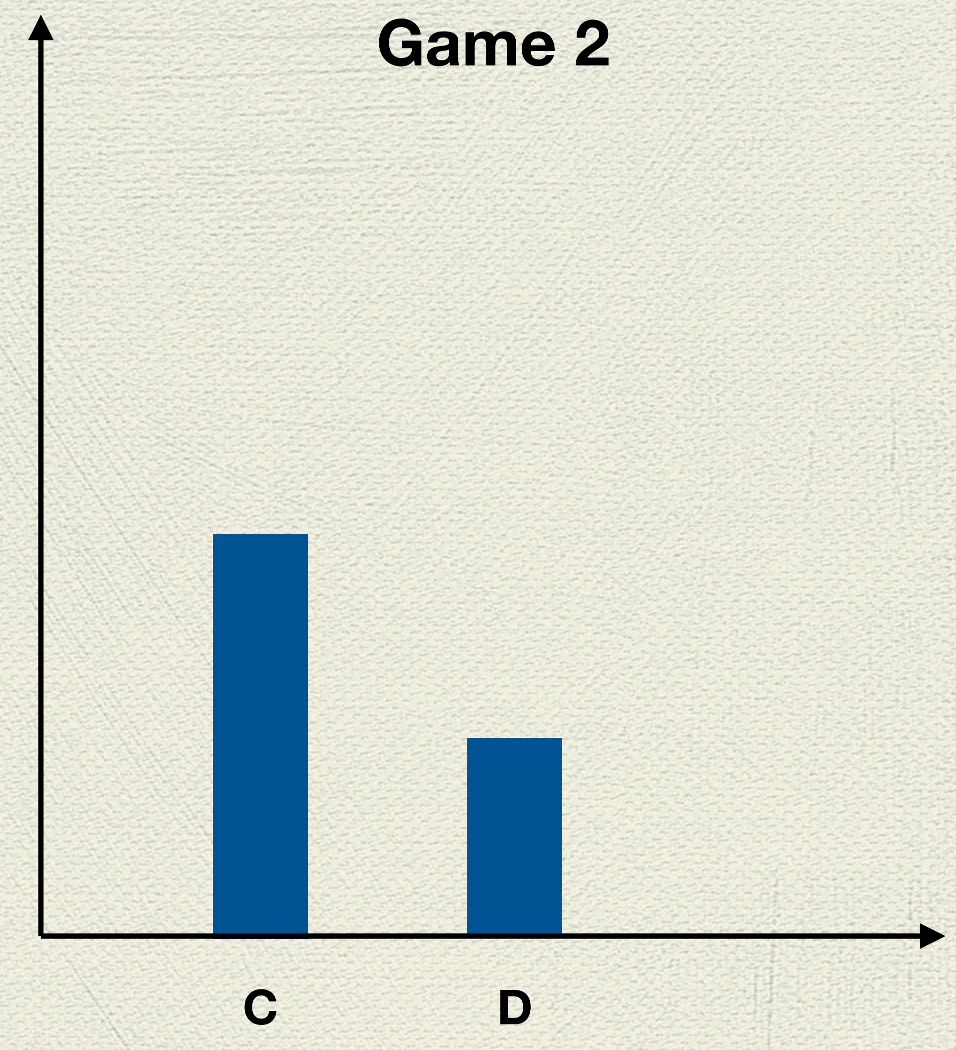
Information and Surprise



$$h_A = h(p_A)$$


$h(p)$
p


Game 1
A
B

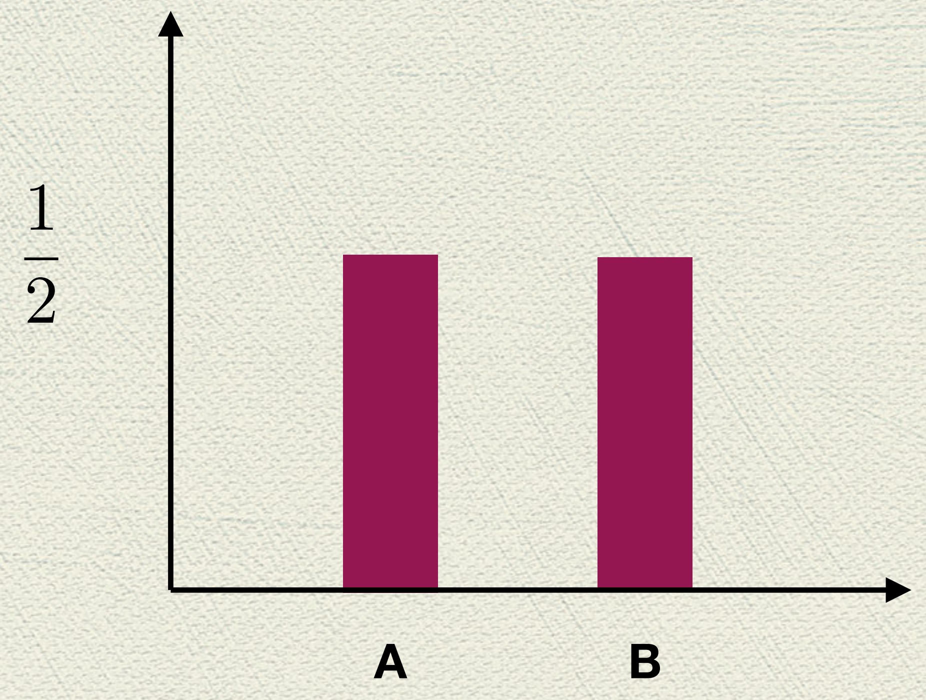

Game 2
C
D

$$h(P_{AC}) = h(P_A) + h(P_C)$$

$$h(P_A P_C) = h(P_A) + h(P_C)$$

$$h(p) = \alpha \log(p)$$

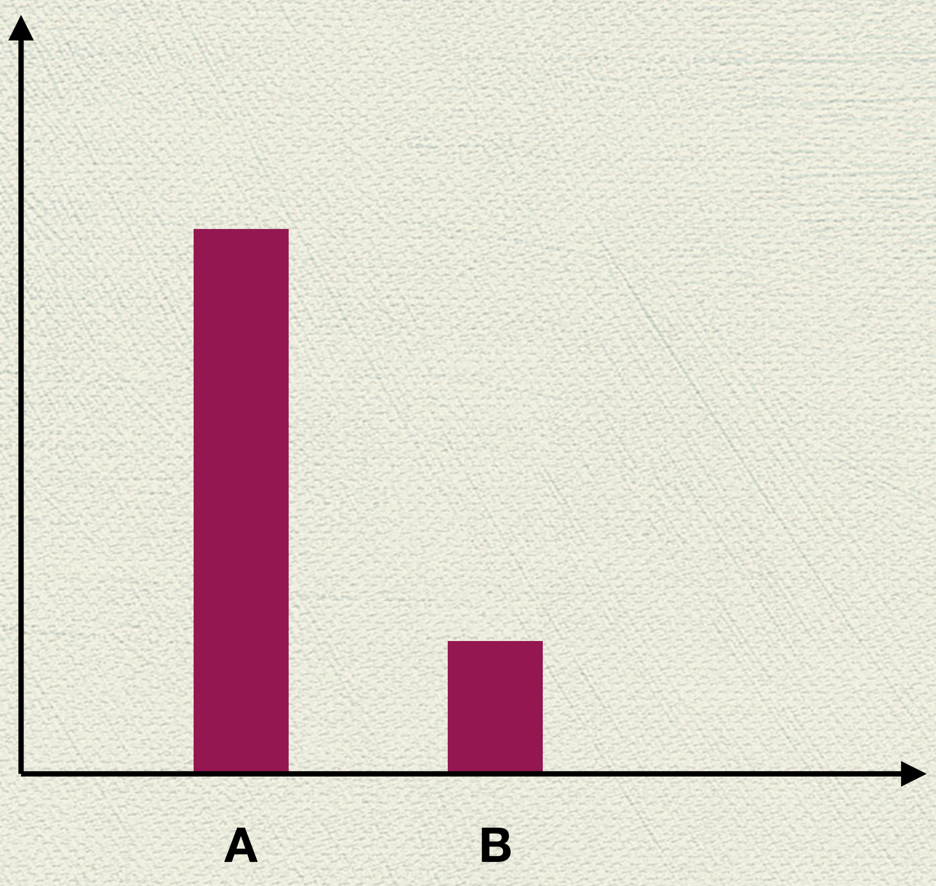
One unit of information



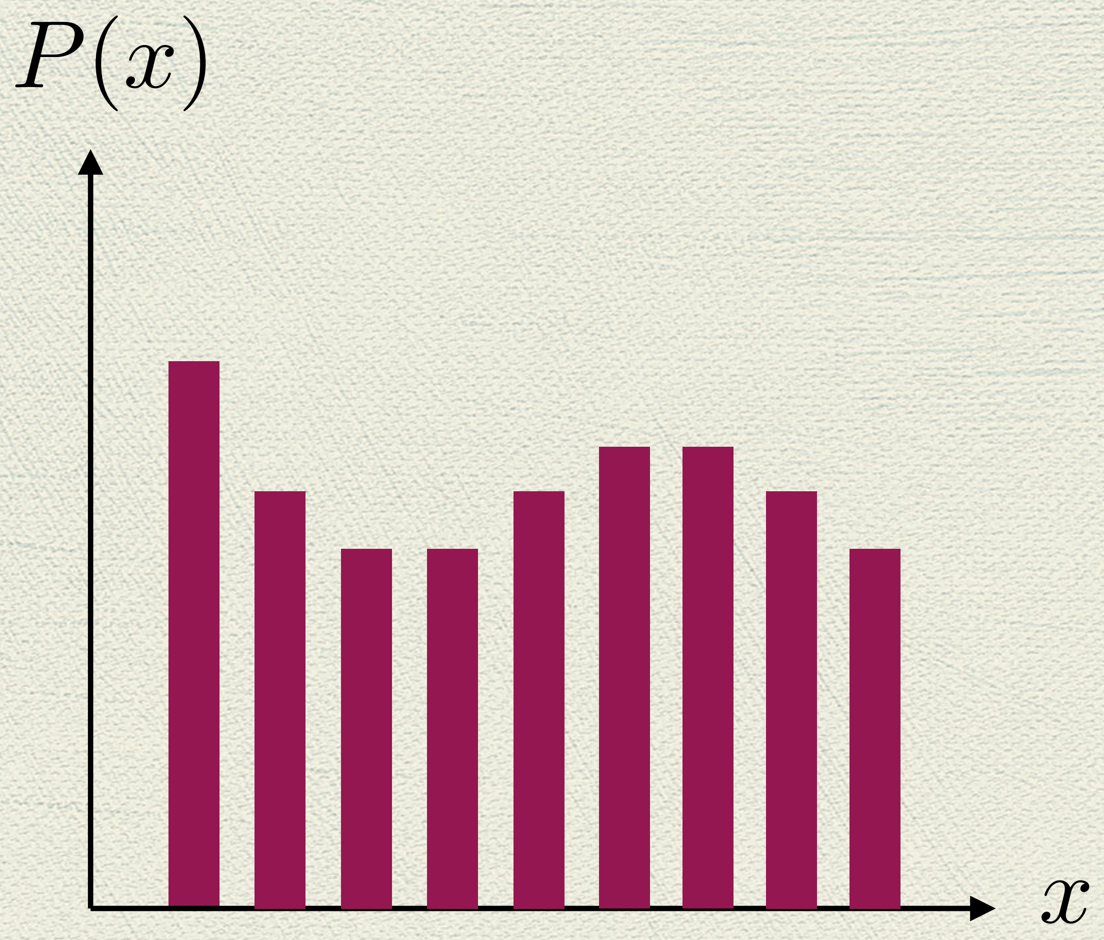
$$h(p) = -\log_2(p)$$

Shannon Entropy

$$H = -P_A log(P_A) - P_B log(P_B)$$



Shannon Entropy



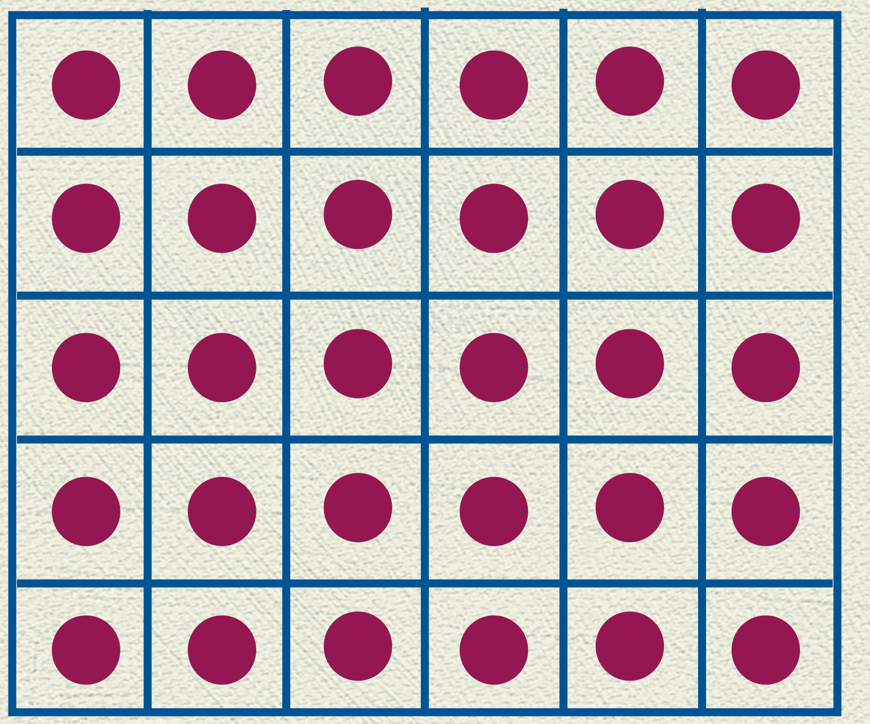
$$H(X) = -\sum_x P_x log(P_x)$$

A useful expression

$$H(X) = -\sum_x P_x log(P_x)$$

$$H = \left\langle \log \frac{1}{p} \right\rangle_p$$

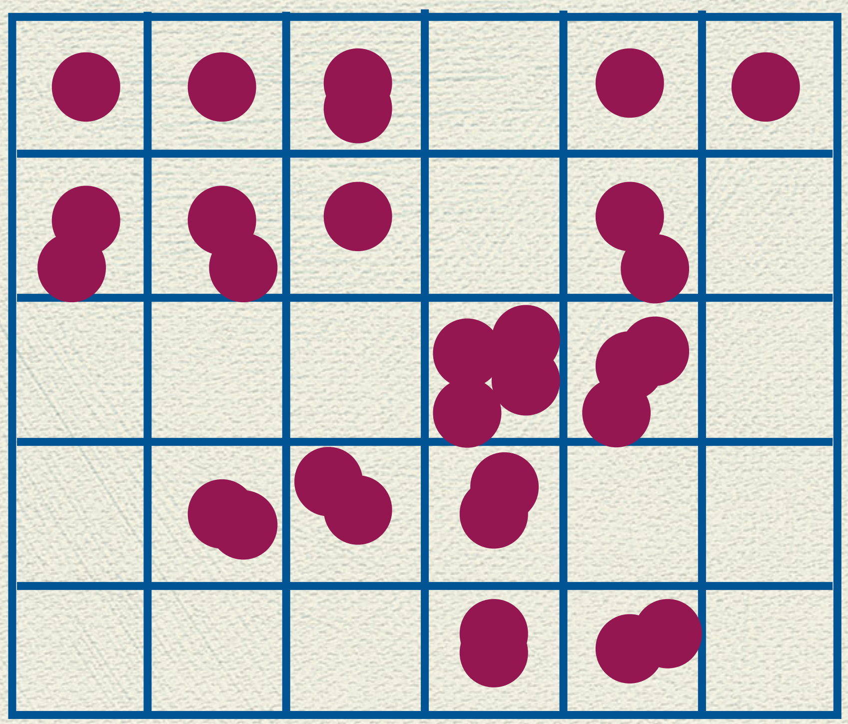
Entropy and Information



A

Zero Entropy

Zero Information



B

High Entropy

High Information

John von Neumann

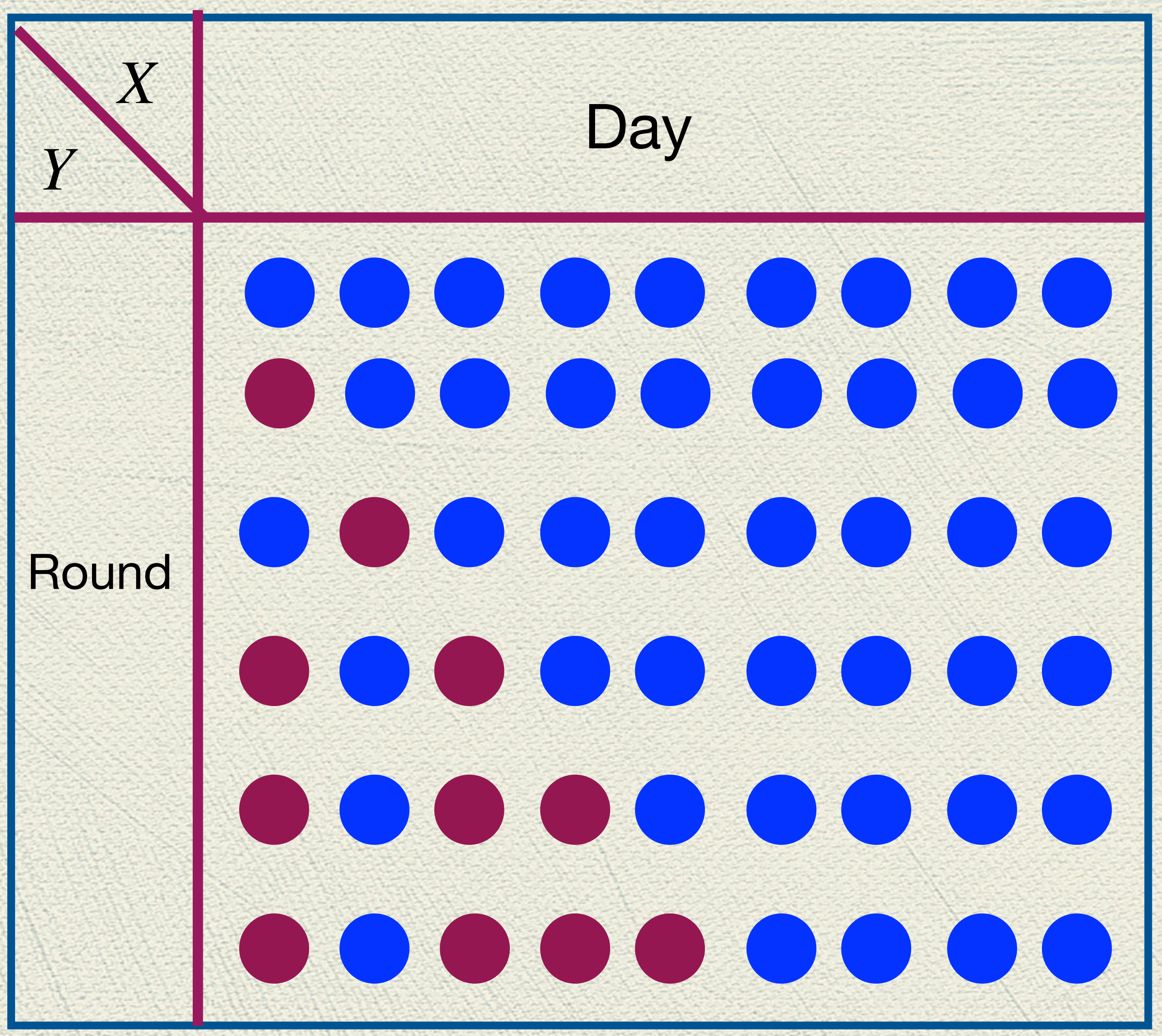
You should call it entropy, for two reasons:

In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name.

In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage.

Joint Information

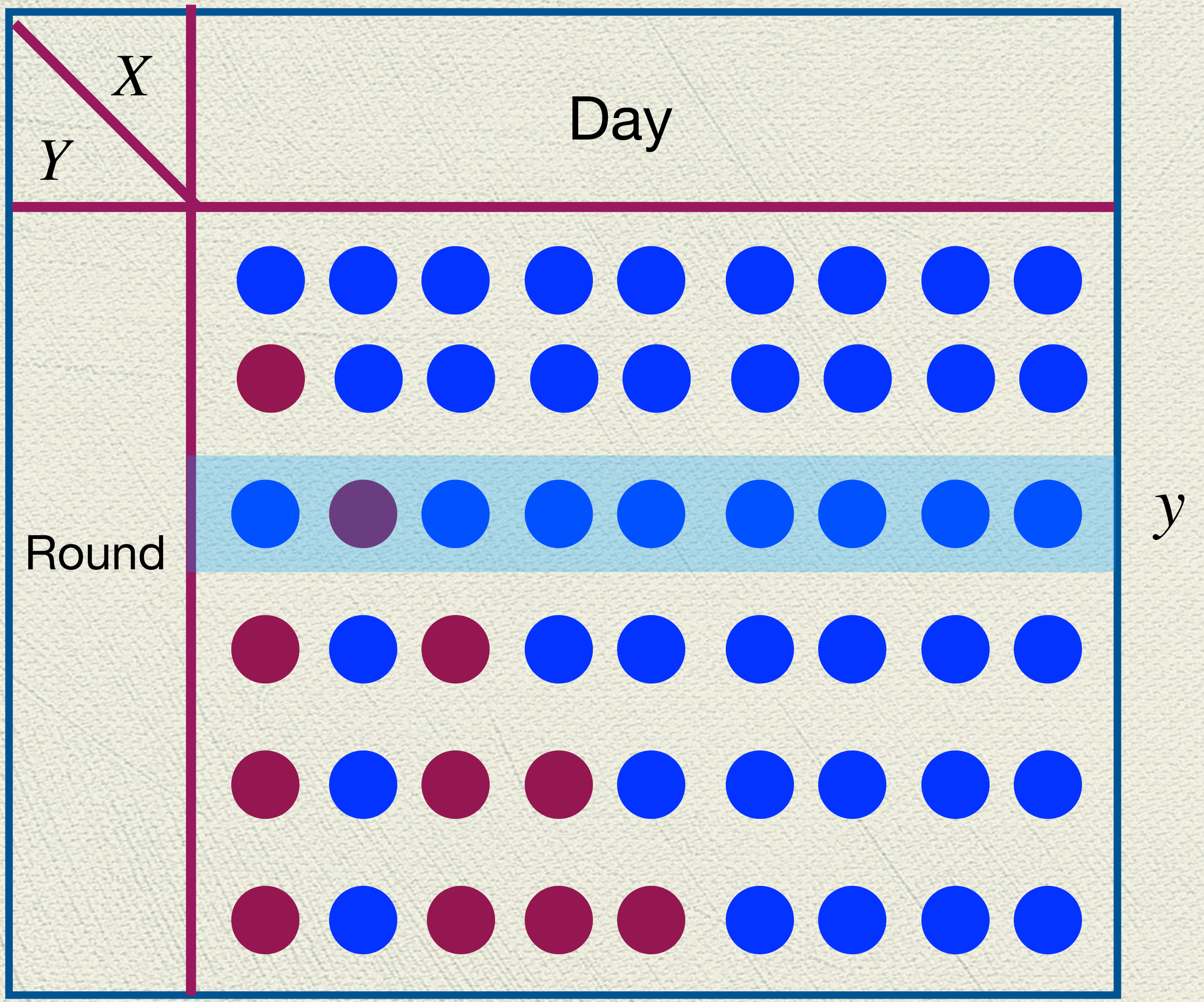
$$H(X, Y) = -\sum_{x,y} p(x, y) \log_2 p(x, y)$$



Conditional Information

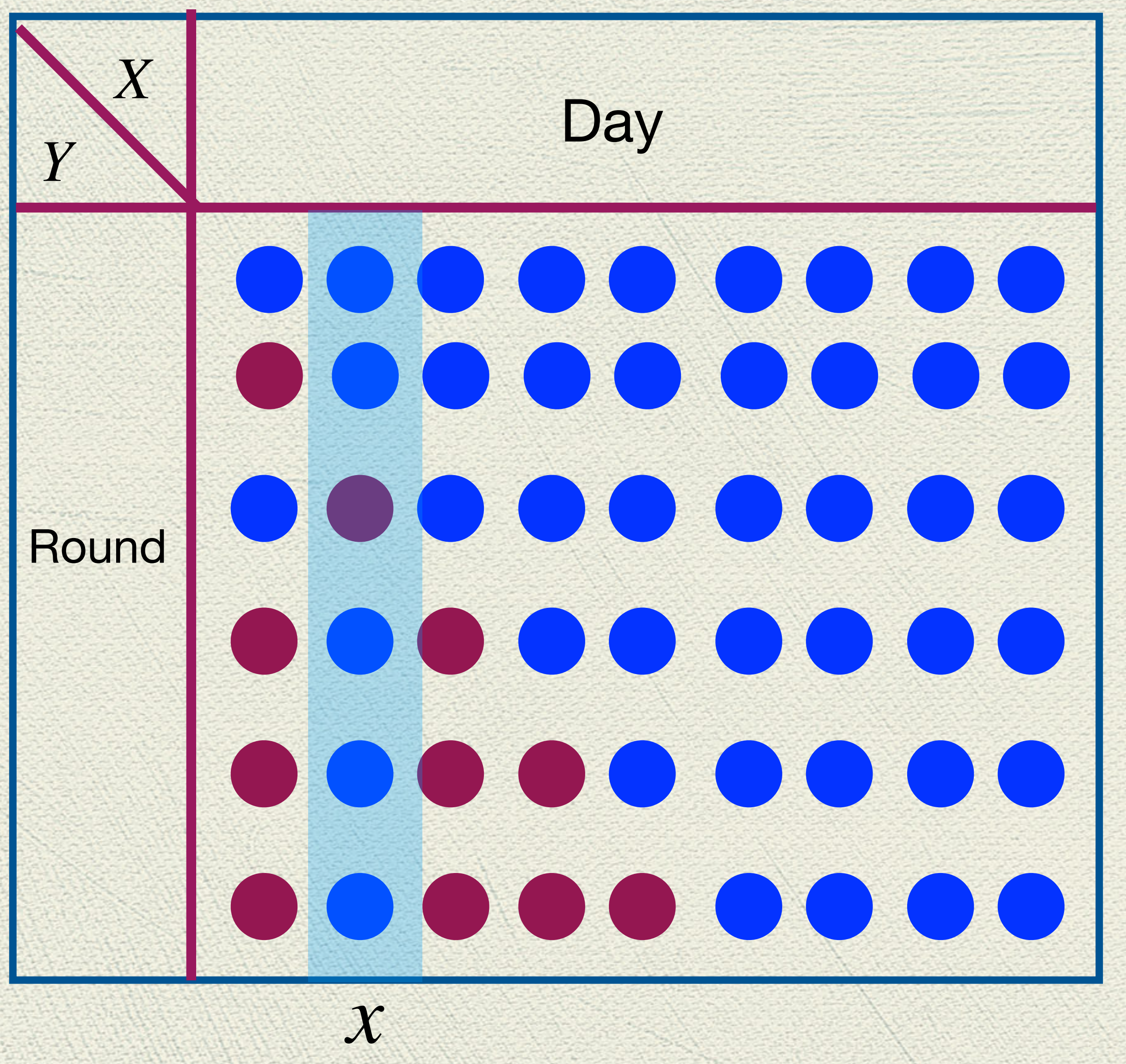
$$H(X|y) := -\sum_{x} p(x|y)\log p(x|y)$$

$$H(X|Y) \geq 0$$



Conditional Information

$H(Y|x)$ $H(Y|X)$



$$H(X) \leq H(X, Y)$$

$$H(Y) \leq H(X, Y)$$

در دو حادثه اطلاعات بیشتری از یک حادثه هست.

An important Inequality

$$ln(x) < x - 1$$

$$\downarrow$$

$$\left\langle \log_2 \frac{q}{p} \right\rangle_p \leq 0 \qquad \forall \; q(x)$$

یک ویرایش دیگر:

$$\left\langle \log_2 \frac{q}{p} \right\rangle_p \leq 0$$

$$H \leq \left\langle \log_2 \frac{1}{q} \right\rangle_p$$

نتیجه یک

$$H \leq \left\langle \log_2 \frac{1}{q} \right\rangle_p$$

$$q(x) = \frac{1}{\Omega} \longrightarrow H(X) \leq \log_2 \Omega$$

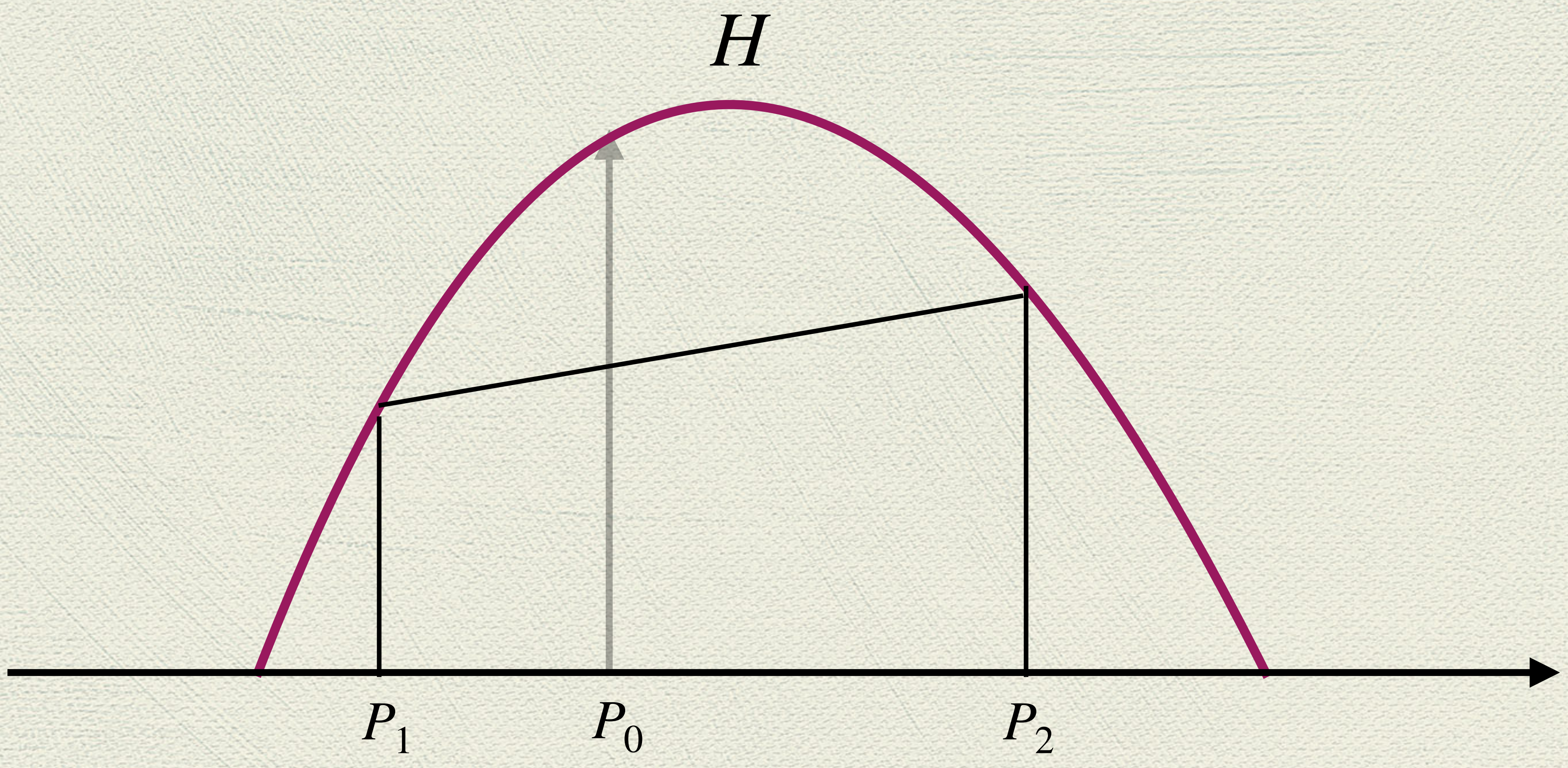
نتیجه دو:

$$H(X,Y) \leq -\sum_{x,y} p(x,y)\log_2 q(x,y)$$

$$q(x,y) = p(x)p'(y) \longrightarrow H(X,Y) \leq H(X) + H(Y)$$

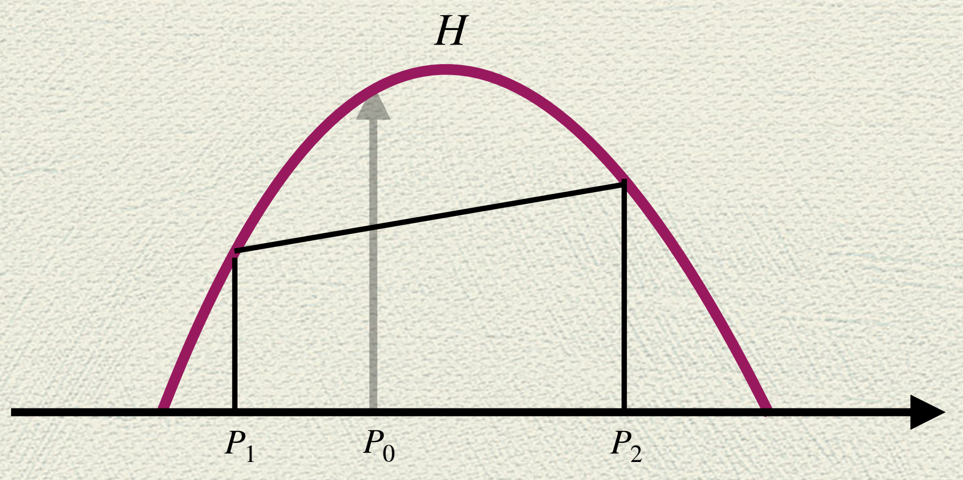
نتیجه سه:

$$\lambda H(P_1) + (1 - \lambda)H(P_2) \leq H(P_0)$$



If you want a maximum, mix it more and more.

If you want a mimum, purify it more and more.



Proof:

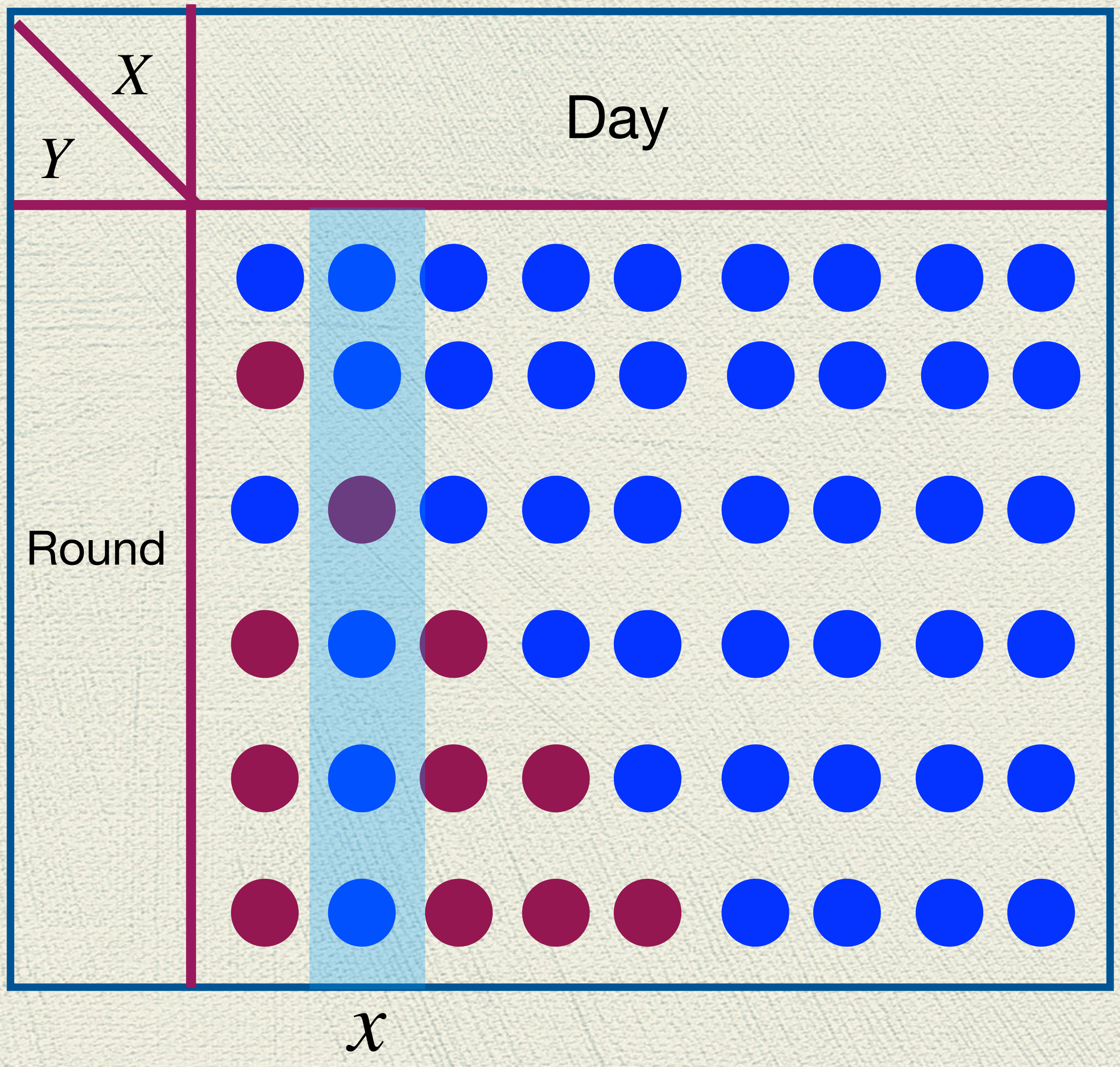
$$\lambda H(P_1) + (1-\lambda)H(P_2) \leq H(P_0)$$

$$H_0 - \lambda H_1 - (1-\lambda)H_2$$

$$= \sum p_0 \log \frac{1}{p_0} - \lambda p_1 \log \frac{1}{p_1} - (1-\lambda)p_2 \log \frac{1}{p_2}$$

$$= \lambda \sum p_1 \log \frac{p_1}{p_0} + (1-\lambda)p_2 \log \frac{p_2}{p_0} \geq 0$$

Mutual Information



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

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

$$I(X:Y) := H(X) + H(Y) - H(X,Y) \geq 0$$

نتیجه چهار:

Strong Subadditivity

$$H(X \mid Y, Z) \leq H(X \mid Y)$$

شرط های بیشتر آنتروپی را کم می کند.

یک معنای دیگر از زیرجمع پذیری قوی:

$$H(X,Y,Z)-H(Y,Z)\leq H(X,Y)-H(Y)$$

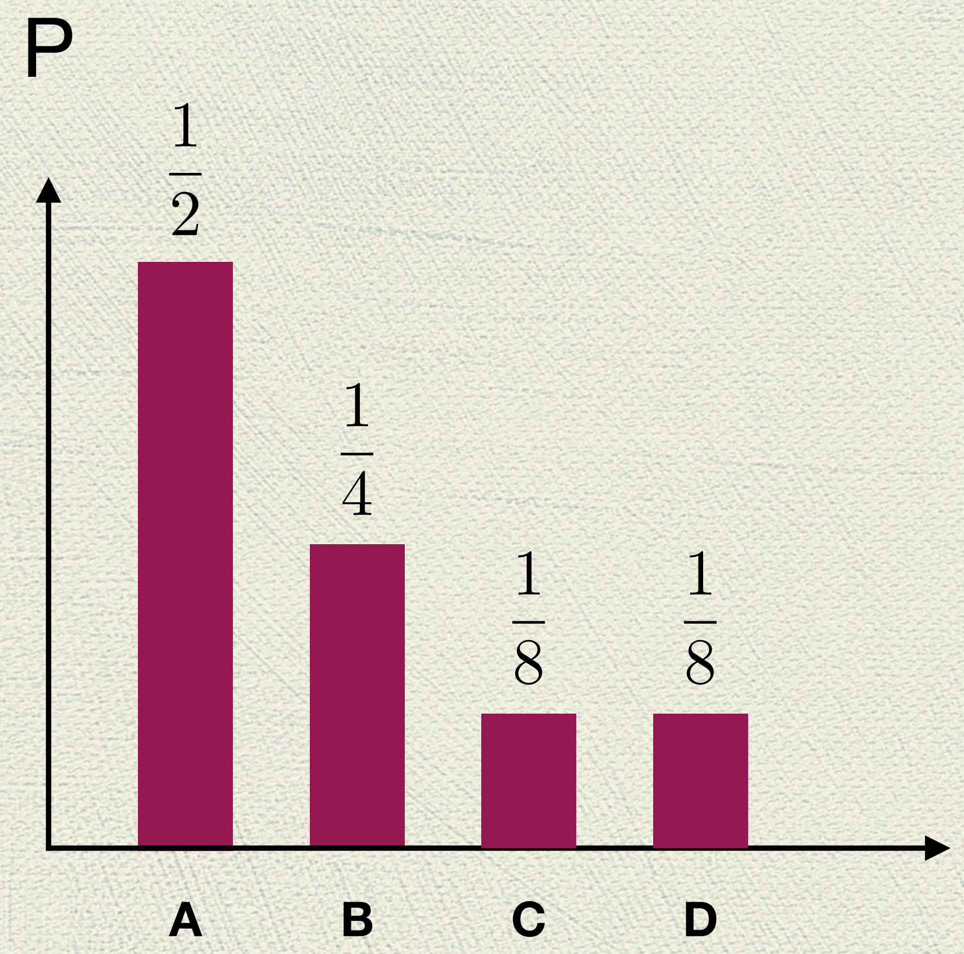
$$H(X,Y,Z)-H(Y,Z)-H(X)\leq H(X,Y)-H(Y)-H(X)$$

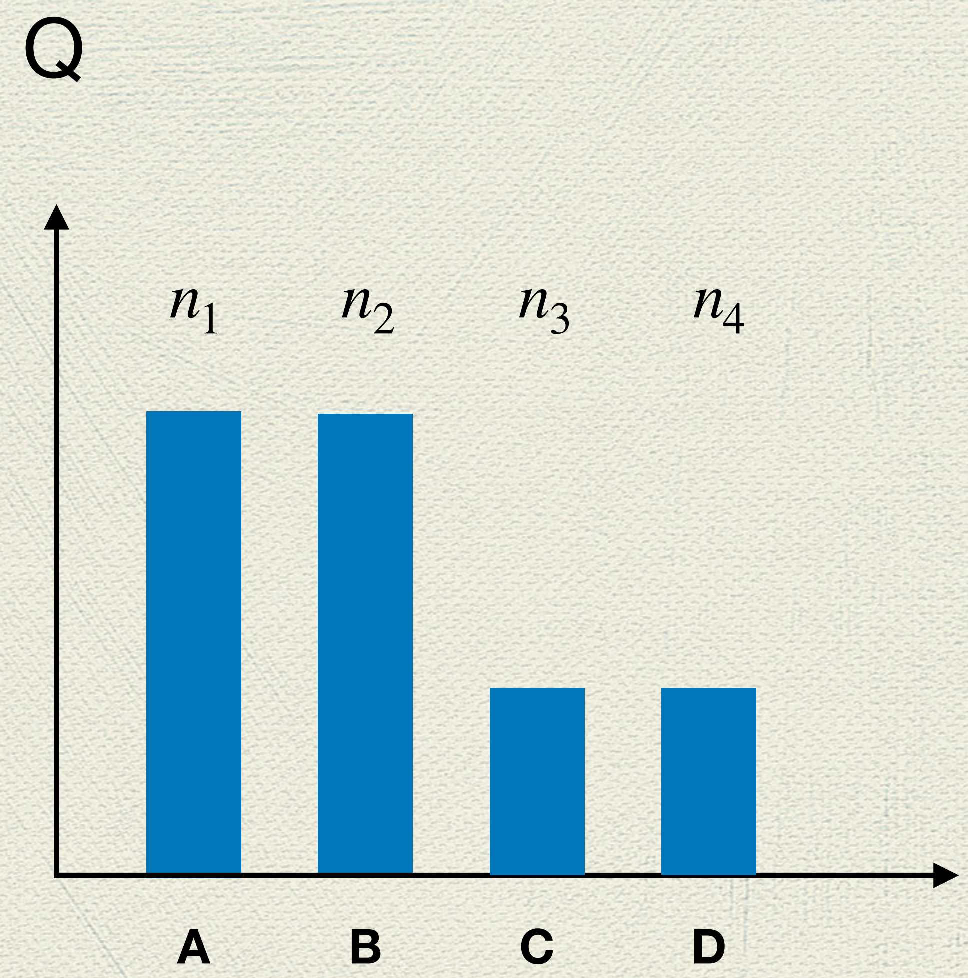
$$I(X:Y,Z)\geq I(X:Y)$$

دانستن Y و Z اطلاعات بیشتری در باره X به ما می دهد تا دانستن تنها Y .

آنتروپی نسبی و تعبیر آن به عنوان فاصله

Relative Entropy





$N_0 = 45$ $N_1 = 55$

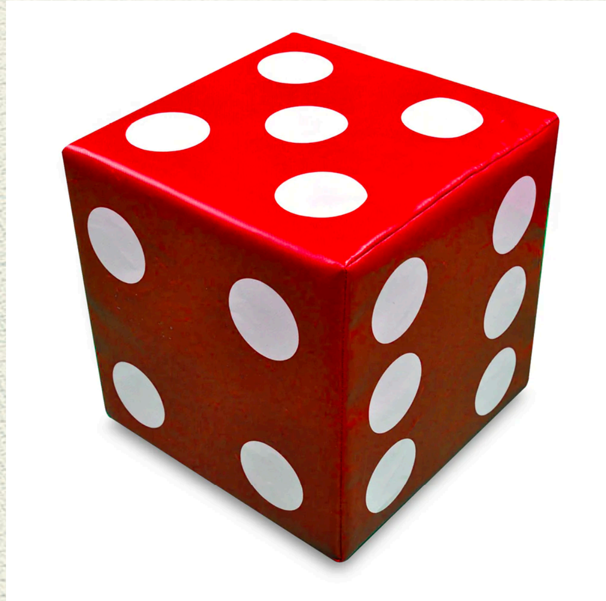
N=100

Guess $\longrightarrow$ $p_0 = ?$ $p_1 = ?$

4555 5445

10000

$p_0 = ?$ $p_1 = ?$



n_1 n_2 n_3 ... n_6

N

p_1 p_2 p_3 p_6

Kullback-Leibler Divergence

$$P_N(p\,|\,q) = \frac{N!}{(Np_1)!(Np_2)!\cdots(Np_k)!} q_1^{Np_1} q_2^{Np_2} \cdots q_k^{Np_k}$$

$$P(p\,|\,q) = 2^{-ND(p\|q)}$$

$$D(p\|q) = \langle \log \frac{p}{q} \rangle_p$$

Relative Entropy and Mutual Information

$$D(p(x,y)\|p(x)q(y)) = I(X:Y)$$

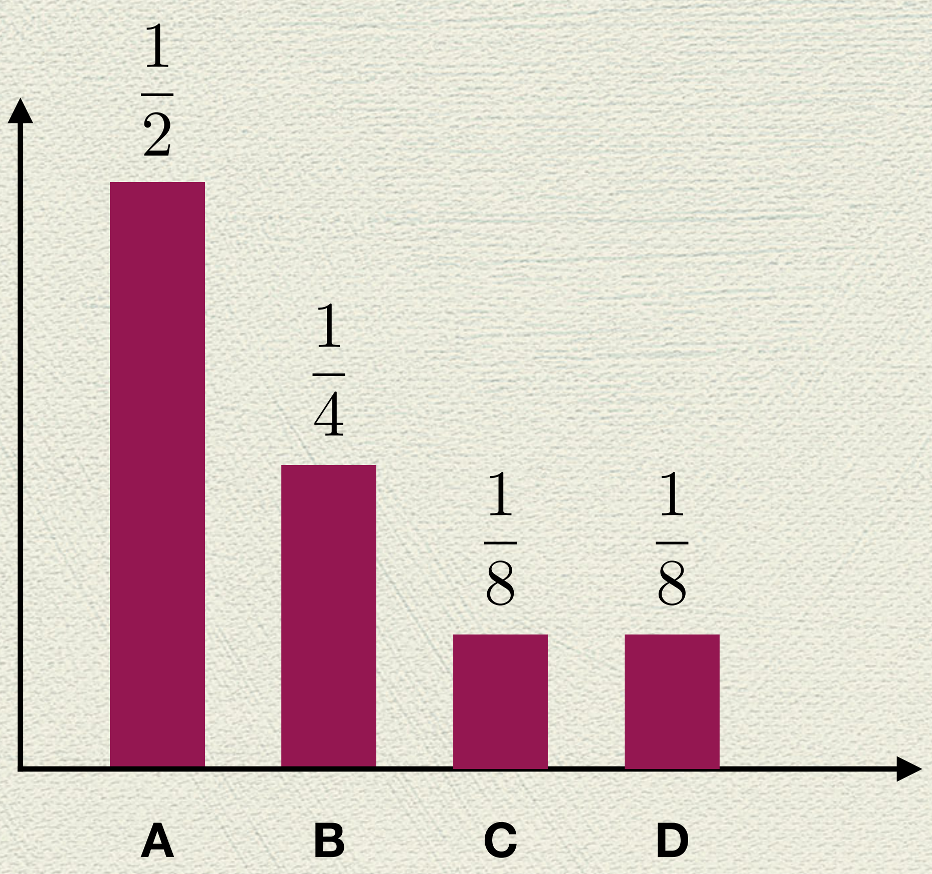

قضیه اول شانون: فشرده سازی اطلاعات

A $\longrightarrow$ 00

B $\longrightarrow$ 01

C $\longrightarrow$ 10

D $\longrightarrow$ 11



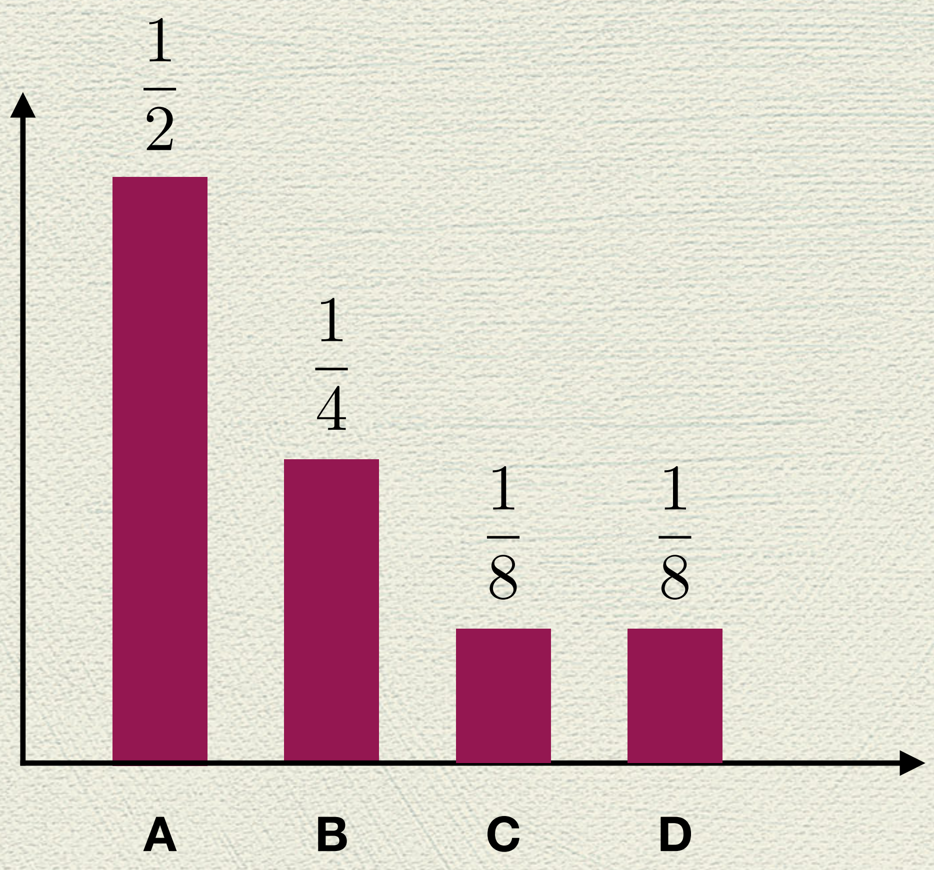
AABABCDABAADCABA $\longrightarrow$ 00000100011011000100001110000100

A → 0

B → 10

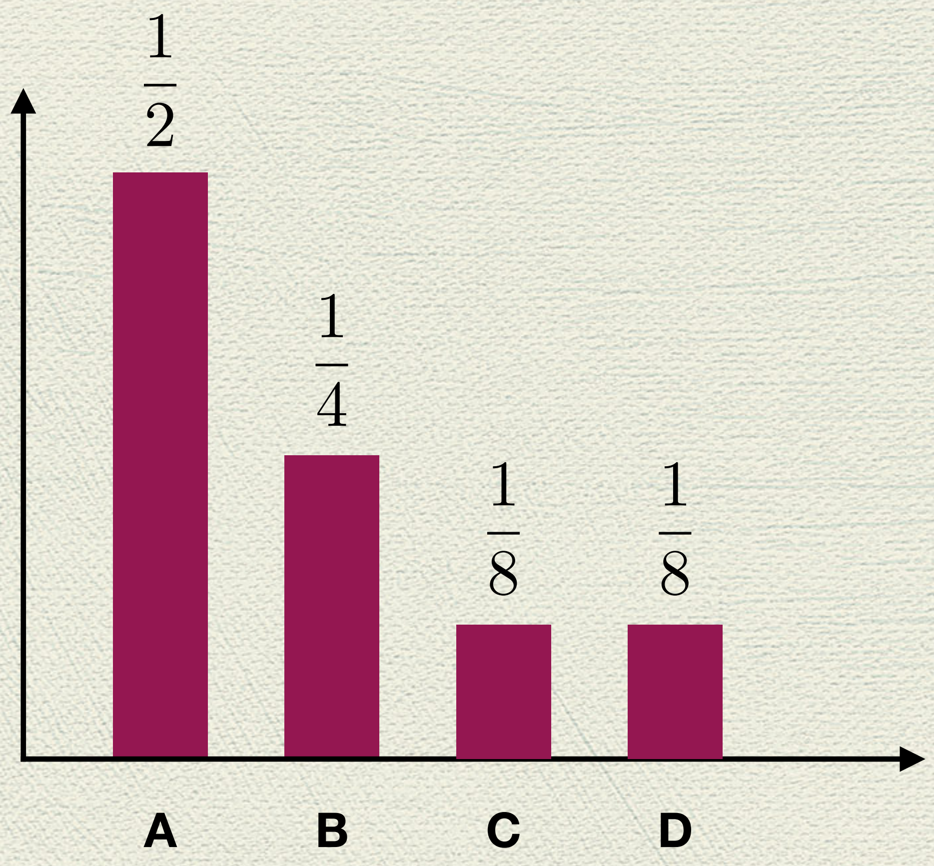
C → 110

D → 111



AABABCDABAADCABA → 00000100011011000100001110000100

AABABCDABAADCABA → 00100110111010001111100100



A ⟶ **0** $\quad -\log(P_A) = 1$

B ⟶ **10** $\quad -\log(P_B) = 2$

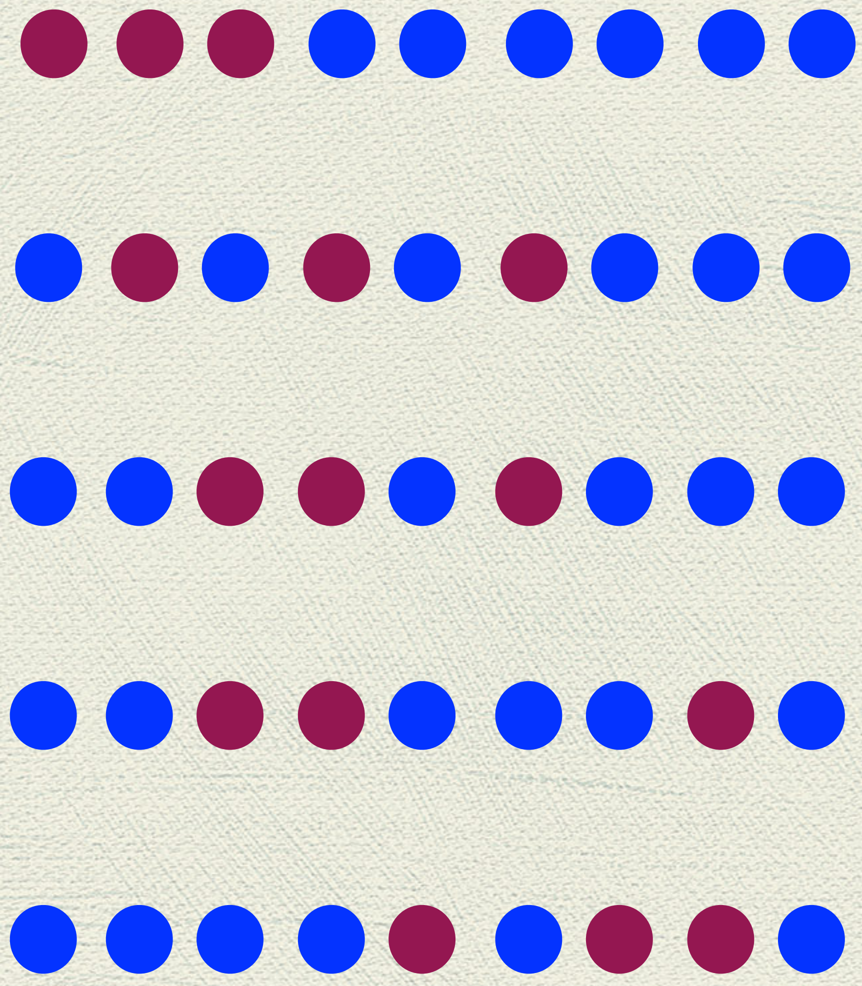
C ⟶ **110** $\quad -\log(P_C) = 2$

D ⟶ **111** $\quad -\log(P_D) = 2$

$$\langle l \rangle = -\sum_i P_i \log P_i = H(X)$$

یک مثال دیگر: یک بازی تکرار شونده.

برای گزارش نتایج یکصد بازی به چه تعداد بیت نیاز داریم؟

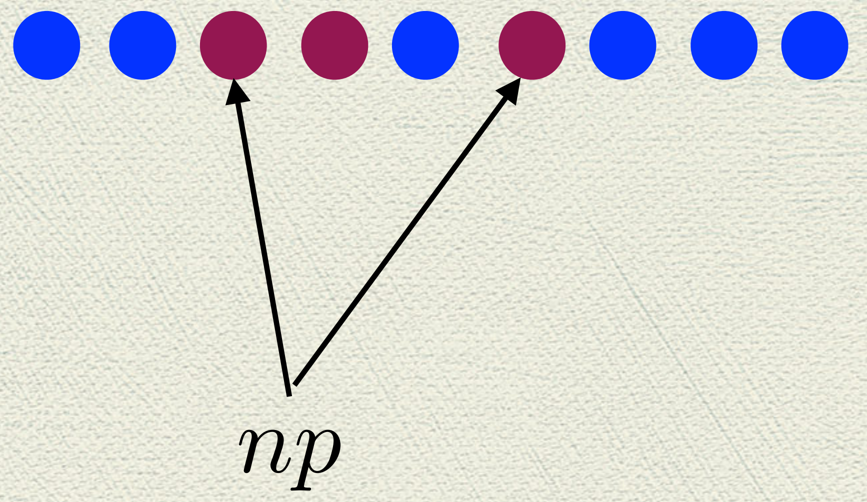


Number of all sequences $= 2^n$

Average number of red balls $= np$

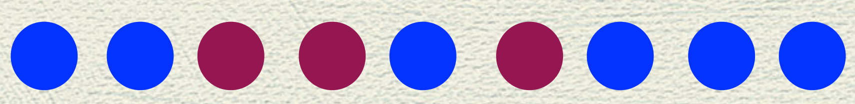
Number of typical sequences $= \binom{n}{np}$

Typical sequences　　رشته های نمونه



تعداد رشته های نمونه

$$\binom{n}{np} = \frac{n!}{(np)!(n-np)!}$$



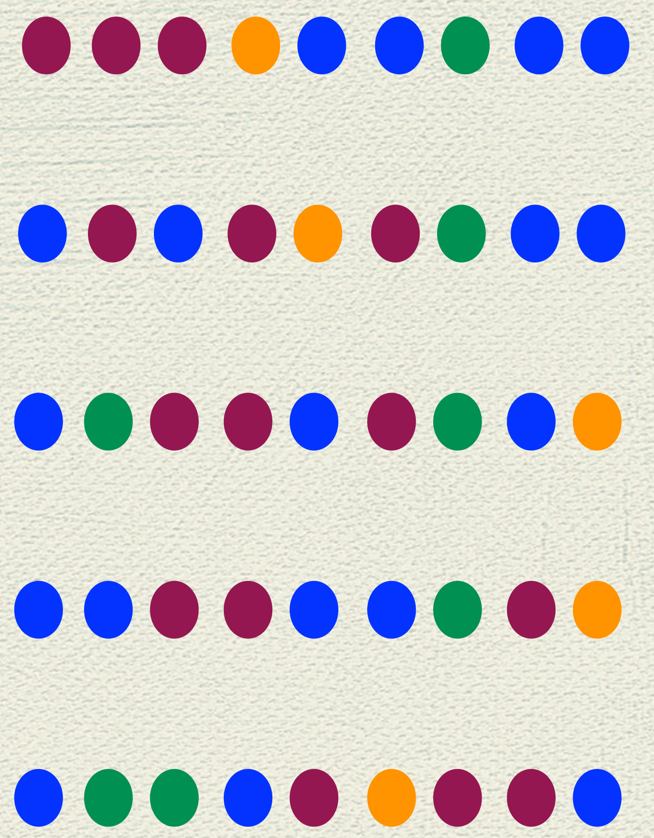
$$\binom{n}{np} = \frac{n!}{(np)!(n-np)!}$$

$$\ln n! = n \ln n - n$$

$$\log \binom{n}{np} = nH(p)$$

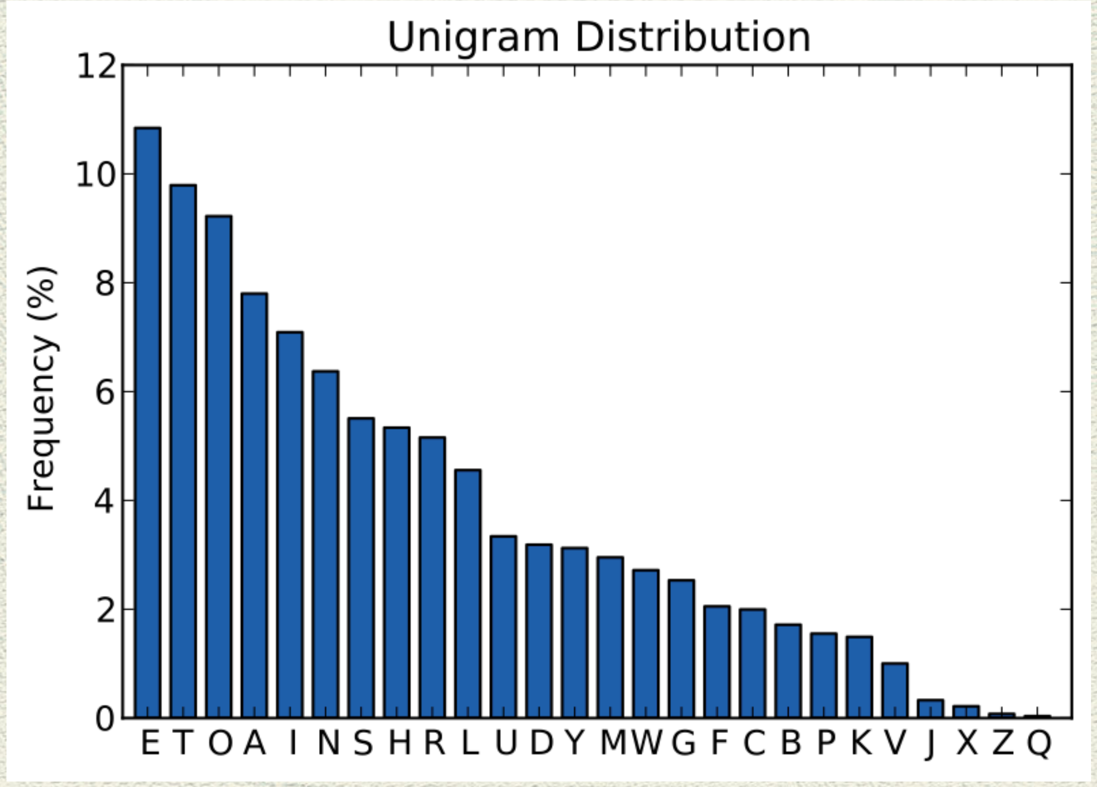
$$\binom{n}{np} = 2^{nH(p)}$$

Number of all sequences $= 2^n$



Number of typical sequences $= 2^{nH(X)}$

A Grand Example



$$H(X) = 4.1$$

The current capacity of Google ~1000 Exabytes = ~1000 billion Gigabyte

$$P(A) = p = \frac{2}{3}$$

$$P(B) = q = \frac{1}{3}$$

A B A B B A B B A A A B B A A B A

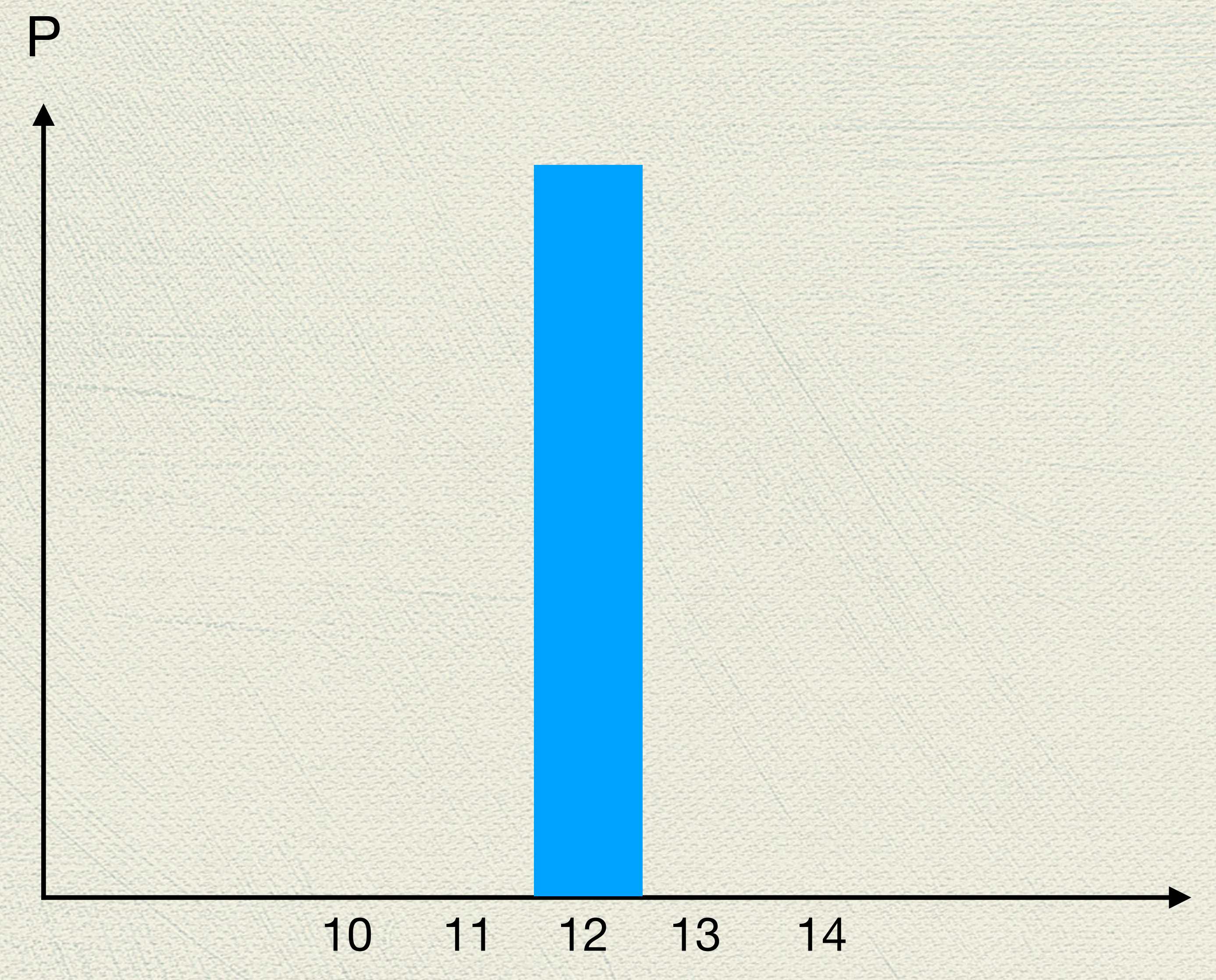
B B A B B A B A B B A B B A A B A

A B A B A A A B B B A A A B A B A

A A A A A A A A B B B A A B B B A

N = ۱۸ = طول رشته

تعریف سخت گیرانه از رشته های متعارف



A B A B B A B B A A A B B A A B A

B B A B B A B A B B A B B A A B A

A B A B A A A B B B A A A B A B A

A A A A A A A A B B B A A B B B A

$$\pi = (\frac{2}{3})^{12} \times (\frac{1}{3})^{6}$$

احتمال اینکه این منبع رشته های متعارف خیلی دقیق تولید کند کم است.

$$\mathscr{P}_0 = \pi \times \binom{18}{12} = 0.1962$$

ABABBABBAAABBAABA ABABBABBAAABBAABA

ABABBABBAAABBAABA ABABBABBAAABBAABA

ABABBABBAAABBAABA ABABBABBAAABBAABA

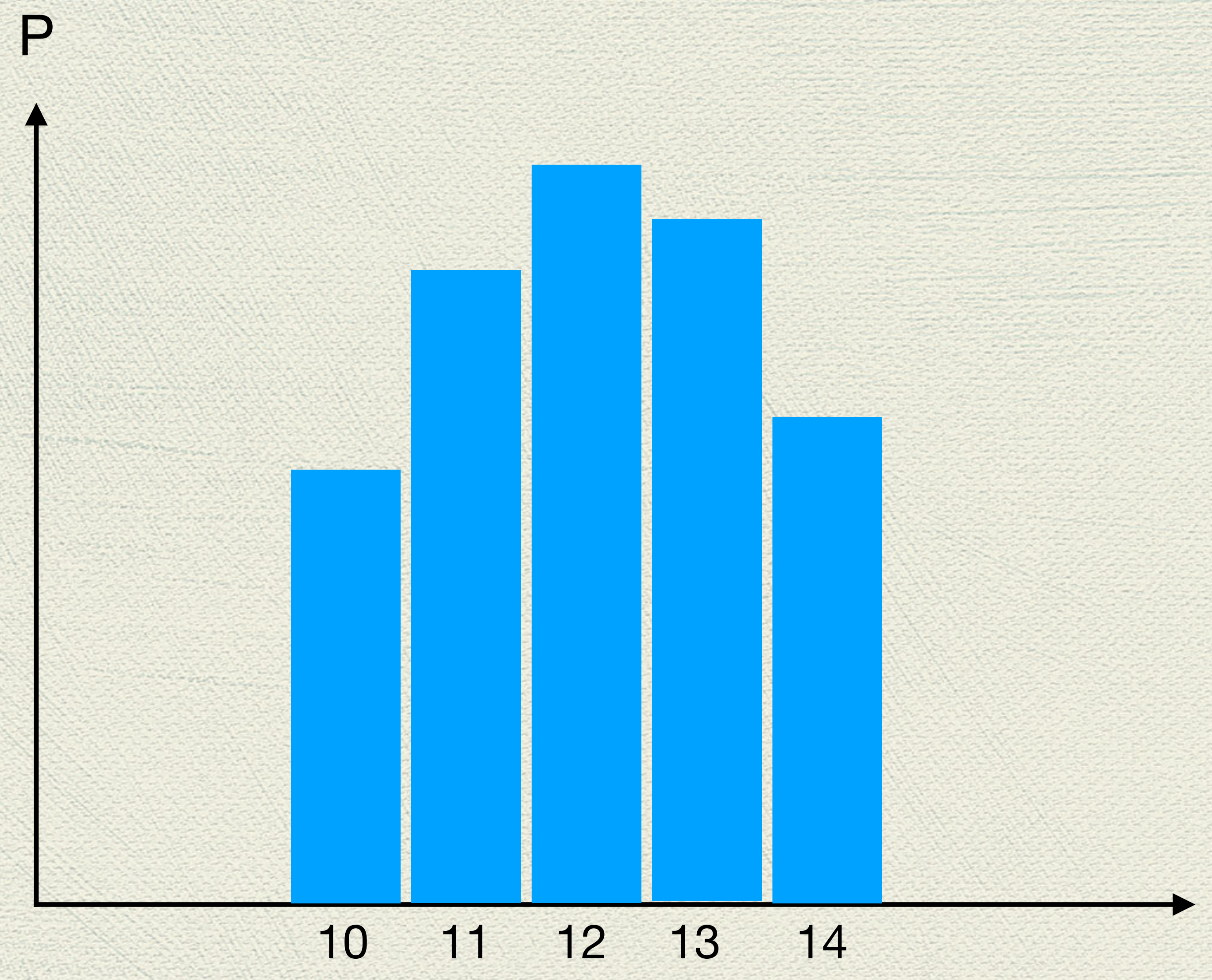
$\mathscr{P}_0 = 0.1399$ طول رشته = ۳۶ = N

$\mathscr{P}_0 = 0.1214$ طول رشته = ۴۸ = N

$\mathscr{P}_0 = 0.1037$ طول رشته = ۶۶ = N

افزایش طول رشته ها فایده ای ندارد.

تساهل و تسامح! یک انحراف معیار را هم قبول می کنیم



$$\sigma = \sqrt{mp(1-p)} = \sqrt{18 \times \frac{2}{3} \times \frac{1}{3}} = 2$$

A B A B B A B B A A A B B A A B A

B B A B B A B A B B A B B A A B A

A B A B A A A B B B A A A B A B A

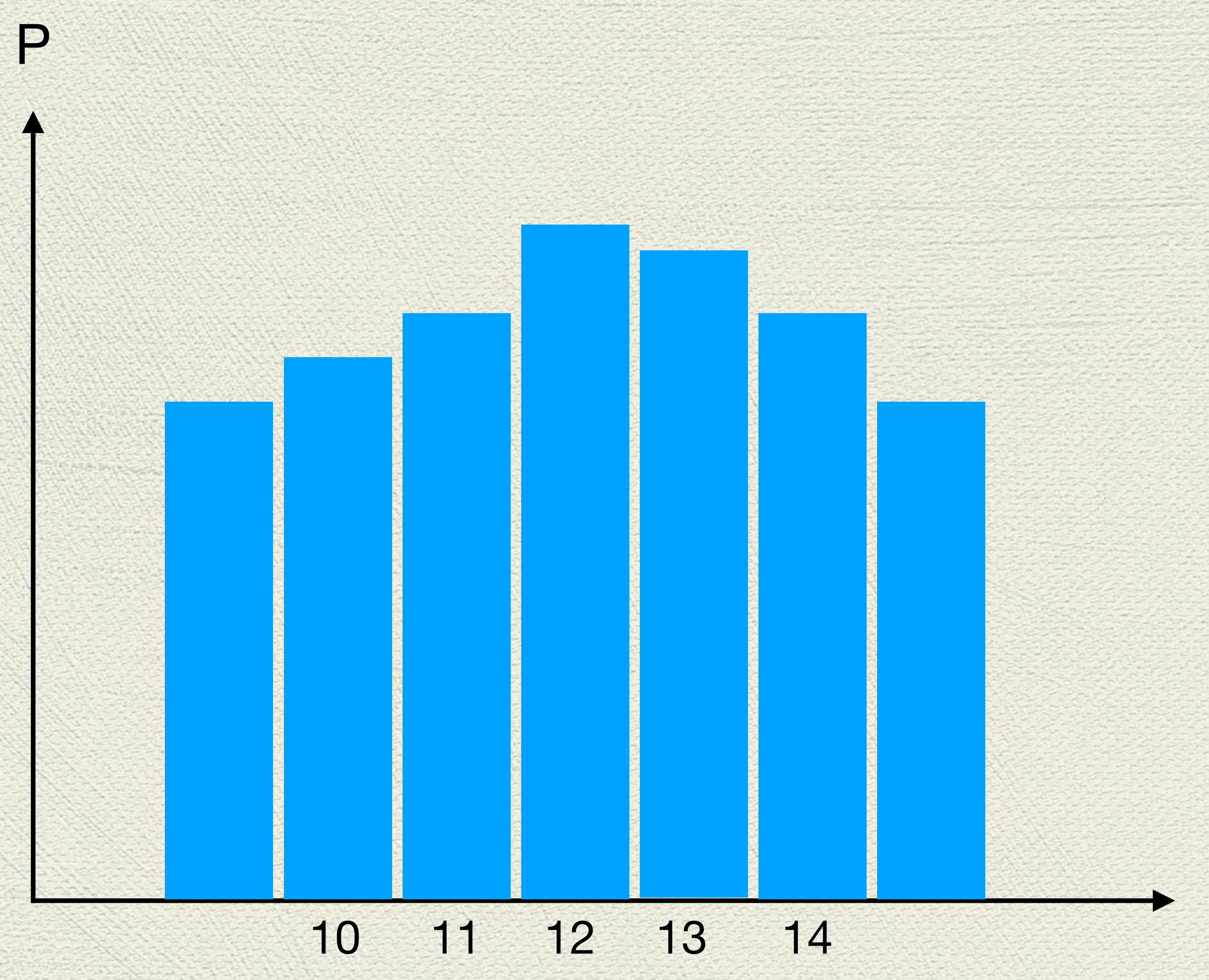
A A A A A A A A B B B A A B B B A

N=18

منبع اغلب اوقات ولی نه همیشه رشته های متعارف تولید می کند.

$$\mathscr{P}_1 = \sum_{x=10}^{14} (\frac{2}{3})^x (\frac{1}{3})^{18-x} \binom{18}{x} = 0.7907$$

تساهل و تسامح و بیشتر! دو انحراف معیار را هم قبول می کنیم



A B A B B A B B A A A B B A A B A

B B A B B A B A B B A B B A A B A

A B A B A A A B B B A A A B A B A

A A A A A A A A B B B A A B B B A

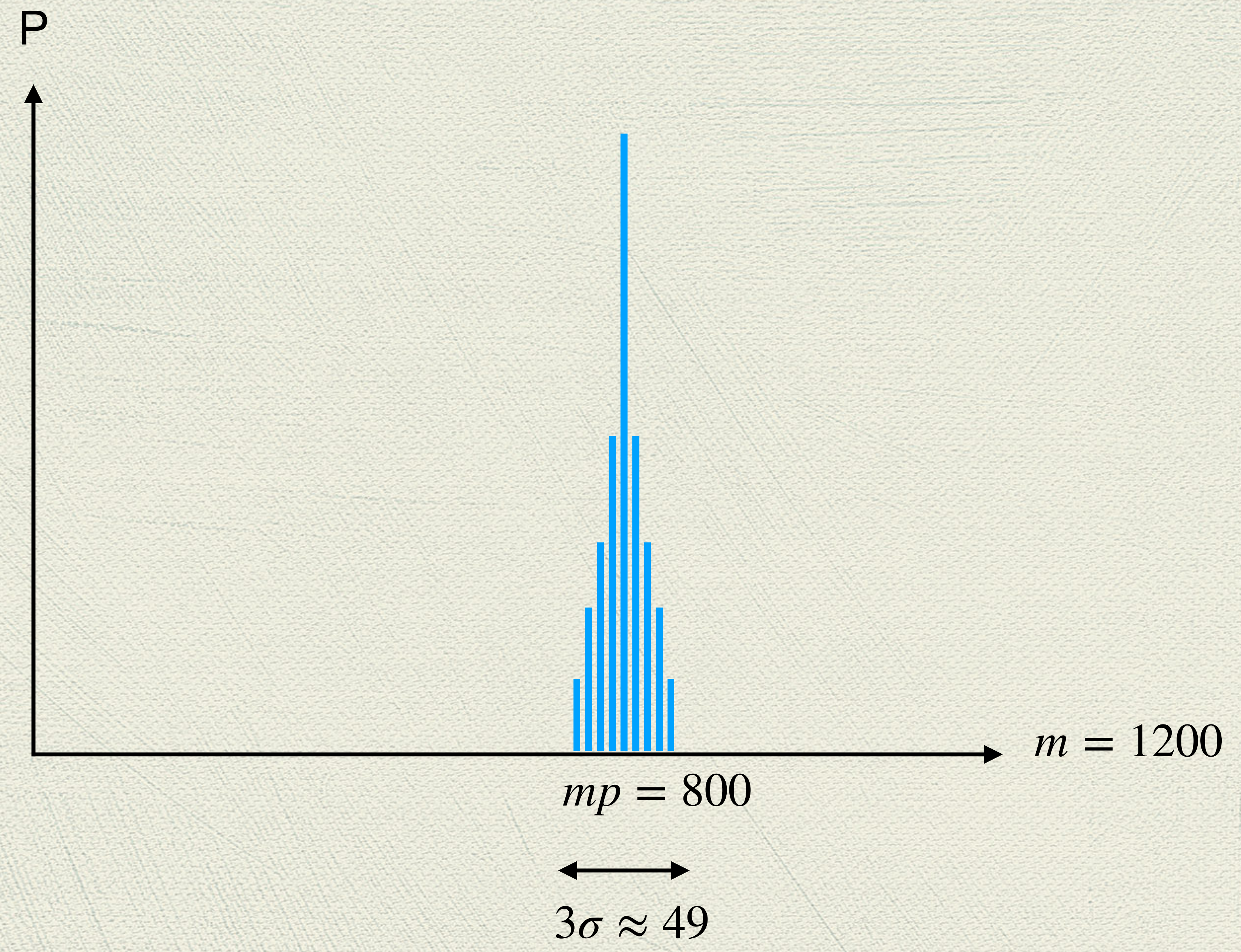
N=18

تقریبا منبع همیشه رشته های متعارف تولید می کند.

$$\mathscr{P}_2 = \sum_{x=8}^{16} (\frac{2}{3})^x (\frac{1}{3})^{18-x} \binom{18}{x} = 0.9788$$

ولی اگر بخواهیم احتمال تولید رشته های نمونه را به ۱ برسانیم، مجبوریم که همه رشته ها را قبول کنیم!

$$\sigma = \sqrt{m(1-p)p}$$



قضیه دوم شانون: ظرفیت کانال های کلاسیک

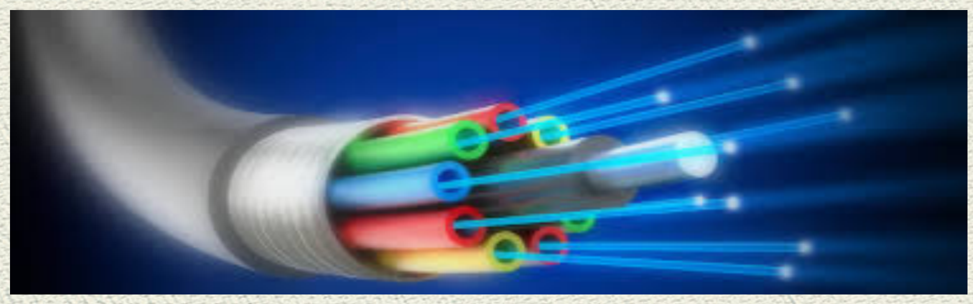
There are always errors

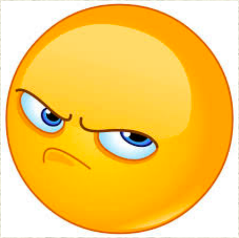
$x \longrightarrow$

$\longrightarrow y$

$$P(x \longrightarrow y)$$

0101000010001





01010010110001

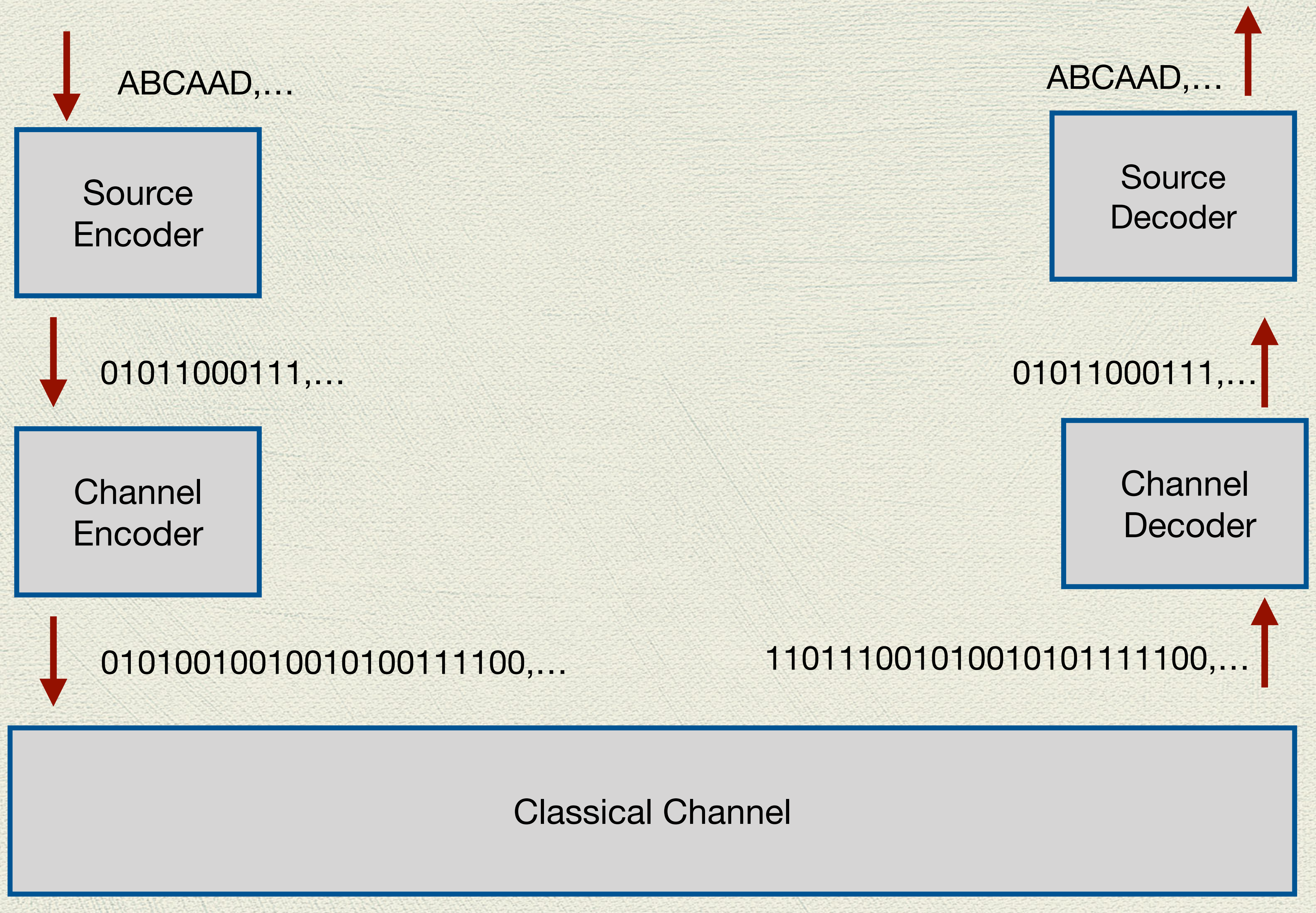
How to correct errors?

0 ⟶ 000

000 ⟶ 0

Encoding

Decoding

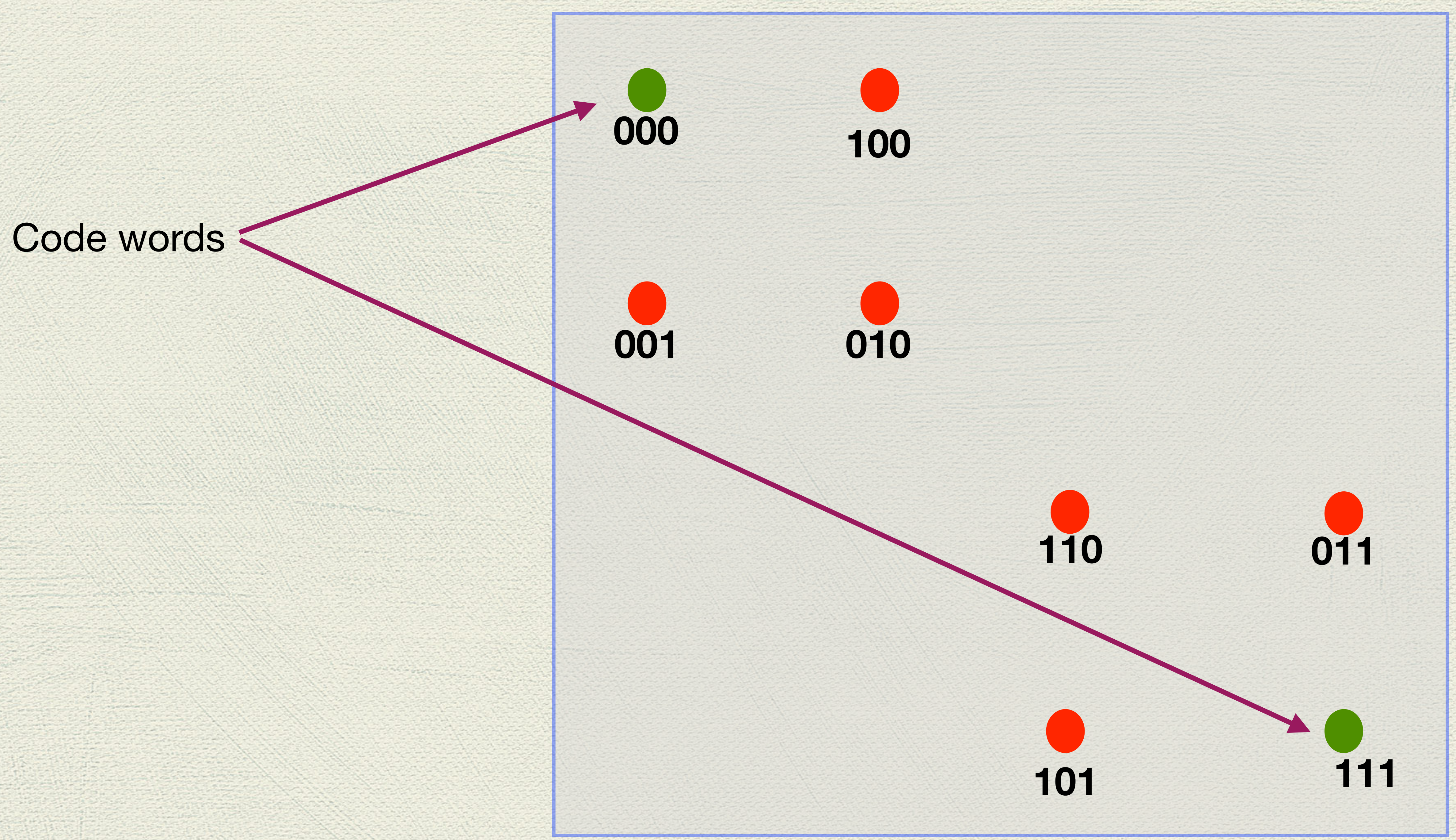

ABCAAD,...
Source Encoder
01011000111,...
Channel Encoder
01010010010010100111100,...
Classical Channel
1101110010100101011111100,...
Channel Decoder
01011000111,...
Source Decoder
ABCAAD,...

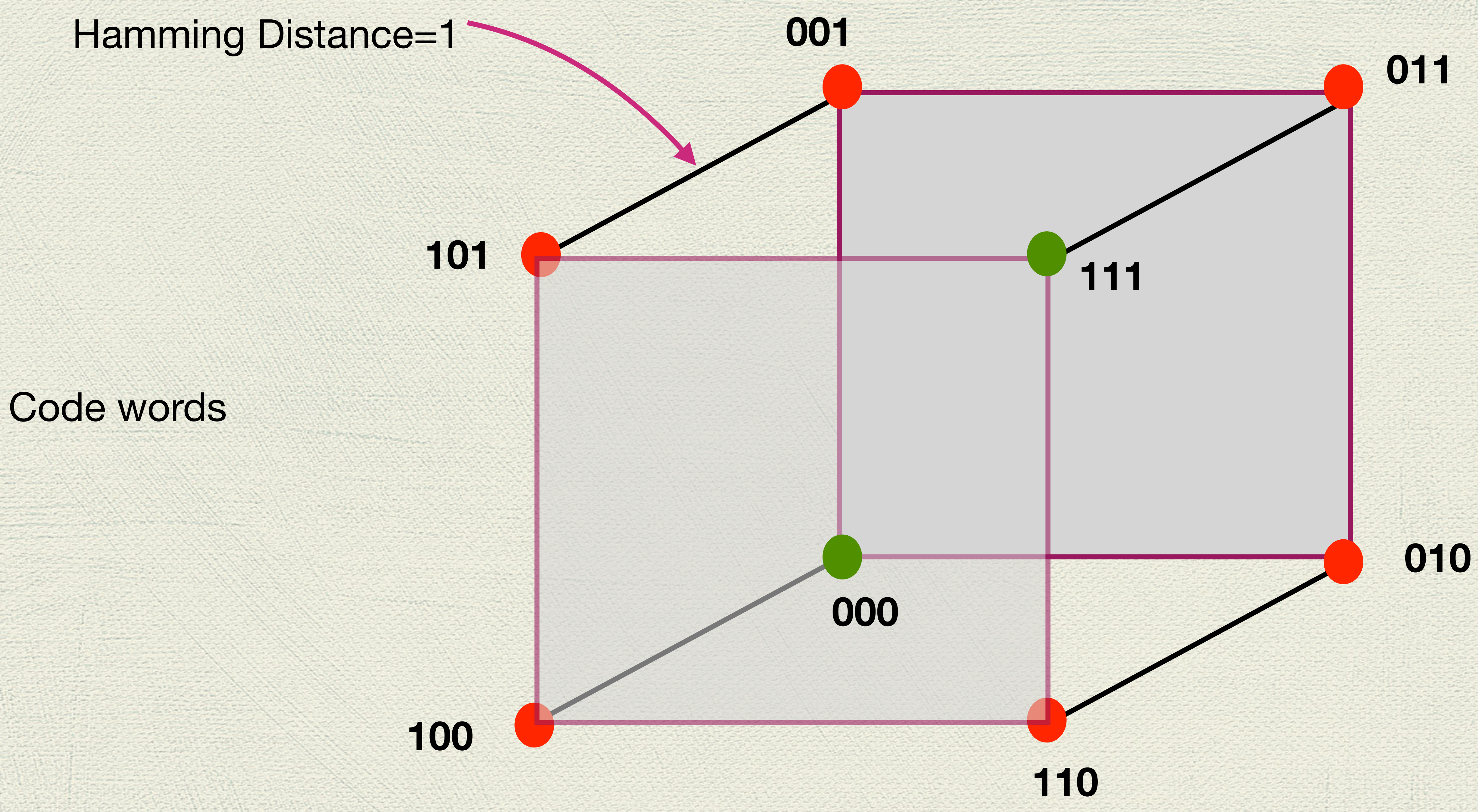
The new error rate

$$P_{error} = 3p^2(1-p) + p^3 \sim 3p^2$$

The Price that we should pay: we reduce the rate.

$$R = \frac{1}{3}$$


Code words
000
100
001
010
110
011
101
111


Hamming Distance=1
Code words
001
011
101
111
000
010
100
110

Lower Probability of Error

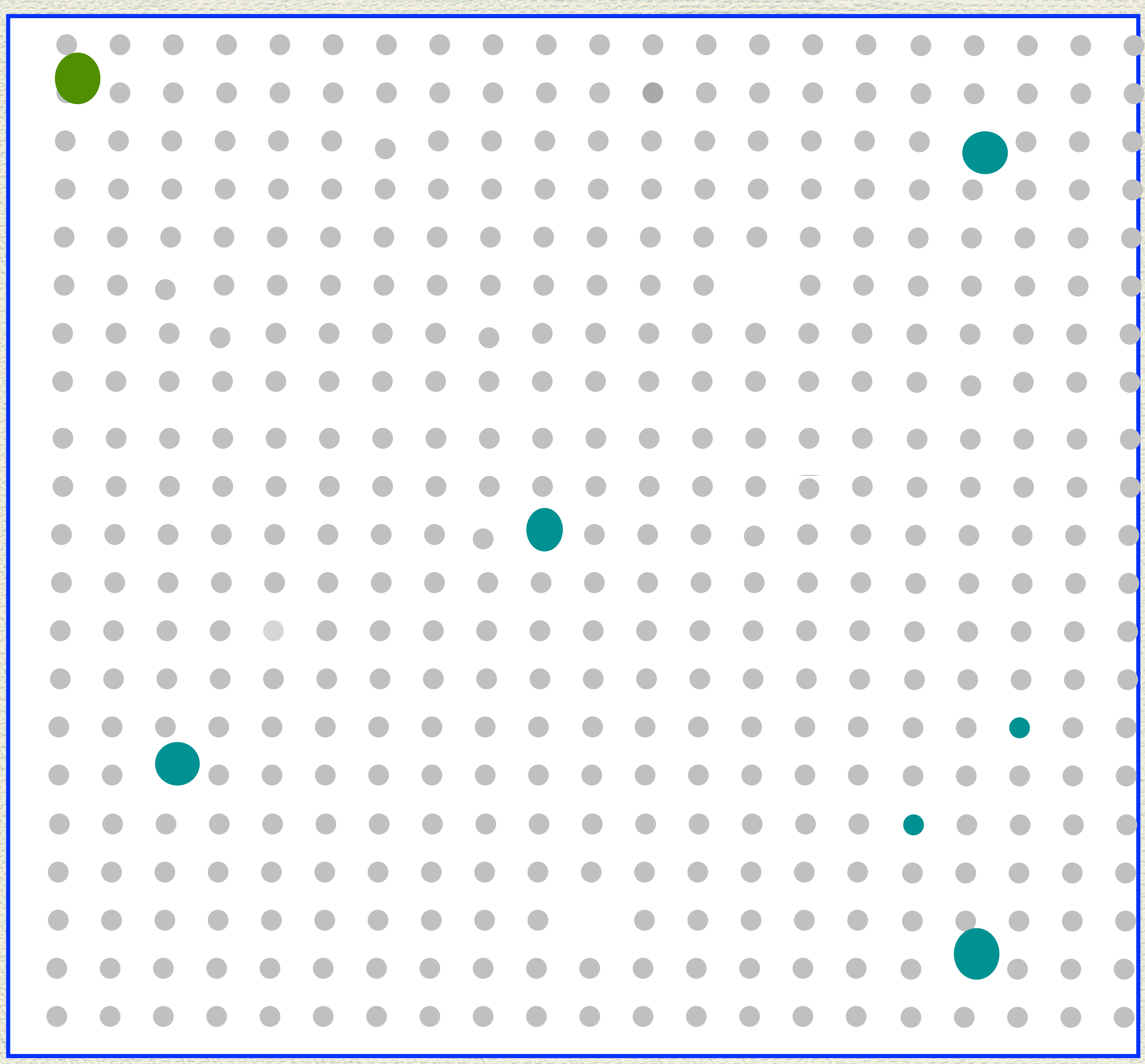
0	$\longrightarrow$	**00000**	00011	00111
1	$\longrightarrow$	**11111**	$P_{error} \sim 10p^3$	

Lower Rate!

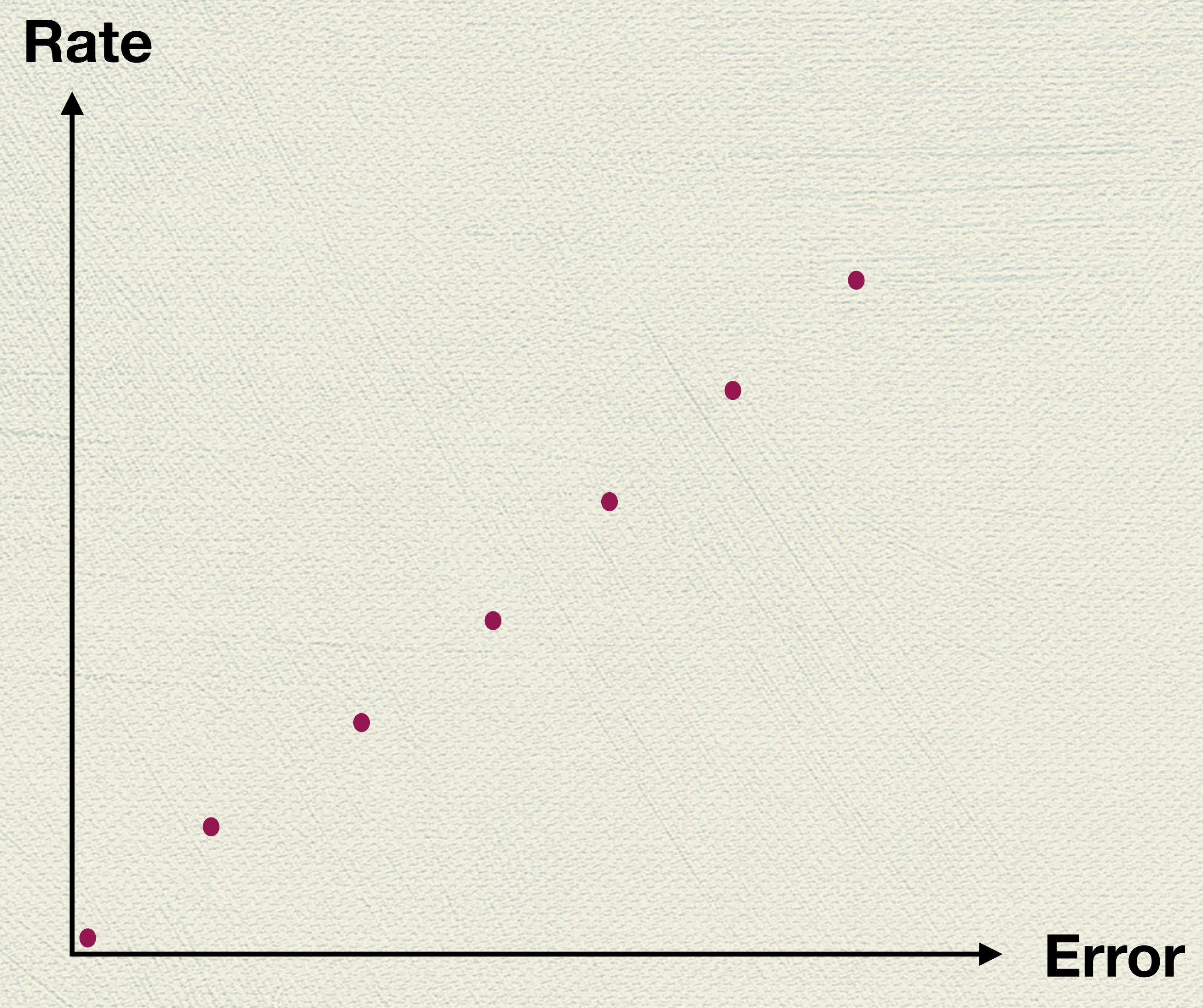
$$R = \frac{1}{5}$$

Probability of Error—>0

Rate—-> 0

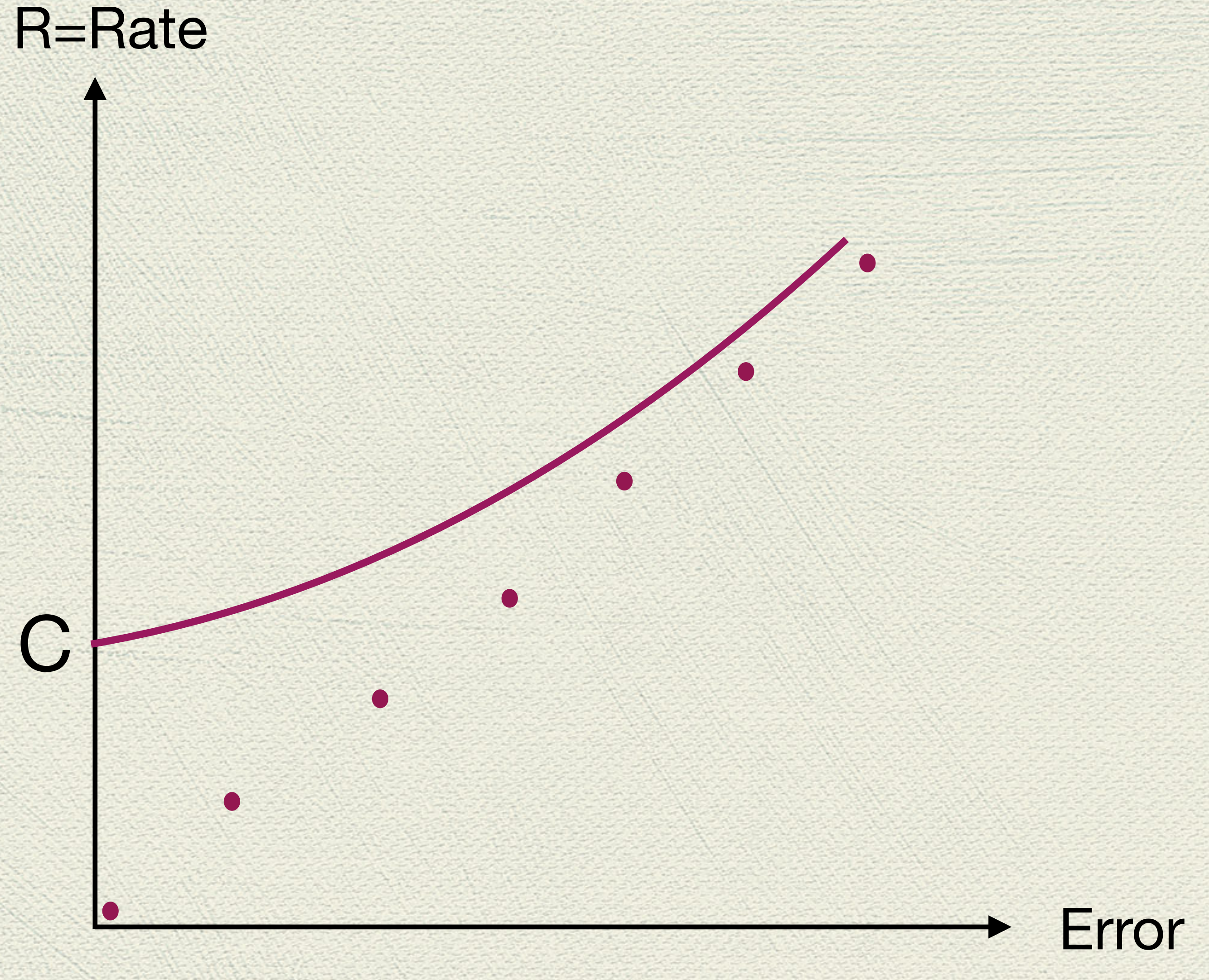


مخابره بدون خطا



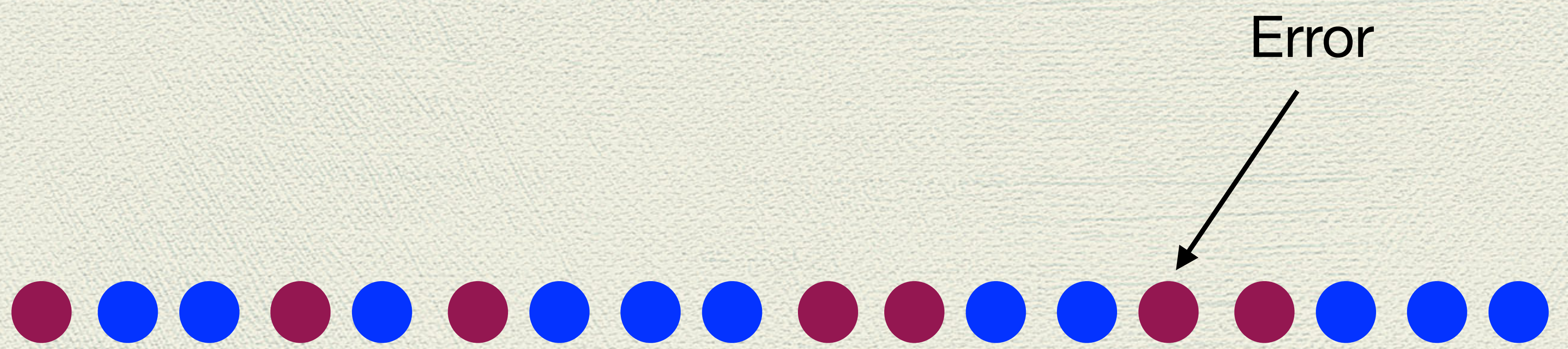
فقط با نرخ بی نهایت کم و در واقع نرخ صفر امکان پذیر است.

مخابره بدون خطا



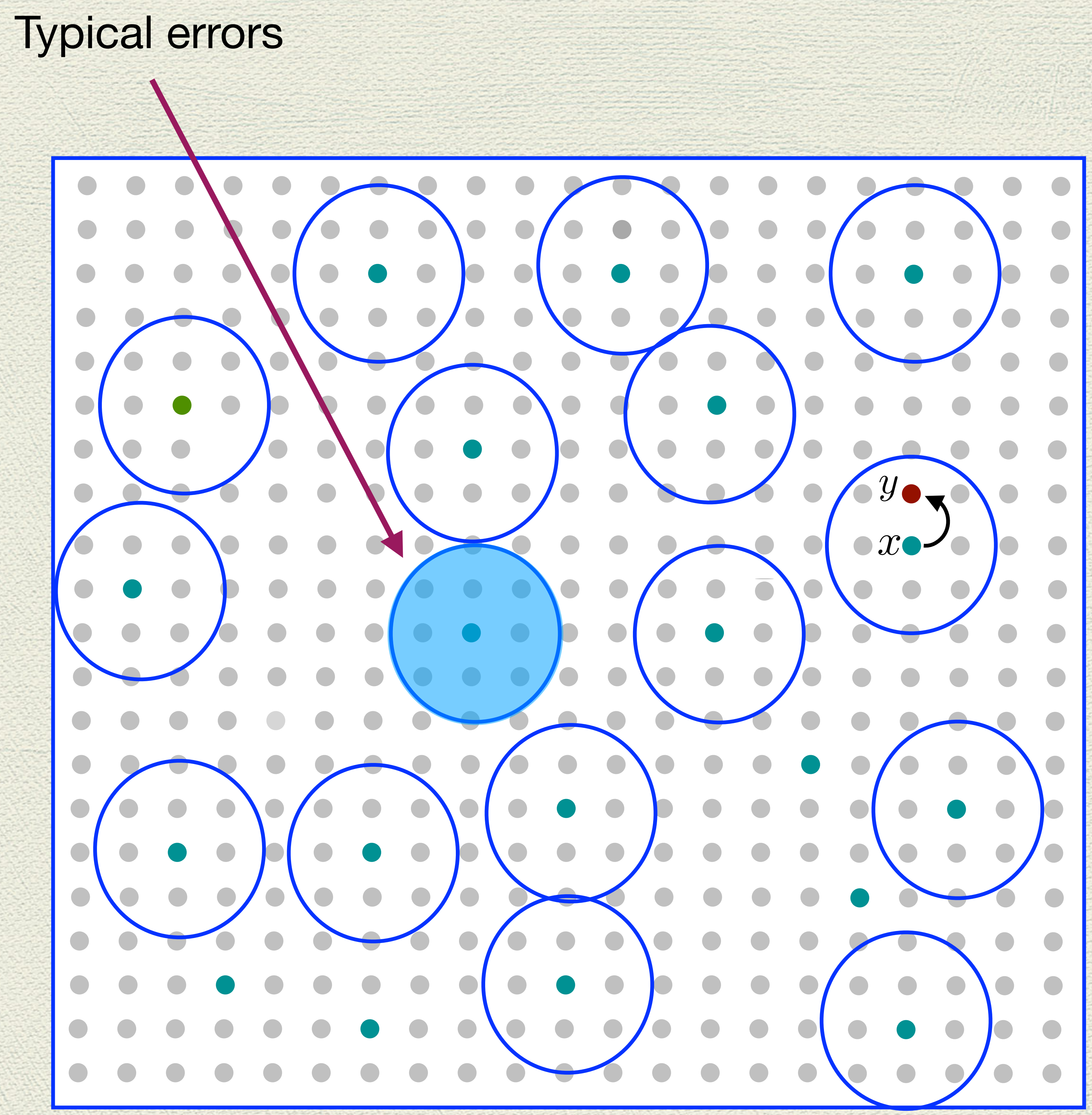
فقط با نرخ بی نهایت کم و در واقع نرخ صفر امکان پذیر است.

Typical Errors رشته خطاهای نمونه



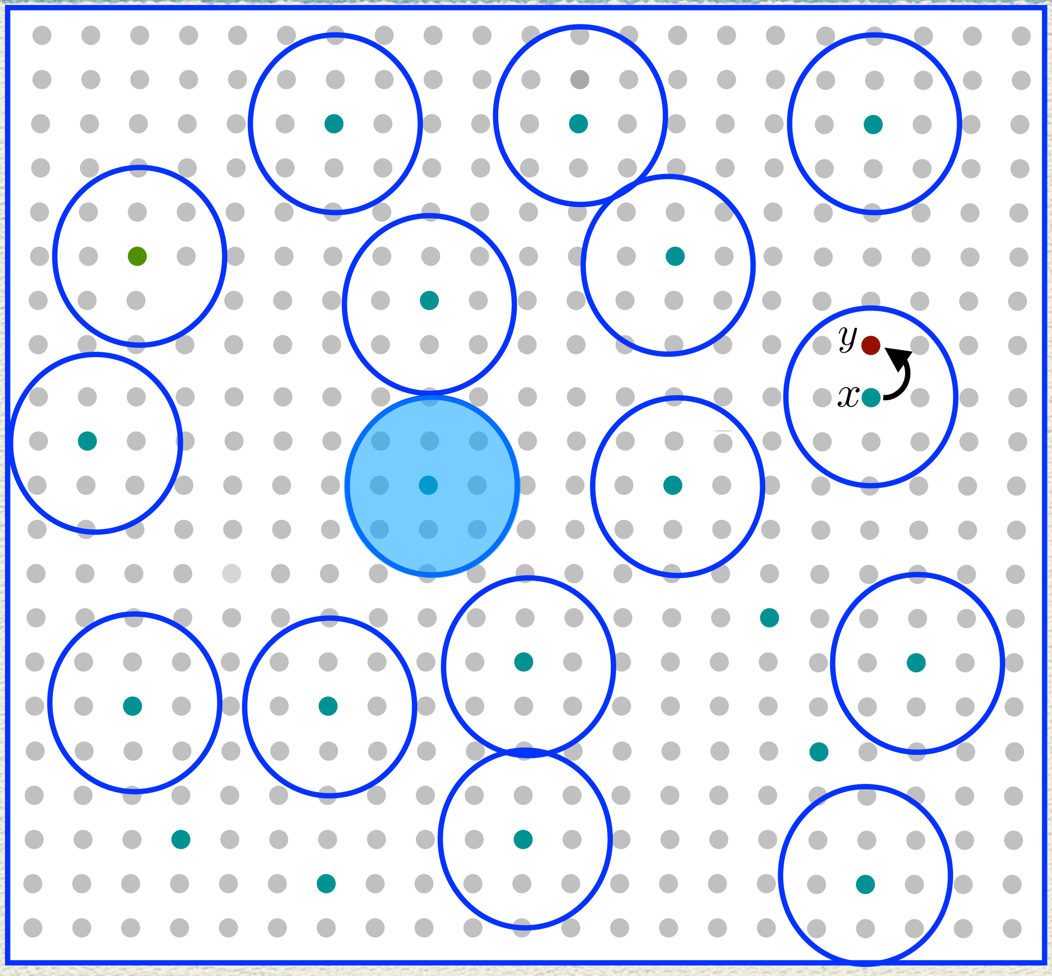
$$m \approx Np$$

$$\binom{N}{m} = \binom{N}{Np} \approx 2^{NH(p)}$$



$$2^k \times 2^{NH(p)} < 2^N$$

Number of code words
=
Number of spheres

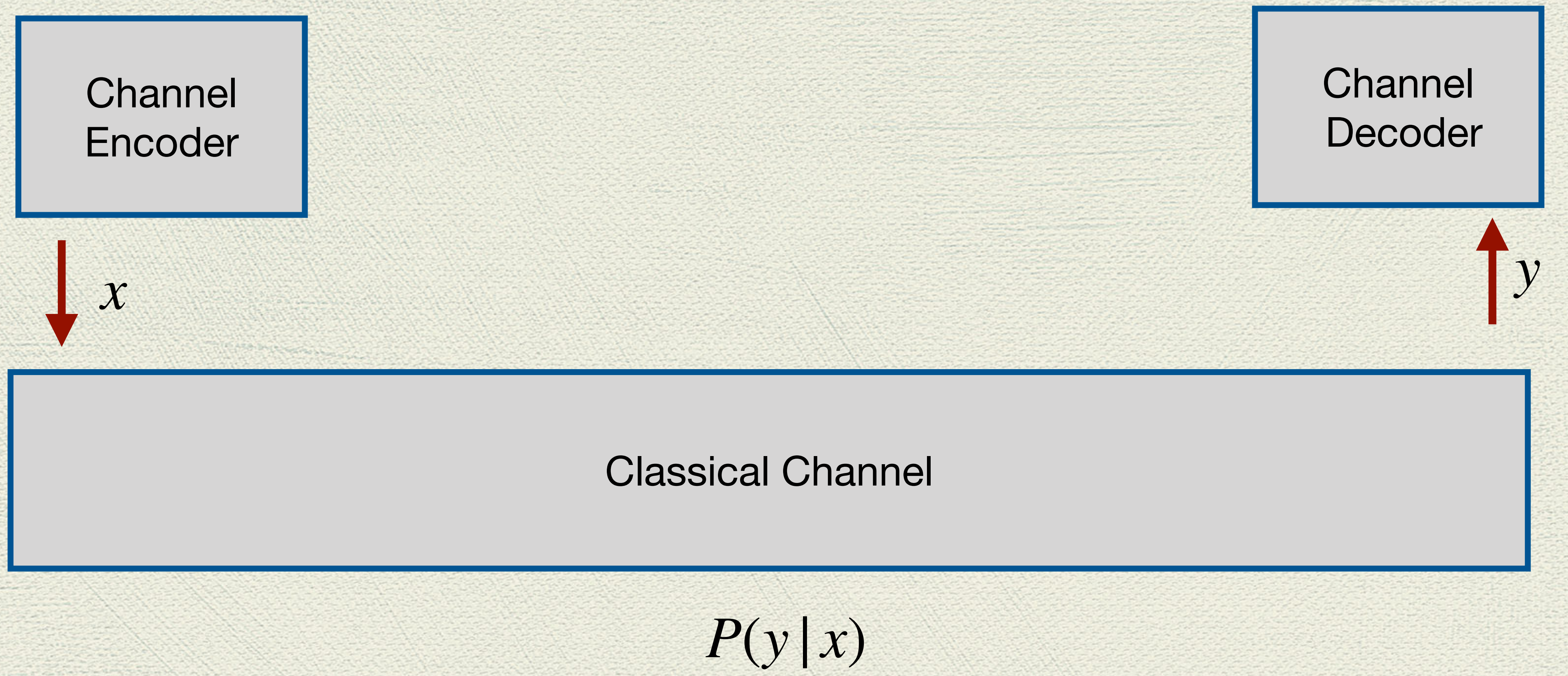


$$2^k \times 2^{NH(p)} < 2^N$$

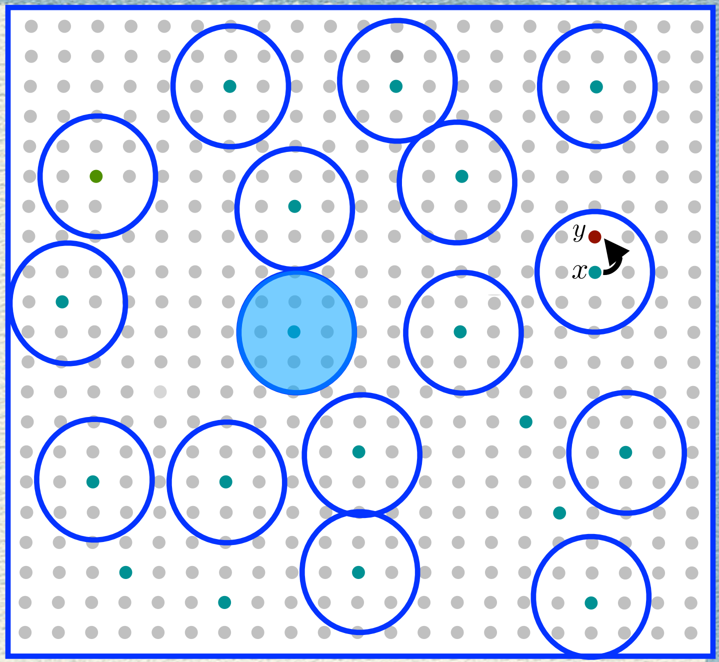
$$k < N(1 - H(p))$$

$$R < 1 - H(p)$$

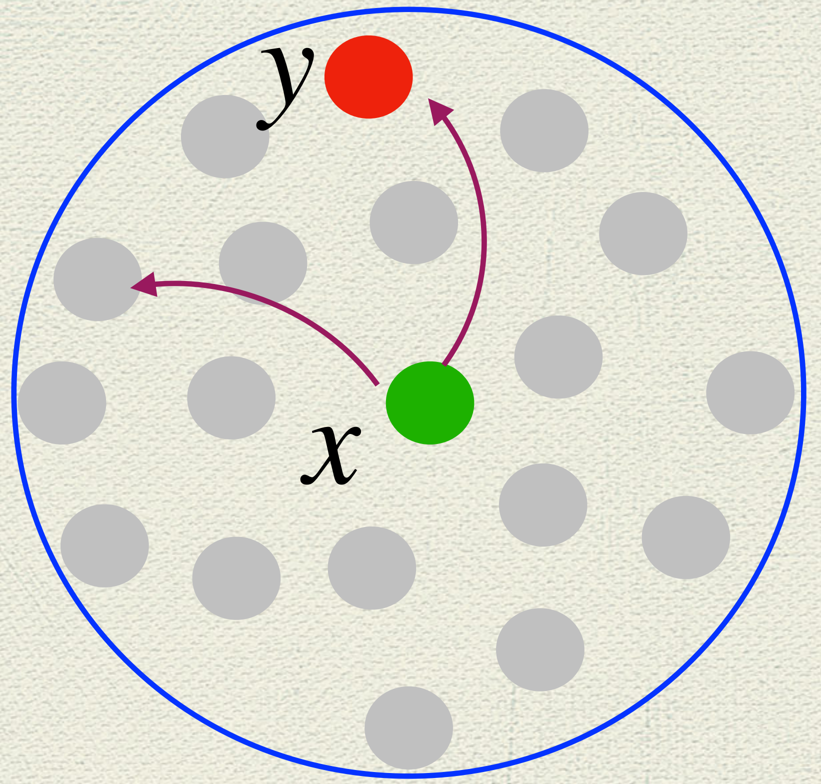
$$C = 1 - H(p)$$


Channel Encoder
x
Channel Decoder
y
Classical Channel
$P(y|x)$

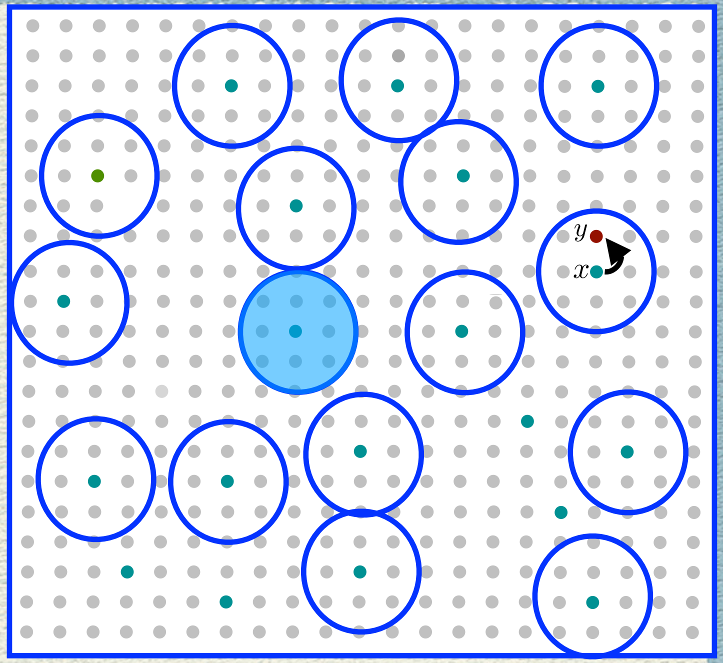
$2^{NH(Y)}$ Total number of possible received words



2^{k} Total number of code-words
=
Total number of spheres



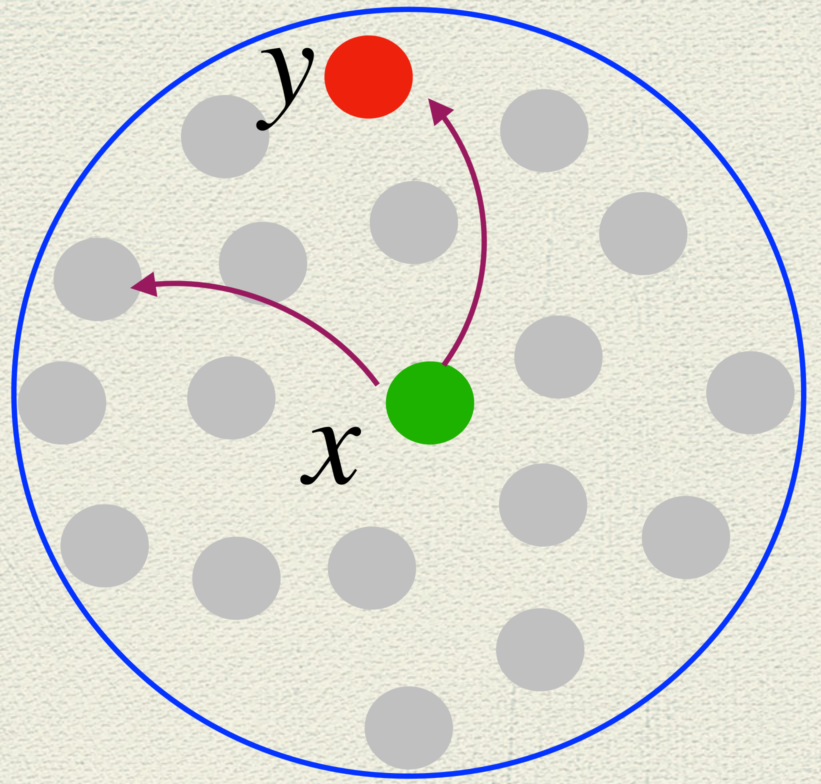
$2^{NH(Y|X)}$ Number of points in a sphere
=
Volume of a sphere



$2^{NH(Y)}$ Total number of possible received words

2^{k} Total number of spheres

$2^{NH(Y|X)}$ Number of points in a sphere



$$2^{k} \times 2^{NH(Y|X)} \leq 2^{NH(Y)}$$

$$\frac{k}{N} \leq H(Y) - H(Y\,|\,X)$$

$$R \leq I(X:Y)$$

$$C = Max \quad I(X:Y)$$

End of part I