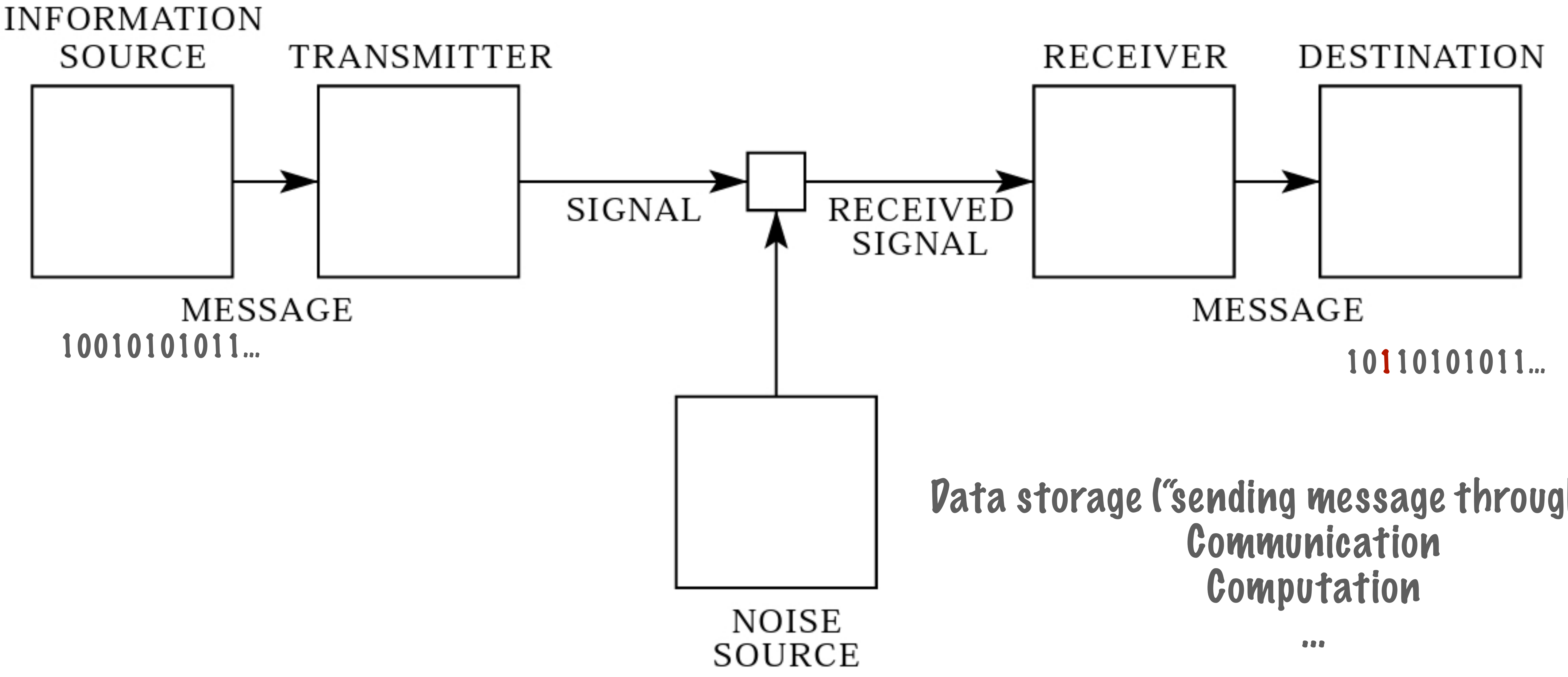
Introduction to Classical Error Correction

Wanqiang Liu, June 20, 2023. wanqiangl@uchicago.edu

Fermilab

THE UNIVERSITY OF CHICAGO

Errors in Classical information



Data storage ("sending message through time")
Communication
Computation
...

Fig. 1 — Schematic diagram of a general communication system.

Warm up: Parity Check Code

- Rule: The coded message always has even number of 1 in a piece string 010…
- If I spot odd number of 1, there must be some error.
- Example: Use 9 bits to encode a piece of information
  - 000 000 000 ✓
  - 001 000 000 ✗
- Q1: How many different code words (“CORRECT MESSAGE”) are allowed? (2^8)
- Q2: How many “flips” at least can make an undetectable error? (2)

[[n, k, d]] code

- n= # of bits used in “a piece of information”
- k= # of logical bits = log_2 (# of code words)
- d= code distance = minimum # of flips to make an undetectable error

The parity check example is [[9, 8, 2]]

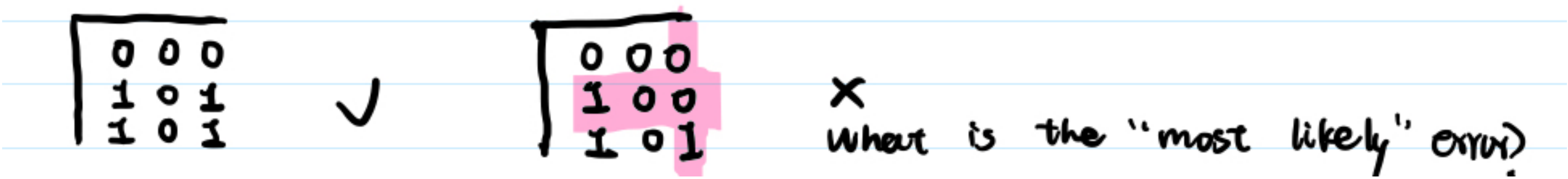
- Q3: If I detect an error in the parity check code, can I know how to correct it?
- No

Make the code correctable!

Rectangular parity check

Rule: If you receive a 9-bit message, stack it into a 3*3 square.

Each column and each row must have even number of 1.



Q1: [[n, k, d]]=? (Hint: how to go from one codeword to another?) [[9,4,4]]

Q2: How many bit flips can make an UNCORRECTABLE error? 2

Another example: Majority Vote

- 111 111 111 ✓
- 000 000 000 ✓
- Anything else is not a codeword, and should be corrected according to the majority. Example: 111 000 111 -> 111 111 111

Q: [[n, k, d]]=?

[[9,1,9]]

What makes a good code?

- Higher k/n: efficient use of bits
- Higher d: can detect more errors
- Efficient decoding
- Examples
  - Hamming code
  - Linear codes…

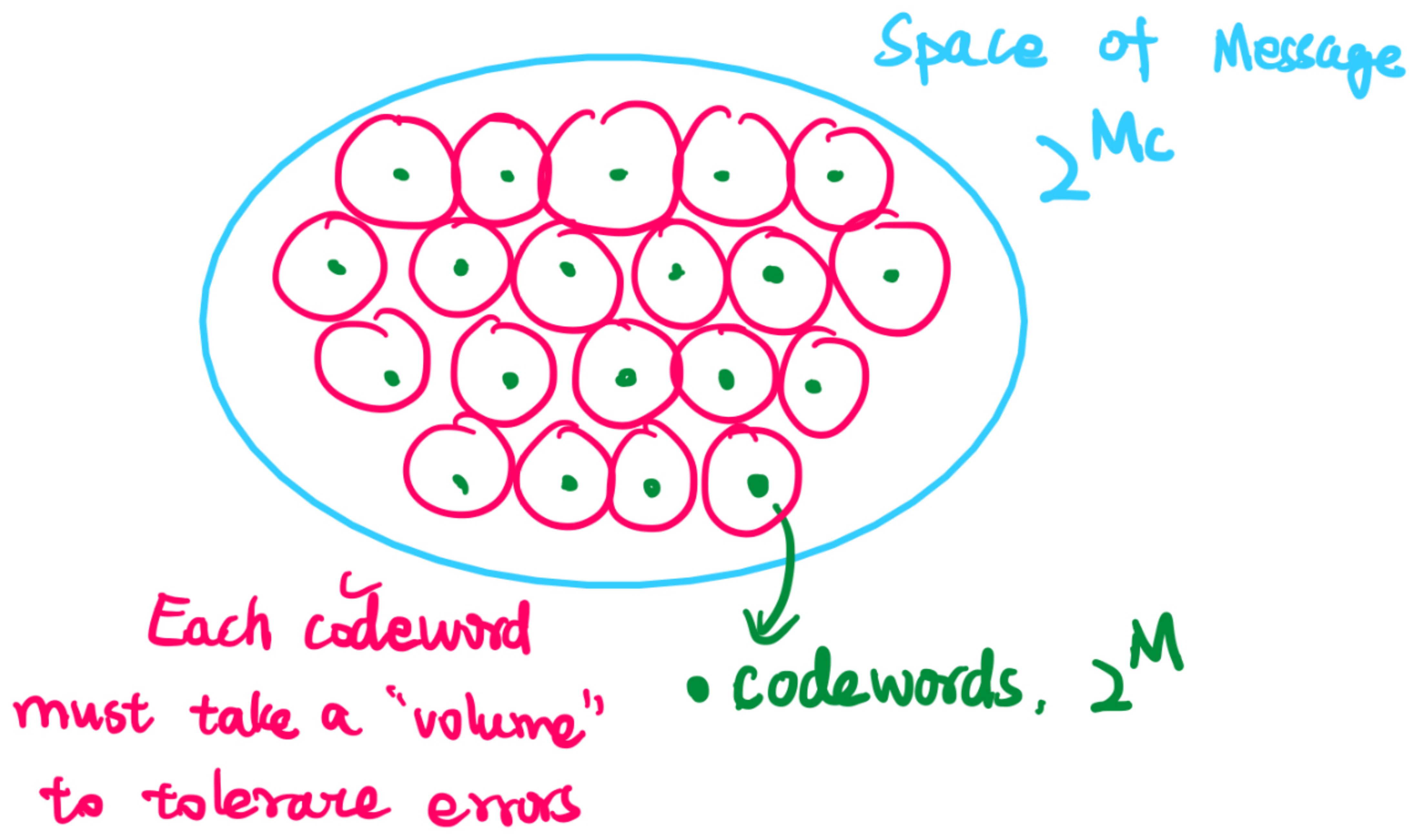
Tradeoff: if the message is "tightly packed" with codewords (high k/n), it's less likely to have room for errors (lower d)

What is the capacity limit? — Classical information theory

Shannon theorem

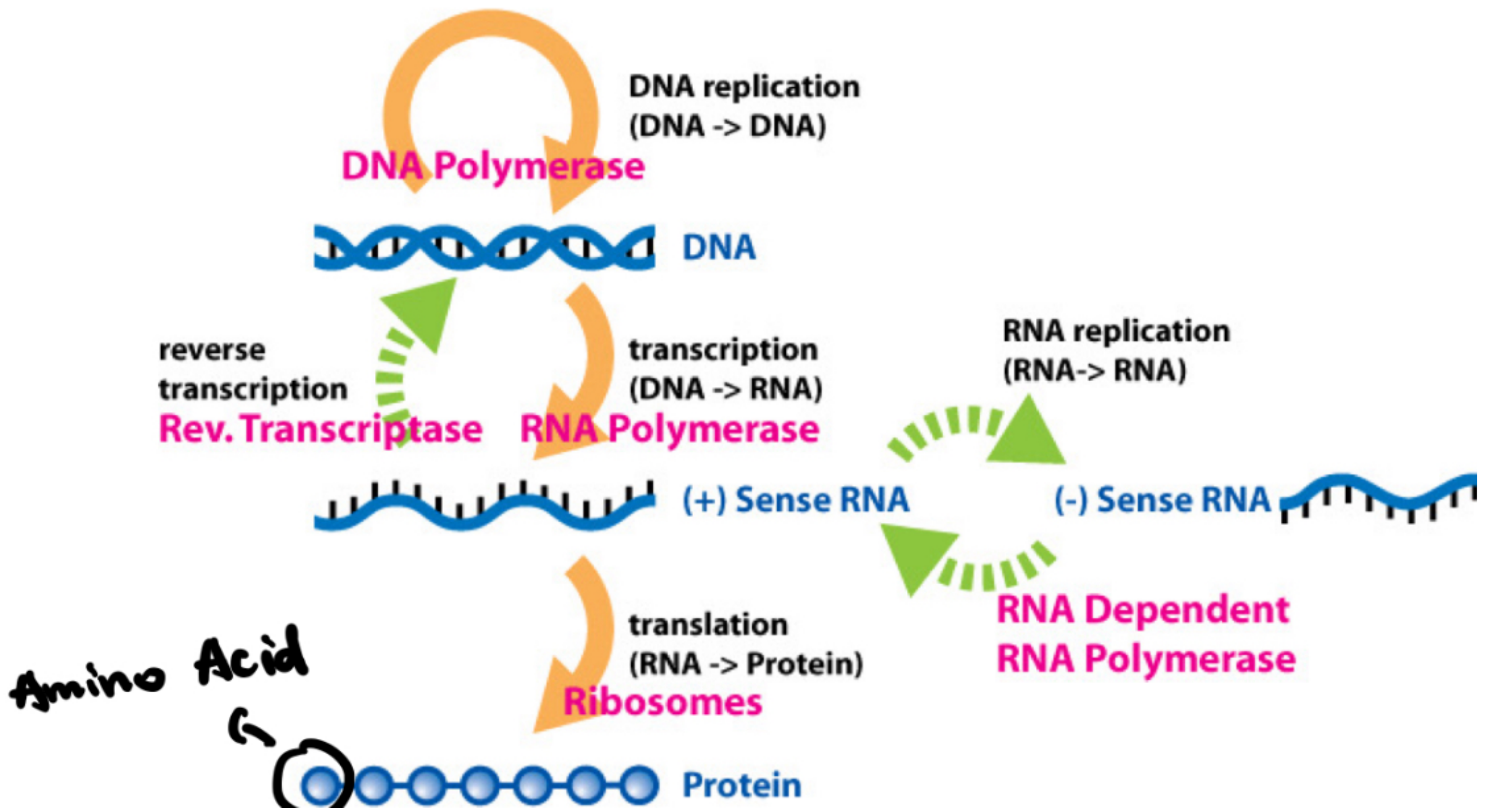
- Length of the coded message to send: M_C
- Length of the coded stat message: M
- For any bit in M_C, the probability to flip is q

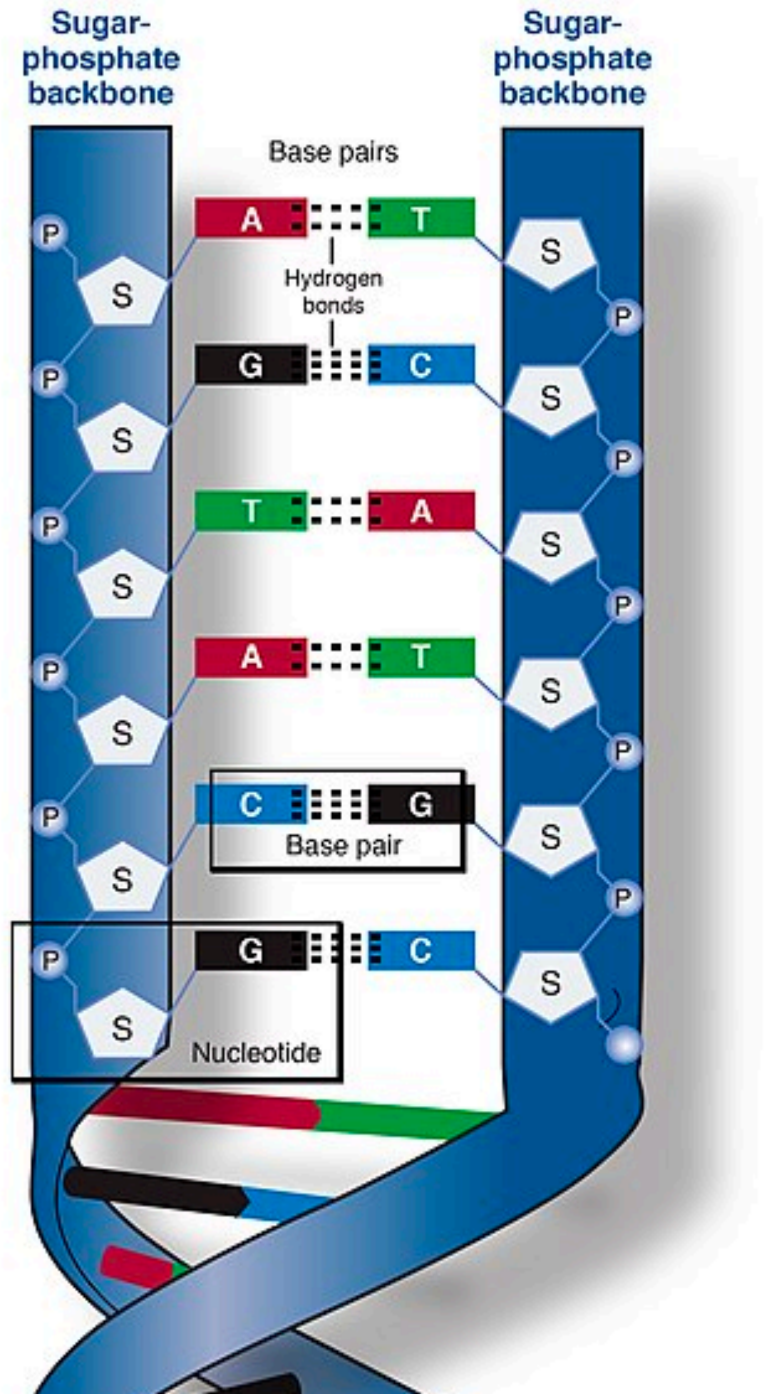
$$\frac{M}{M_C} \leq 1 - q[\log_2(\frac{1}{q}) + (1-q)\log_2(\frac{1}{1-q})]$$

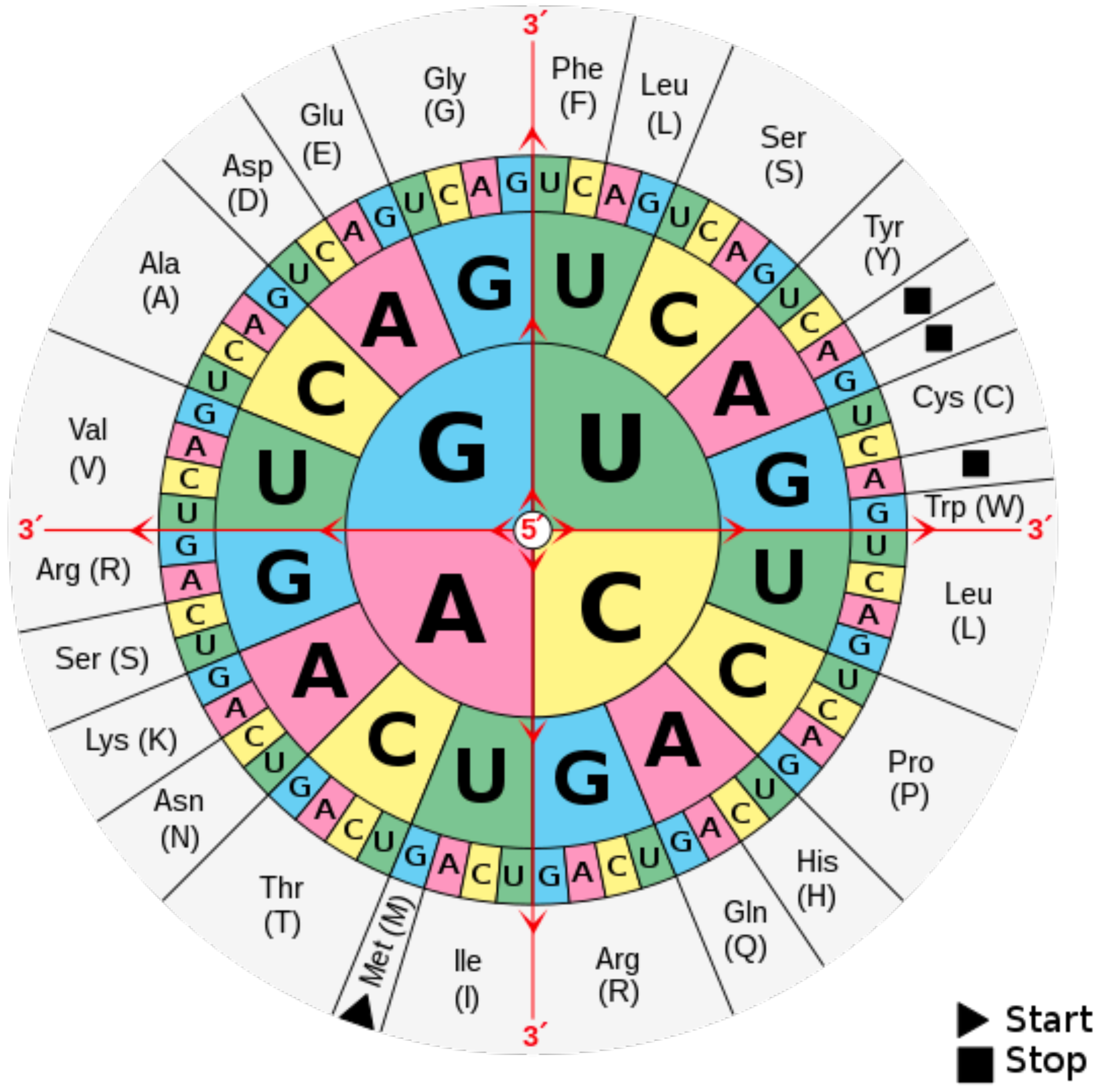


Feynman Lectures on Computation, Chapter 4

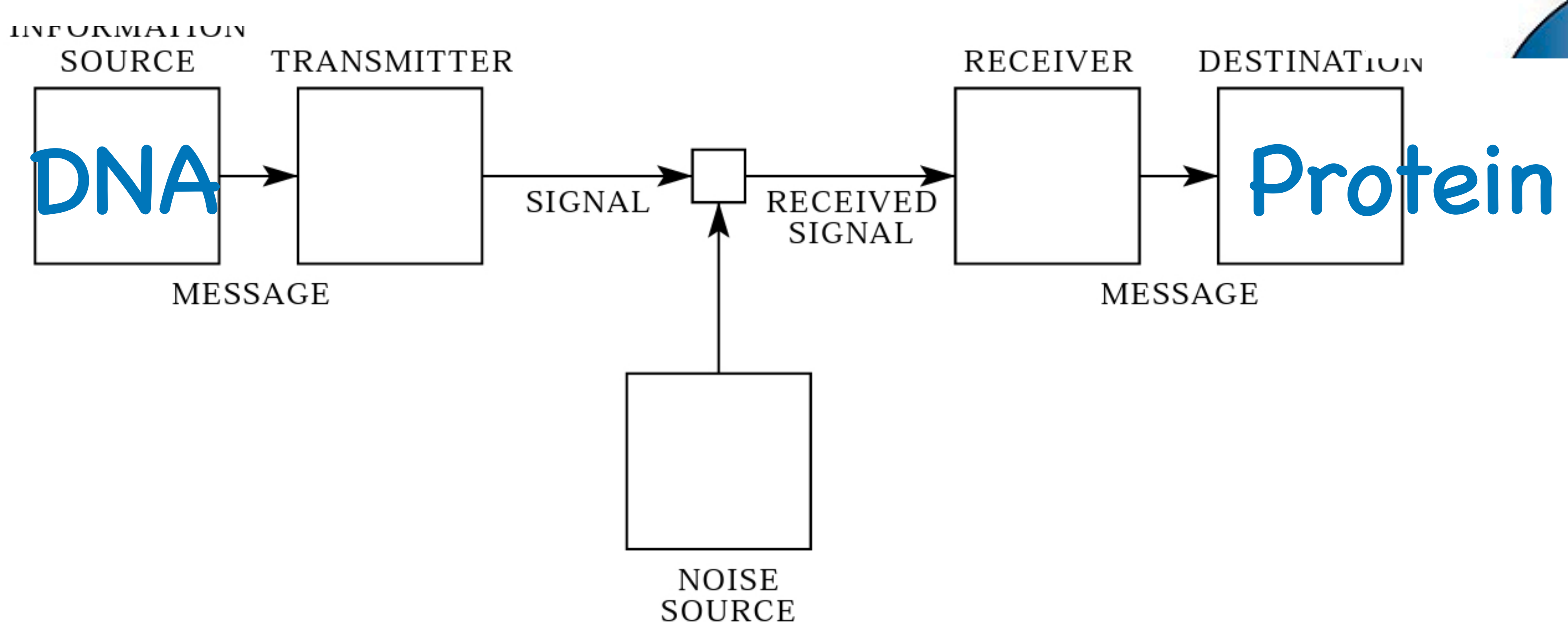
Error correction in nature – DNA!







RNA codon table



4^3 DNA/RNA codons for 20 amino-acids +1 stop codon

Redundancy – room to tolerate errors!

Q: Using Shannon's theorem, estimate what error rate q is tolerated in the DNA-to-protein process

Hint (you can disagree with this)

- Shannon: $\frac{M}{M_C} \leq 1 - q[\log_2(\frac{1}{q}) + (1-q)\log_2(\frac{1}{1-q})]$
  - LHS: $\frac{M}{M_C} \sim \frac{\log_4(21)}{3}$
  - RHS: $\log_2 \to \log 4$, because there are 4 base pairs.
- Look up the true error rates in nature. How does the estimate compare?