KO codes

Ashok Vardhan Makkuva (UIUC)

Joint work with Xiyang Liu, Mohammad Vahid Jamali,
Hessam Mahdavifar, Sewoong Oh, Pramod Viswanath

Outline

- Motivation

Outline

- Motivation
- Learning codes

Outline

- Motivation
- Learning codes
- KO codes: novel neural codes
  - KO codes, *ICML 2021*

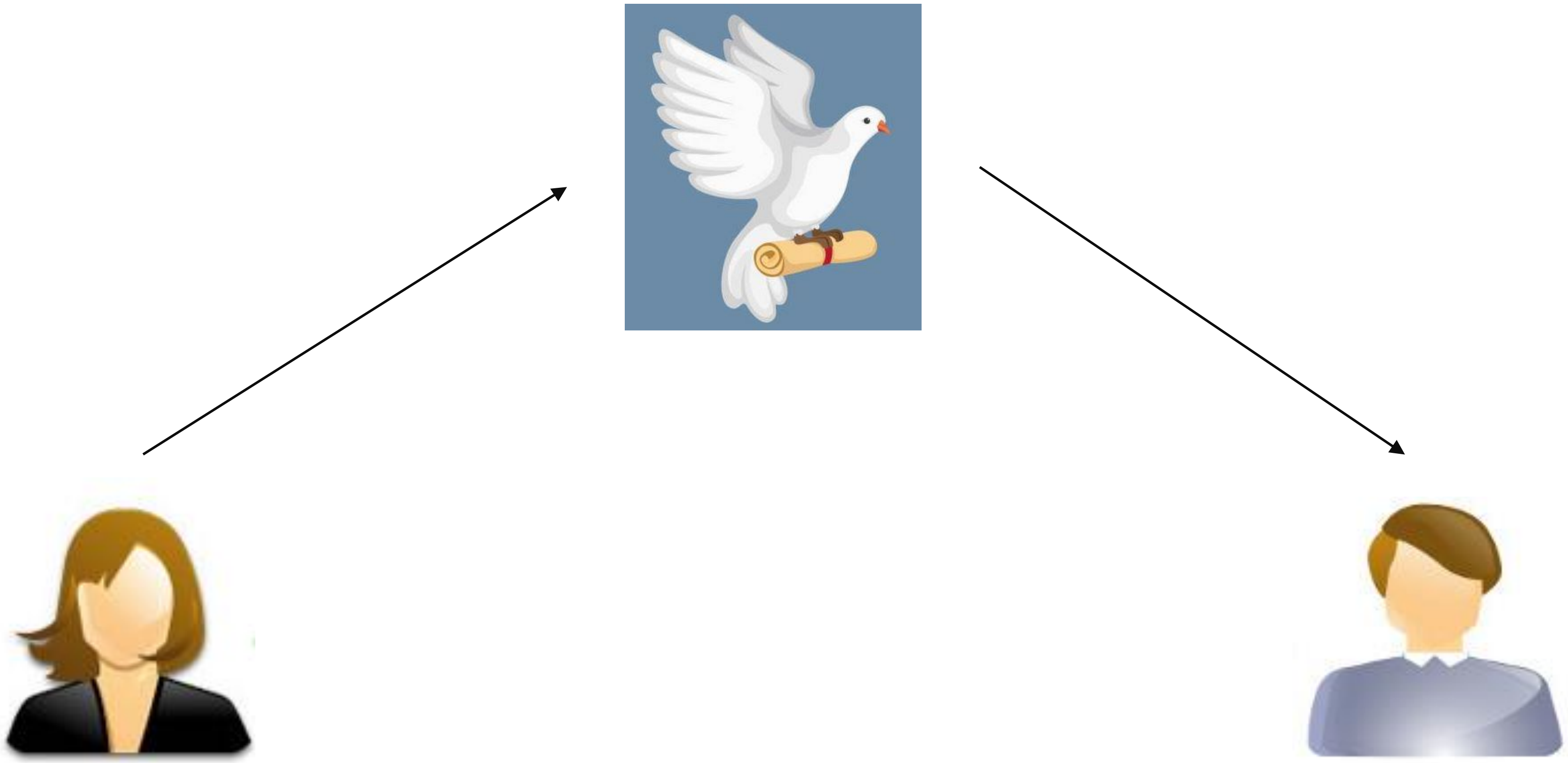
Outline

- Motivation
- Learning codes
- KO codes: novel neural codes
  - KO codes, *ICML 2021*
- Future directions

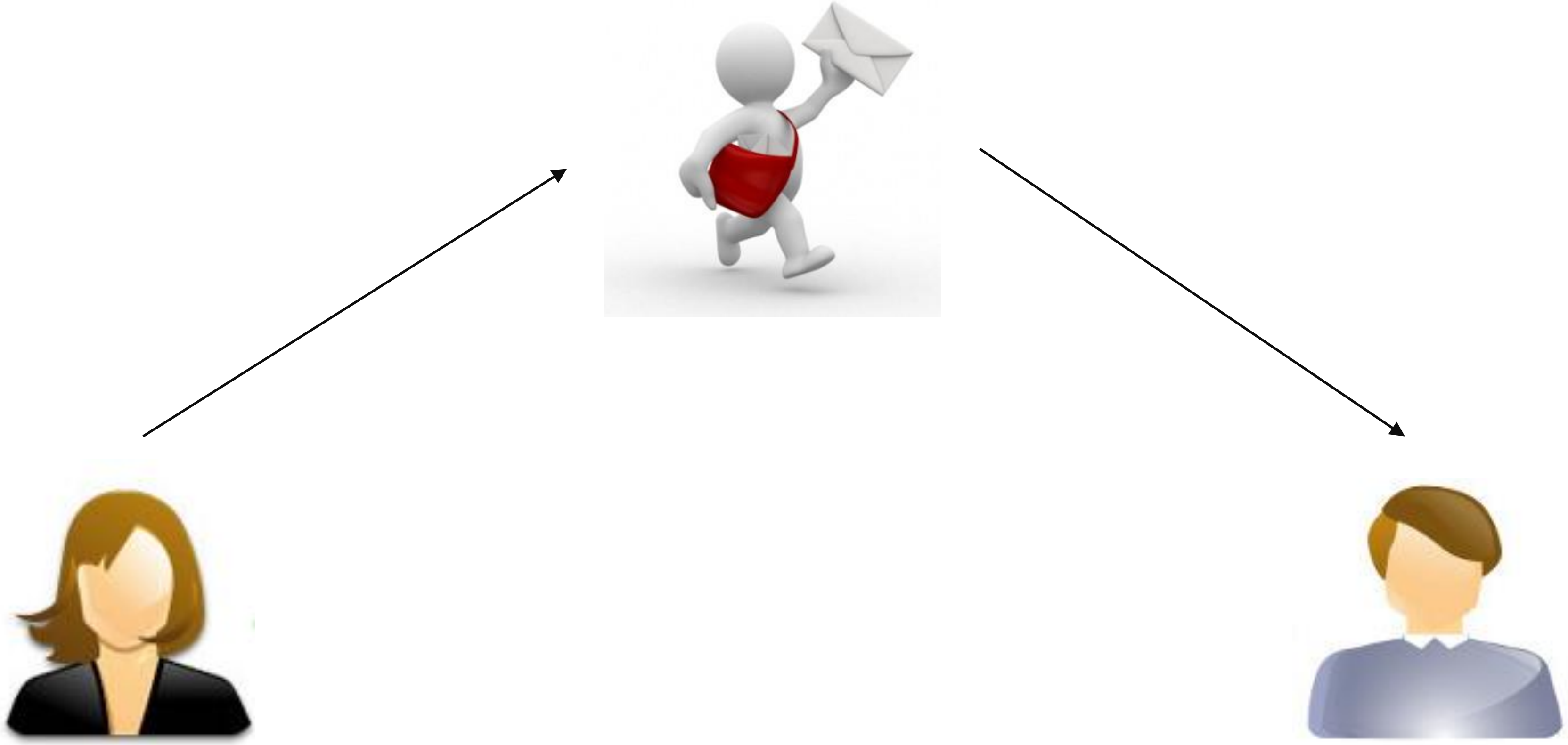
Age of Information

Once upon a time…

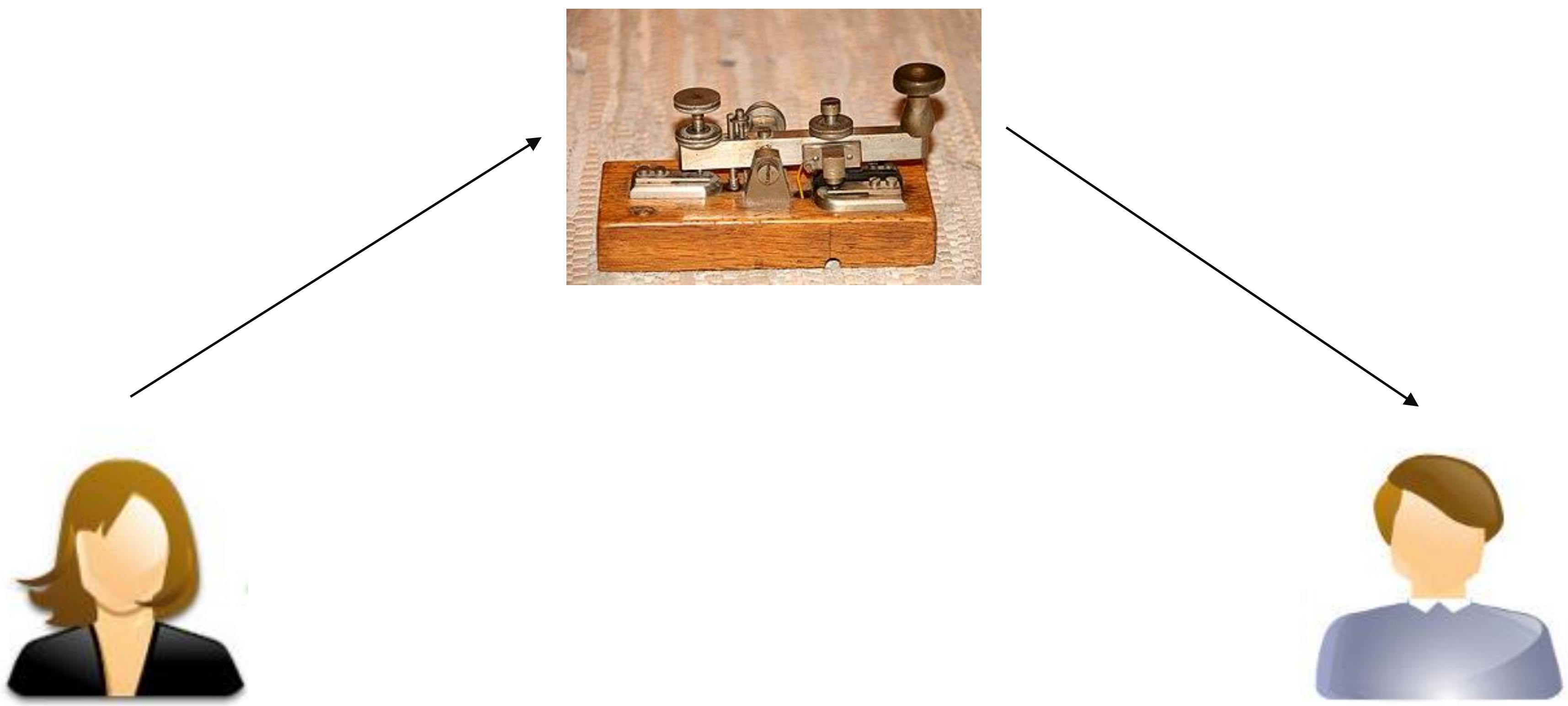
3000 BC



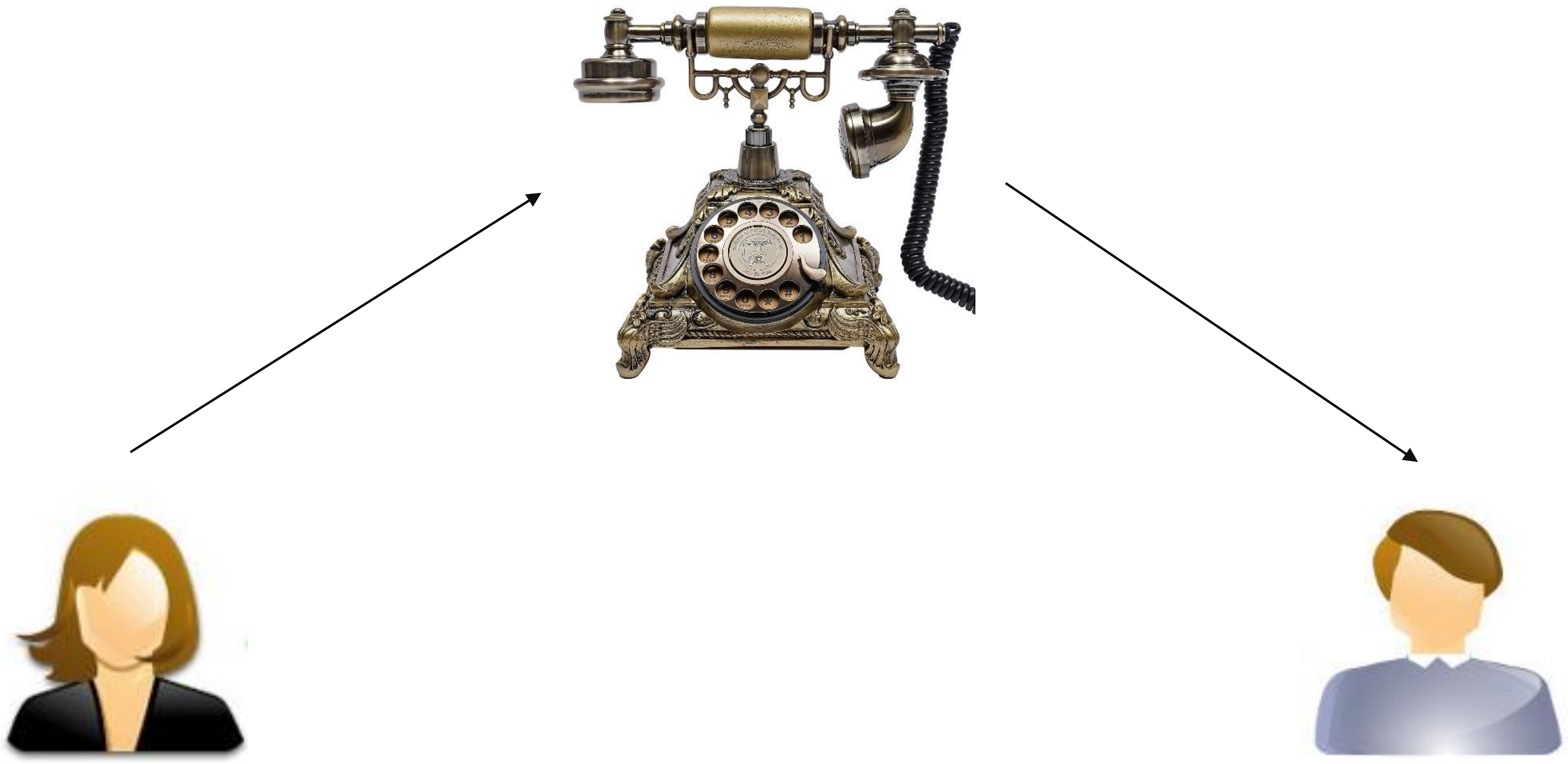
500 BC



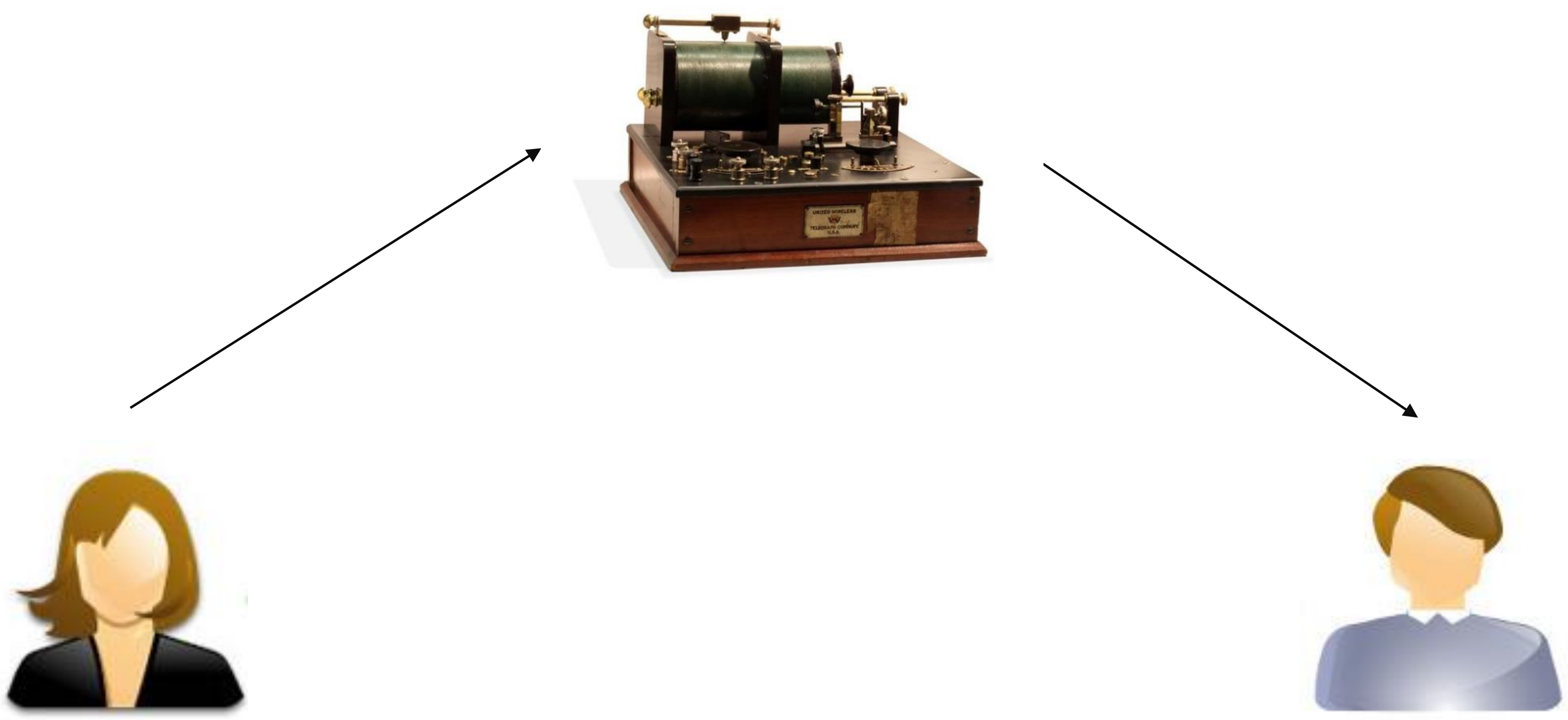
Early 19th century



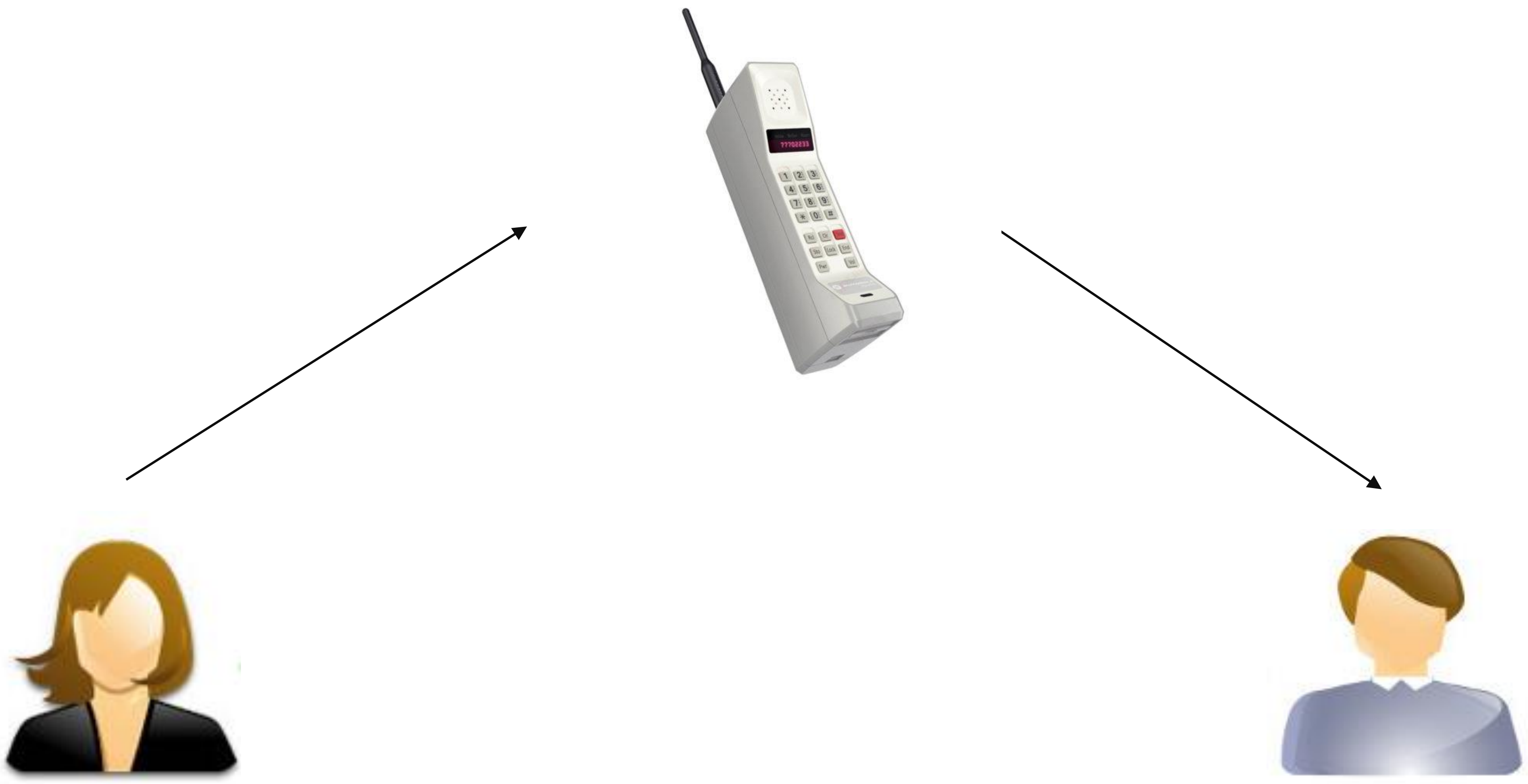
Late 19th century



Early 20th century

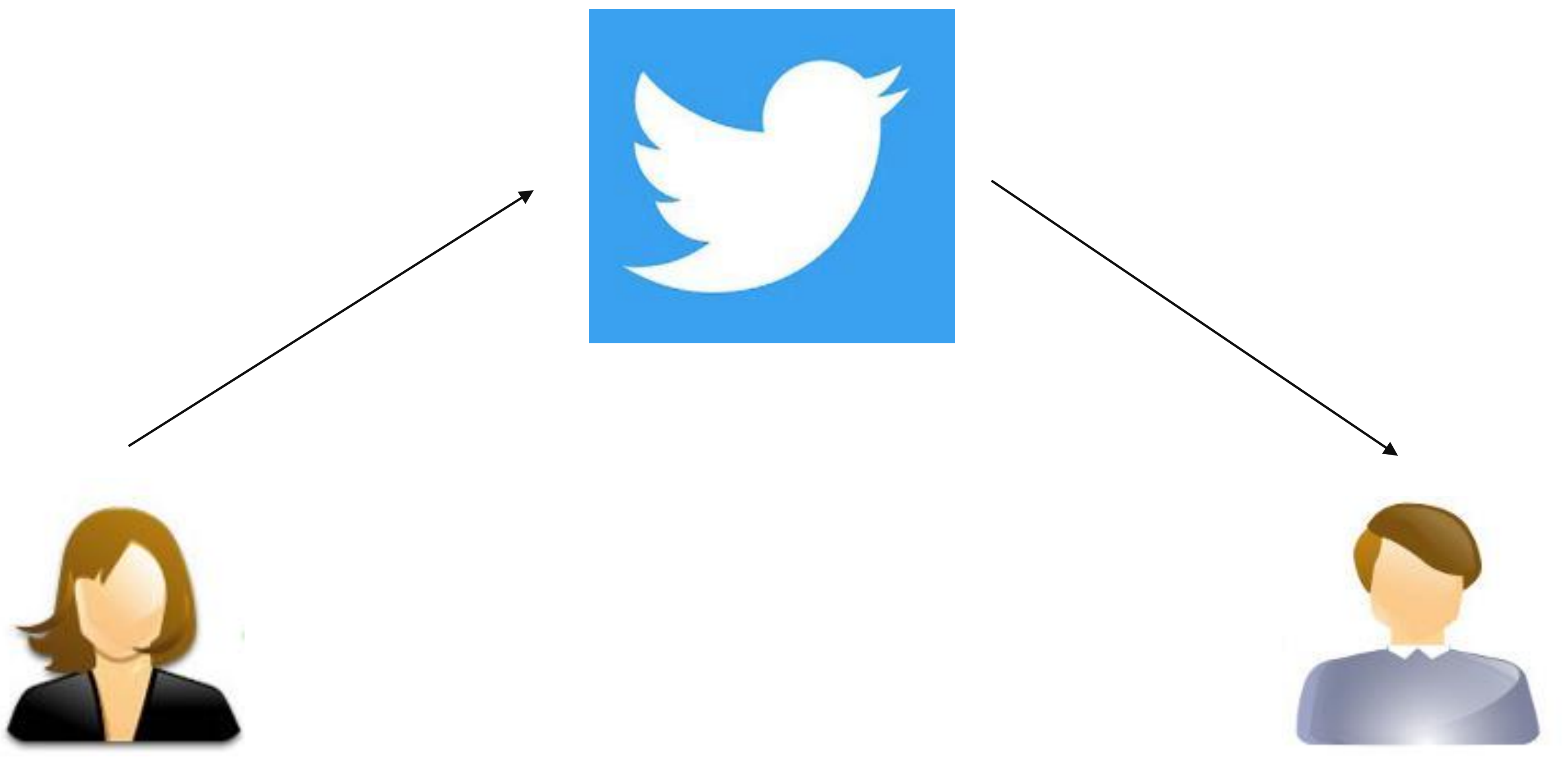


1974

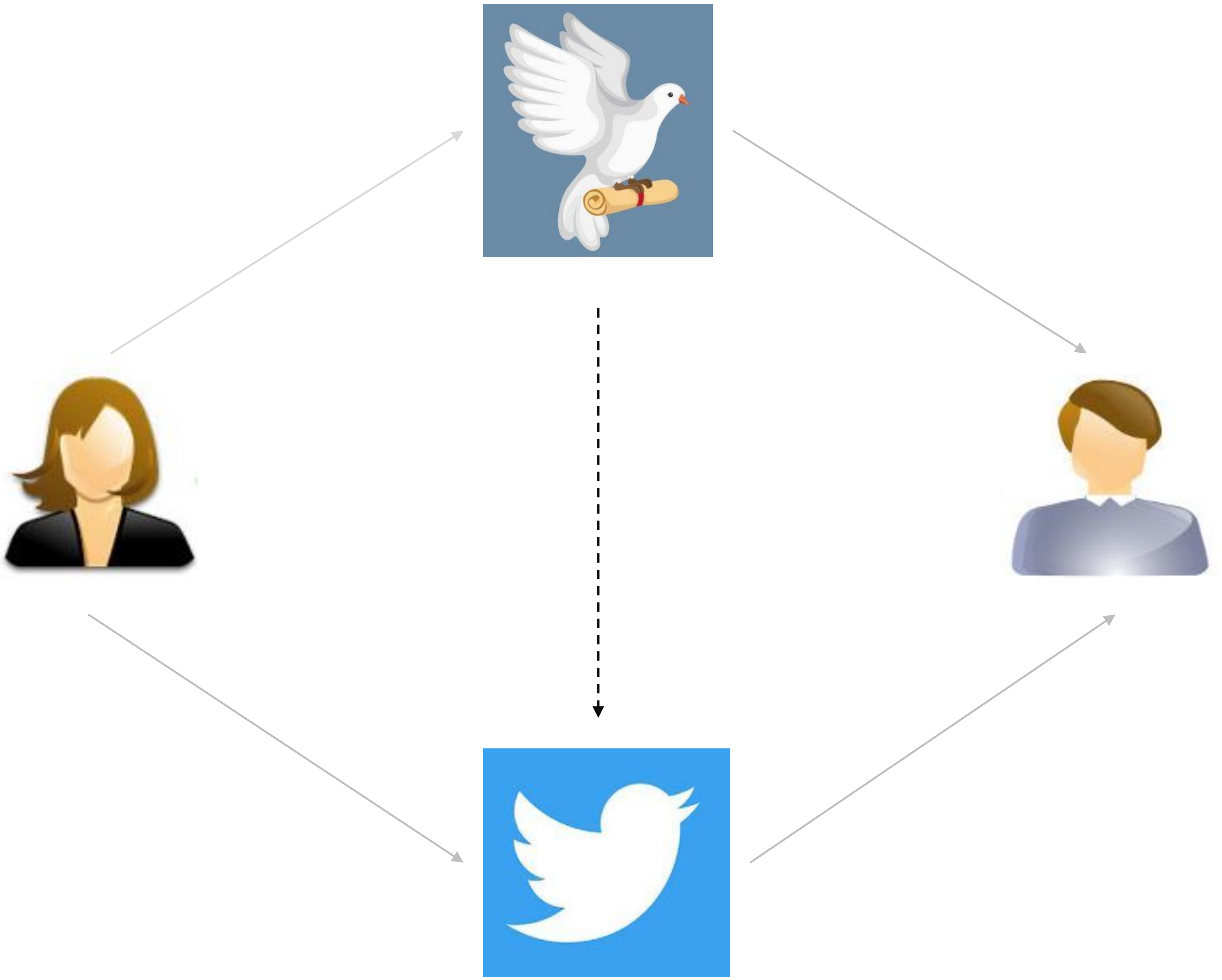


21^{st} century

21st century

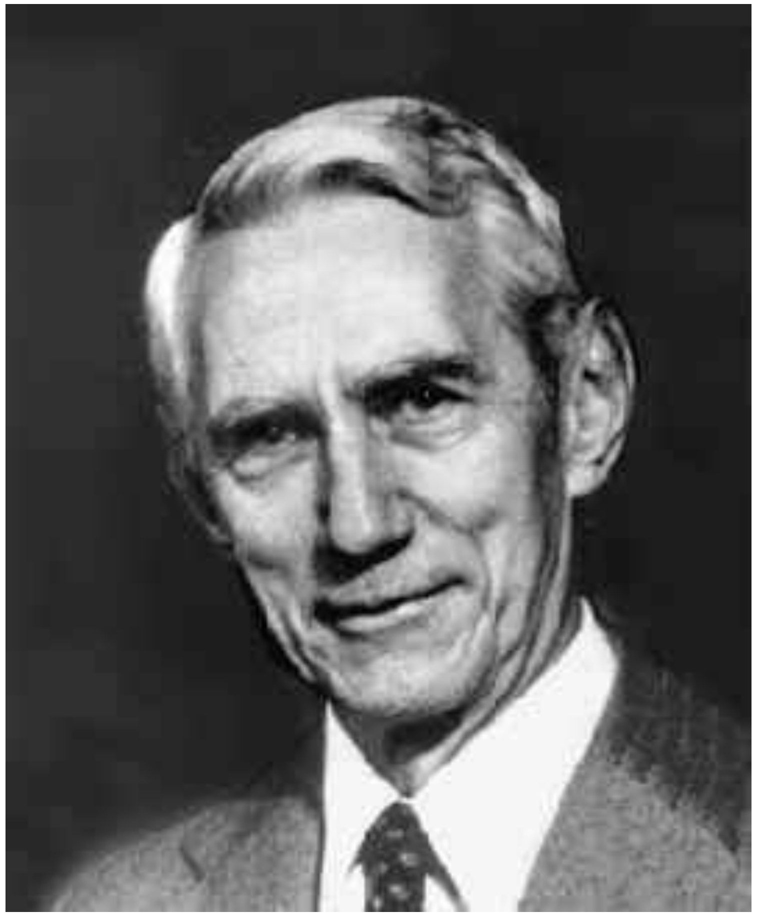


A communication odyssey

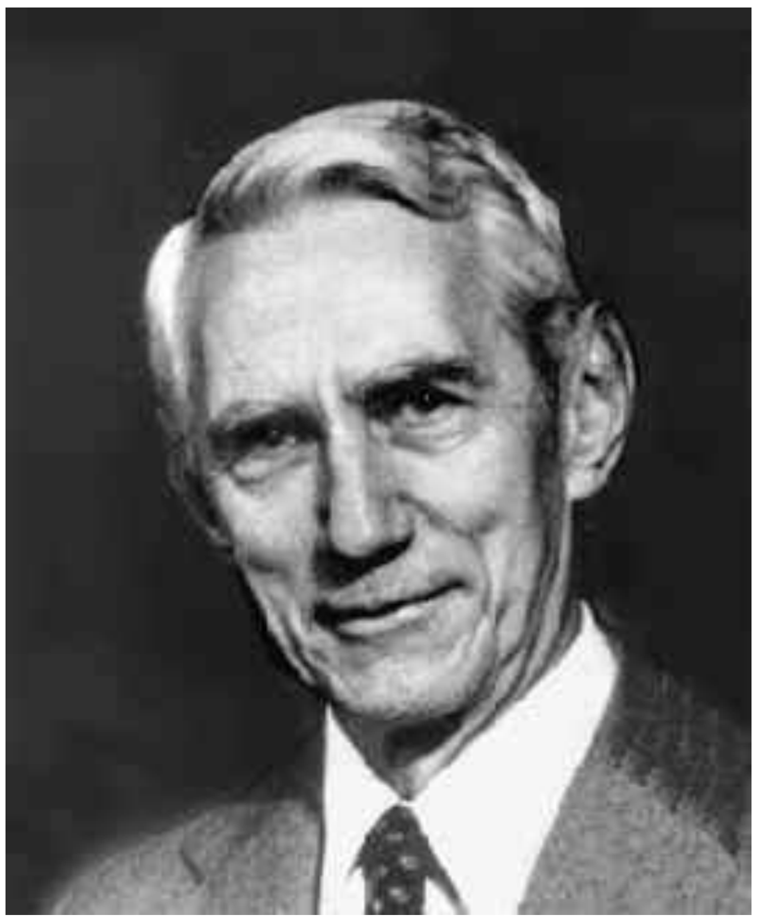


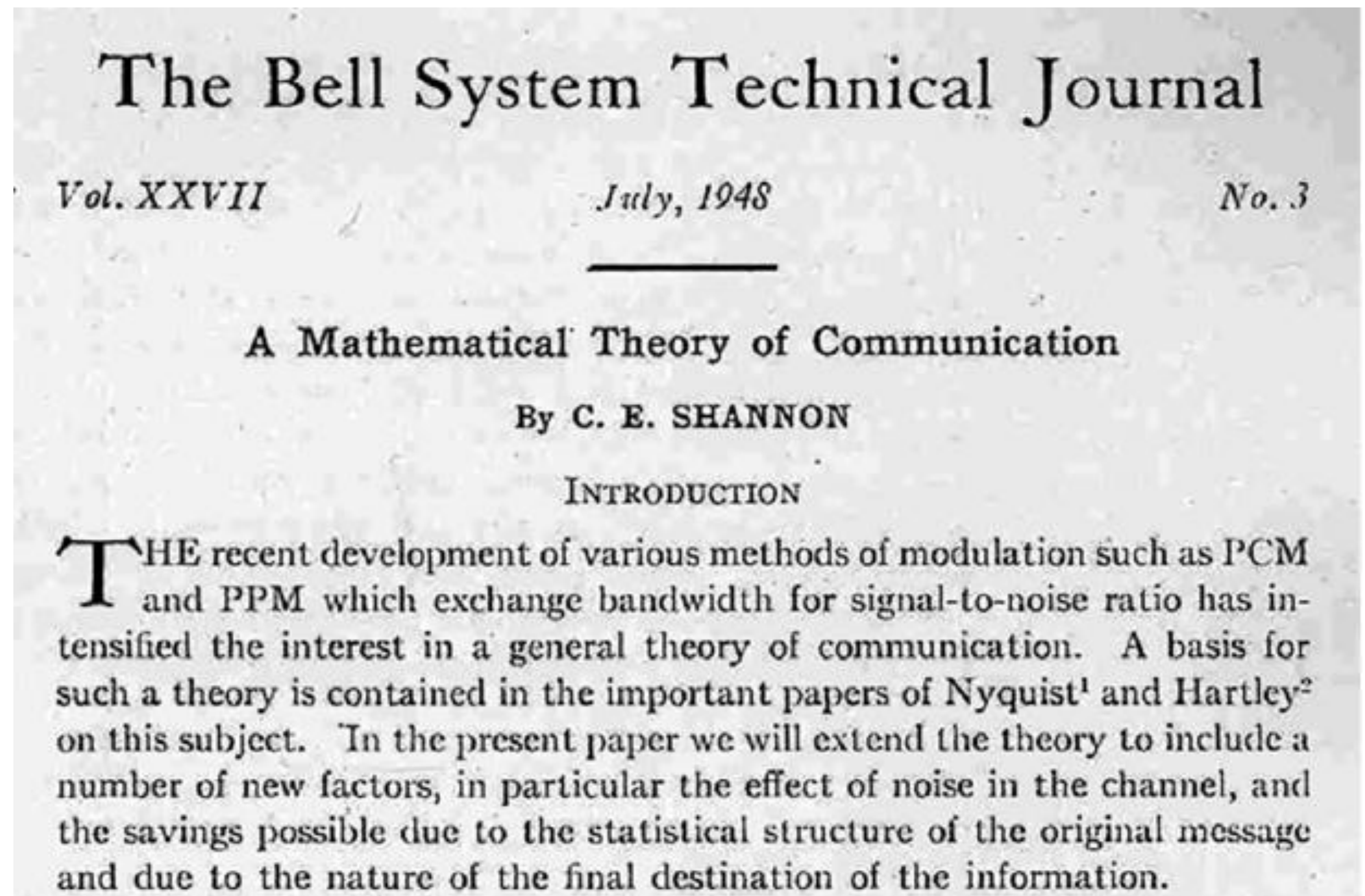
1948: Then there was light

1948: Then there was light



1948: Then there was light





The Bell System Technical Journal

Vol. XXVII *July, 1948* *No. 3*

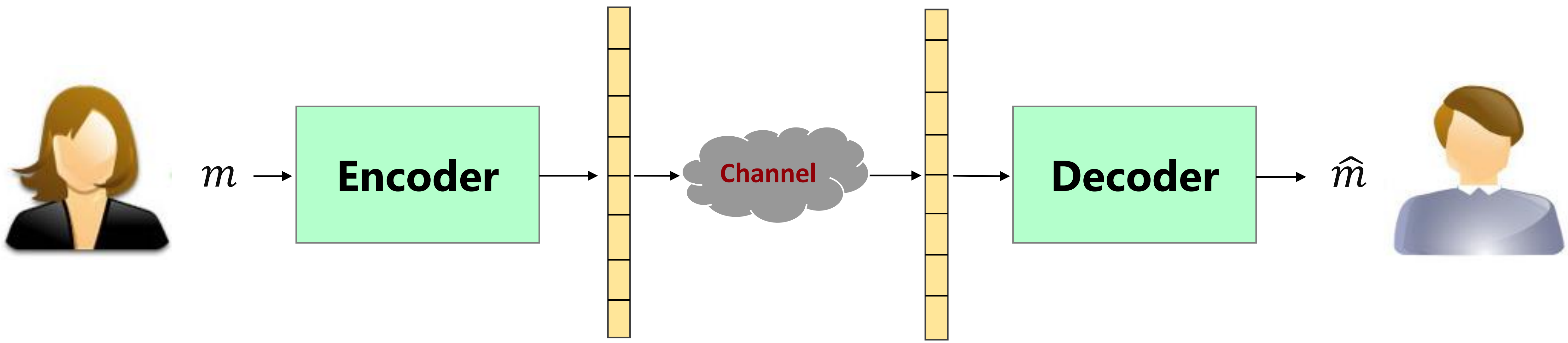
A Mathematical Theory of Communication

By C. E. SHANNON

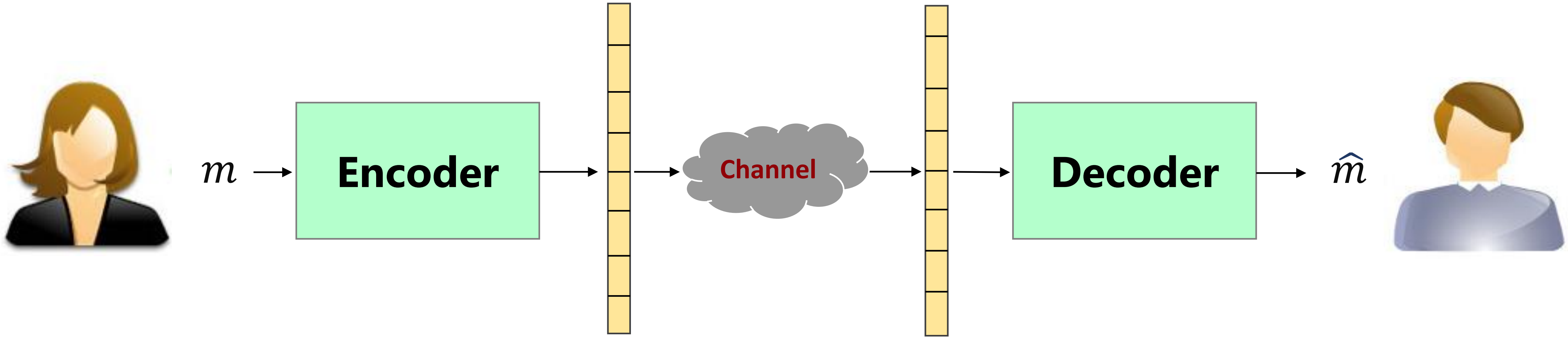
INTRODUCTION

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.

Codes: a mathematical lens



Codes: a mathematical lens



Code = (Encoder, Decoder)

Communication codes

- **AWGN channel**
  - Precise performance metrics

Communication codes

- **AWGN channel**
  - Precise performance metrics

- Challenge: Space of (encoders, decoders) very large
  - $\text{Rate} = \frac{1}{2},\ k = 100$: 2^{100} codewords in 200 dimensional space

Communication codes

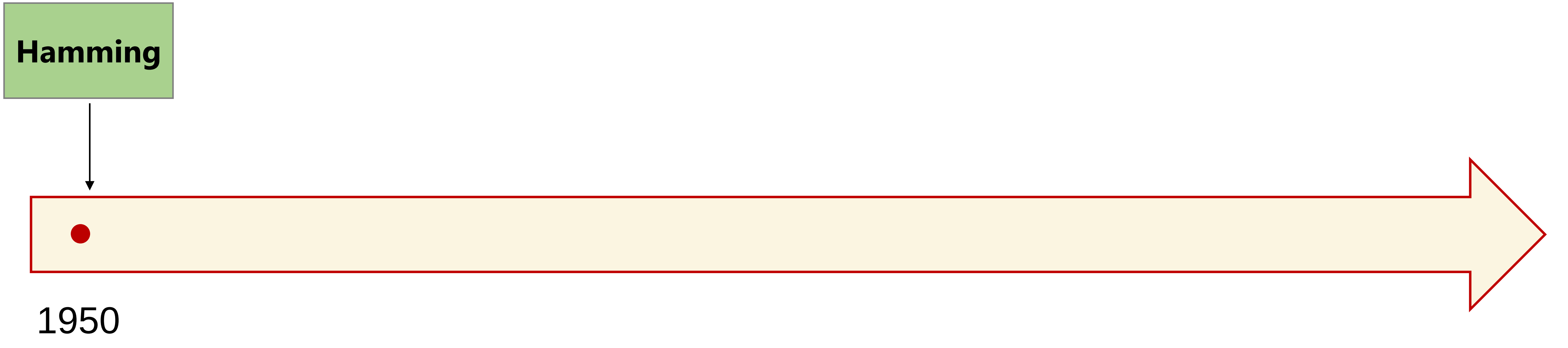
- **AWGN channel**
  - Precise performance metrics

- Challenge: Space of (encoders, decoders) very large
  - $\text{Rate} = \frac{1}{2},\ k = 100$: 2^{100} codewords in 200 dimensional space

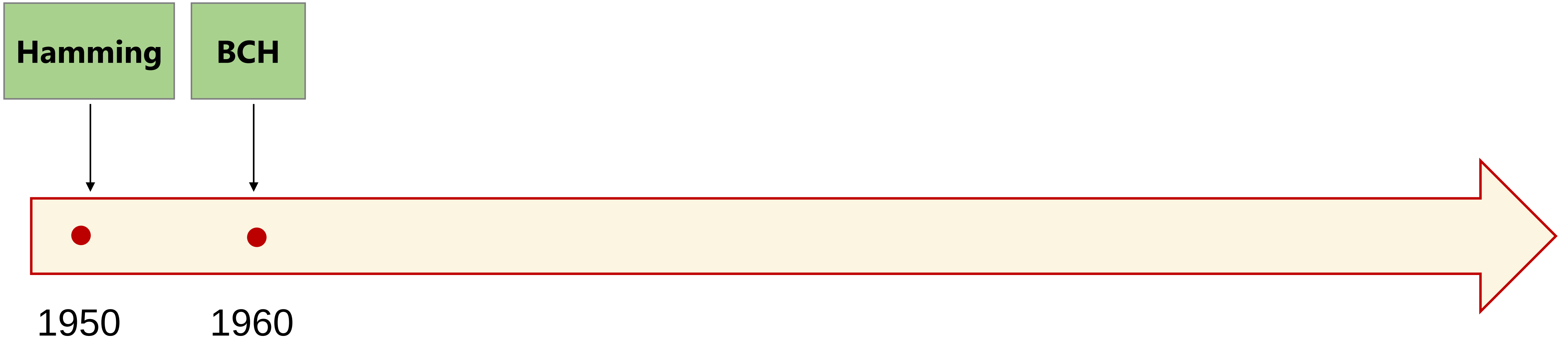
- Information theory, Coding theory, Comm. theory

Landmark codes: AWGN

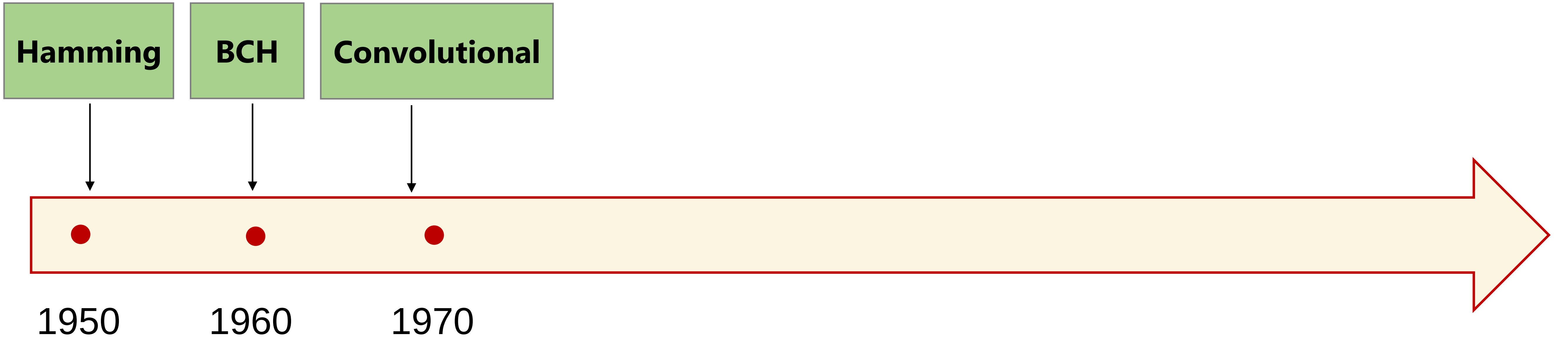
Landmark codes: AWGN



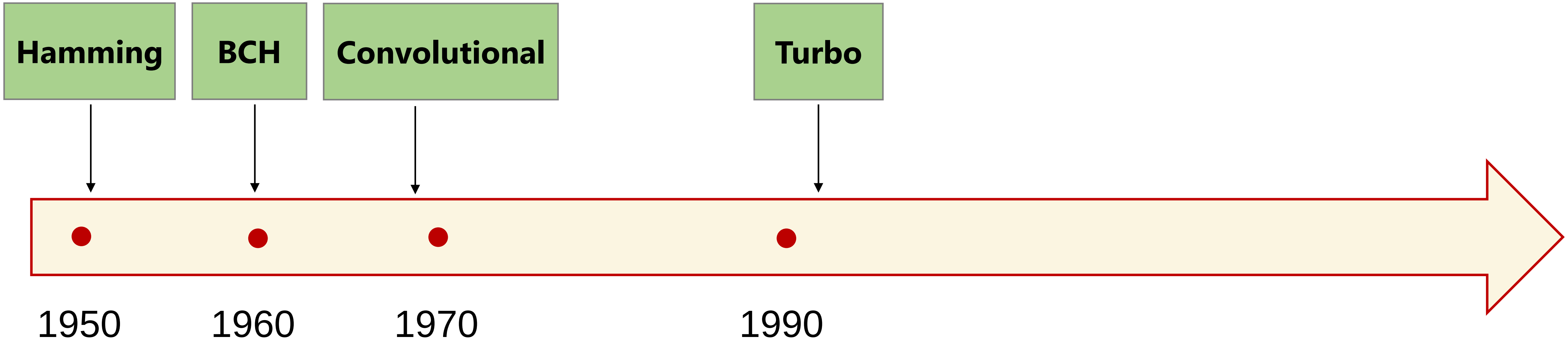
Landmark codes: AWGN



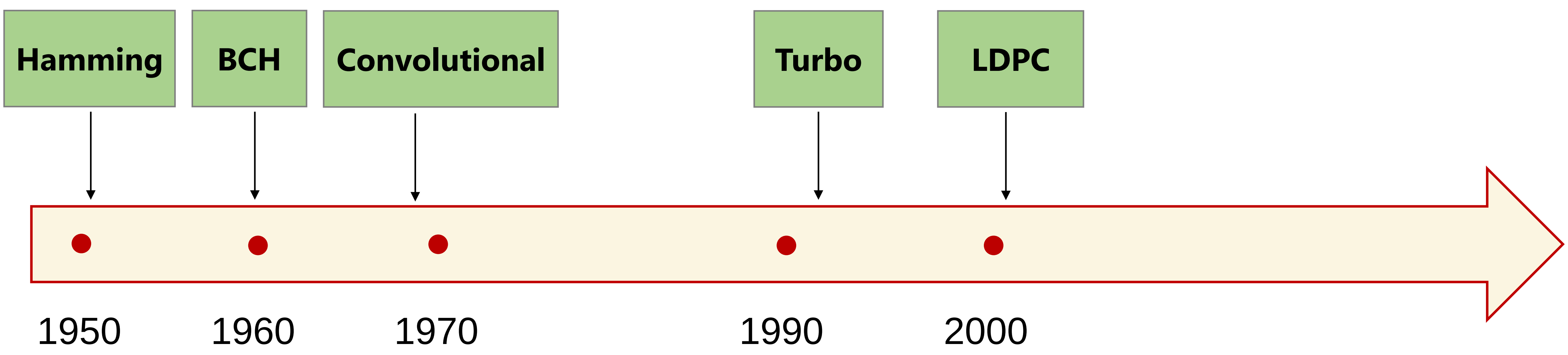
Landmark codes: AWGN



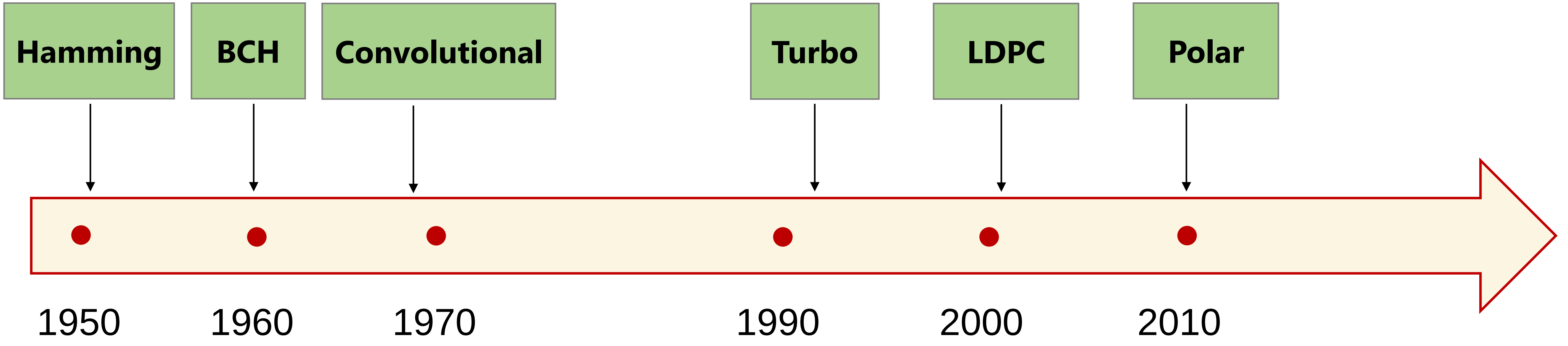
Landmark codes: AWGN



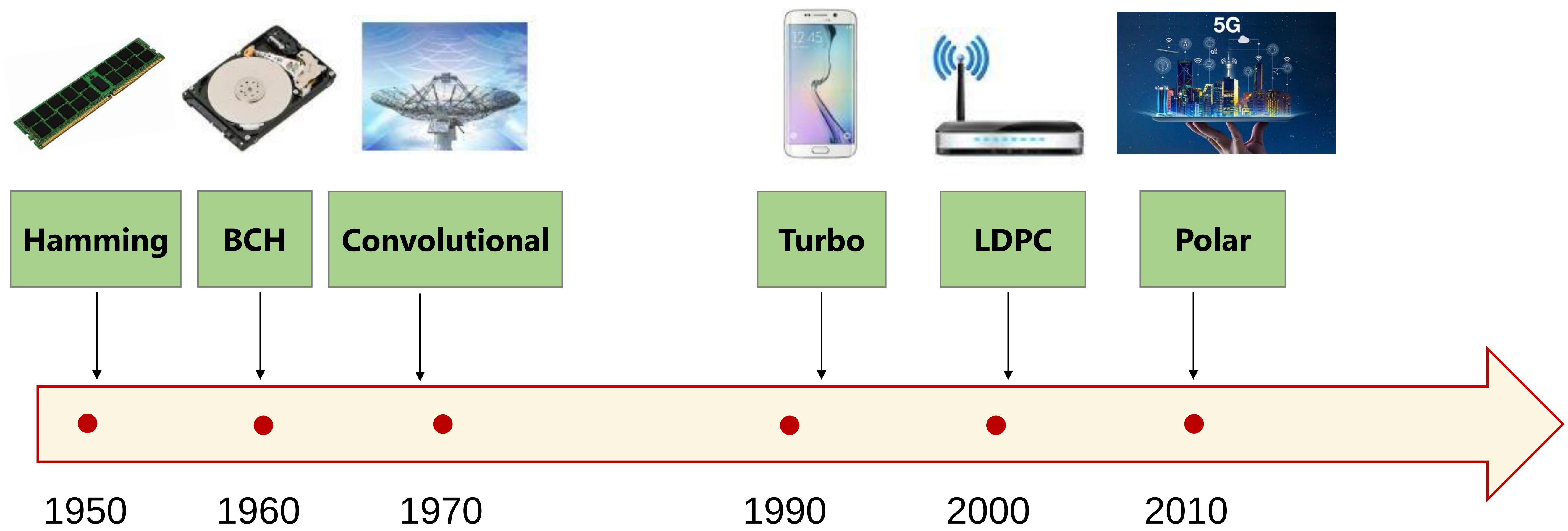
Landmark codes: AWGN



Landmark codes: AWGN



Huge practical impact

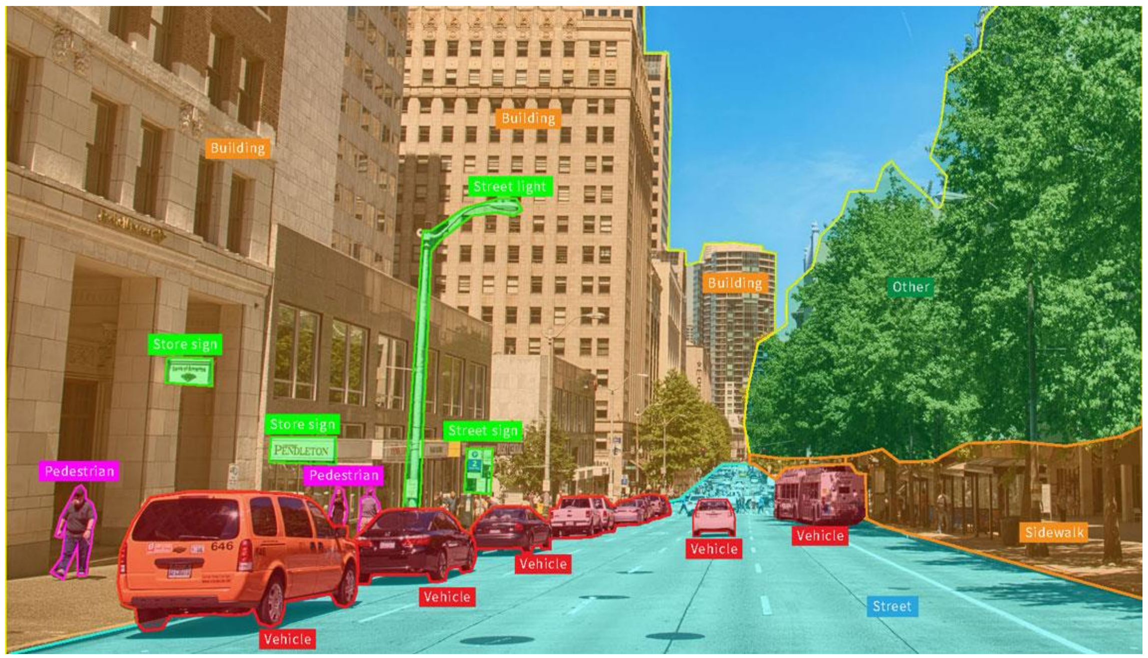


Discovery of codes

- Human eureka moments
  - Sporadic

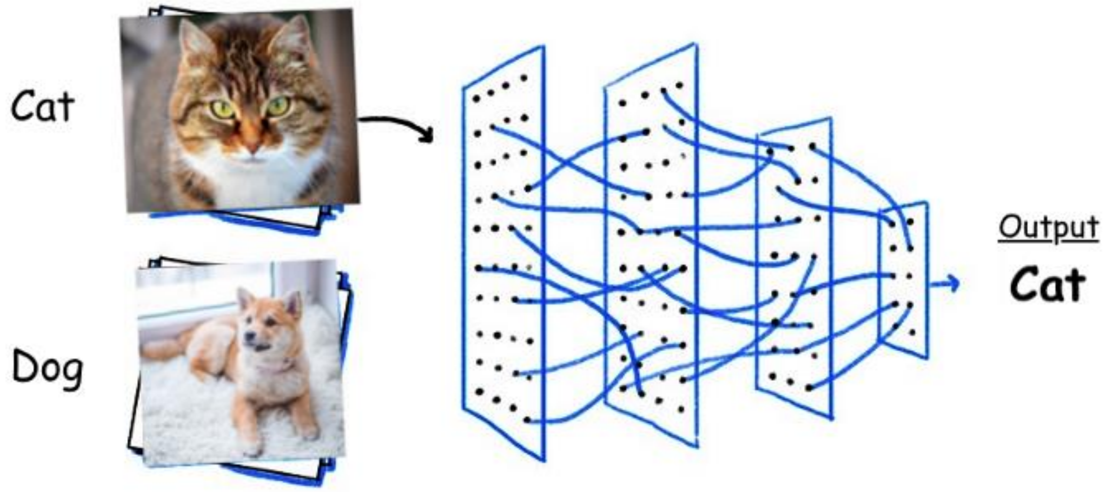
- **Goal:** Automate the discovery

Deep learning (DL)

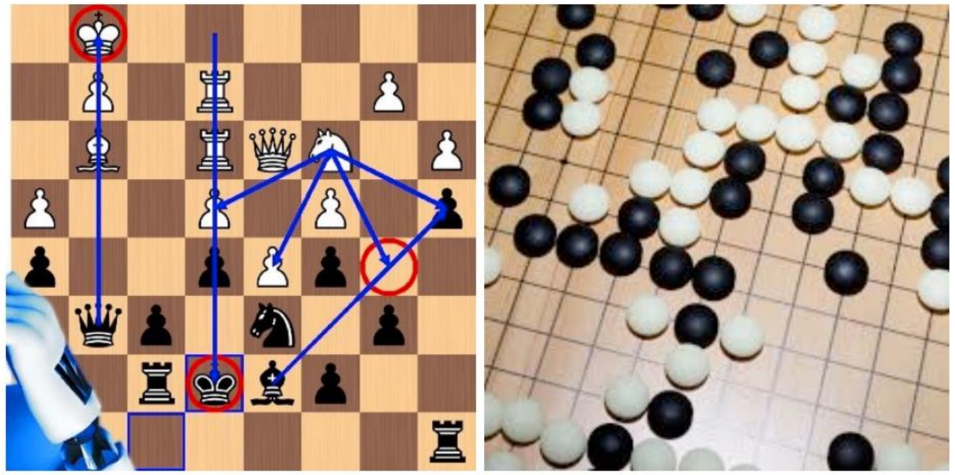


DL success story

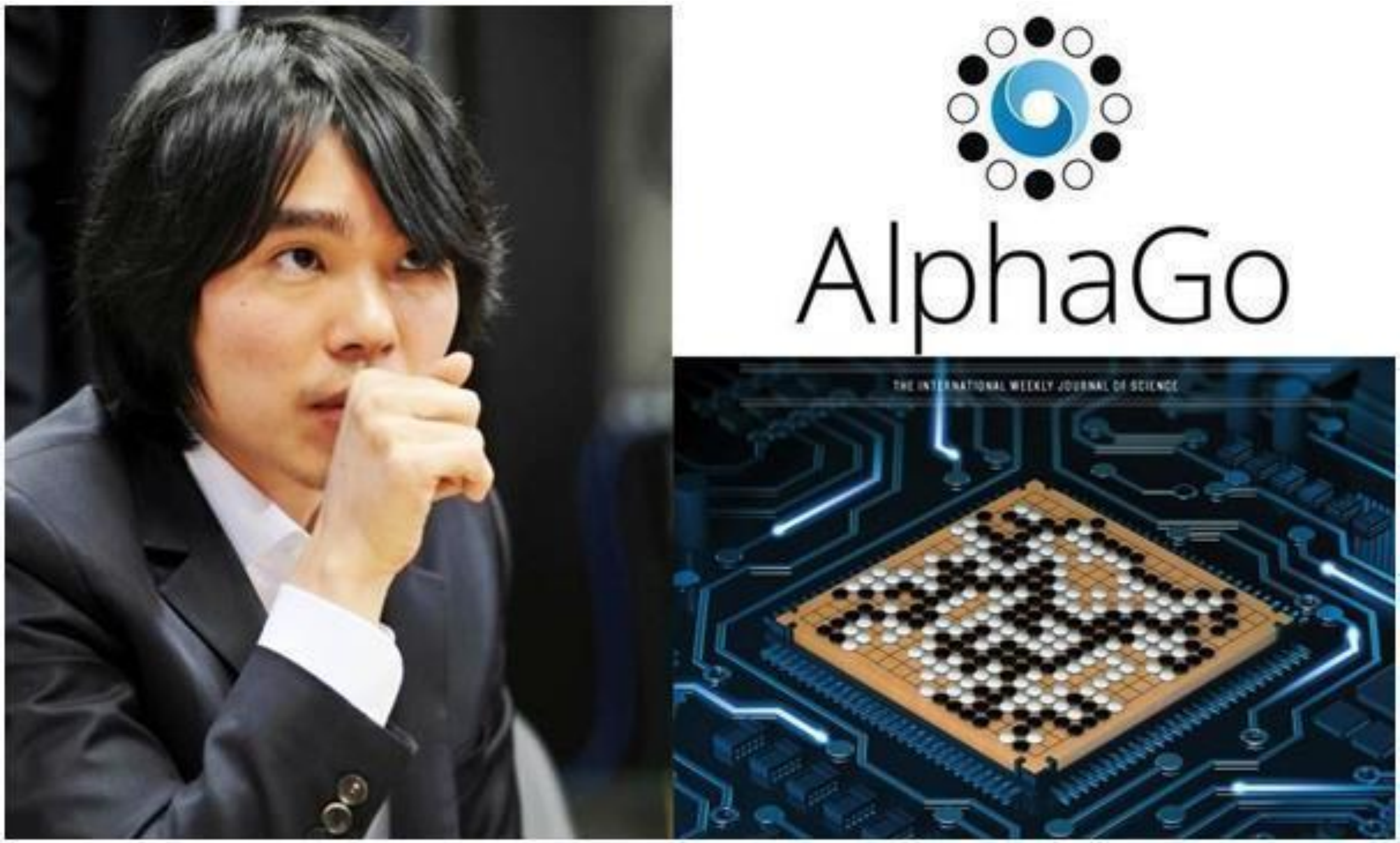
- Model deficiency
  - No analytical model
  - **AlexNet**



- Algorithm deficiency
  - Clear model
  - Space of algorithms large
  - **AlphaGo**



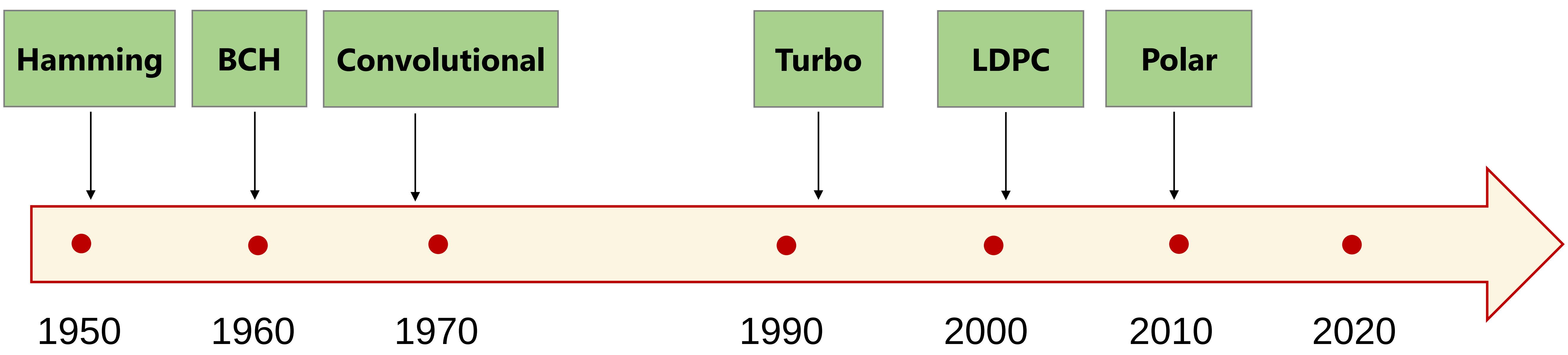
Breakthroughs of DL



Main goal

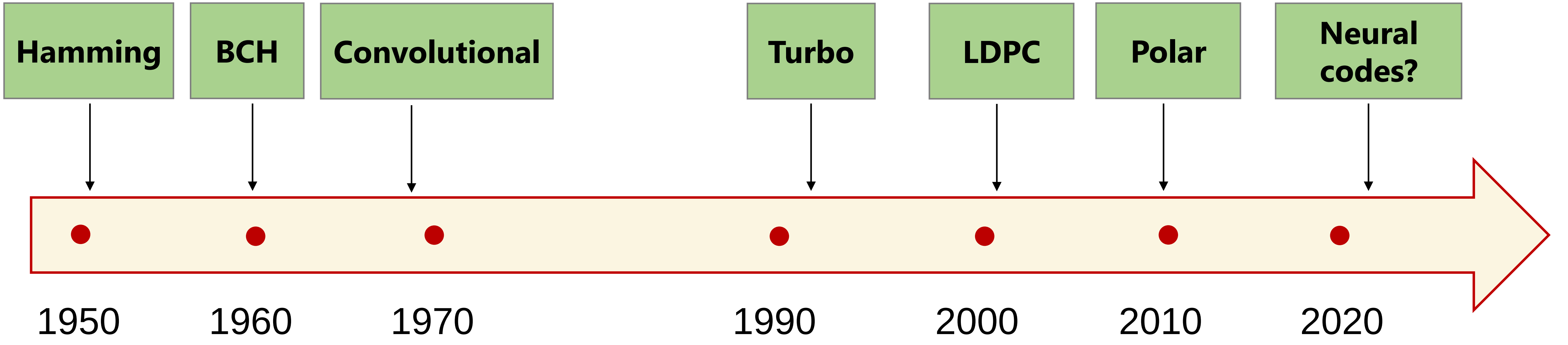
Can we automate the search for codes via DL?

Main goal



Hamming
BCH
Convolutional
Turbo
LDPC
Polar
1950
1960
1970
1990
2000
2010
2020

Main goal



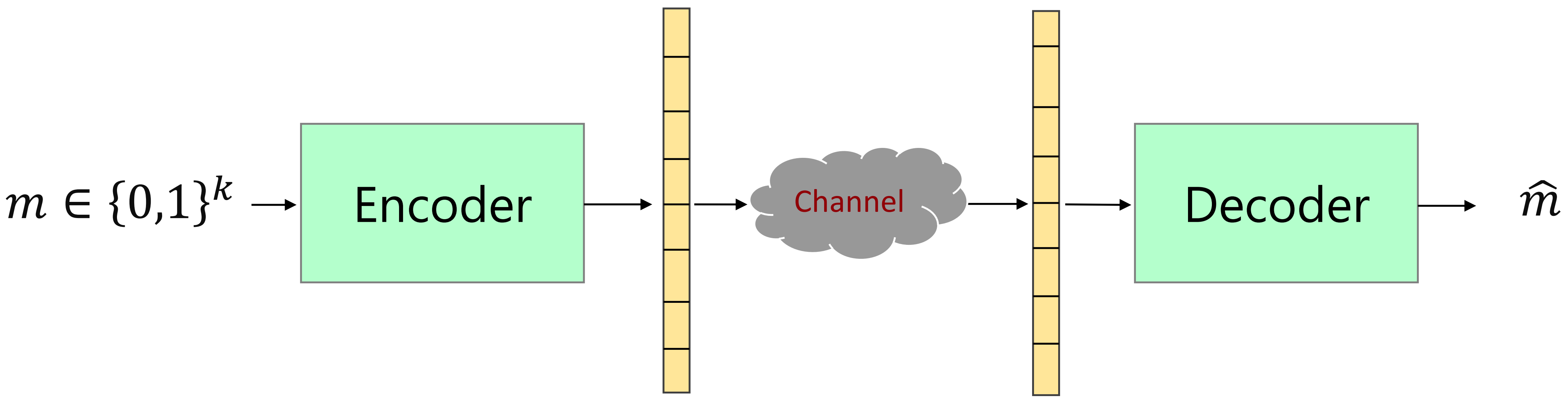
Hamming
BCH
Convolutional
Turbo
LDPC
Polar
Neural codes?
1950
1960
1970
1990
2000
2010
2020

Agenda

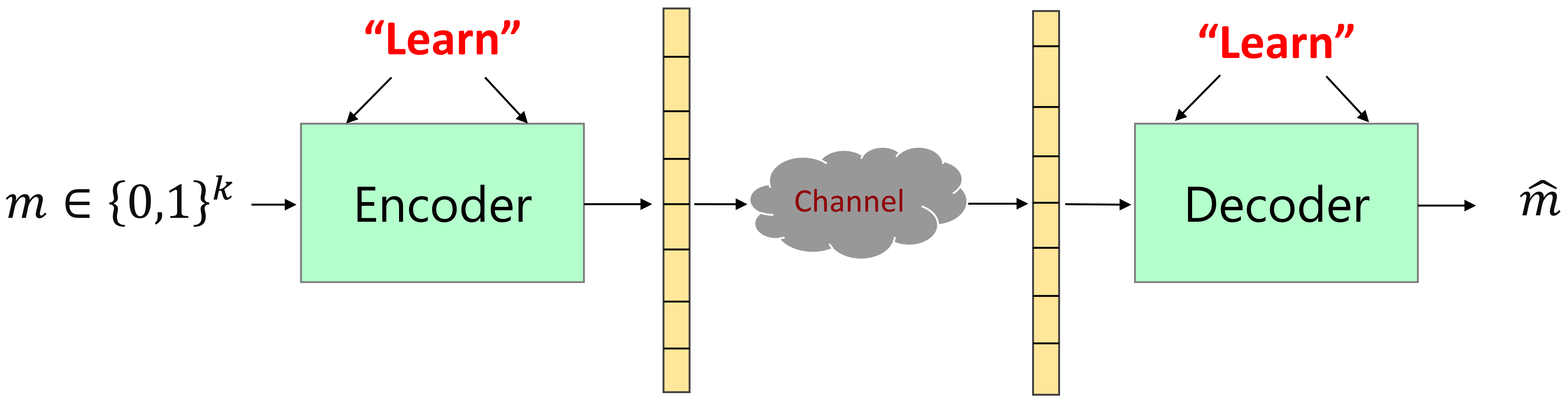
- New (deep learning) tools for classical problems
  - New state-of-the-art codes
  - Inherent practical value

- Insight into deep learning methods
  - Communication framework as a lens

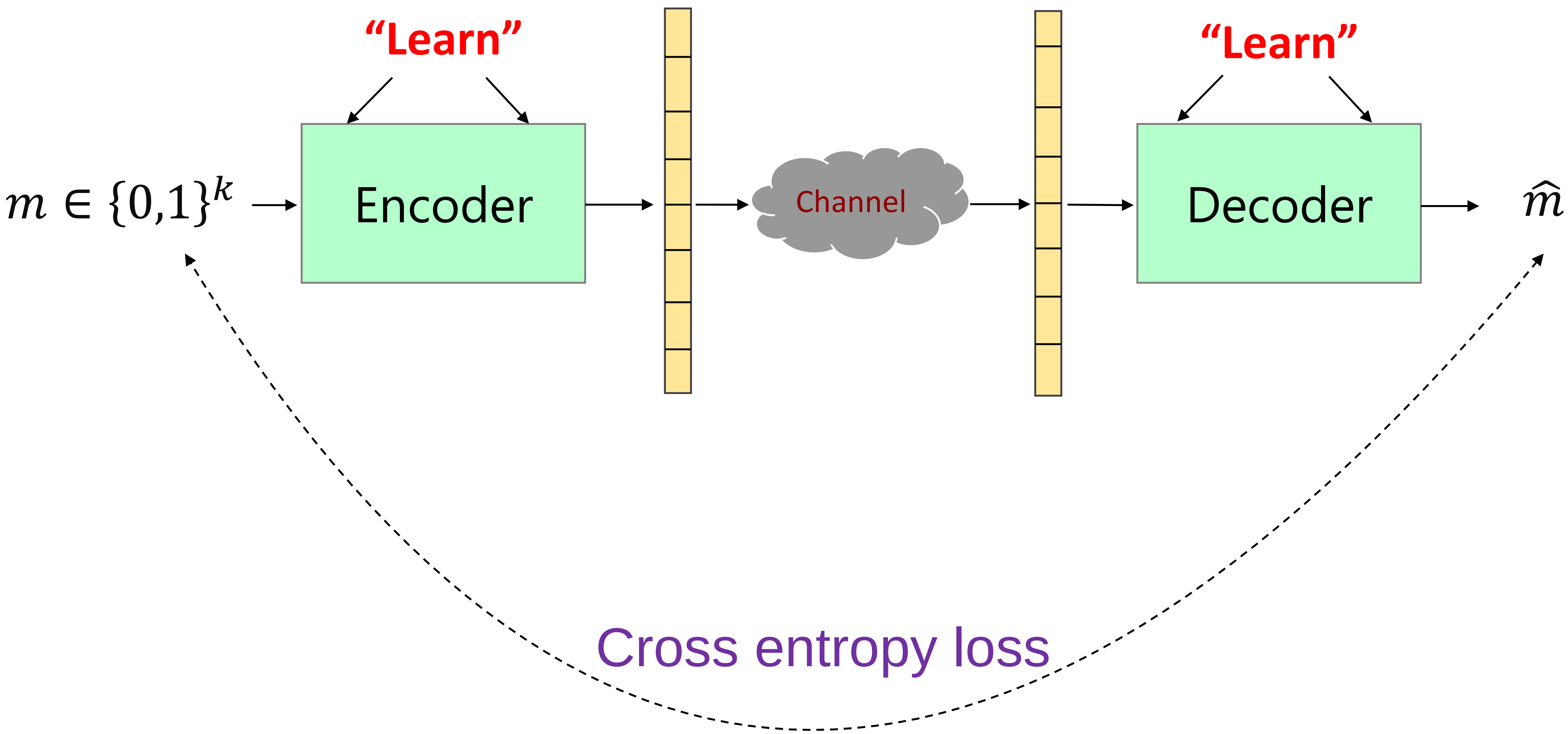
Learning a new code



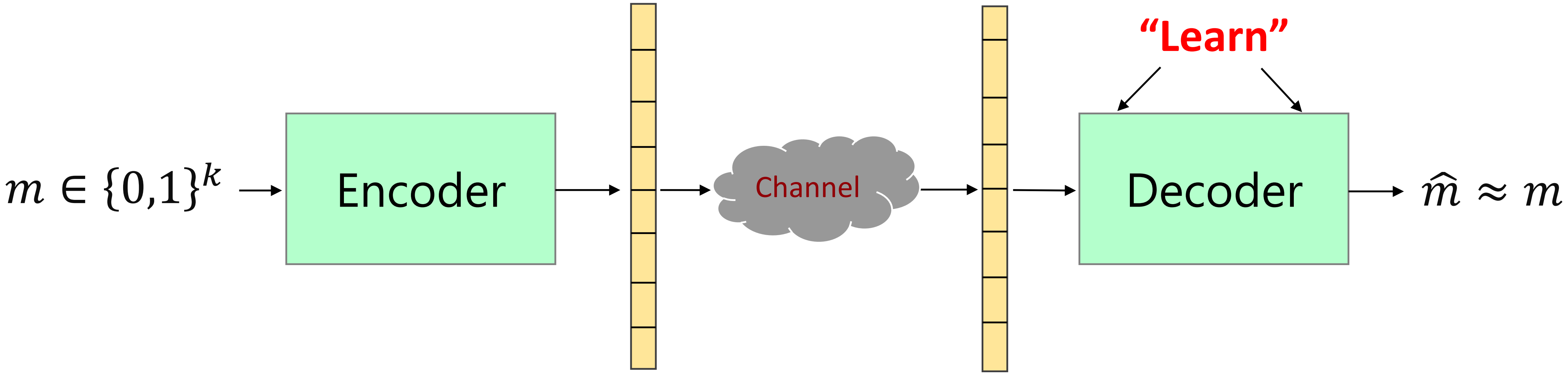
Learning a new code



Learning a new code



Learning to decode



Vast literature

- Supervised learning
  - Nachmani et al., 2016
  - Gruber et al. 2017
  - Cammerer et al., 2017
  - Nachmani et al., 2018
  - Kim et al., 2018a;b
  - Vasic et al., 2018
  - Teng et al., 2019
  - Nachmani & Wolf, 2019
  - Buchberger et al., 2020
  - Chen & Ye, 2021

- Reinforcement learning
  - Carpi et al., 2019
  - Habib et al., 2020
  - Doan et al., 2020

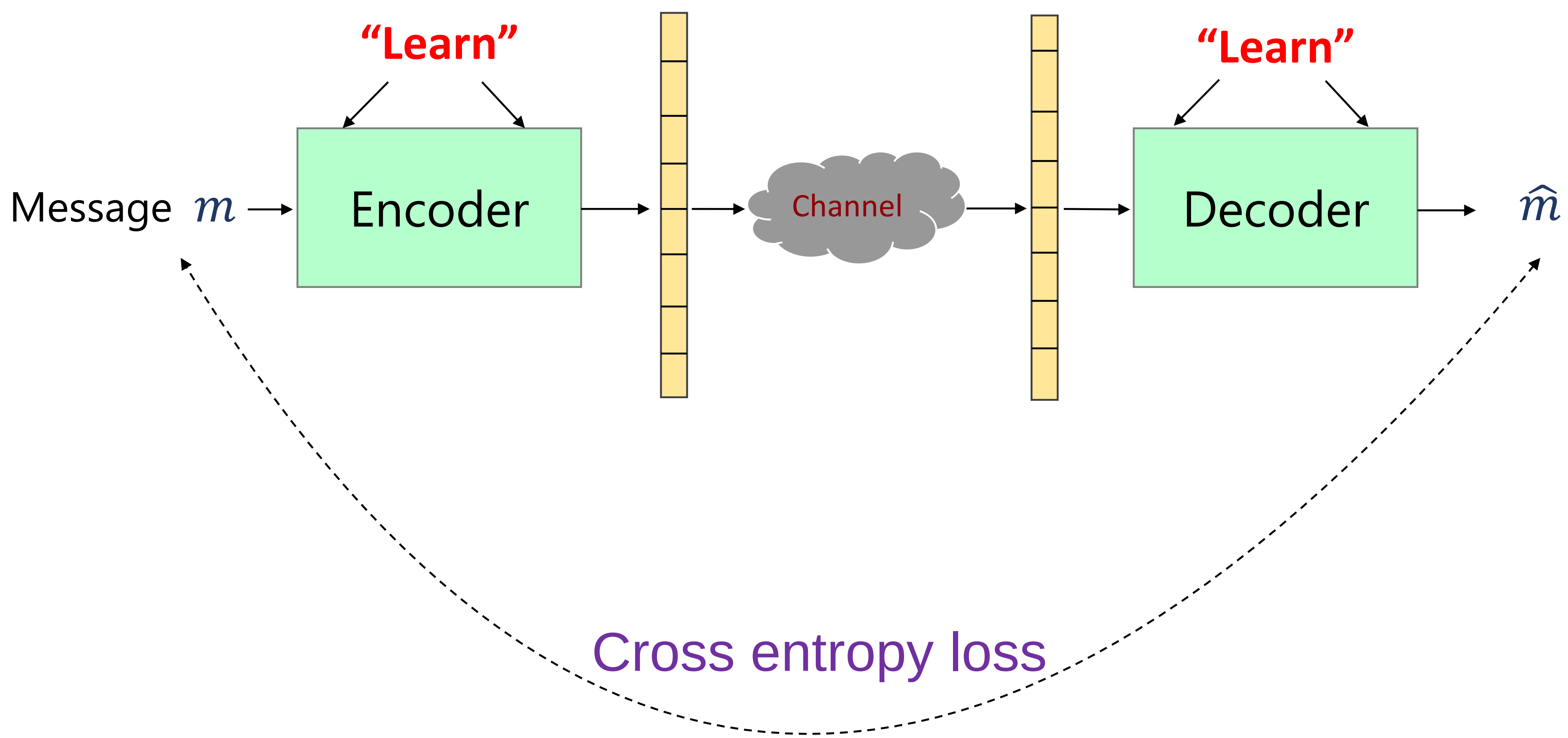
Learning to decode: summary

- Fix the encoding

- DL decoders learn state-of-the-art decoders
  - Convolutional codes: Viterbi, BCJR, dynamic programming
  - Turbo codes: BCJR
  - RM & Polar codes: Successive Cancellation

- Clever architectural choices
  - Recurrent neural networks $\longleftrightarrow$ dynamic programming

Learning a new code



Code structure

Code structure

- Linear and binary: Classical codes

Code structure

- Linear and binary: Classical codes

- Non-linear and real valued: **Neural networks (NNs)**

  - Fully connected NNs worse than repetition codes (Jiang et. al ‘19)
  - Still need a structure

Imparting structure

- Capitalize on state-of-the-art codes

Imparting structure

- Capitalize on state-of-the-art codes
- **How?**
- **What class of codes?**

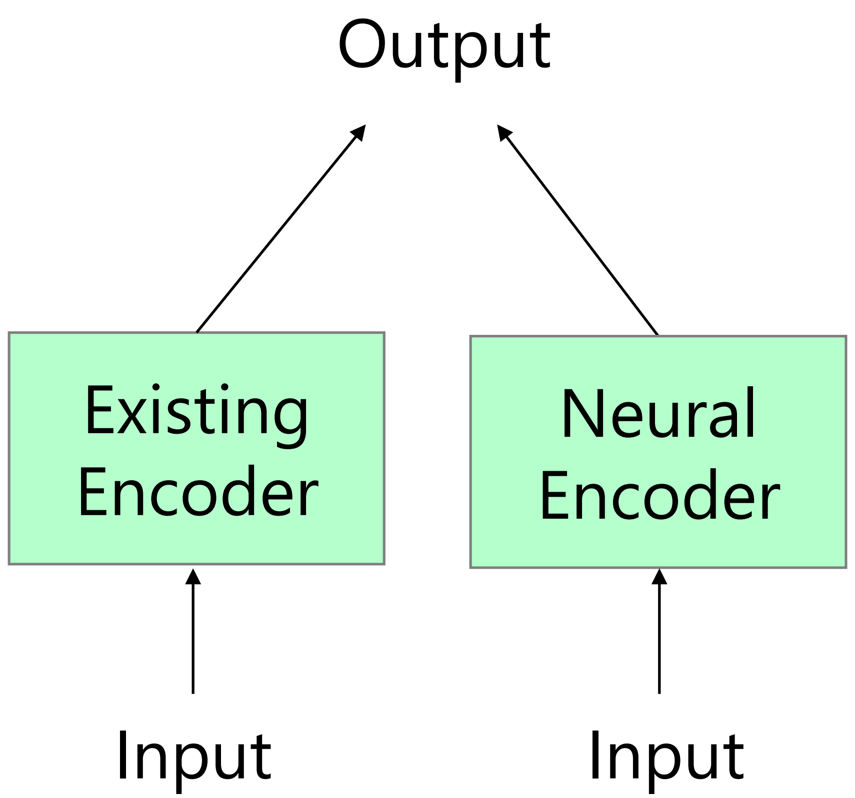
Imparting structure

- Capitalize on state-of-the-art codes

- **How?**

Neural Augmentation

- (Welling '20)



Imparting structure

- Capitalize on state-of-the-art codes

- How?
  - Neural Augmentation

- **What class of codes?**

Taxonomy of codes

Sequential codes
Convolutional and
Turbo codes.

Graphical codes
LDPC codes.

Algebraic codes
Reed-Solomon, BCH,
Reed-Muller and Polar codes.

Taxonomy of codes

Sequential codes
Convolutional and
Turbo codes.

Graphical codes
LDPC codes.

Algebraic codes
Reed-Solomon, BCH,
Reed-Muller and Polar codes.

Jiang et al, 2019

Algebraic codes

Sequential codes
Convolutional and
Turbo codes.

Graphical codes
LDPC codes.

Algebraic codes
Reed-Solomon, BCH,
Reed-Muller and Polar codes.

Reed-Muller and Polar codes

Sequential codes
Convolutional and
Turbo codes.

Graphical codes
LDPC codes.

Algebraic codes
Reed-Solomon, BCH,
Reed-Muller and Polar codes.

Reed-Muller (RM) codes

- **Classical**
  - Muller, 1954
  - Efficient decoder by Reed, 1954

- **Recent Interest**
  - Polar codes
  - Achieve capacity (very recent!)

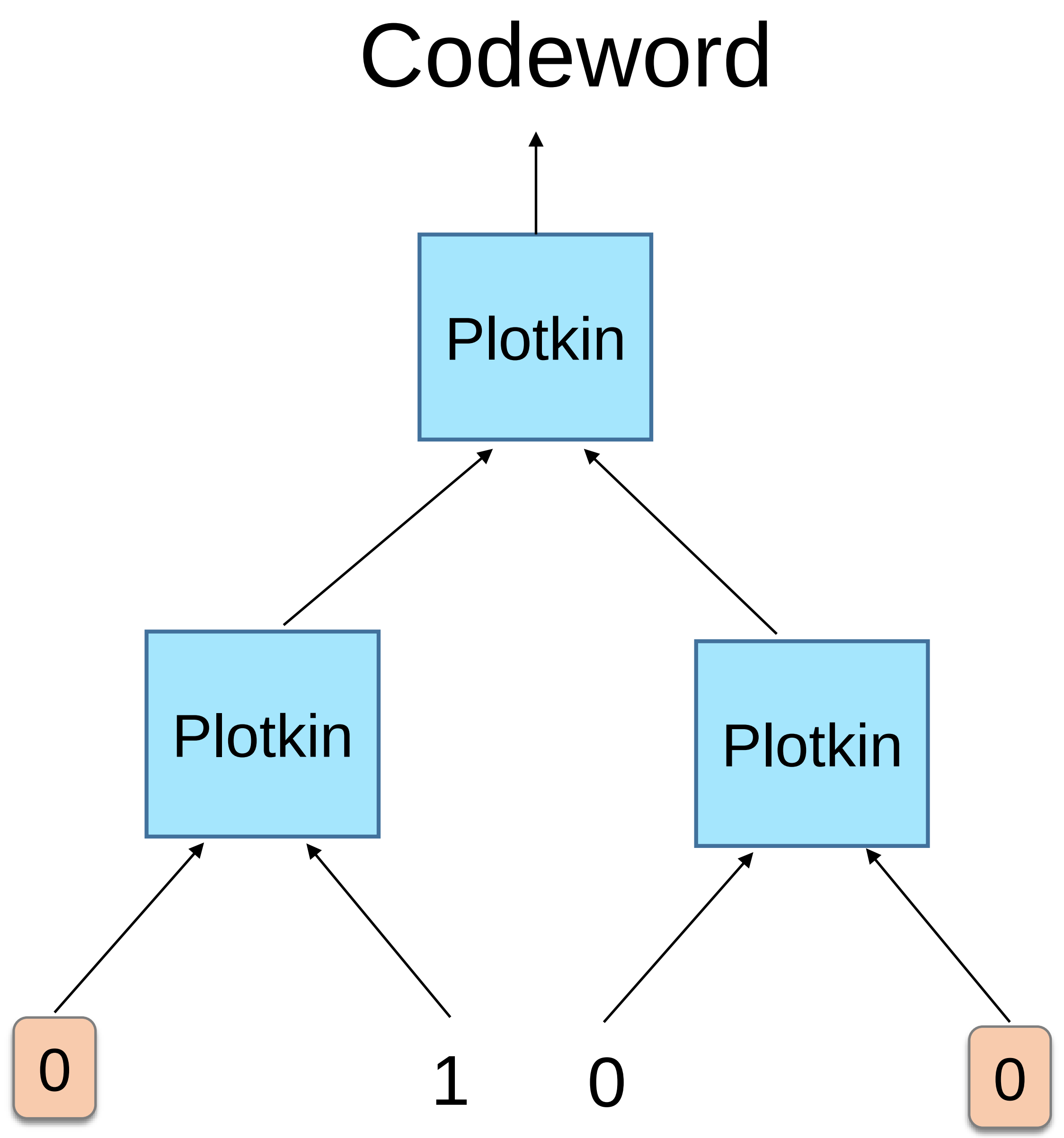
Polar codes

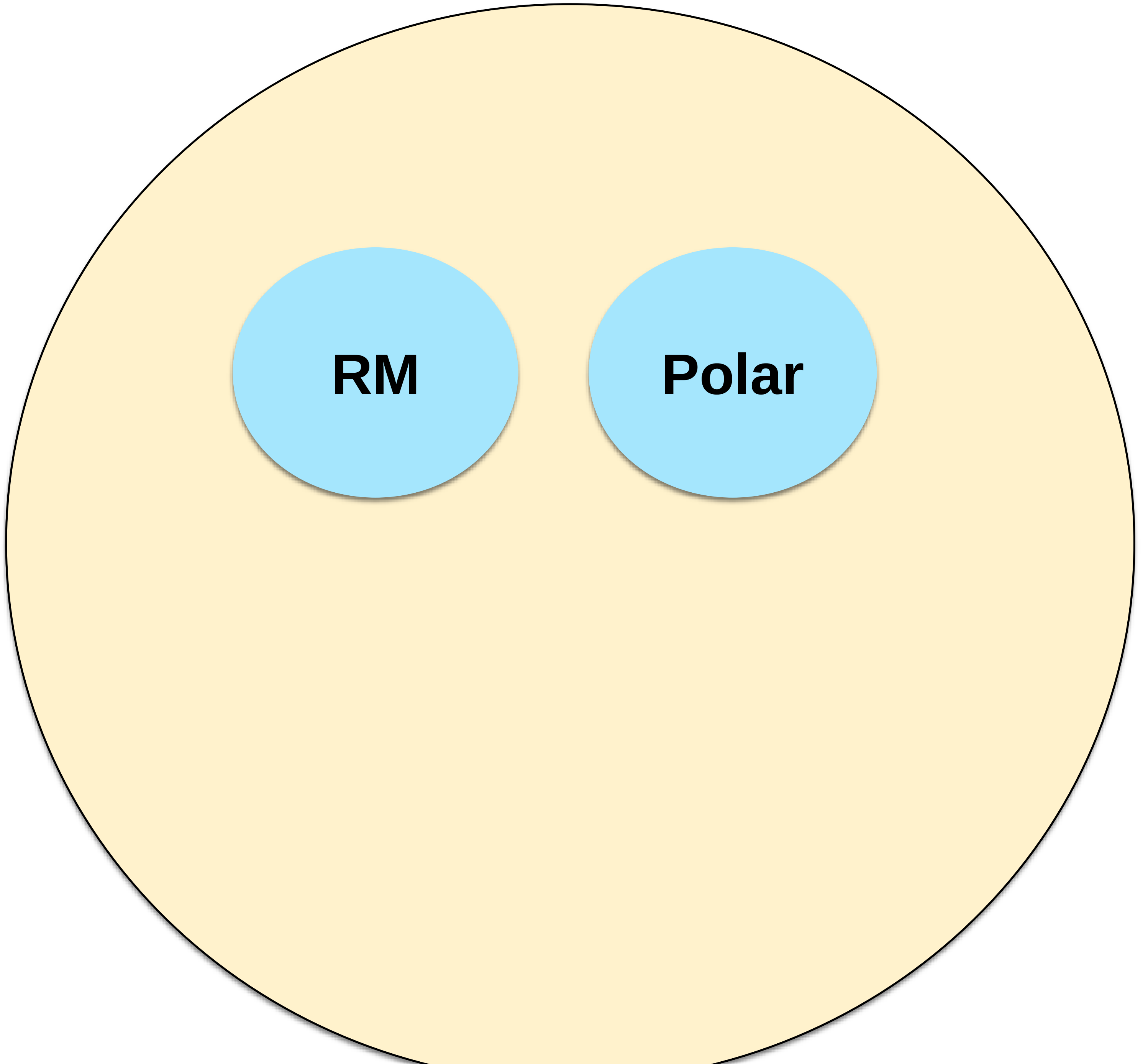
- Arikan, 2009
- **First codes proven to achieve capacity**
- Recent interest: 5G

RM and Polar

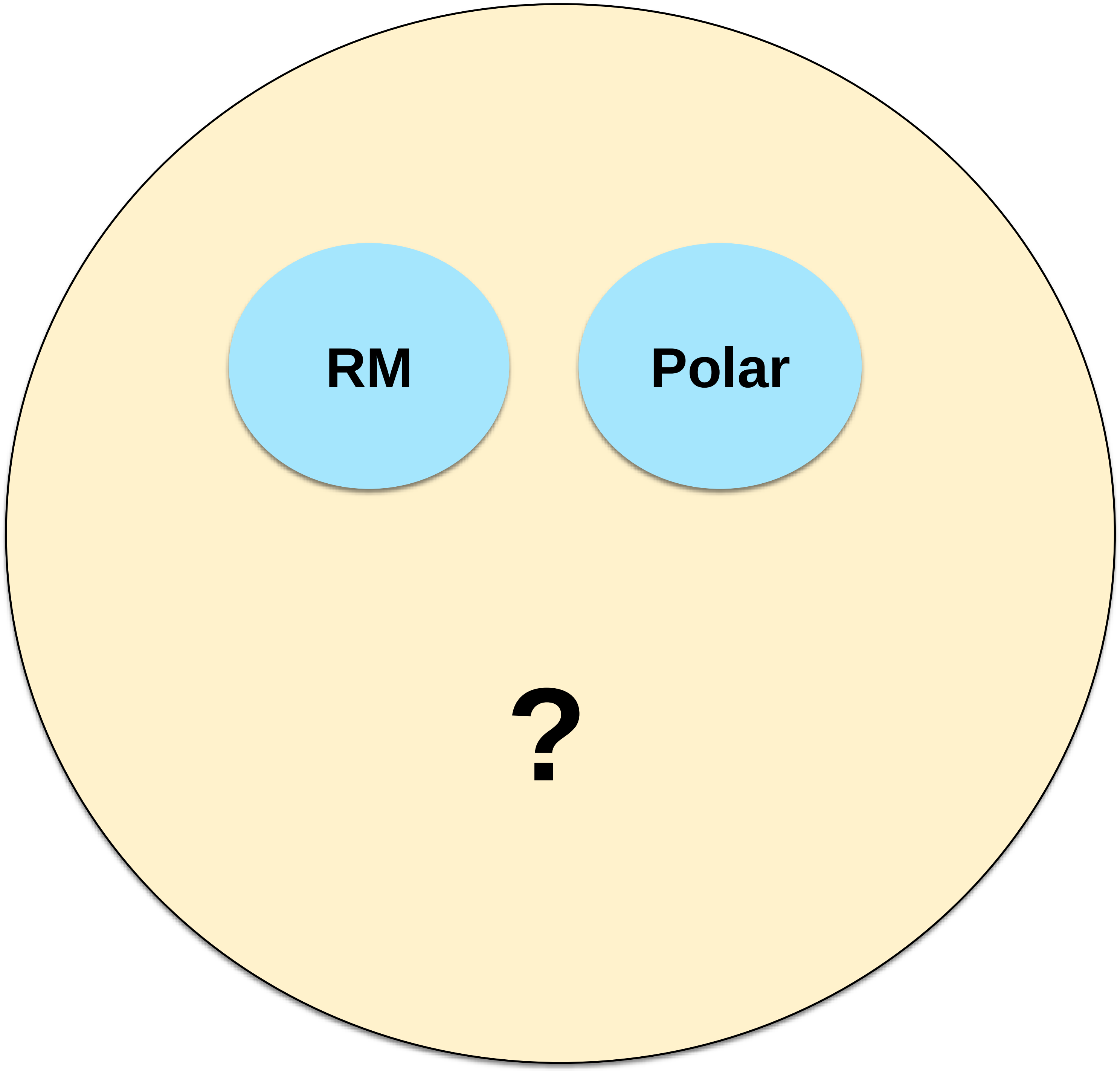
- Common structure
- **Kronecker Operation on the Plotkin transform**

Structure: Kronecker Operation (KO)

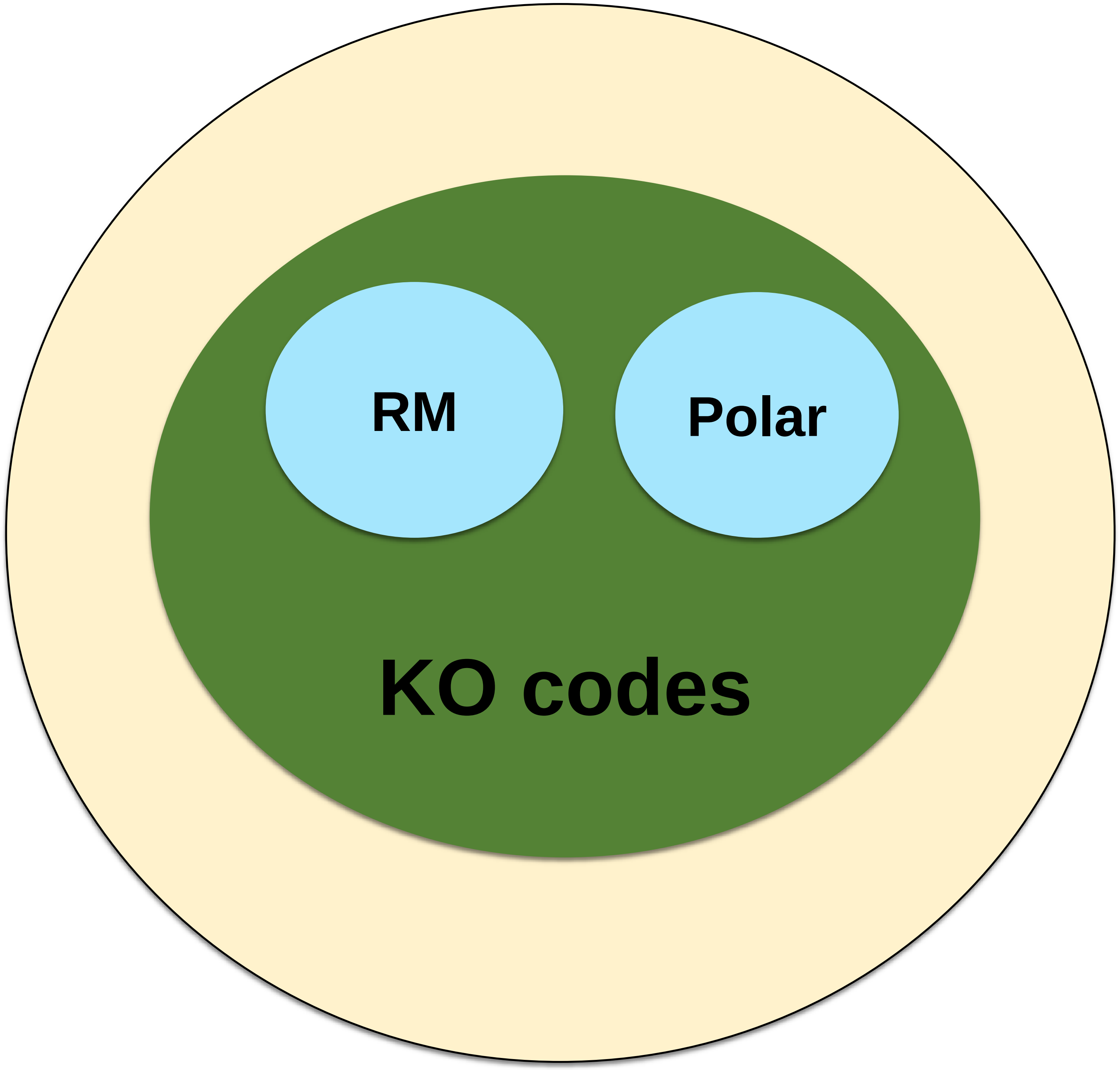



RM
Polar

More codes?



KO codes



KO codes

A. V. Makkuva*, X. Liu*, M. V. Jamali, H. Mahdavifar, S. Oh, and P. Viswanath, “KO codes: inventing nonlinear encoding and decoding for reliable wireless communication via deep-learning,” *in Proceedings of the 38th International Conference on Machine Learning (ICML), 2021.*

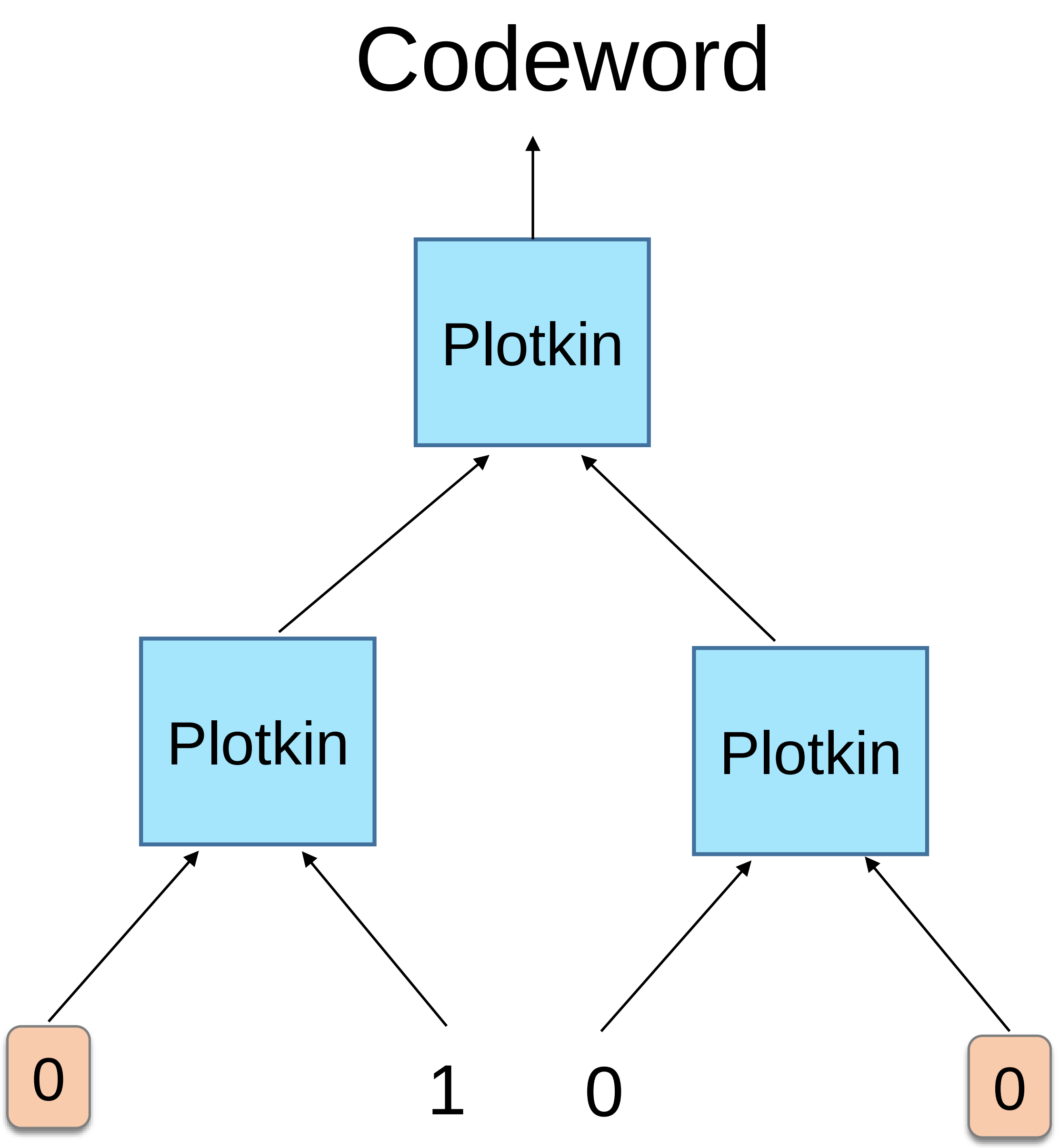
KO codes: An overview

- Novel family of neural codes
- Outperform both RM and Polar in certain regimes
- Fascinating properties

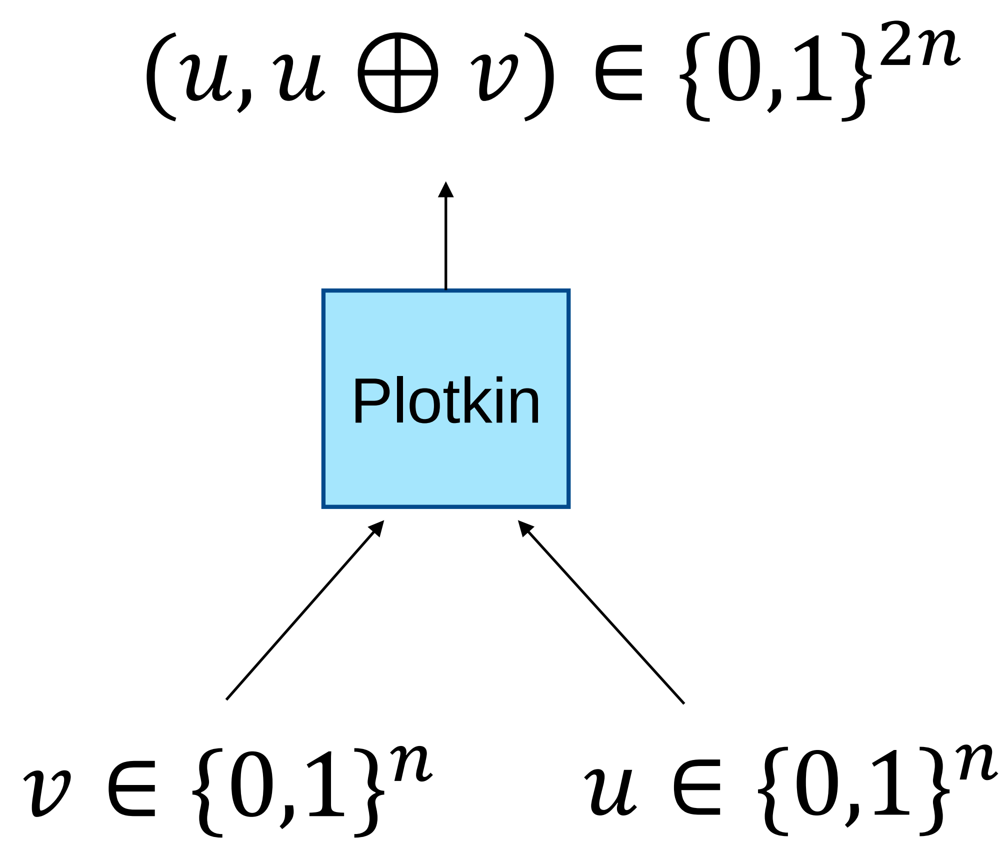
KO codes: An overview

- **Novel family of neural codes**
- Outperform both RM and Polar in certain regimes
- Fascinating properties

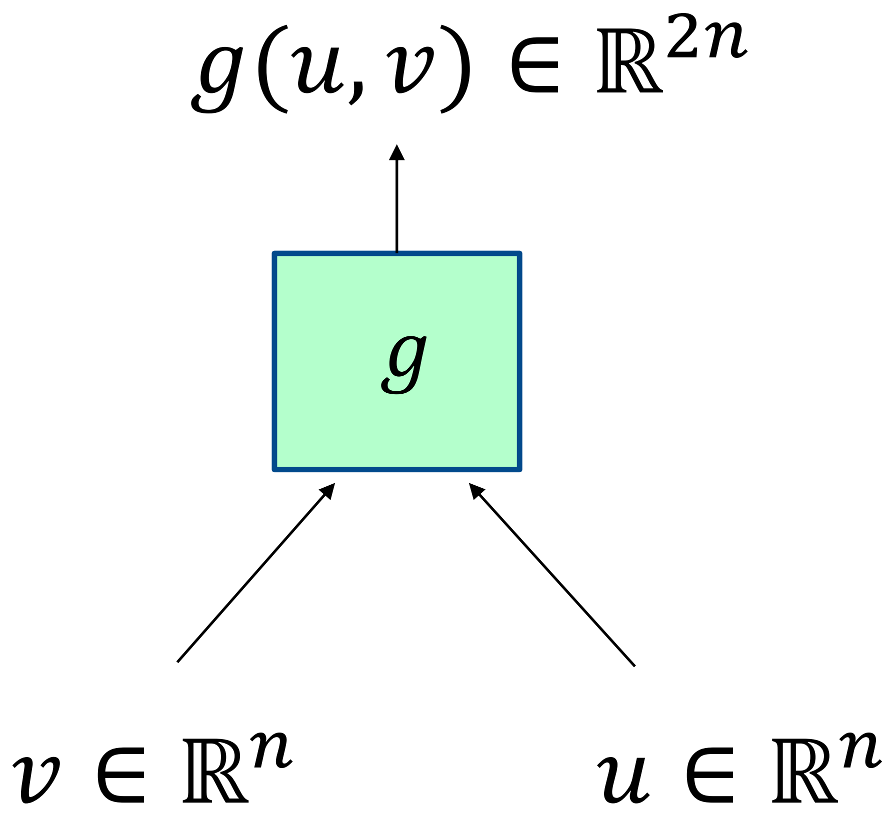
Encoding: RM and Polar



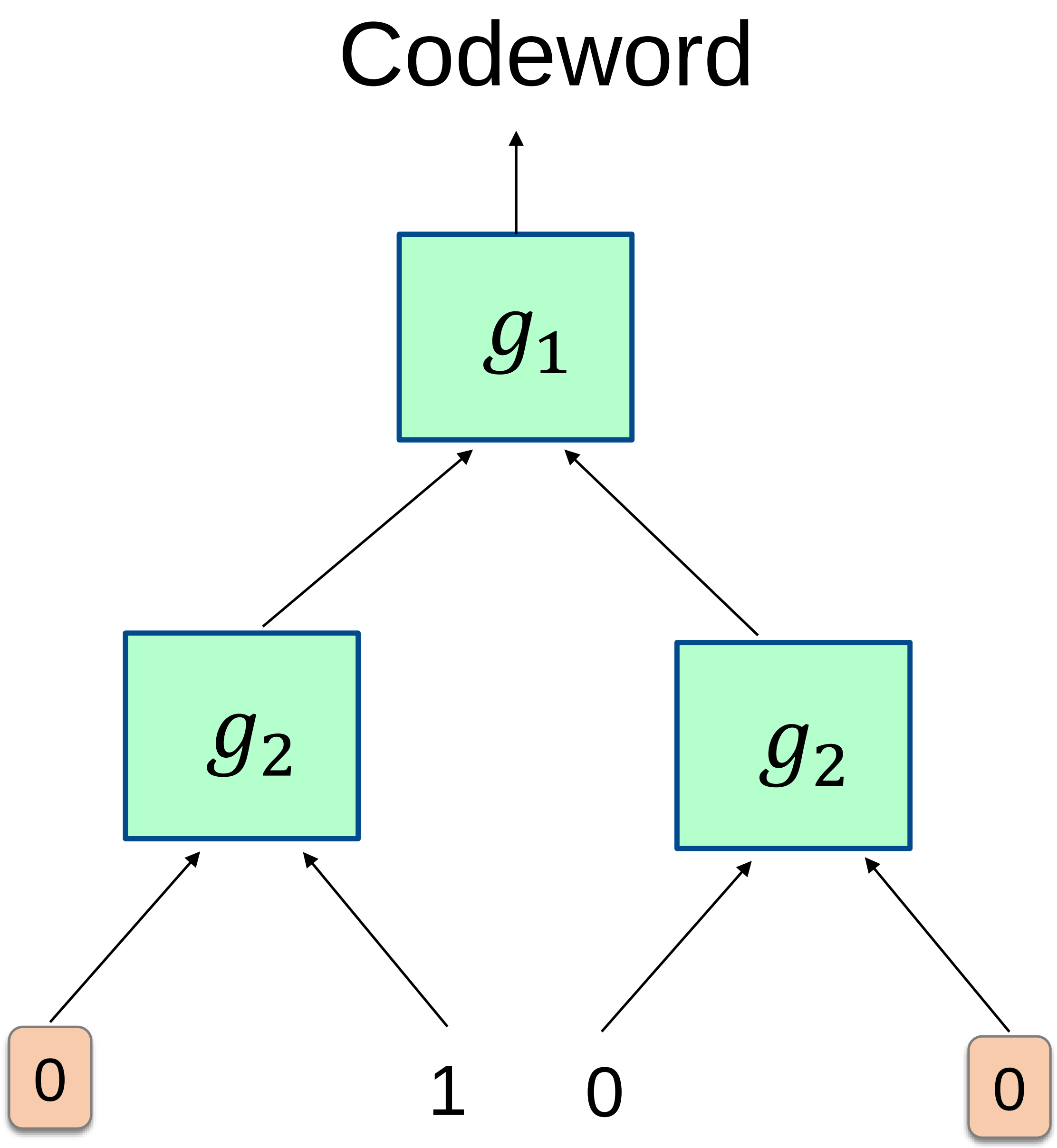
Plotkin mapping



KO neural network



KO encoder

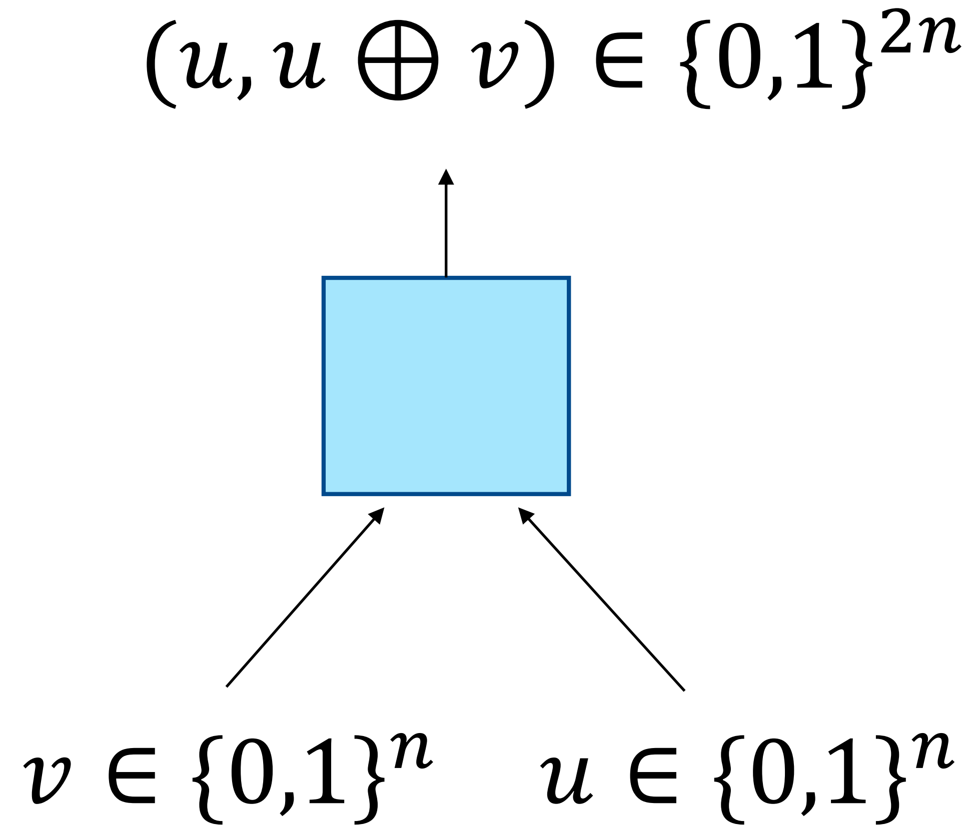


Decoder

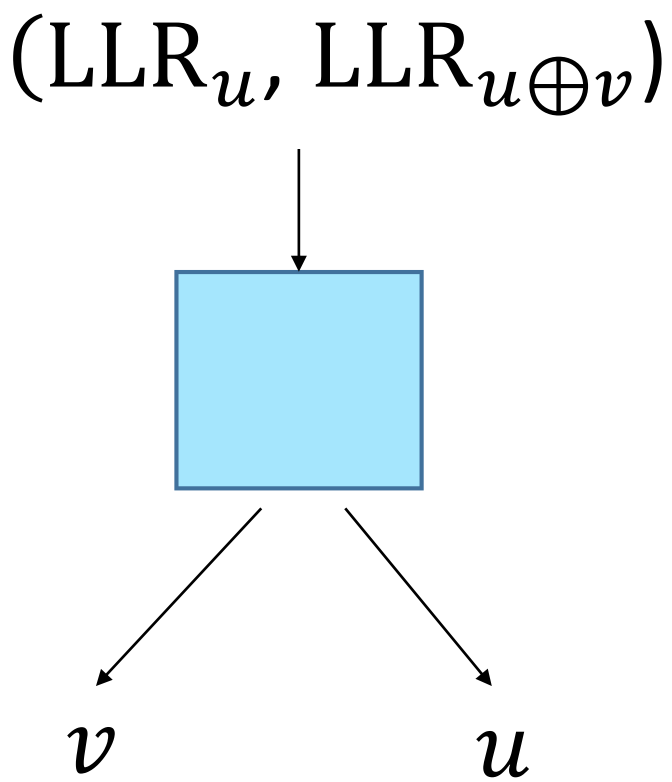
- Matching decoder for KO encoder?

- **Dumer's decoder / Successive Cancellation (SC)**

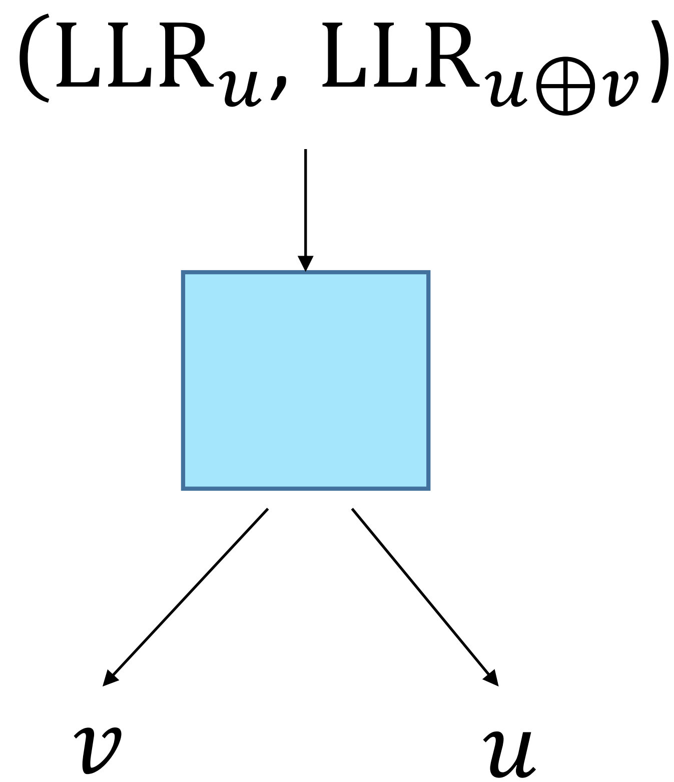
Plotkin revisited



Decoding

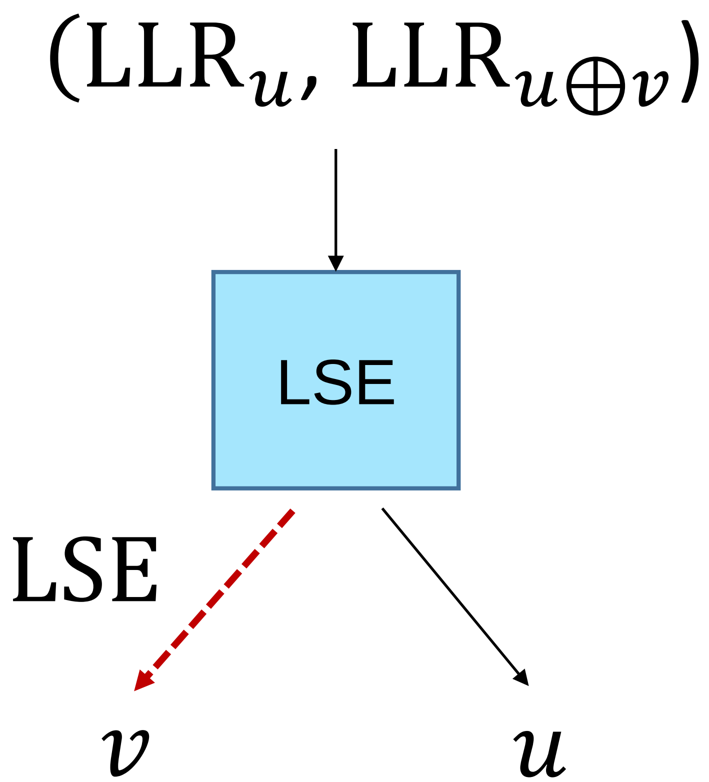


SC decoder



Dumer, 2004-06
Arikan, 2009

SC decoder

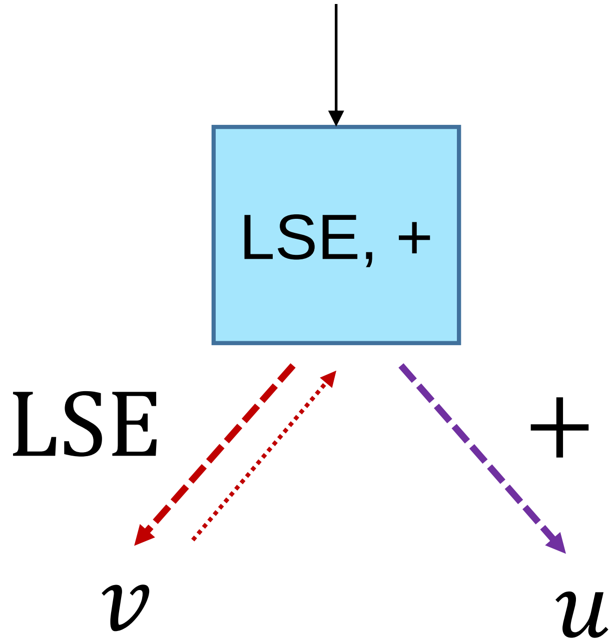


First Decode v

Dumer, 2004-06
Arikan, 2009

SC decoder

$(\mathrm{LLR}_u, \mathrm{LLR}_{u \oplus v})$

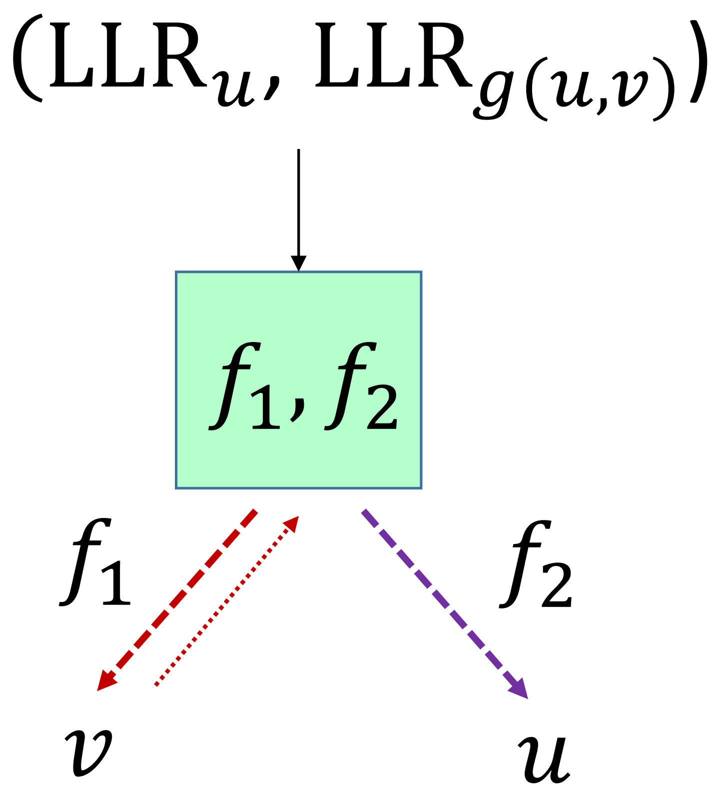


First Decode v

Next Decode u

Dumer, 2004-06
Arikan, 2009

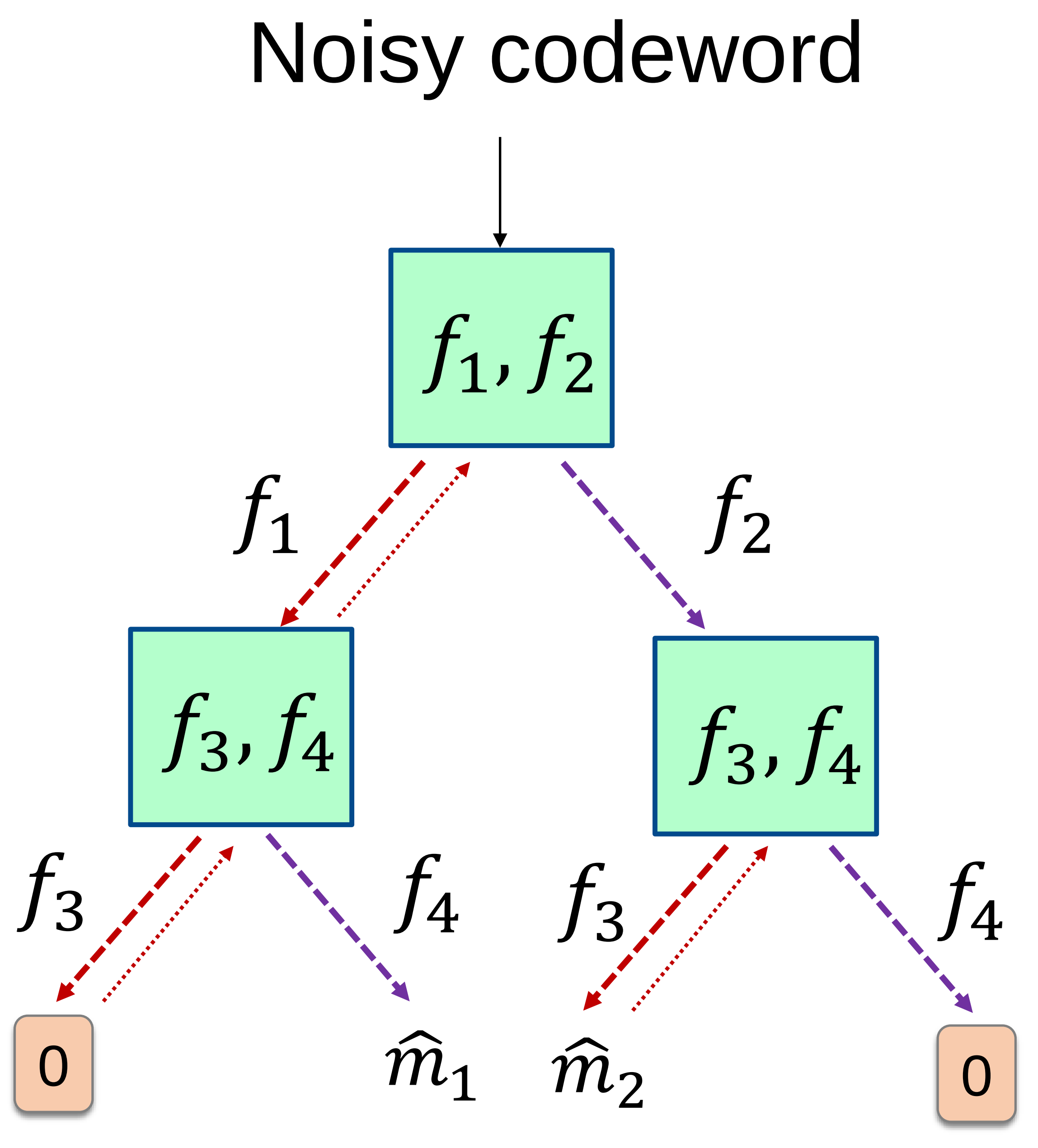
KO decoder



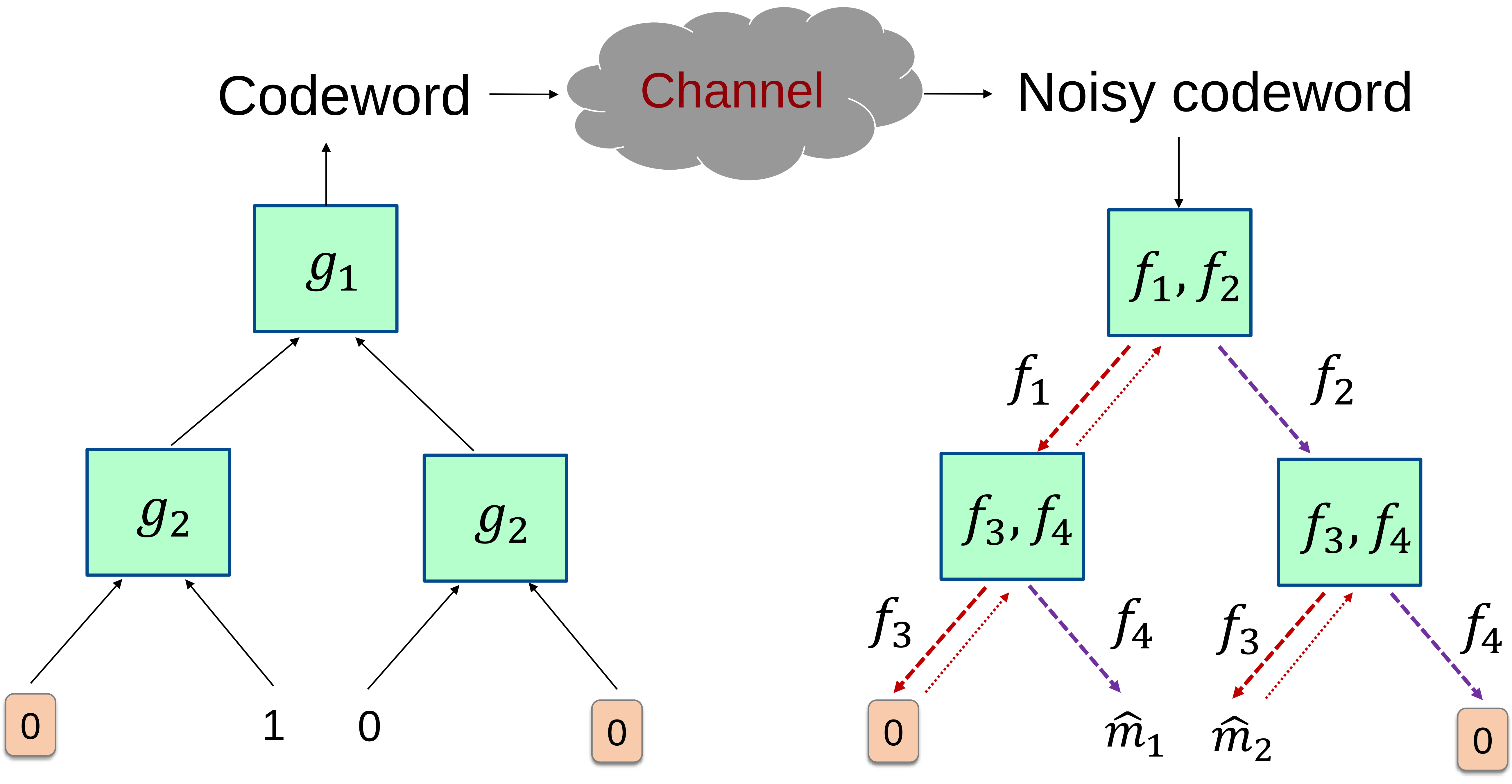
First Decode v

Next Decode u

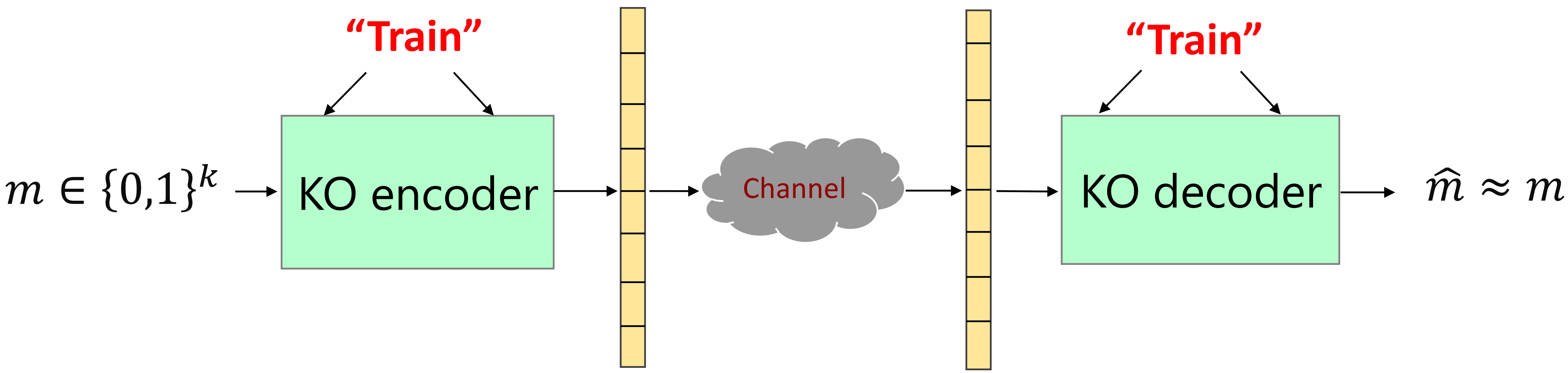
KO decoder



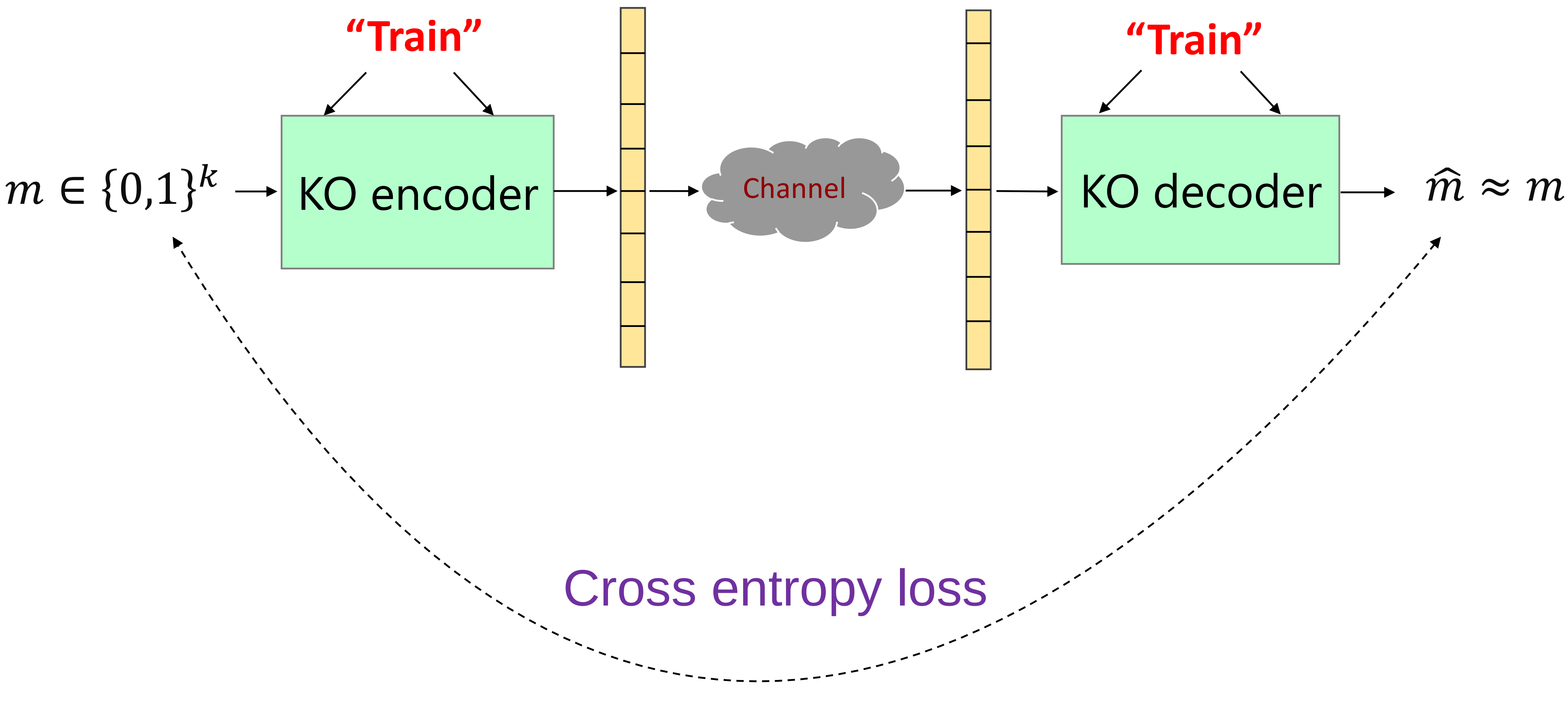
KO (encoder, decoder)



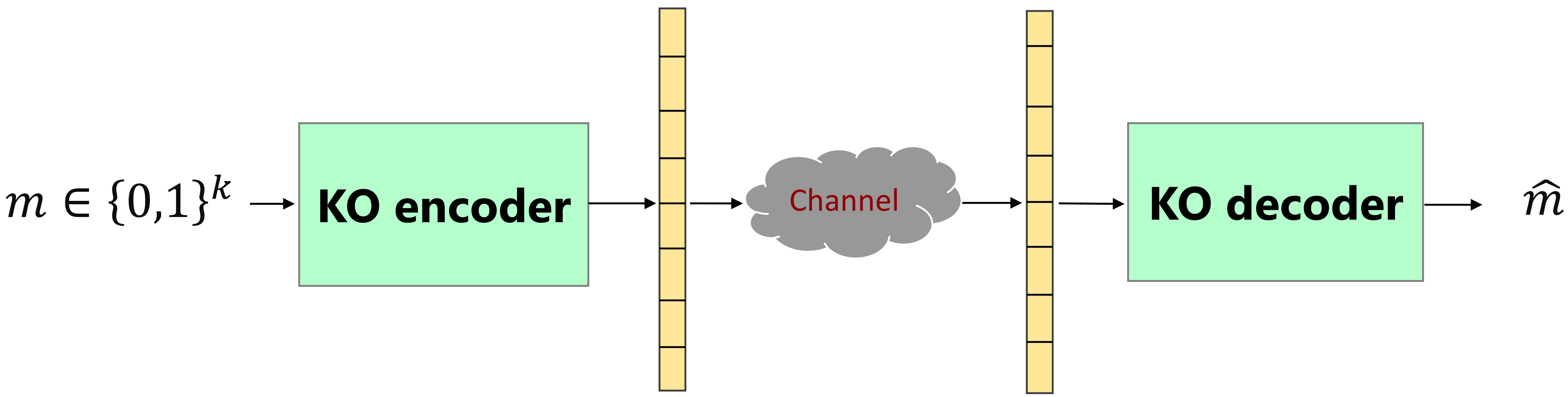
Training KO Codes



Training KO Codes



Testing



Performance metrics

- Reliability

- Computational complexity

Baselines

- KO codes vs. RM codes

- KO codes vs. Polar codes

Setup

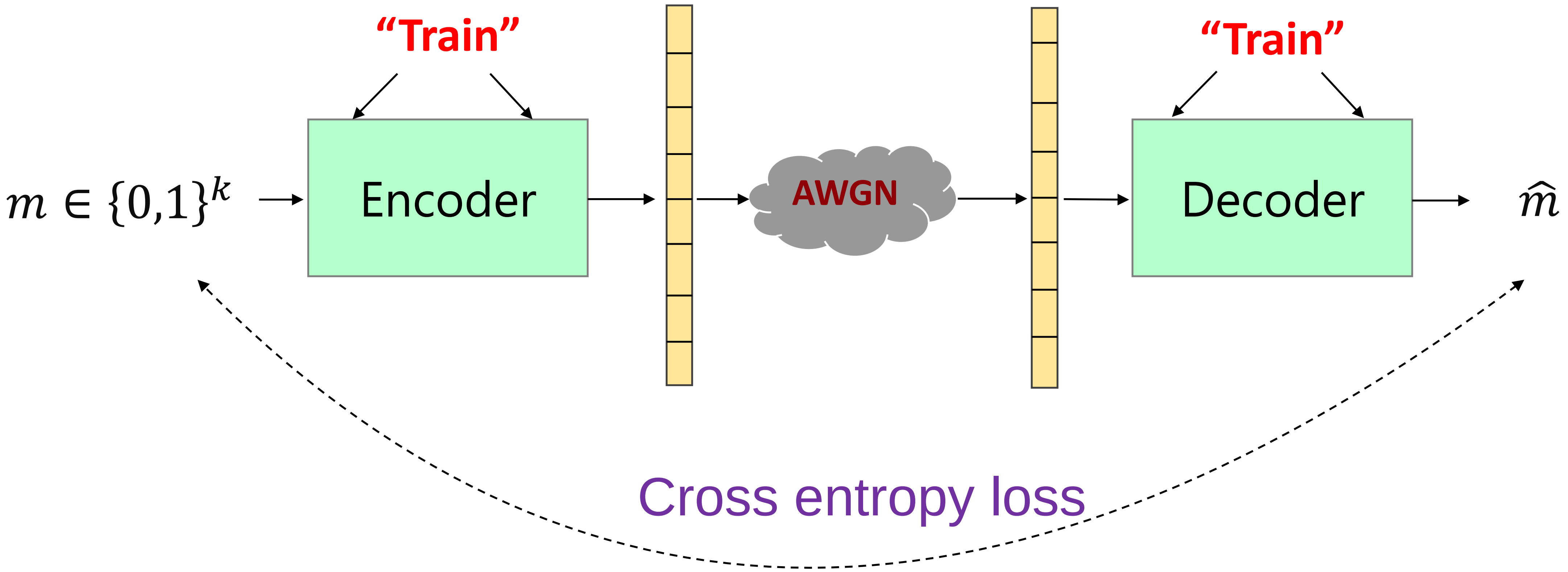
- Train and test on the same channel
  - AWGN
- Robustness: Train and test on different channels
  - Rayleigh fading

Setup

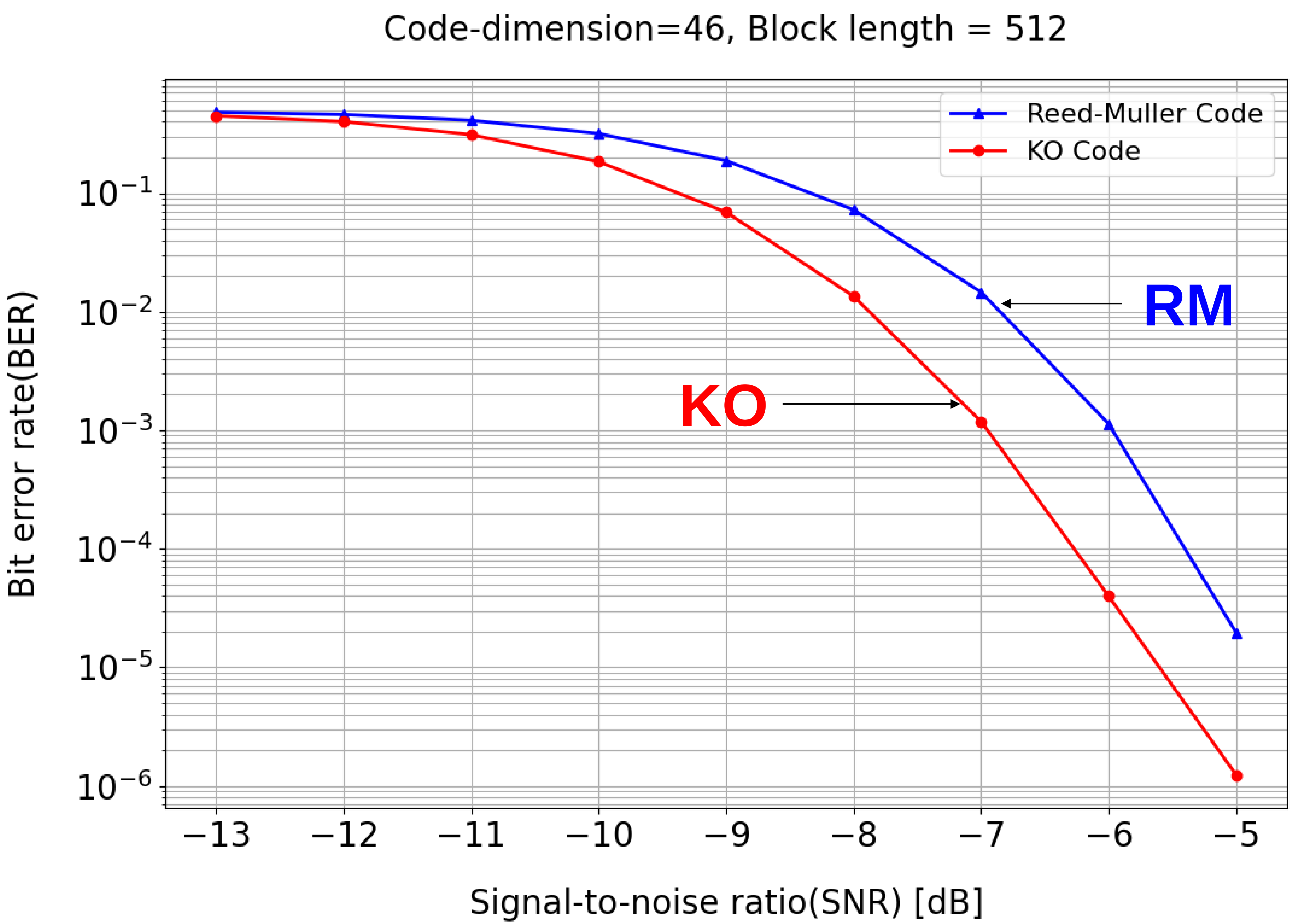
- Train and test on the same channel
  - AWGN
- Robustness: Train and test on different channels
  - Rayleigh fading

Setup #1: AWGN

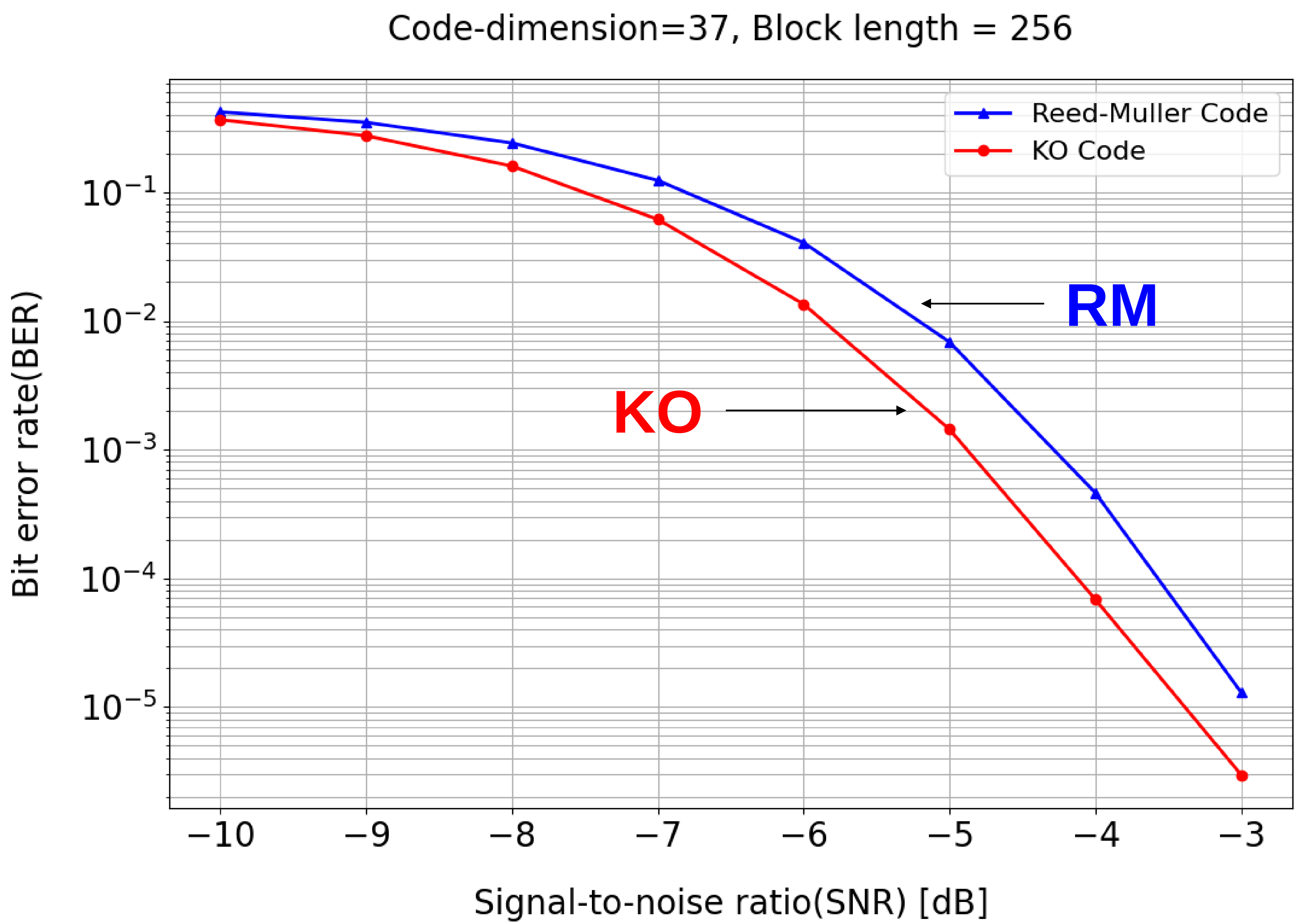
- Train and test on AWGN



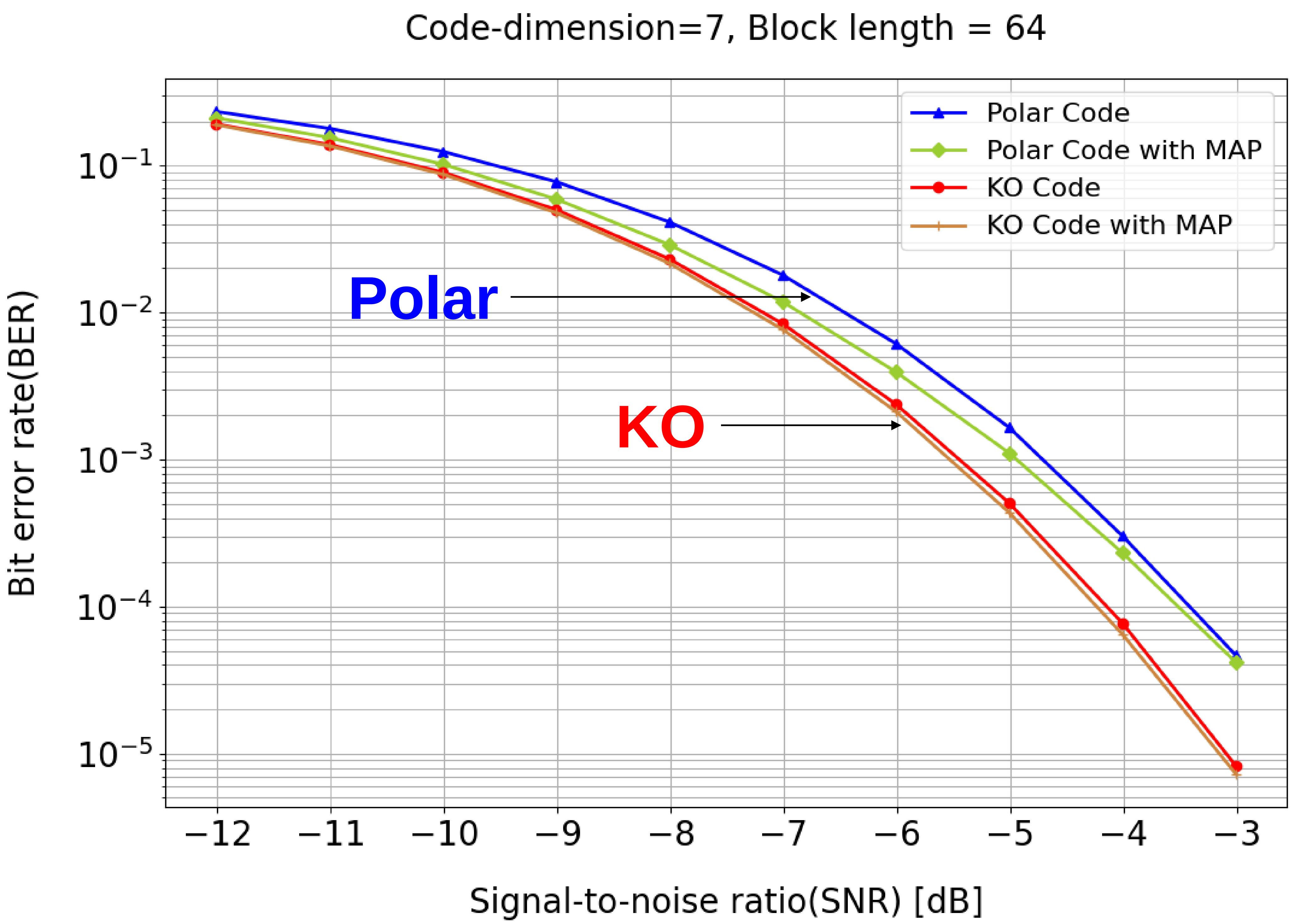
KO codes beat RM



KO codes beat RM

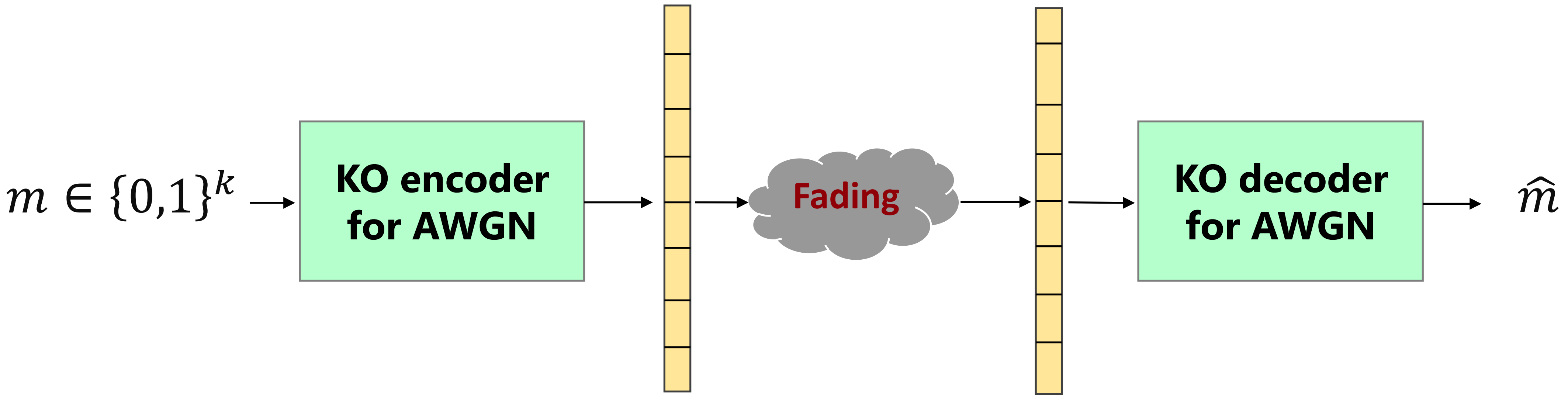


KO beats Polar

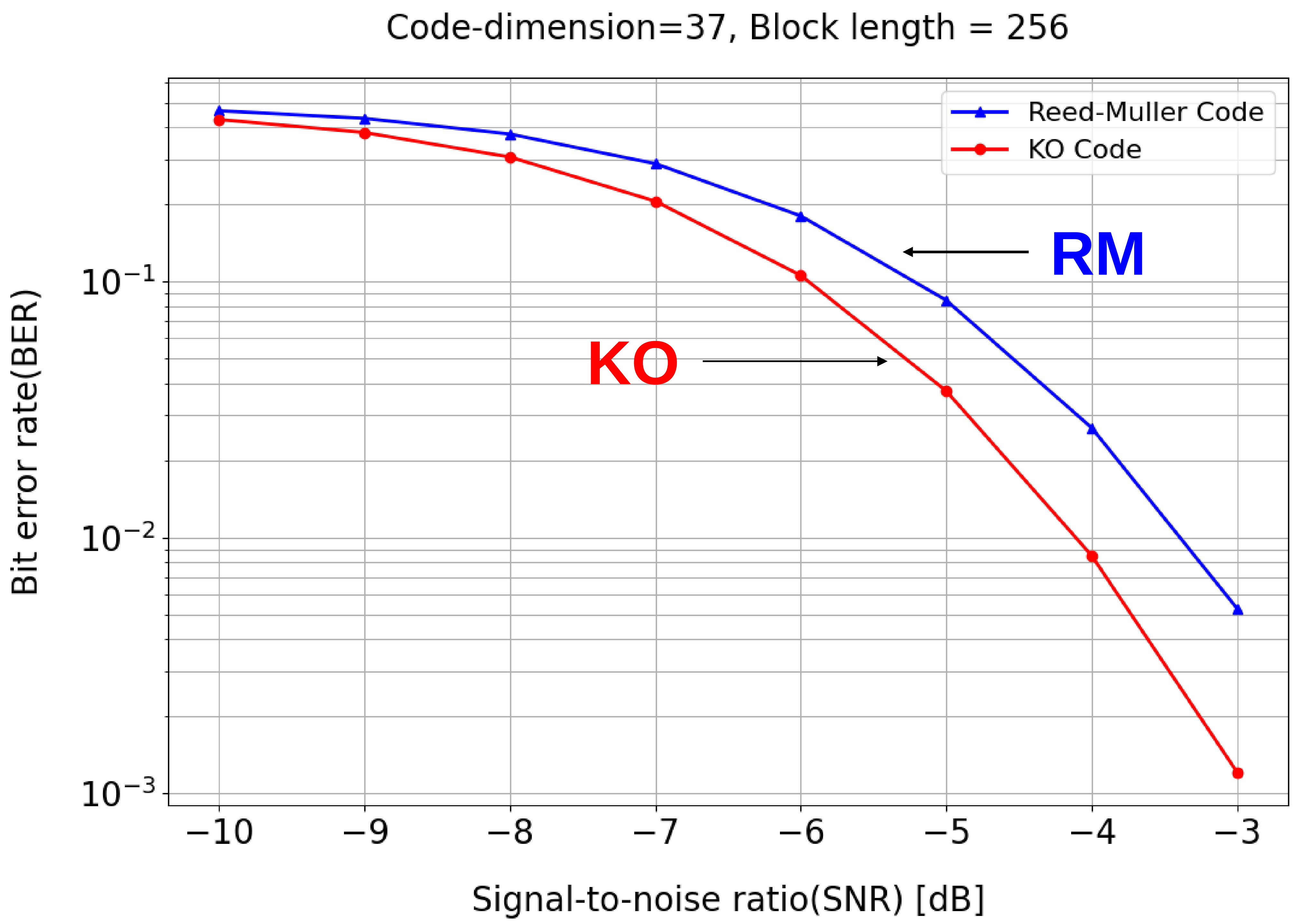


Setup #2: Robustness

- Train on AWGN → Test under Rayleigh fading



Robustness: Fading channel



Setup

- Train and test on the same channel
  - AWGN
- Robustness: Train and test on different channels
  - Rayleigh fading

Complexity

- Computational complexity: $O(n \log n)$
  - ➢ KO codes ≈ RM codes
- Number of operations
  - ➢ RM codes (11k) ≪ KO codes (550k)

Complexity

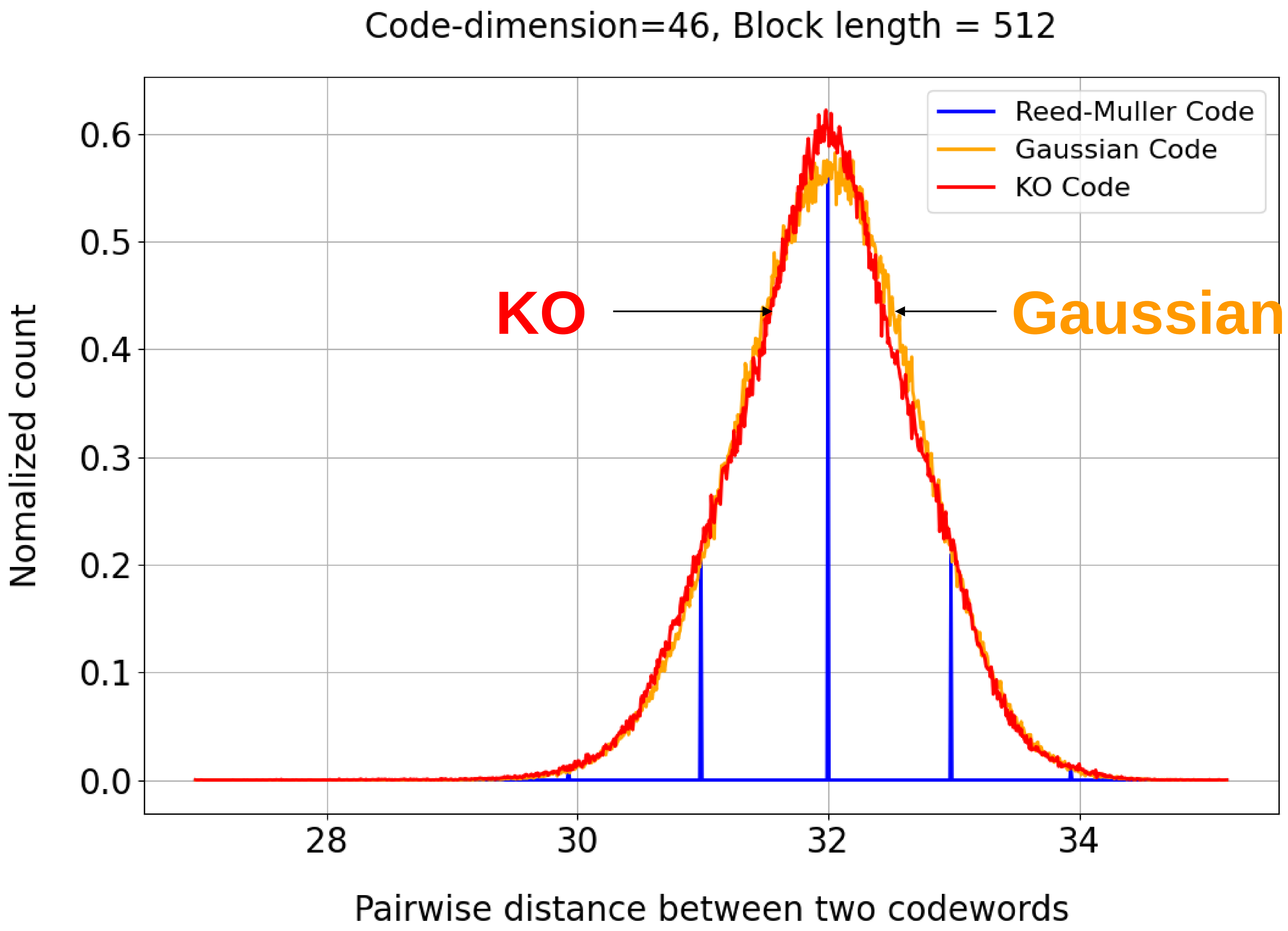
- Computational complexity: $O(n \log n)$
  - KO codes $\approx$ RM codes
- Number of operations
  - RM codes (11k) $\approx$ Tiny KO (44k) $\ll$ KO codes (550k)

Why are KO codes good

Why are KO codes good

- Surprising resemblance to Gaussian codes!

Gaussian like!



Future directions

- Training with complex decoding algorithms
  - Recursive Projection Aggregation (RPA)
  - SC + list decoder

Future directions

- Training with complex decoding algorithms
  - Recursive Projection Aggregation (RPA)
  - SC + list decoder

- Learning the frozen bits: Liao et al, 2020

Future directions

- Training with complex decoding algorithms
  - Recursive Projection Aggregation (RPA)
  - SC + list decoder
- Learning the frozen bits: Liao et al, 2020
- Commercialization
  - Hardware implementation

Long-term directions

- Discover new coding structures
  - Recursive: this work
  - Graph: LDPC
  - Sequential: convolutional
  - What is the best structure?

- Theoretical analysis

- Beyond point-to-point: Network coding

Collaborators


Any
Question

La Fin

Thank you!