HD Computing in Communication

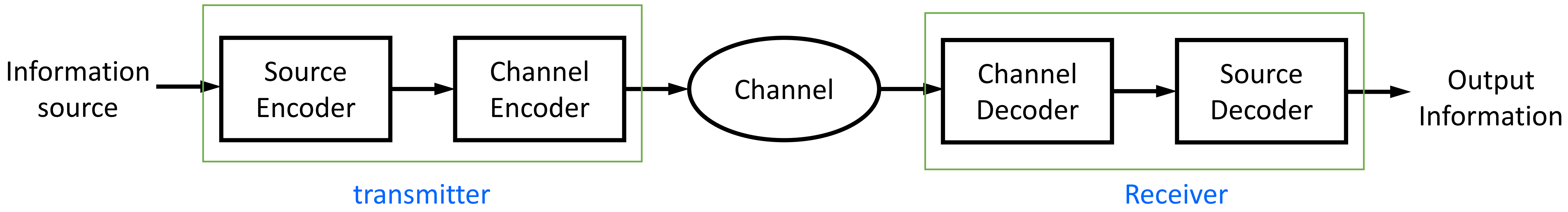
Ping-Chen Huang

Nov. 17, 2021

Outline

- Overview of communication problems
- Principles related to HD computing in communication
  - Law of large numbers & random codes
  - Orthogonality via random projection
- Applications of HD computing in communication
- AI/Machine learning in communication

Main Ideas from Shannon's paper



- Treat the information source as stochastic
- Separate the transmitter/receiver design into two sub-problems – "source-channel separation"

Why the madness of having two competing encoders?

Recall Module 7 (Fritz's lecture)

Mapping data to vector spaces

Source coding (remove redundancy in data)

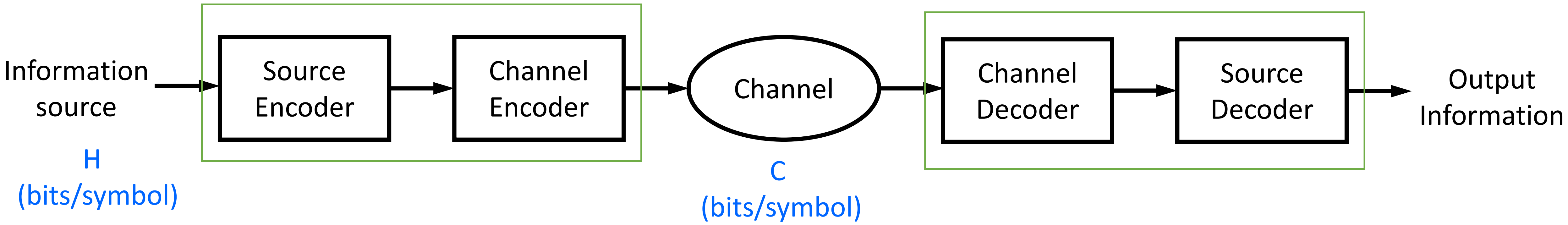
Data lie in subspace (SS)	*Learning method*	*Coordinates in SS*
Linear low-D SS	PCA	Axes of covariance matrix
Nonlinear low-D SS	Manifold learning	location on manifold
Clusters	Cluster analysis	Cluster number (+ loc.)
Union of lin. low-D SS	Sparse coding	Axes of Indep. Comp.
Union of nonlin. Low-D SS	Manifold learning	Manifold number + loc.

Vector encoding of the new coordinates

Feature local: a neuron's activity encodes a coordinate: PCA, ICA,…

Distributed: values of a coordinate are represented by many neurons:
VSA

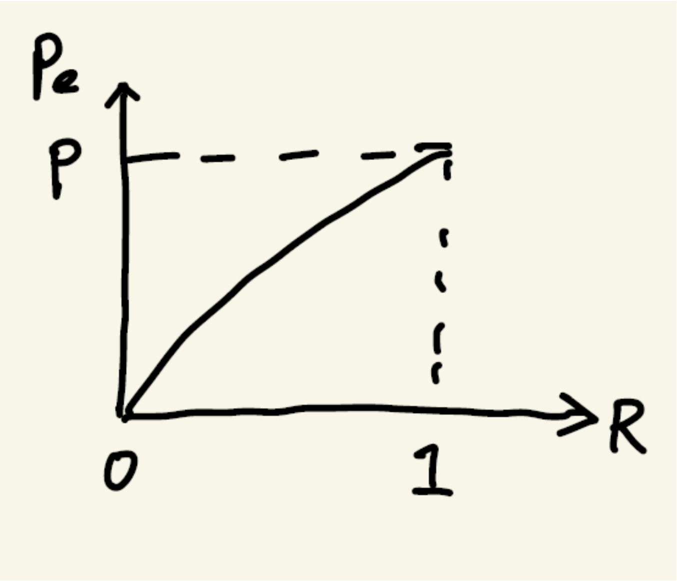
Main ideas from Shannon’s paper (cont’d)



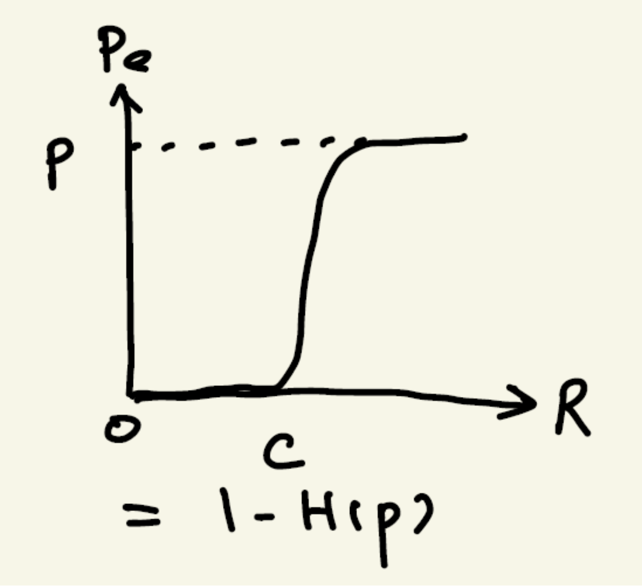
- Nice answers to both questions:
  i. Min. bits to represent a source = H (source entropy) bits/symbol
  ii. Max. rate of reliable communication over a channel = C (capacity) bits/use
- Source can be reliably communicated over a channel iff $H < C$

Surprise from this theory

- We can communicate with arbitrarily small error rates without operating at zero-rate



Repetition code

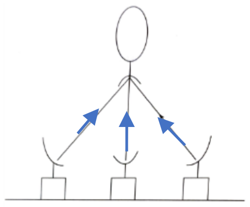


Capacity achieving code

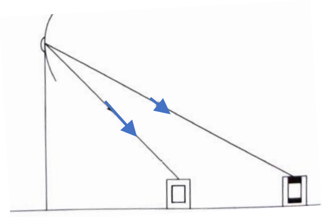
Eg. for binary symmetric channel with bit flip probability p

Open questions from Shannon’s paper

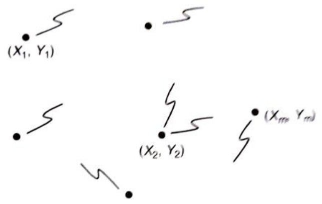
- How to design codes?
  - ⇨Source codes: Huffman, Lemple-Ziv (77, 78)
  - ⇨Channel codes: Turbo, LDPC, Polar, etc.
- Network Communication?
  - ⇨Distributed source compression: Slepian-Wolf Theory
  - ⇨Distributed communication
    - Multiple-Access Channel (MAC)
    - Broadcast Channel (BC)
    - General network communication



Multiple Access

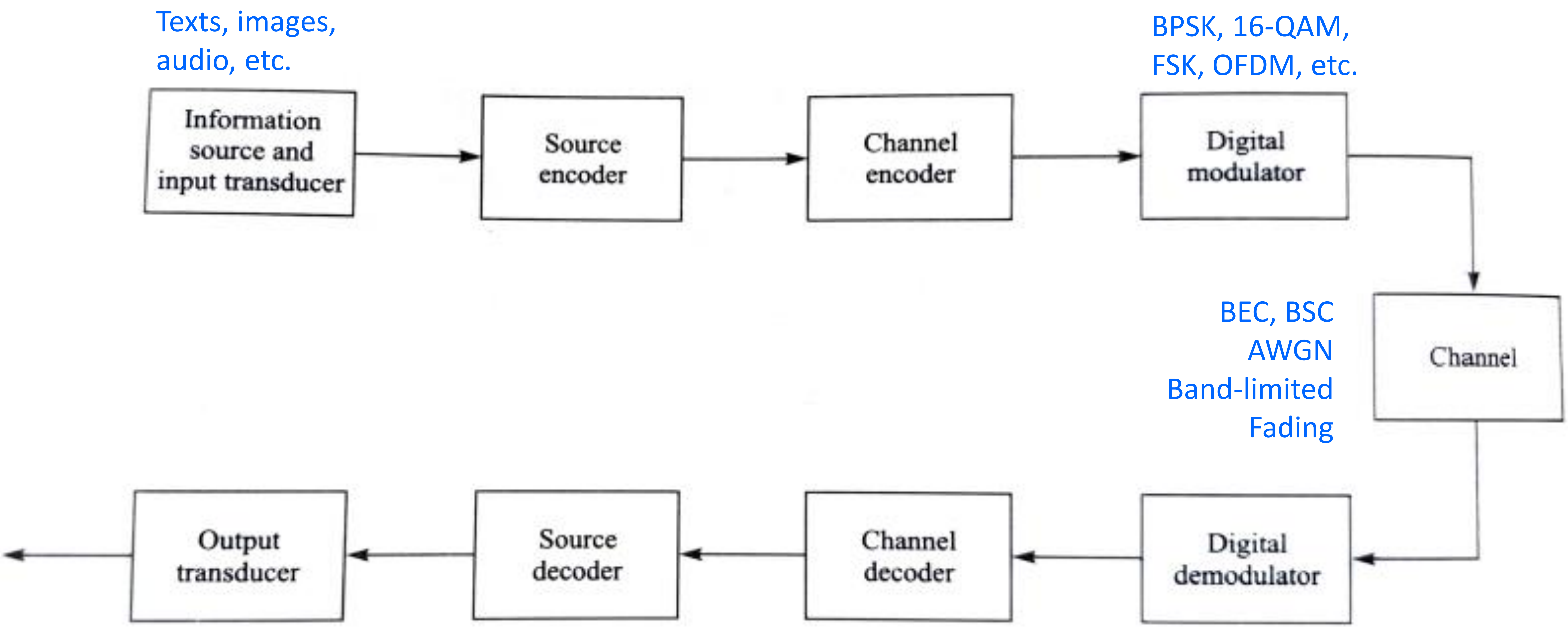


Broadcast



General

Elements in digital communication systems



Entropy and mutual information

- The entropy of a discrete random variable

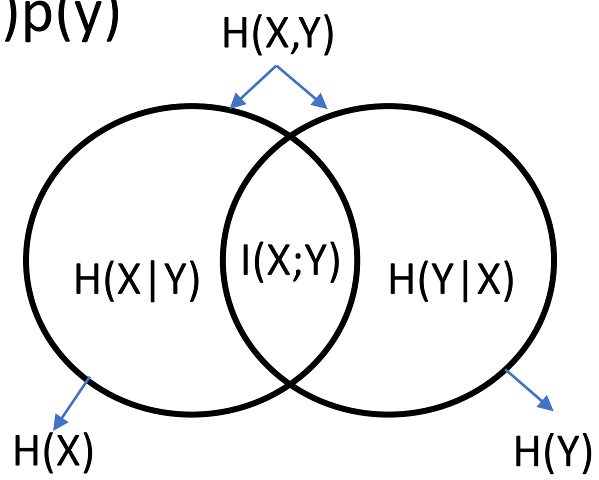
$$H(X) = -\sum_{x \in \mathcal{X}} p(x) log\, p(x)$$

- The joint entropy of a pair of discrete random variables (X,Y) with joint distribution p(x,y)

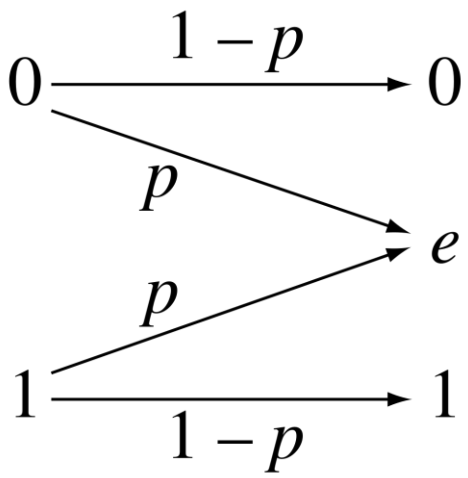
$$H(X,Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) log\, p(x,y)$$

- Mutual information is the relative entropy between p(x,y) and p(x)p(y)

$$\begin{aligned} I(X;Y) &= \sum_{x,y} p(x,y) log \frac{p(x,y)}{p(x)p(y)} \\ &= H(X) + H(Y) - H(X,Y) \\ &= H(Y) - H(Y|X) \end{aligned}$$



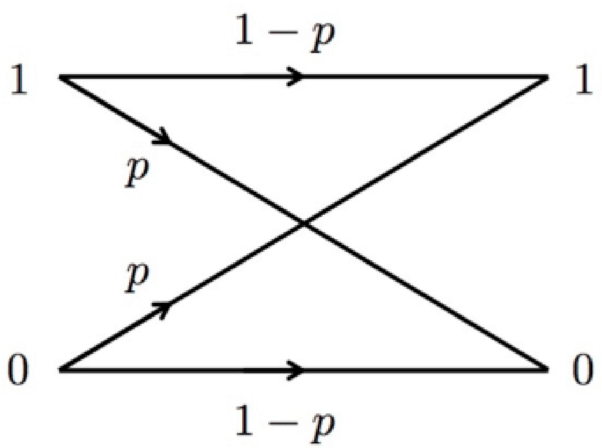
Basic Channel Models



Binary Erasure Channel (BEC)

$$C = 1 - p$$

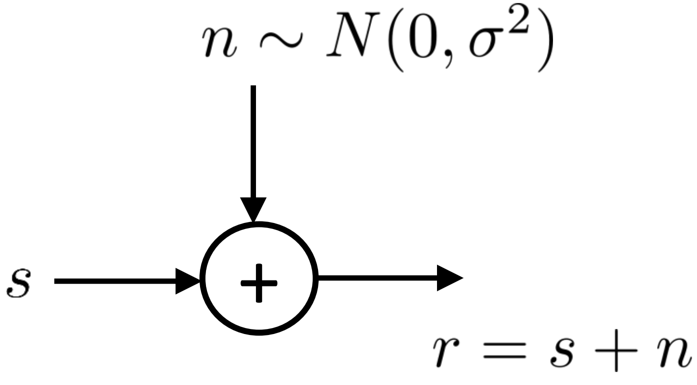
[bits/use]



Binary Symmetric Channel (BSC)

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

[bits/use]



Additive White Gaussian Noise Channel (AWGN)

$$C = \frac{1}{2} log(1 + \frac{p}{\sigma^2})$$

[bits/use]

Asymptotic Equipartition Property (AEP)

- The analog of the law of large numbers in information theory:

  If X_1, X_2, … are i.i.d. ~ $p(x)$, then

  $$-\frac{1}{n} log\, p(X_1, X_2, ..., X_n) \rightarrow H(X) \quad \textit{in probability.}$$

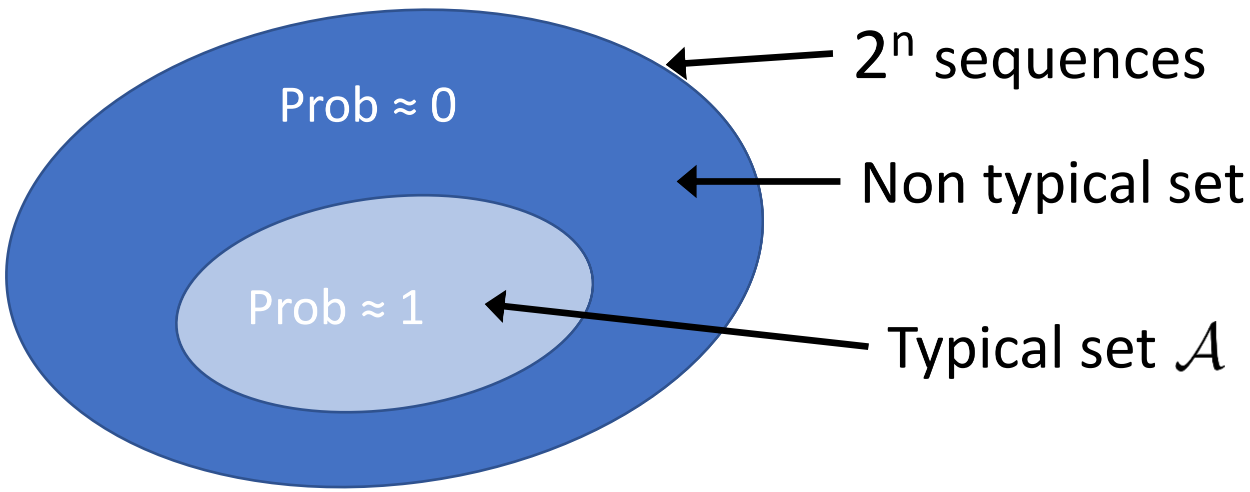
- AEP holds when the source has memory (Markov source)

Typical Sequence

- Example: X={a,b}, p(a) = p, p(b)=1-p, where p ≠ 0.5

Look at blocks of length *n*, we expect roughly *np* a's and *n(1-p)* b's in the sequences will occur with highest probability ⇦ typical sequence $\mathbf{x}$

$$p(\mathbf{x}) = p^{np}(1-p)^{n(1-p)} = 2^{-nH(X)}$$



$$p(\mathcal{A}) \to 1$$

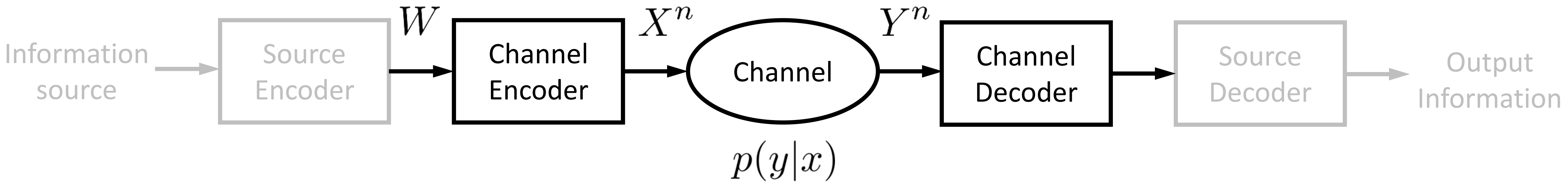
$$\Rightarrow |\mathcal{A}| \to \frac{1}{p(\mathbf{x})}$$

Source Coding Theorem:

Min # bits to represent the source:

$$\frac{nH(X)}{n} = H(X) \text{ bits/symbol}$$

Channel Coding Setup

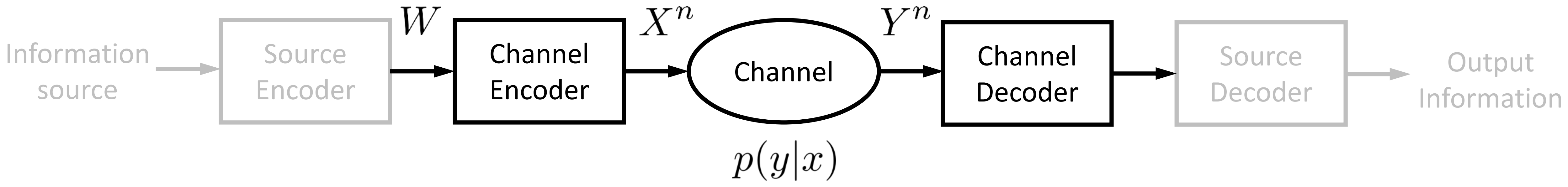


- W is uniform in $\{1, 2, \cdots, 2^{nR}\}$
  - n: block length
  - R: rate [bits/usage]
- Up to which R can we still reliably communicate over the channel?

Channel Coding Theorem:

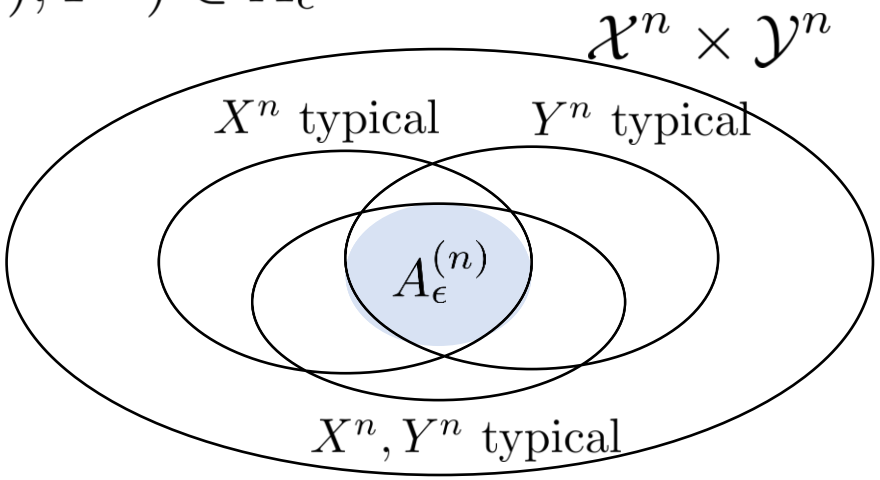
For any $R < C$, there exists a sequence of $(2^{nR}, n)$ codes, such that $Pr(W \neq \hat{W}) \to 0$ as $n \to \infty$

Proof of the Channel Coding Theory using random codebooks and the joint typical decoder



- Coder: i.i.d. randomly chosen according to $p(x)$
- Decoder: Choose unique w such that $(X^n(w), Y^n) \in A_\epsilon^{(n)}$

$$Pr(A_\epsilon^{(n)}) \to 1 \text{ as } n \to \infty$$



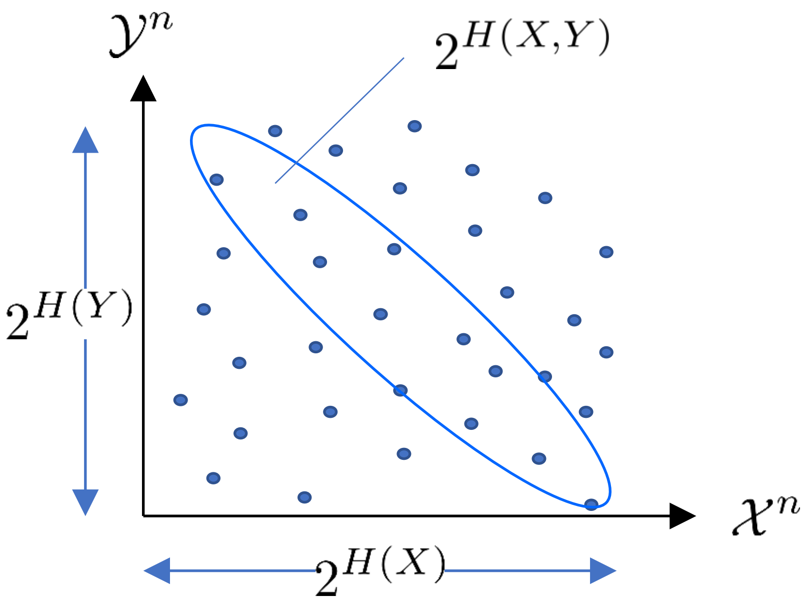
Proof of the Channel Coding Theory (cont'd)

- Assume W=1 is transmitted, there are two error events:
  - $E_0 := (X^n(1), Y^n) \notin A_\epsilon^{(n)}$
  - $E_i := (X^n(i), Y^n) \in A_\epsilon^{(n)}$ for $i \geq 2$

$$Pr(\mathcal{E}|W=1) = Pr(E_0 \cup E_1 \cup \cdots \cup E_{2^{nR}}|W=1)$$

$$\leq \cancel{Pr(E_0|W=1)} + \sum_{i=2}^{2^{nR}} \boxed{Pr(E_i|W=1)}$$

$$\frac{2^{nH(X,Y)}}{2^{n(H(X)+H(Y))}} = 2^{-nI(X;Y)}$$



Proof of the Channel Coding Theory (cont'd)

- $Pr(\mathcal{E}|W=1) \leq (2^{nR}-1)\cdot 2^{-nI(X;Y)}$

  $\Rightarrow$If $R < \max_{p(x)} I(X;Y) := C$, then $Pr(\mathcal{E}) \to 0$ as $n \to \infty$.

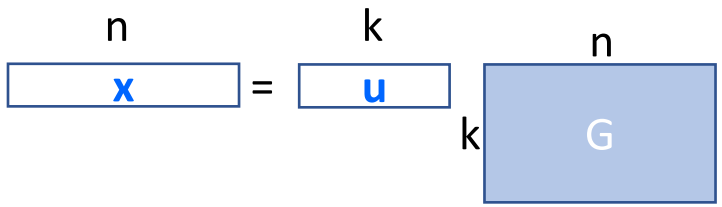
- $P(\mathcal{E}) = E_{\mathcal{C}}(Pr(\mathcal{E}|\mathcal{C}))$

  $\Rightarrow \forall n,\ \exists \mathcal{C}_n$ such that $Pr(\mathcal{E}|\mathcal{C}_n) \to 0$

The proof is *non-constructive* – it shows a good code exists, but does not give the explicit encoding scheme.

Random Codes in Reality

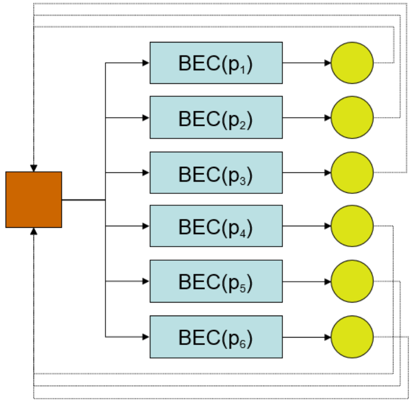
- **Linear block codes** with random generator matrix
  - $\mathbf{x} = \mathbf{u}G$
  - $G_{ij} \in \{0, 1\}$ i.i.d. with 0 and 1 equally likely



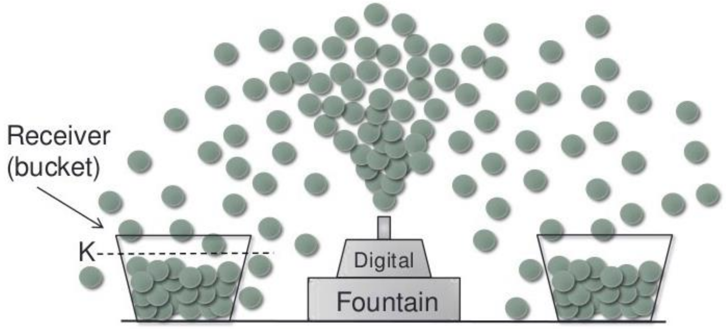
- For BEC and BSC channels, there exists a G_n that approaches capacity.
- Although non-structured G is easy to encode, it's hard to decode:
  - $\min ||\mathbf{u}G - \mathbf{y}||$
  - Decoding complexity $\sim O(n^3)$ (NP-hard)

Random Codes in Reality

- **Fountain codes** are a new class of random linear codes with
  - Sparse generator matrix
  - Decoding complexity $O(n\log(n/\delta))$ with decoding probability $>1-\delta$
- Approaches capacity for erasure channels with unknown erasure probabilities
  - Sender sends a stream of encoded bits.
  - Receivers collect bits until they are reasonably sure that they can recover the content from the received bits, and send STOP feedback to sender.
  - Automatic adaptation: Receivers with larger loss rate need longer to receive the required information.
- Practical realizations
  - L-T code
  - Raptor code
  - Online code



Credit: [A. Shokrollahi]



[credit: Jose Lopes]

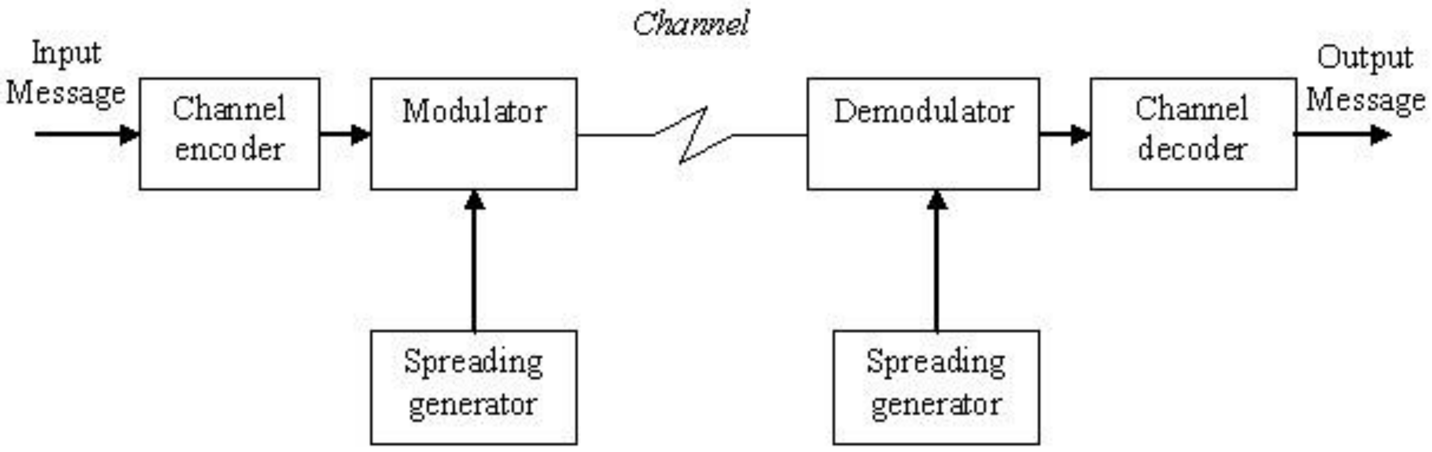
Remarks about channel codes

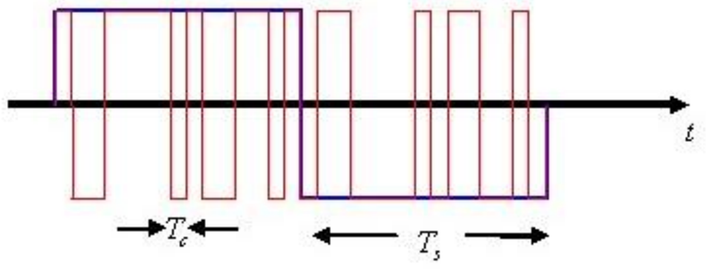
- Important considerations
  - Capacity achieving
  - Efficient decoding
    - ⇨Find codes with good structures of G (e.g. LDPC code, R-S code)
    - ⇨Suboptimal decoding schemes (iterative decoding)
- Current 5G standard
  - LDPC code + iterative message passing (decoding complexity $O(n^2)$-$O(n)$)
  - Polar code + successive cancellation (decoding complexity $O(n\log n)$)
- HD computing offers error correction, but its power lies beyond this

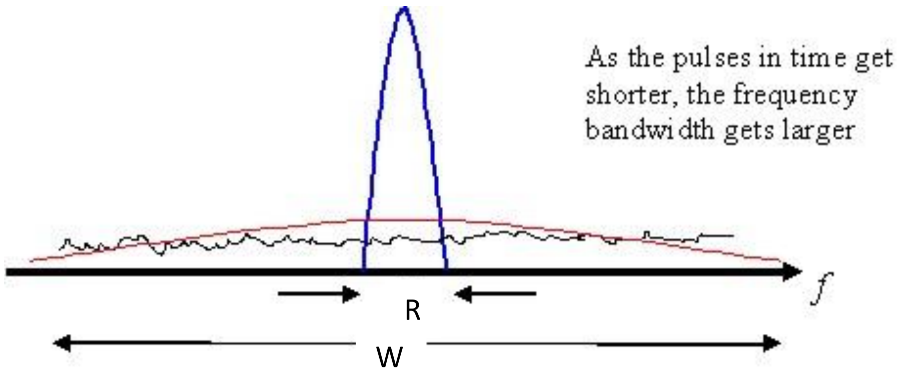
Outline

- Overview of communication problems
- Principles related to HD computing in communication
  - Law of large numbers & random codes
  - Orthogonality via random projection
- Applications of HD computing in communication
- Machine learning in communication

Direct Sequence Spread Spectrum





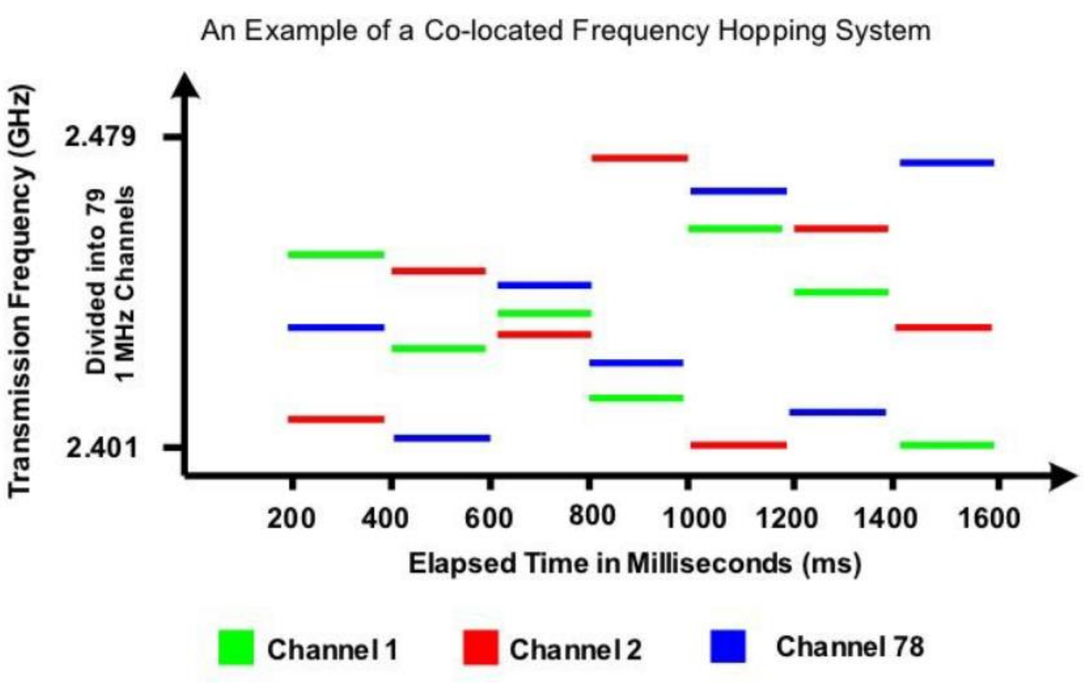


- Applications
  - Overcome jammers
  - Message privacy
  - Multiple access (CDMA)

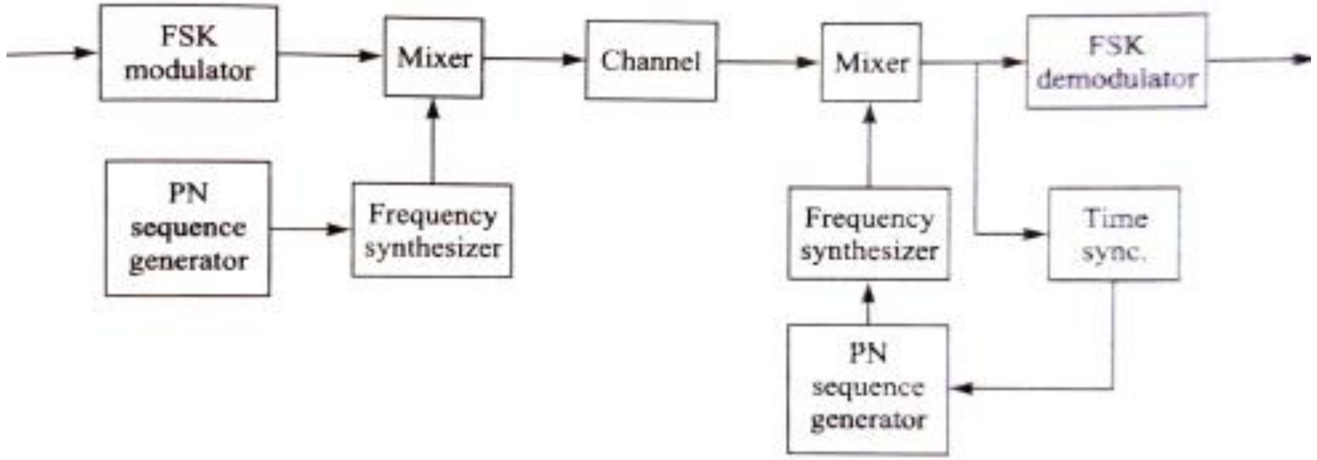
Bandwidth expansion factor (processing gain) $\frac{W}{R} = \frac{T_s}{T_c}$

$$\frac{E_b}{J_0} = \frac{2W/R}{J_{av}/P_{av}}$$

Frequency Hopping Spread Spectrum

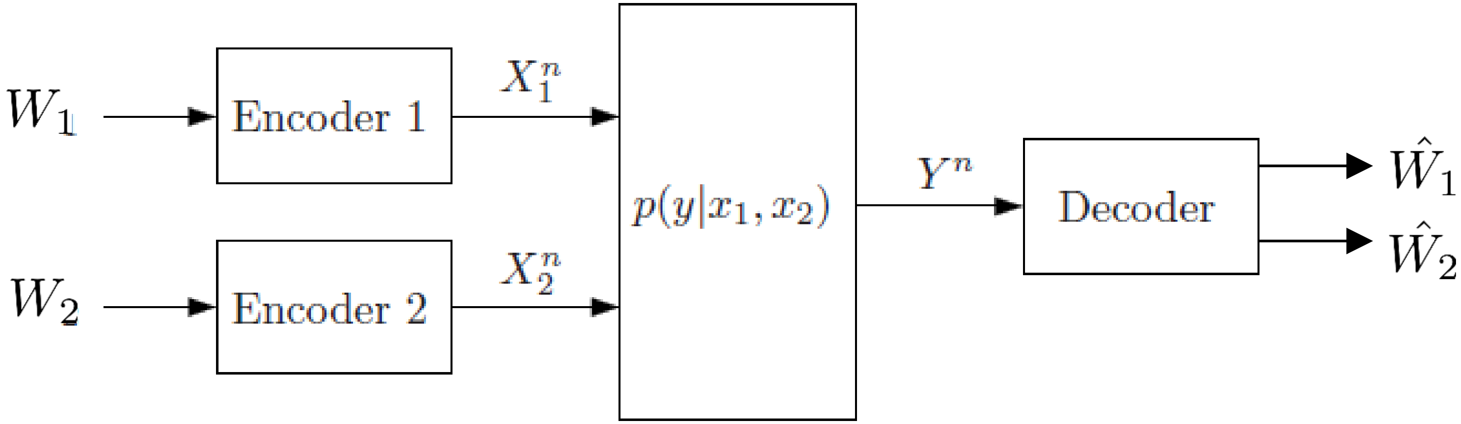


https://www.slideshare.net/HILDA519/spread-spectrum-modulation



[Credit: Proakis]

Capacity of the multiple access channel



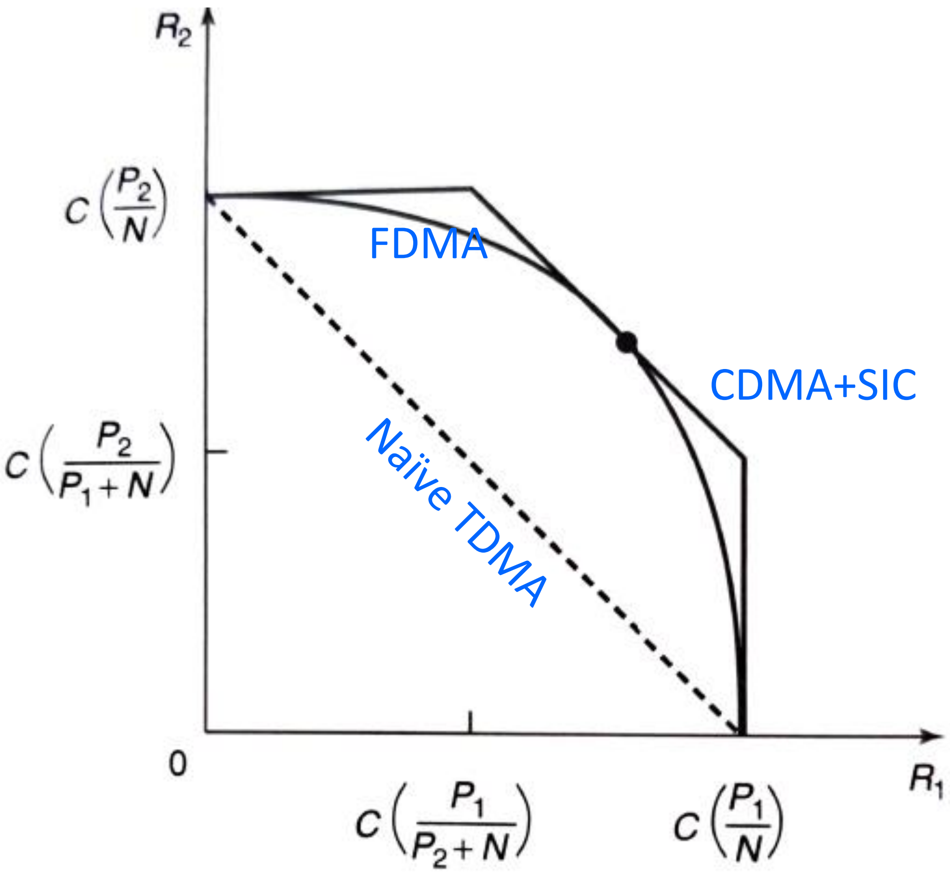
For Gaussian channels

$$Y = X_1 + X_2 + Z \text{ where } Z \sim N(0, \mathcal{N})$$

$$R_1 \leq I(X_1;Y|X_2) \leq \frac{1}{2}log(1+\frac{P_1}{N})$$

$$R_2 \leq I(X_2;Y|X_1) \leq \frac{1}{2}log(1+\frac{P_2}{N})$$

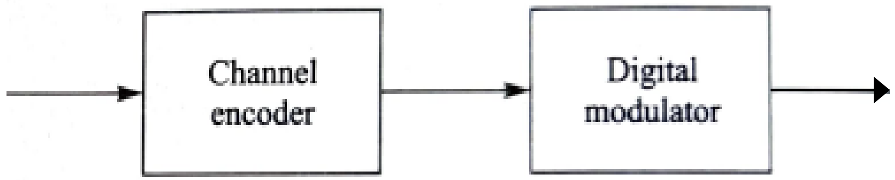
$$R_1 + R_2 \leq I(X_1,X_2;Y) \leq \frac{1}{2}log(1+\frac{P_1+P_2}{N})$$

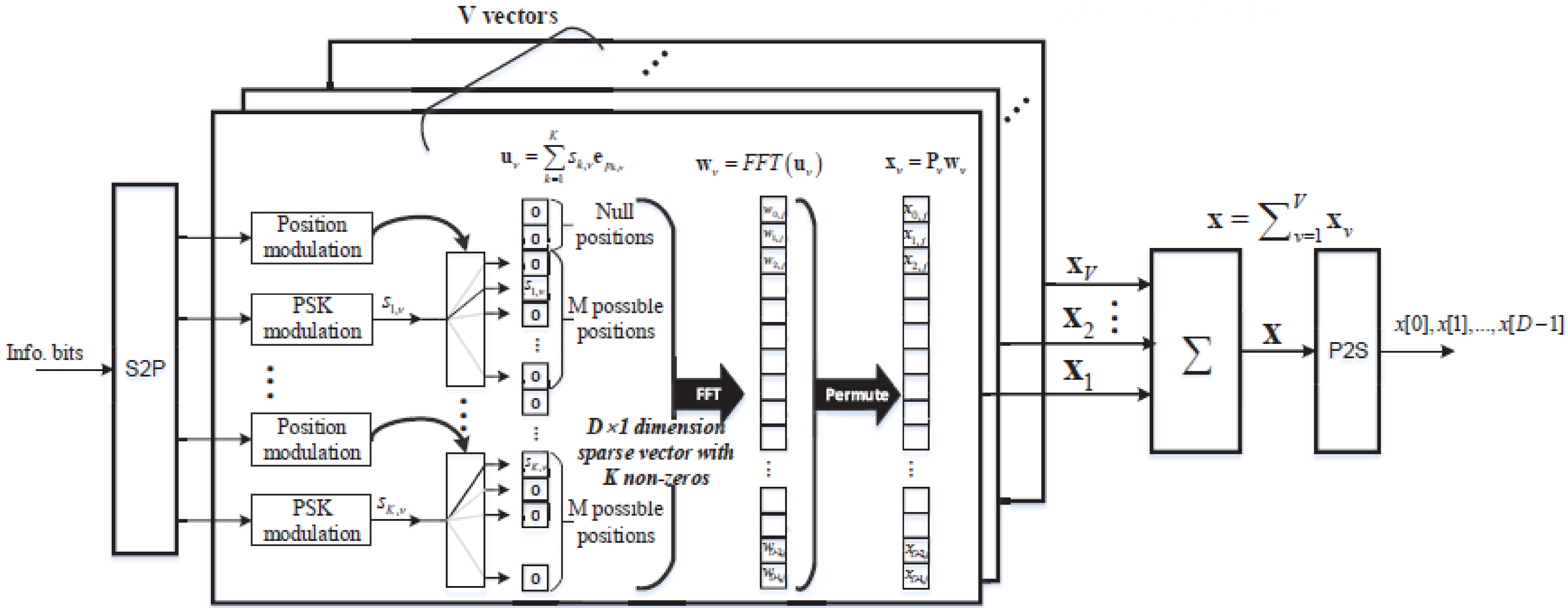


Outline

- Overview of communication problems
- principles related to HD computing in communication
  - Law of large numbers & random codes
  - Orthogonality via random projection
- Applications of HD computing in communication
- Machine learning in communication

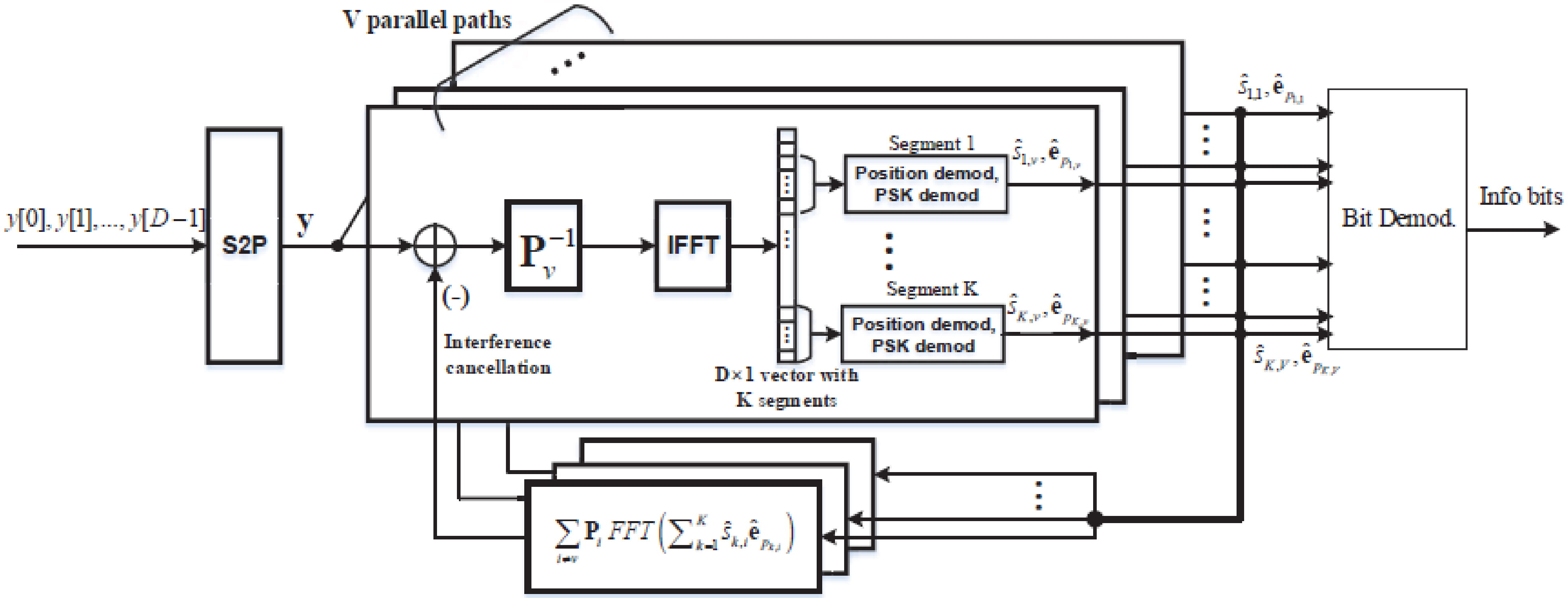
Hyperdimensional Modulation





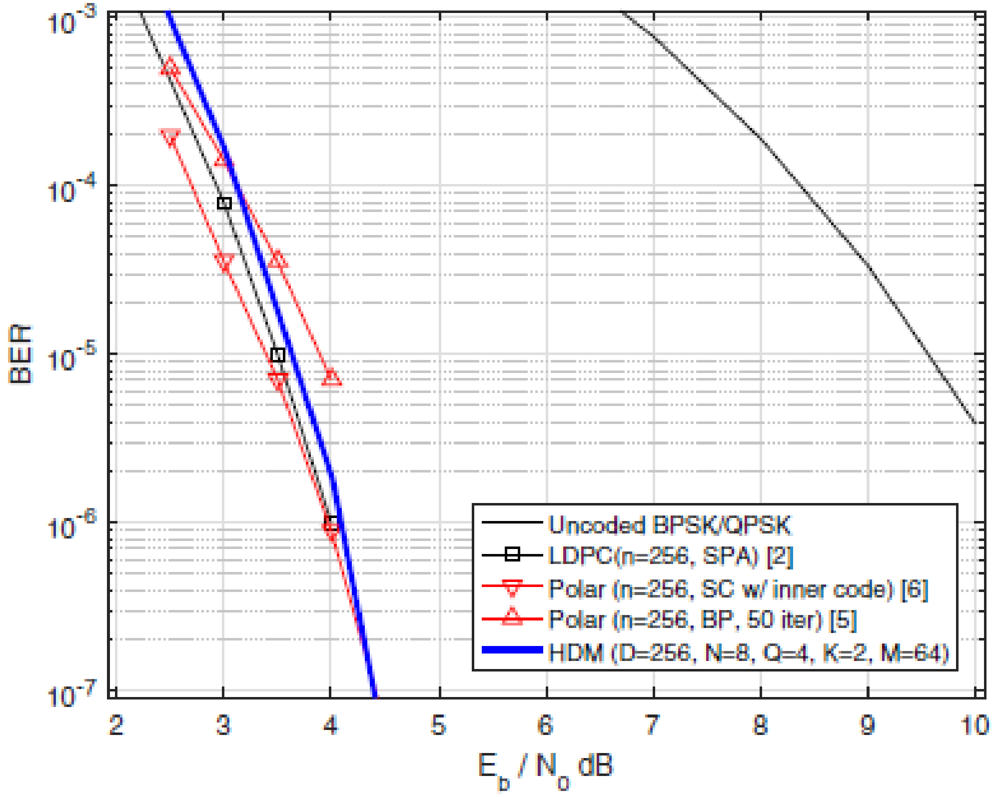
Kim, Hun-Seok. "HDM: Hyper-dimensional modulation for robust low-power communications." *2018 IEEE International Conference on Communications (ICC)*.

Hyperdimensional Modulation



Kim, Hun-Seok. "HDM: Hyper-dimensional modulation for robust low-power communications." *2018 IEEE International Conference on Communications (ICC)*.

Hyperdimensional Modulation



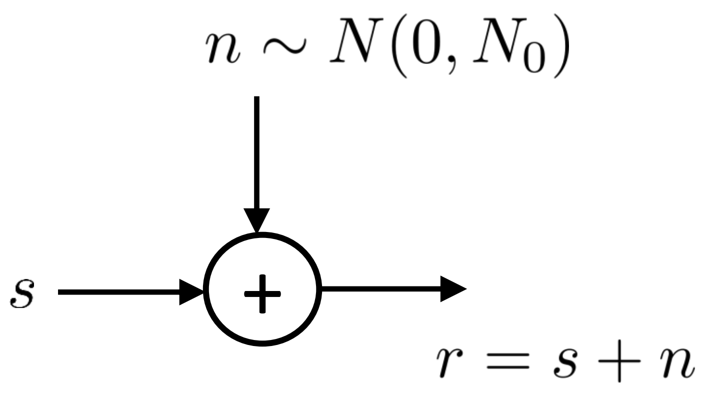
	HDM	LDPC [4]	LDPC [2]	Polar [7]
D or blk. length	256	256	256	256
Avg. I_{iter}	2.5	5	2.7	6
Operations per bit	280	$518^{*\circ}$	$127^{*+\circ}$	336^{*}

* Not including demodulation (soft-decision) complexity.
\+ $tanh$ function is counted as a single operation.
° Message passing complexity ignored.

Kim, Hun-Seok. "HDM: Hyper-dimensional modulation for robust low-power communications." *2018 IEEE International Conference on Communications (ICC)*.

SNR and Eb/N_0

- Continuous-time AWGN channel
  - Signal power = P
  - Noise PSD = N_0
  - Bandwidth = W
  - SNR = P/N_0W



- Discrete-time AWGN channel
  - Energy per symbol Es= P/W
  - Noise energy = variance of noise = N_0
  - SNR in discrete time = Es/N_0 = P/N_0W = SNR in continuous time
- Eb/N_0 = SNR/R=(Es/N_0)/(# of transmissions/# of info bits)

Application to sensing problems – Example 1

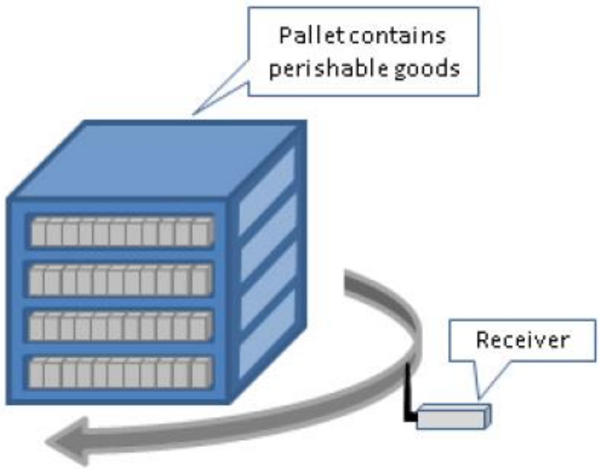
- Assign a binary vector to each temp range: eg. v0:[0-10], v1:[10-20], v2: [20-30], …
- Receiver receives the sum of the vectors sent by all sensors **s=s1+s2+….**
- Binary Query: any inventory in a temp range i?

$$s \sim v_i$$

- Proportion query: How much inventory is in temp rang i?

$$z(a_1 v_1 + a_2 v_2 + \cdots + a_M v_M) = s$$

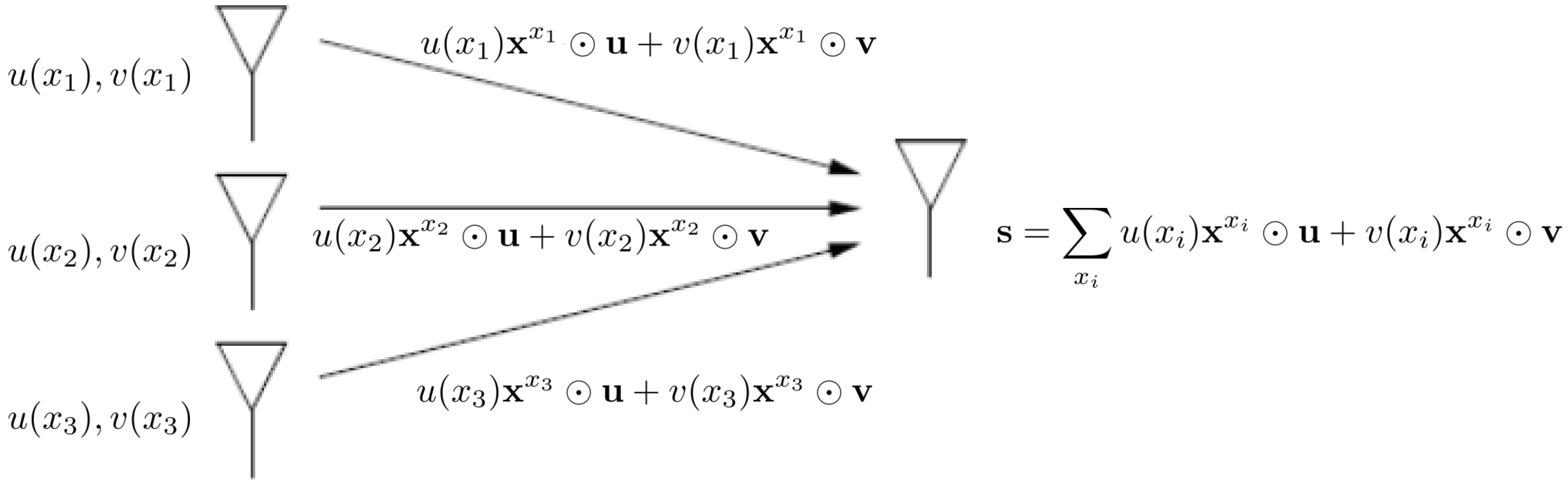
Linear least square regression to find ai



Process the single sum vector **s** without decoding individual sensor's reading

Jakimovski, Predrag, et al. "Collective communication for dense sensing environments." *Journal of Ambient Intelligence and Smart Environments* (2012).

Application to sensing problems – Example 2



- x_i is the position id in integers
- $u(x_i)$ and $v(x_i)$ are different measurements (eg. sound or gas sensor readings)
- **x**, **u**, **v** are complex HD vectors for position, u sensor, v sensor

Application to sensing problems – Example 2

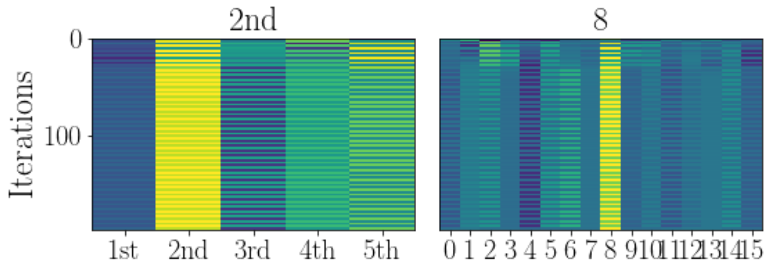
Patterns of 16 Sensors' Readings

1st [1, 1, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 0]
2nd [0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1]
3rd [0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0]
4th [1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1]
5th [0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]

Pattern # Position #

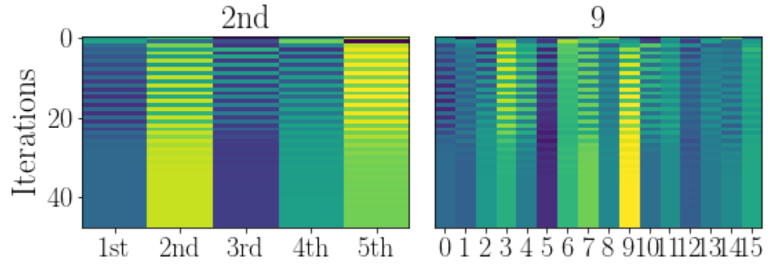
2nd Pattern (original)

[0 0 0 1 0 1 0 1 0 1 0 0 1 0 0 1]



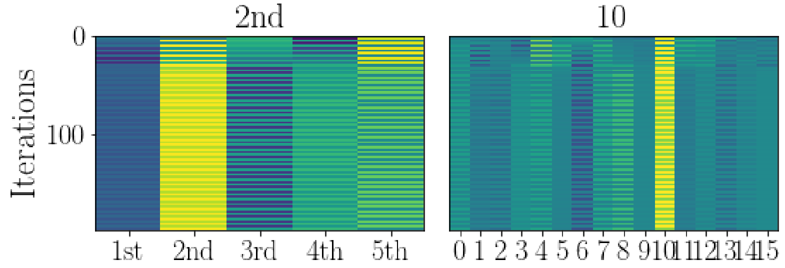
2nd Pattern (R shift +1)

[1 0 0 0 1 0 1 0 1 0 1 0 0 1 0 0]



2nd Pattern (R shift +2)

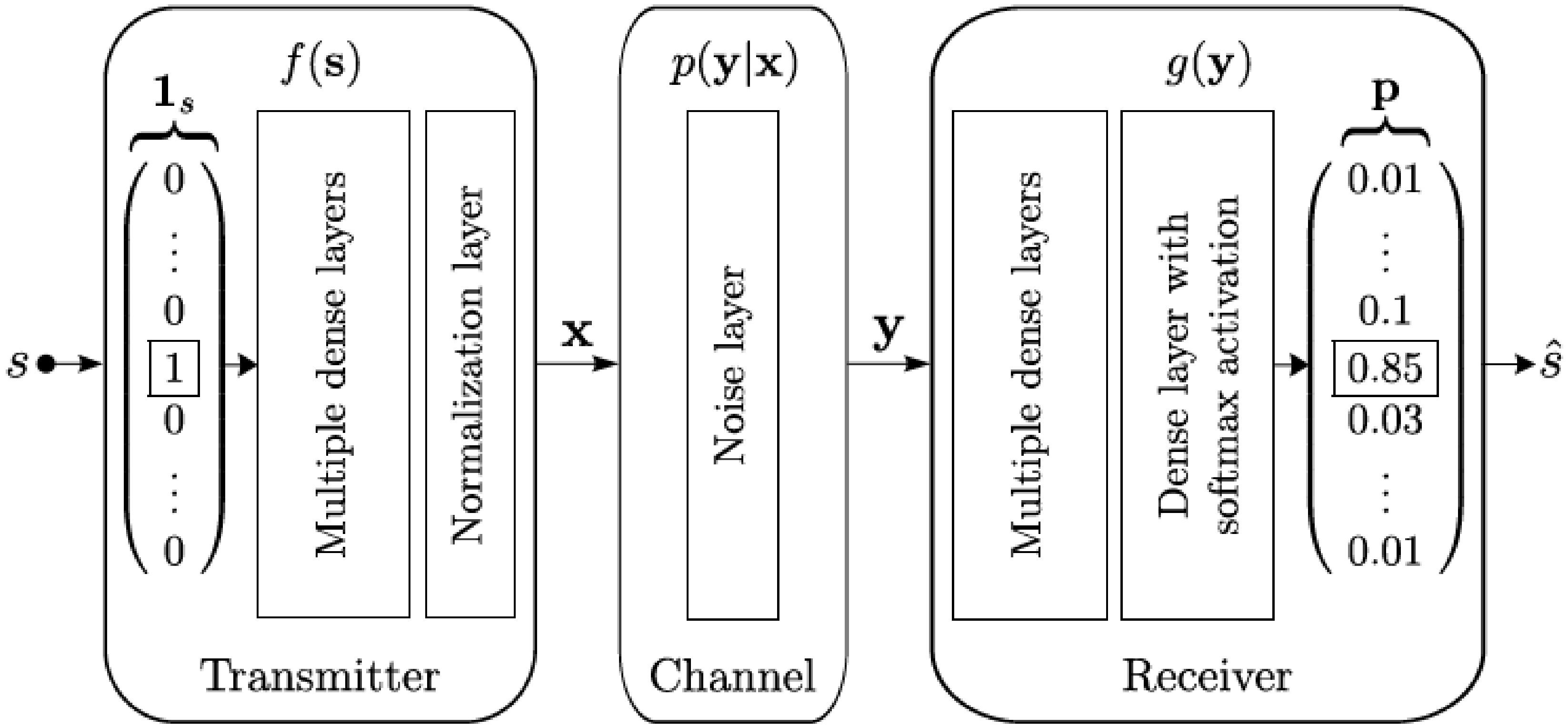
[0 1 0 0 0 1 0 1 0 1 0 1 0 0 1 0]



Outline

- Overview of communication problems
- Principles related to HD computing in communication
  - Law of large numbers & random codes
  - Orthogonality via random projection
- Applications of HD computing in communication
- AI/Machine learning in communication

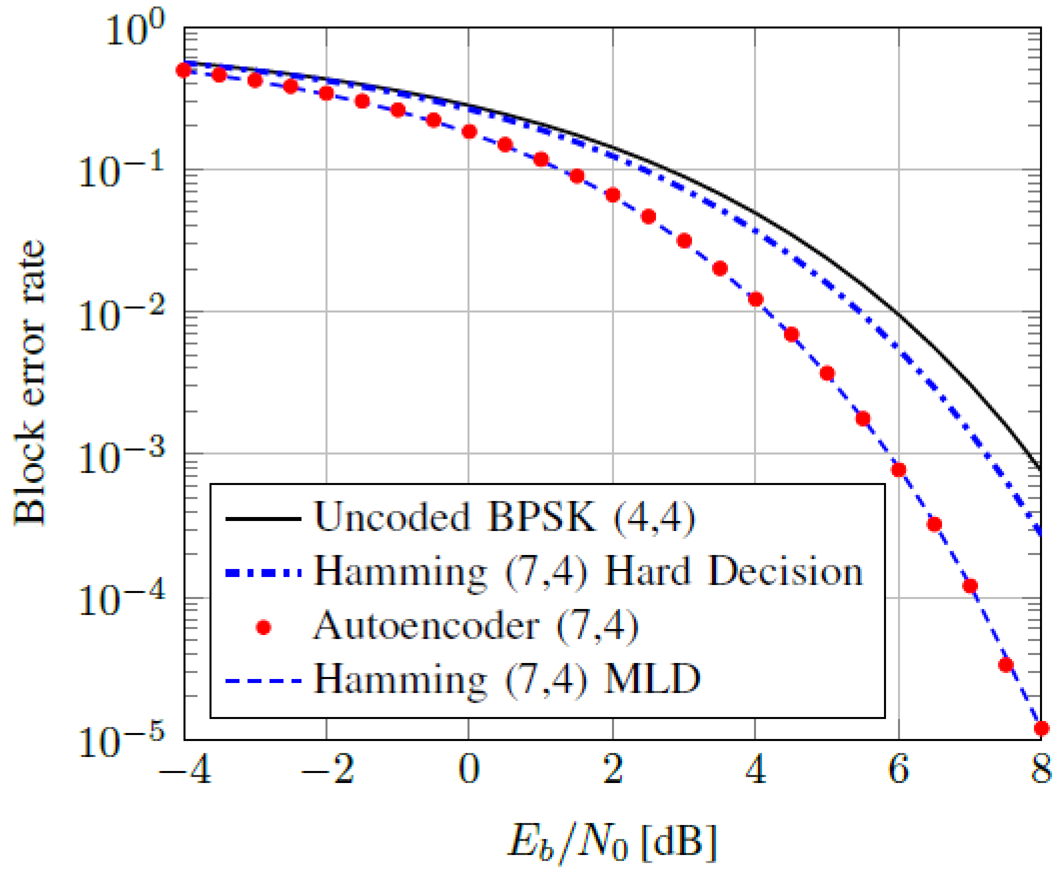
Communication system as an autoencoder



Loss function to minimize : block level error rate

O'shea, Timothy, and Jakob Hoydis. "An introduction to deep learning for the physical layer." *IEEE Transactions on Cognitive Communications and Networking* (2017).

Communication system as an autoencoder



(a)

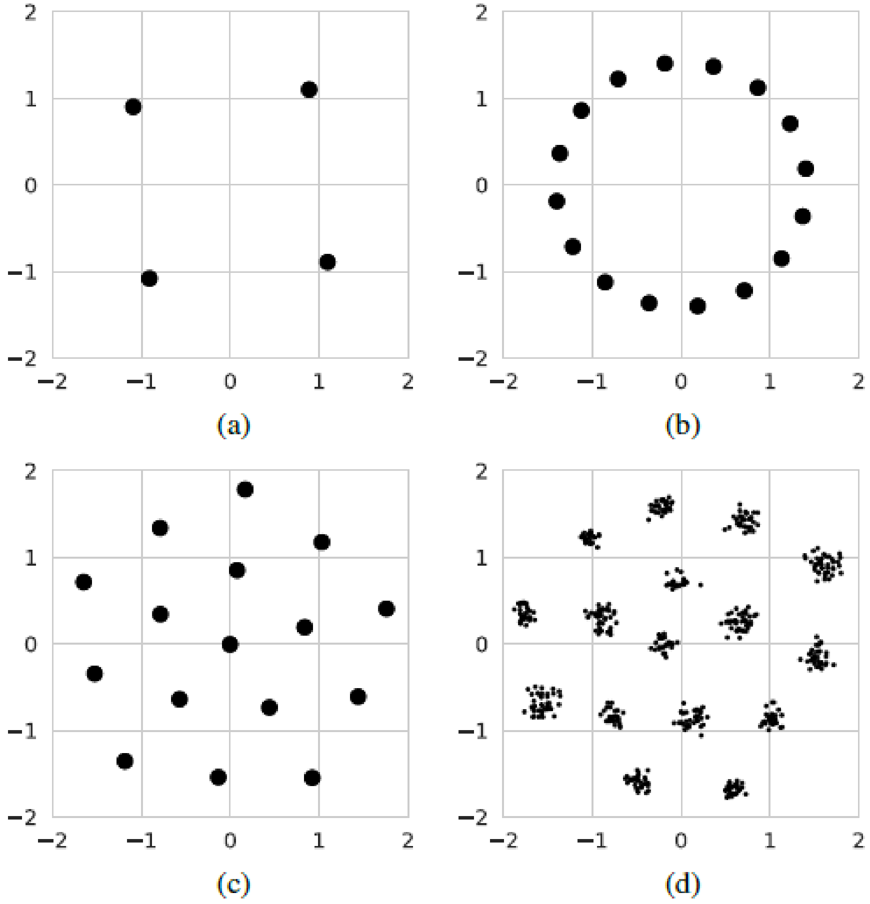


Figure 4: Constellations produced by autoencoders using parameters (n, k): (a) $(2, 2)$ (b) $(2, 4)$, (c) $(2, 4)$ with average power constraint, (d) $(7, 4)$ 2-dimensional t-SNE embedding of received symbols.

O’shea, Timothy, and Jakob Hoydis. "An introduction to deep learning for the physical layer." *IEEE Transactions on Cognitive Communications and Networking* (2017).

Learning to decode

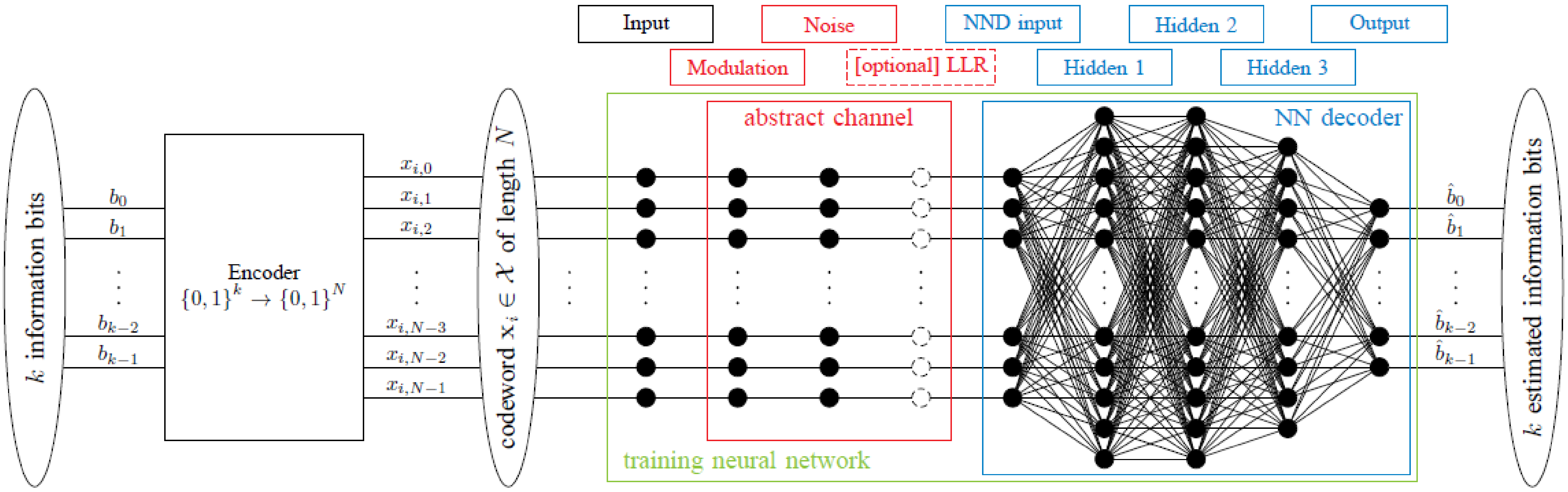
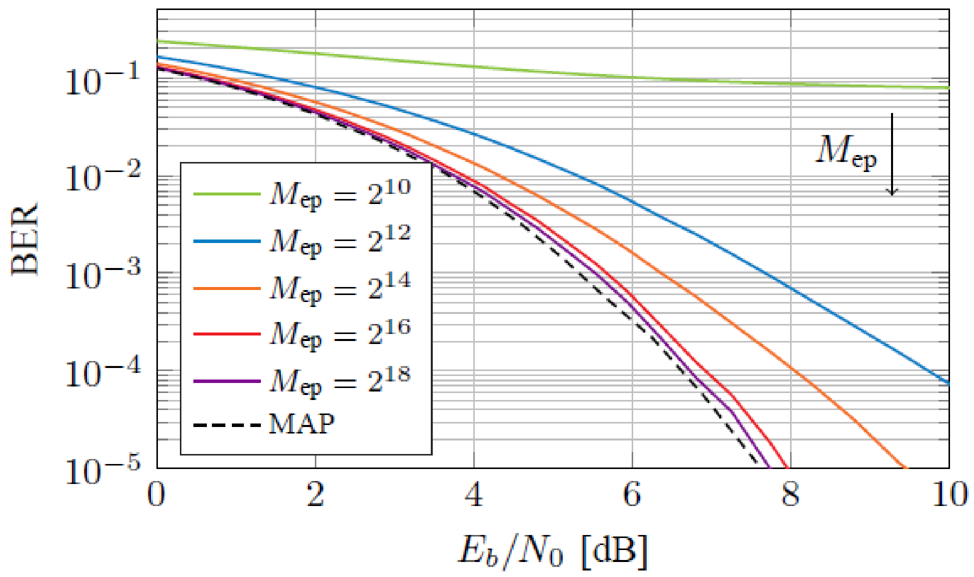


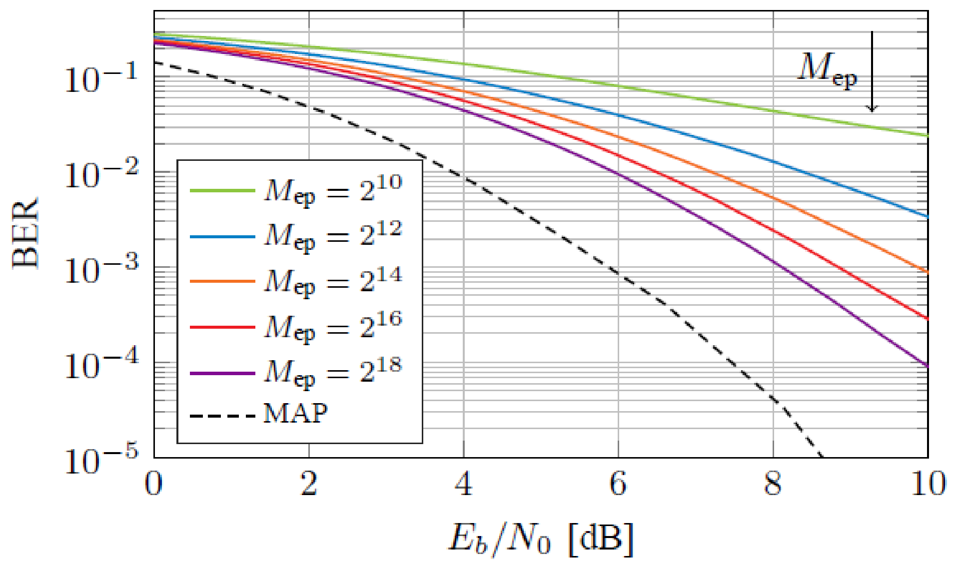
Fig. 1: Deep learning setup for channel coding.

Gruber, Tobias, et al. "On deep learning-based channel decoding." *2017* IEEE *Annual Conference on Information Sciences and Systems (CISS)*.

Learning to decode



(a) Polar Code

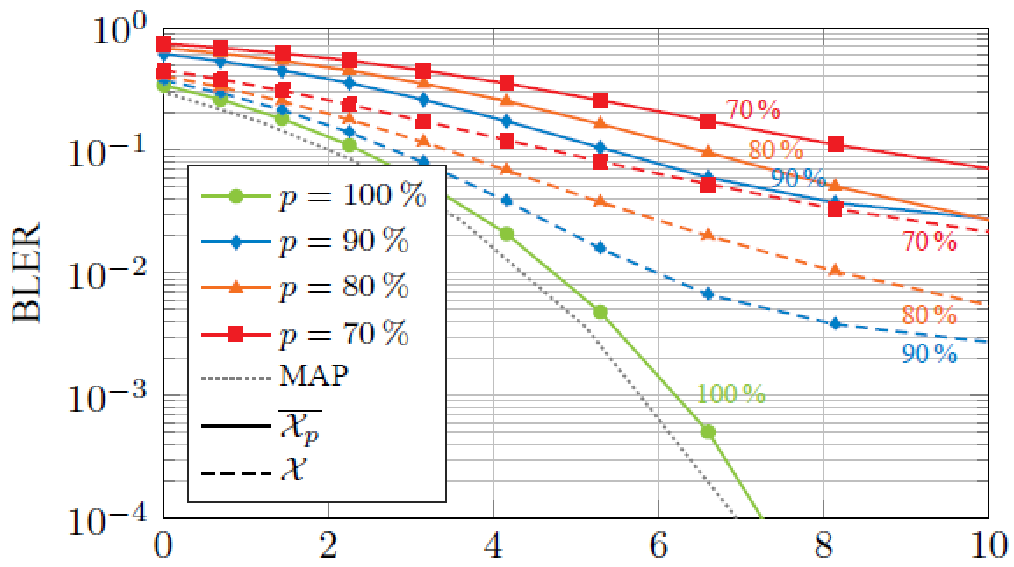


(b) Random Code

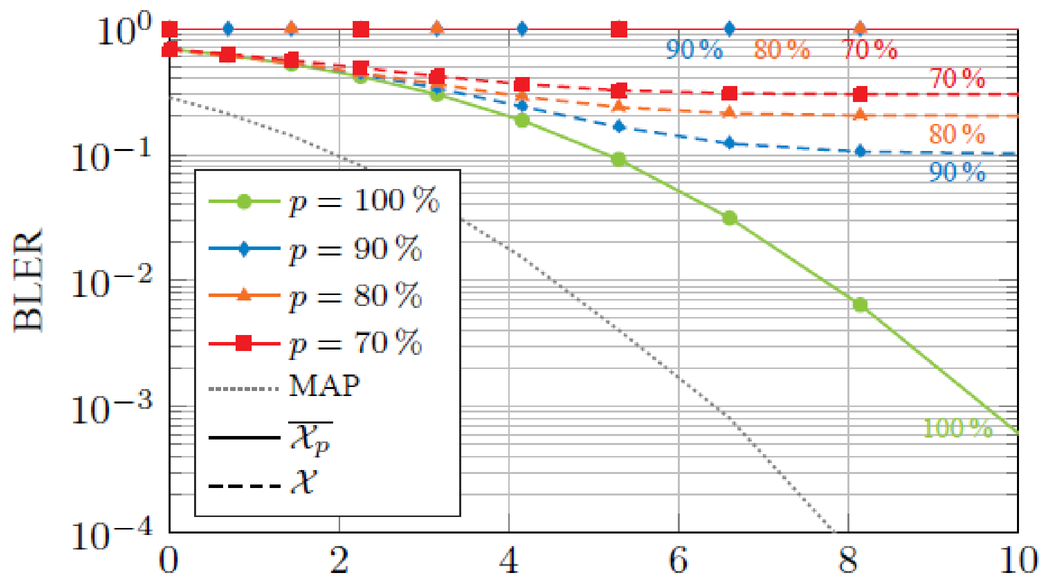
- Structured codes are easier to learn than random codes

Gruber, Tobias, et al. "On deep learning-based channel decoding." *2017* IEEE *Annual Conference on Information Sciences and Systems (CISS)*.

Learning to decode



(a) 16 bit-length Polar Code ($r = 0.5$)

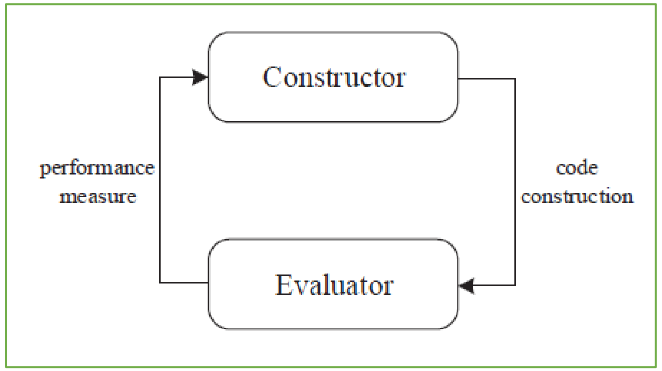


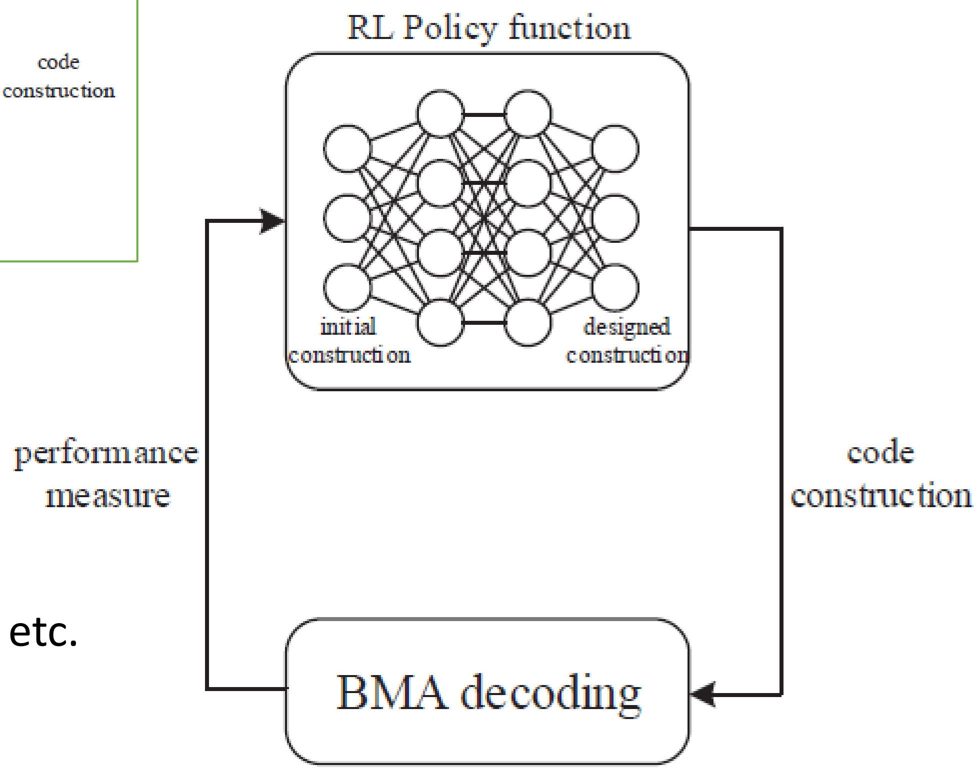
(b) 16 bit-length Random Code ($r = 0.5$)

- Provide a subset of the entire set during training (p%)
- Able to generalize to codewords that it has never seen during training for structure codes, but not for random codes

Gruber, Tobias, et al. "On deep learning-based channel decoding." *2017* IEEE *Annual Conference on Information Sciences and Systems (CISS)*.

Learning to construct codes





Hamming distance,
Decoding threshold, etc.

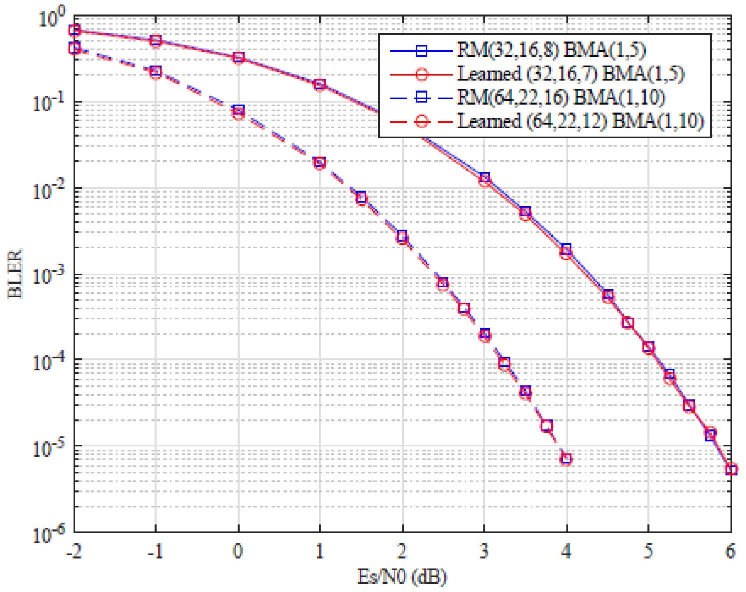


Fig. 5: Performance comparison between learned linear block codes and RM codes

Huang, L., Zhang, H., Li, R., Ge, Y., & Wang, J. (2019). AI coding: Learning to construct error correction codes. *IEEE Transactions on Communications*

Conclusions

- The power of HD computing lies in encoding structures, learning, and classification.
- New applications drive communication from traditional point-to-point transmission paradigm into a variety of new problems

⇨Opportunities for HD computing

Acknowledgements: Connor Bybee, Denis Kleyko, Chris Kymn

References

- Cover, Thomas M. Elements of information theory. John Wiley & Sons, 1999.
- Proakis, J., and Masoud Salehi. "Digital Communication Fifth edition." 2008.
- El Gamal, Abbas, and Young-Han Kim. *Network information theory*. Cambridge university press, 2011.