Neural Distributed Data Compression and Communication

New solutions to old problems in information theory.

Ezgi Ozyilkan

PhD Defense

Dec 18, 2025

Neural Distributed Data Compression and Communication

New solutions to old problems in information theory.

Ezgi Ozyilkan

Joint work with:

- Jona Ballé (NYU), Fabrizio Carpi (→Samsung) and Elza Erkip (NYU)

PhD Defense

Dec 18, 2025

"[...] in first half of 2022, video accounted for a hefty 65.93% of total volume over the internet."

– Sandvine's Global Internet Phenomena Report, Jan. 2023

Representative internet video

Source: Vimeo

"[...] in first half of 2022, video accounted for a hefty 65.93% of total volume over the internet."

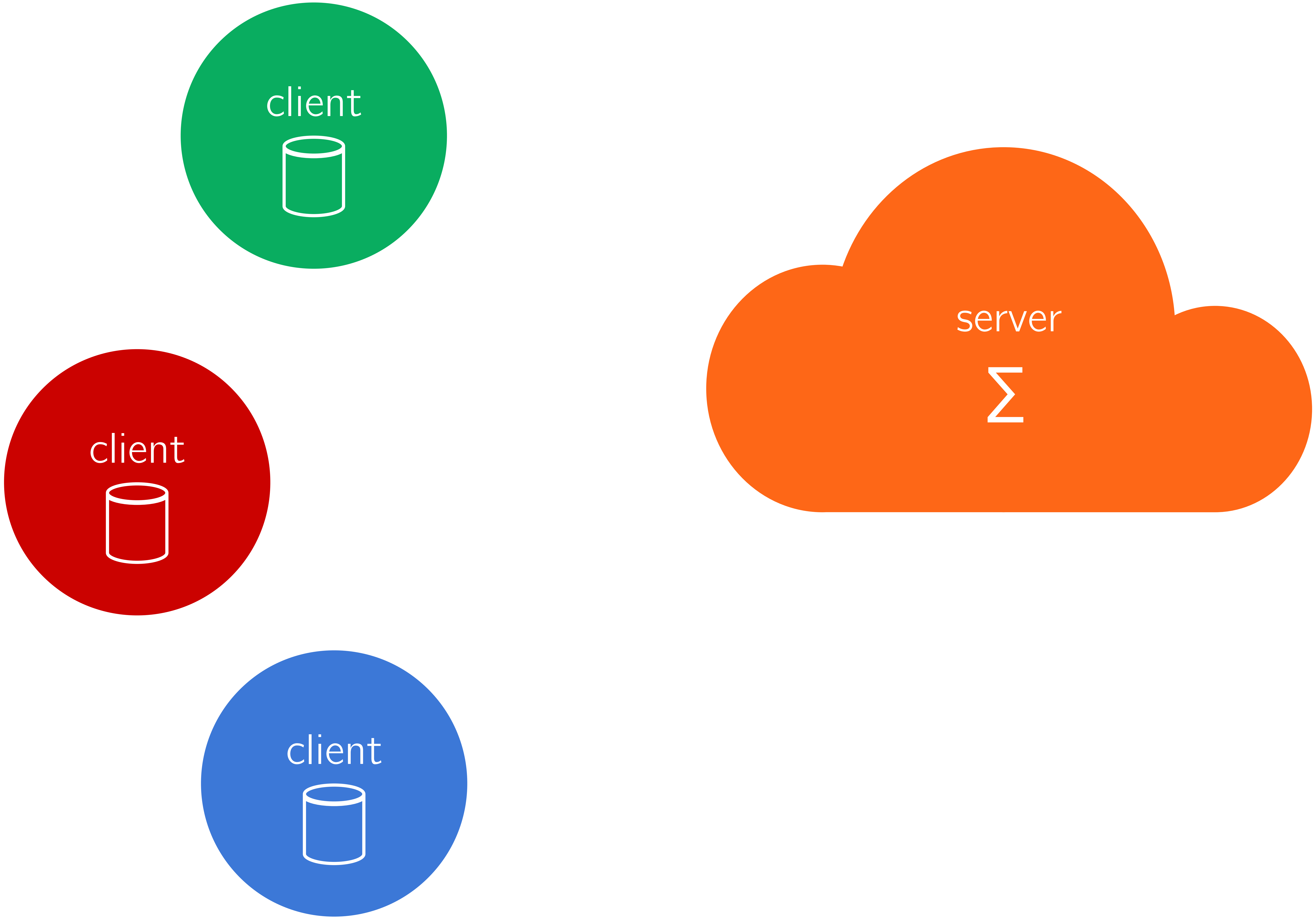
– Sandvine's Global Internet Phenomena Report, Jan. 2023

Representative internet video

Source: Vimeo

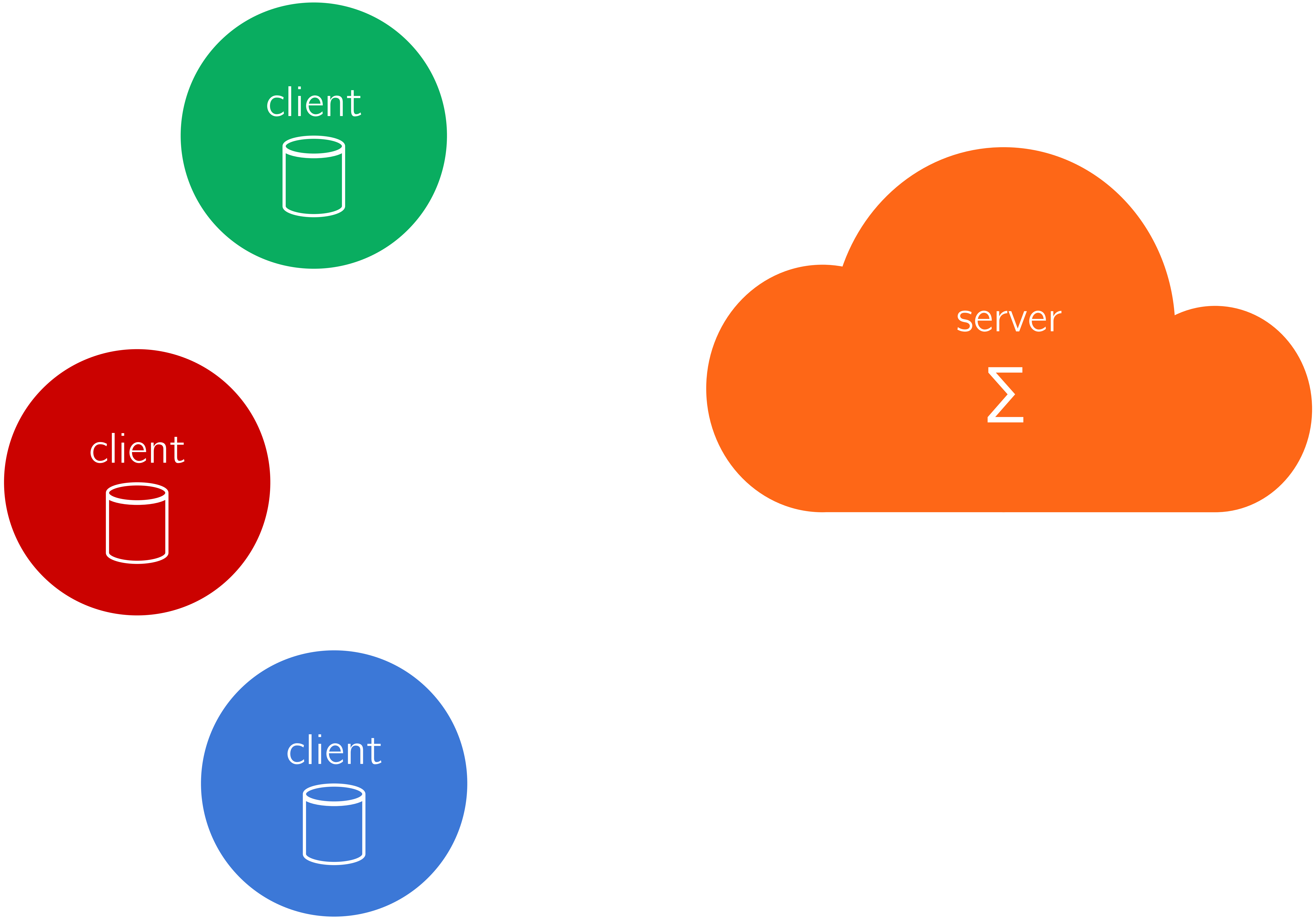
Example: Federated learning

e.g., next-word prediction



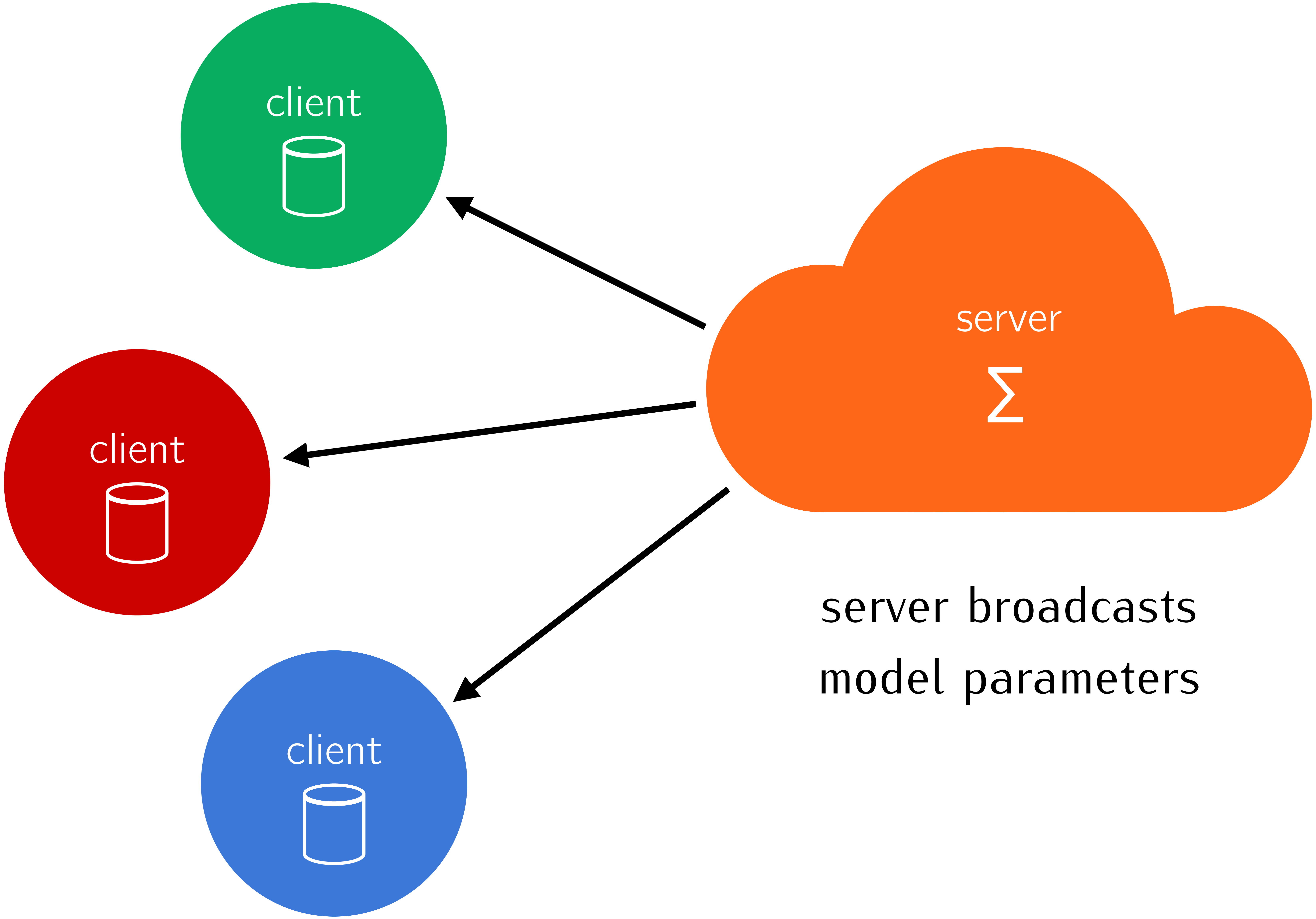
Example: Federated learning

e.g., next-word prediction



Example: Federated learning

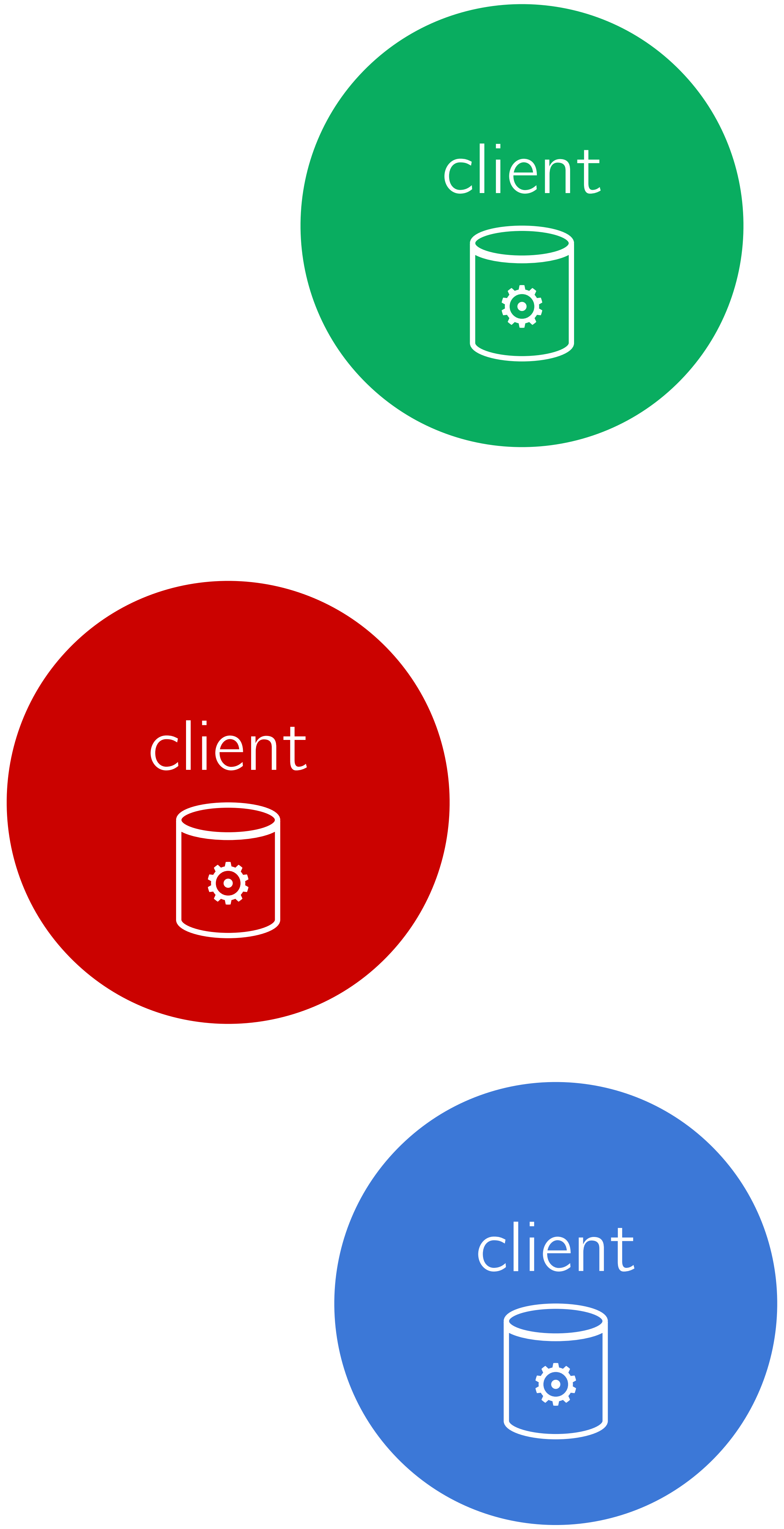
e.g., next-word prediction

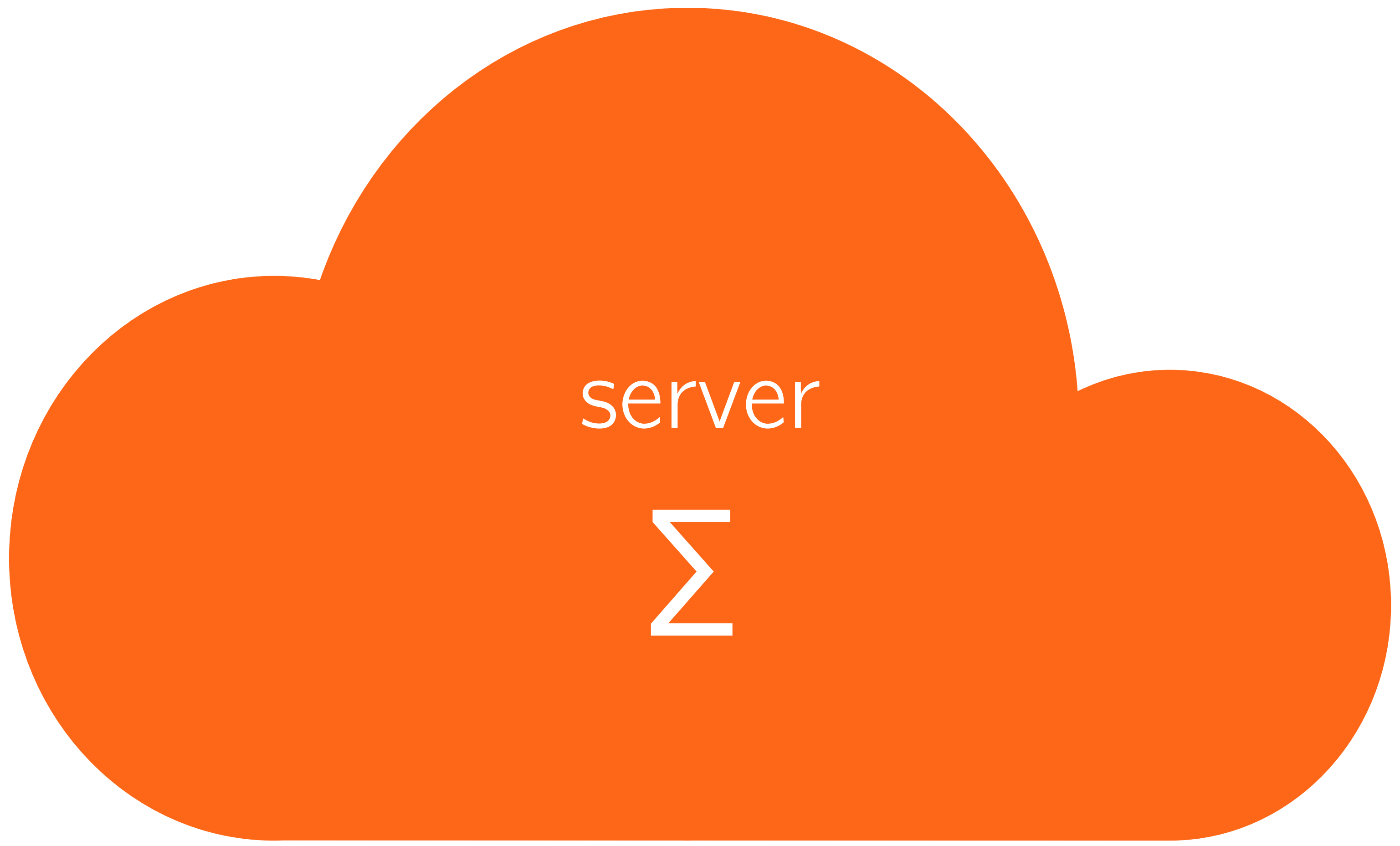


server broadcasts model parameters

Example: Federated learning

e.g., next-word prediction

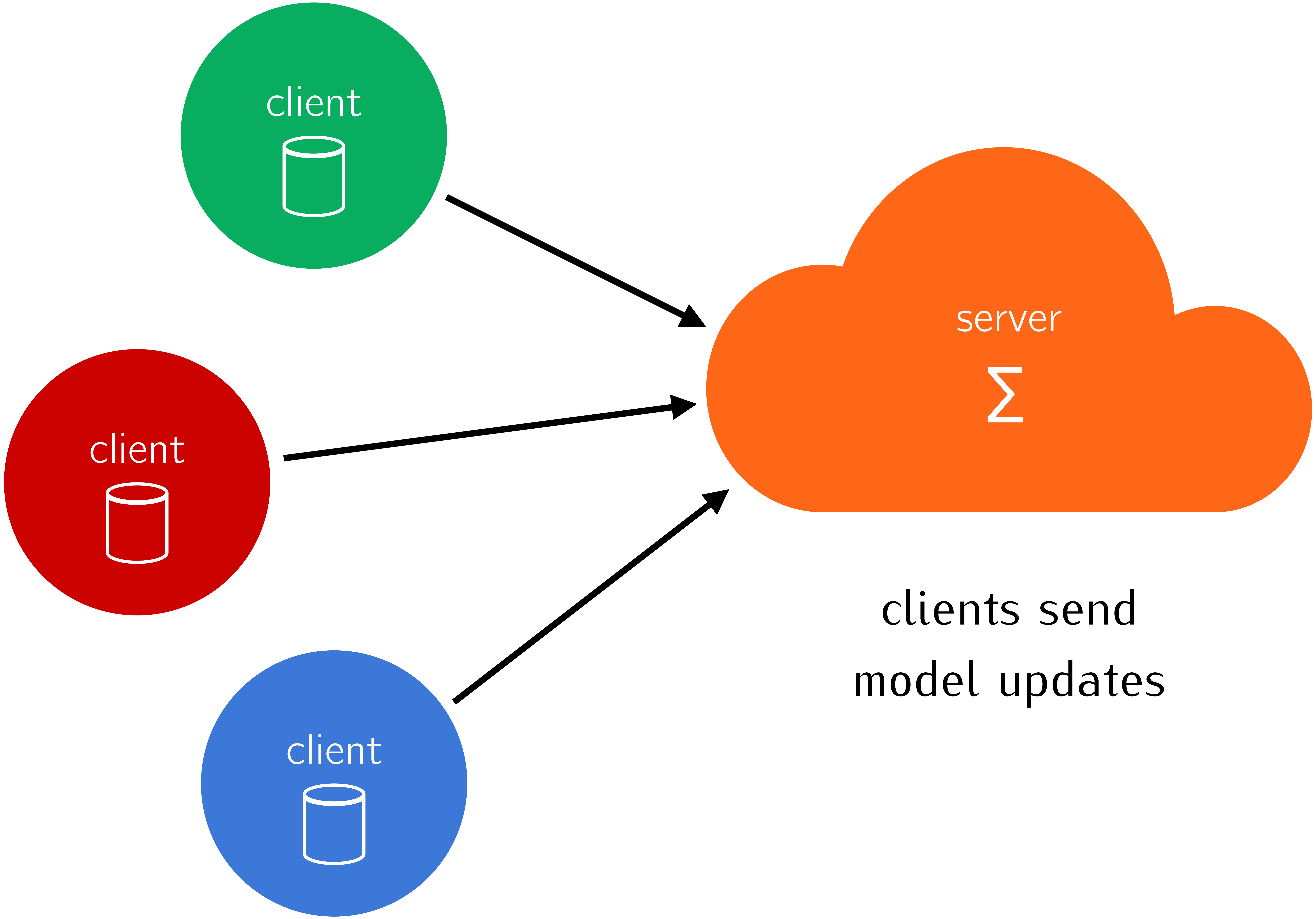




clients update their models based on local data

Example: Federated learning

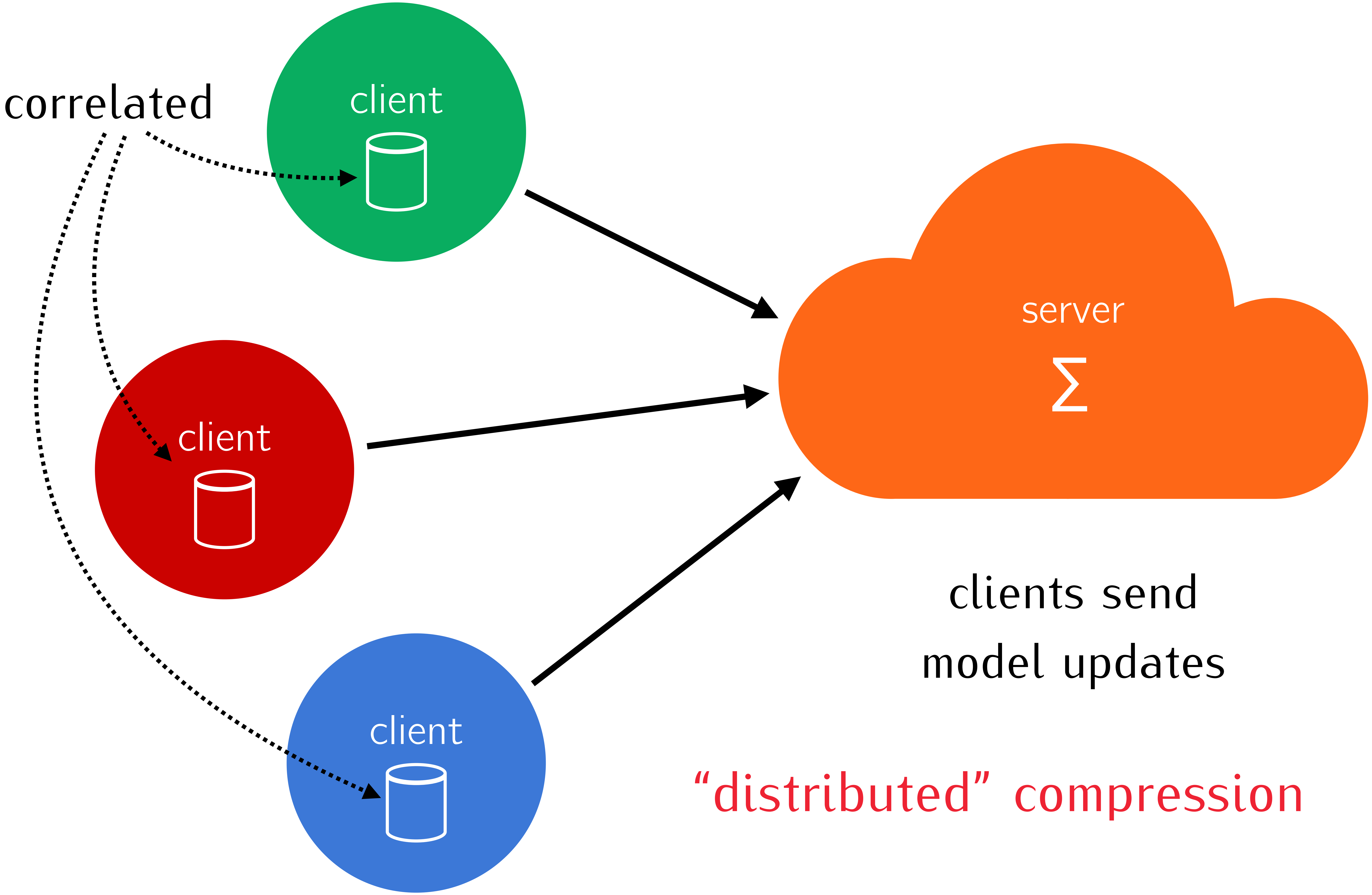
e.g., next-word prediction



clients send model updates

Example: Federated learning

e.g., next-word prediction



clients send model updates

“distributed” compression

Toy example for distributed compression

- Suppose X and Y are equiprobable 3-bit binary words.

Toy example for distributed compression

- Suppose X and Y are equiprobable 3-bit binary words.
- Let correlation pattern be such that $d_{hamming}(X, Y) \leq 1$.

Toy example for distributed compression

- Suppose X and Y are equiprobable 3-bit binary words.
- Let correlation pattern be such that $d_{hamming}(X, Y) \leq 1$.
- If Y is available at both encoder-decoder, describe X using 2 bits.
  - Realize that there are only 4 possibilities for $X + Y$, $\{000; 001; 010; 100\}$

Toy example for distributed compression

- Suppose X and Y are equiprobable 3-bit binary words.
- Let correlation pattern be such that $d_{hamming}(X, Y) \leq 1$.
- If Y is available at both encoder-decoder, describe X using 2 bits.
  - Realize that there are only 4 possibilities for $X + Y$, $\{000; 001; 010; 100\}$
- What if Y is **"only" available at decoder** ?
- X <u>can still be described using only 2 bits</u> !!

Toy example (continued)

- Realization:

Toy example (continued)

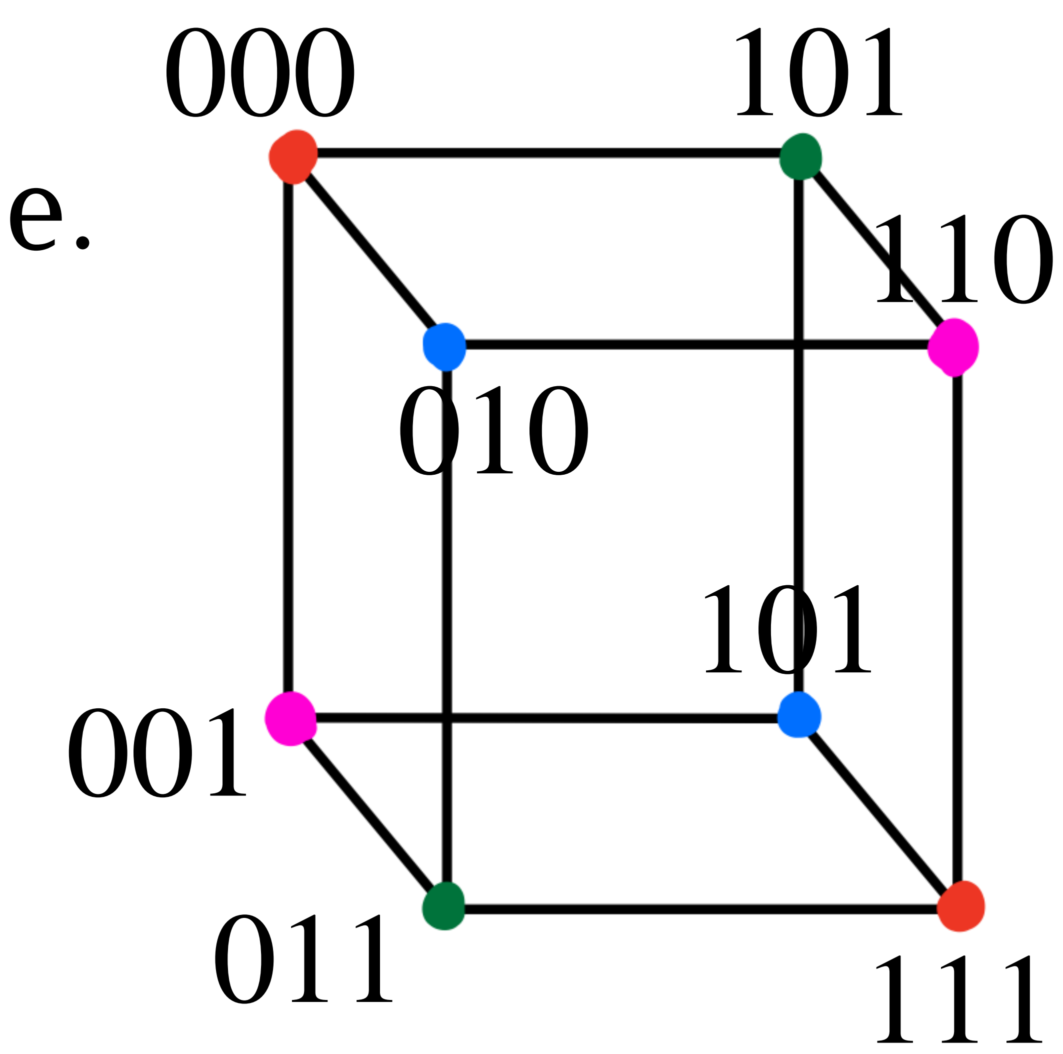
- Realization:
  - Wasteful to spend bits in differentiating X between 'distant' codewords.

Toy example (continued)

- Realization:
  - Wasteful to spend bits in differentiating X between ‘distant’ codewords.
  - Group 8 possible values of X into 4 groups (“**binning**”):
  - $B_0 = \{000; 111\}$, $B_1 = \{001; 110\}$, $B_2 = \{010; 101\}$, $B_3 = \{011; 100\}$
  - Send the index of the bin (or coset).

Toy example (continued)

- Realization:
  - Wasteful to spend bits in differentiating X between 'distant' codewords.
  - Group 8 possible values of X into 4 groups ("**binning**"):
  - $B_0 = \{000; 111\}$, $B_1 = \{001; 110\}$, $B_2 = \{010; 101\}$, $B_3 = \{011; 100\}$
  - Send the index of the bin (or coset).
  - Resolve the uncertainty with Y by checking Hamming distance.



“[...] despite the existence of potential applications, the conceptual importance of distributed compression has not been mirrored in practical data compression.”

50th year Commemorative Special Issue of Trans. on Information Theory

Verdú (IEEE Trans. on Information Theory, 1998)

“[...] despite the existence of potential applications, the conceptual importance of distributed compression has not been mirrored in practical data compression.”

50th year Commemorative Special Issue of Trans. on Information Theory

Data-driven methods may help here!

Verdú (IEEE Trans. on Information Theory, 1998)

Outline

1. **Review:**

   "Learning" data compression via *Nonlinear Transform Coding*

Outline

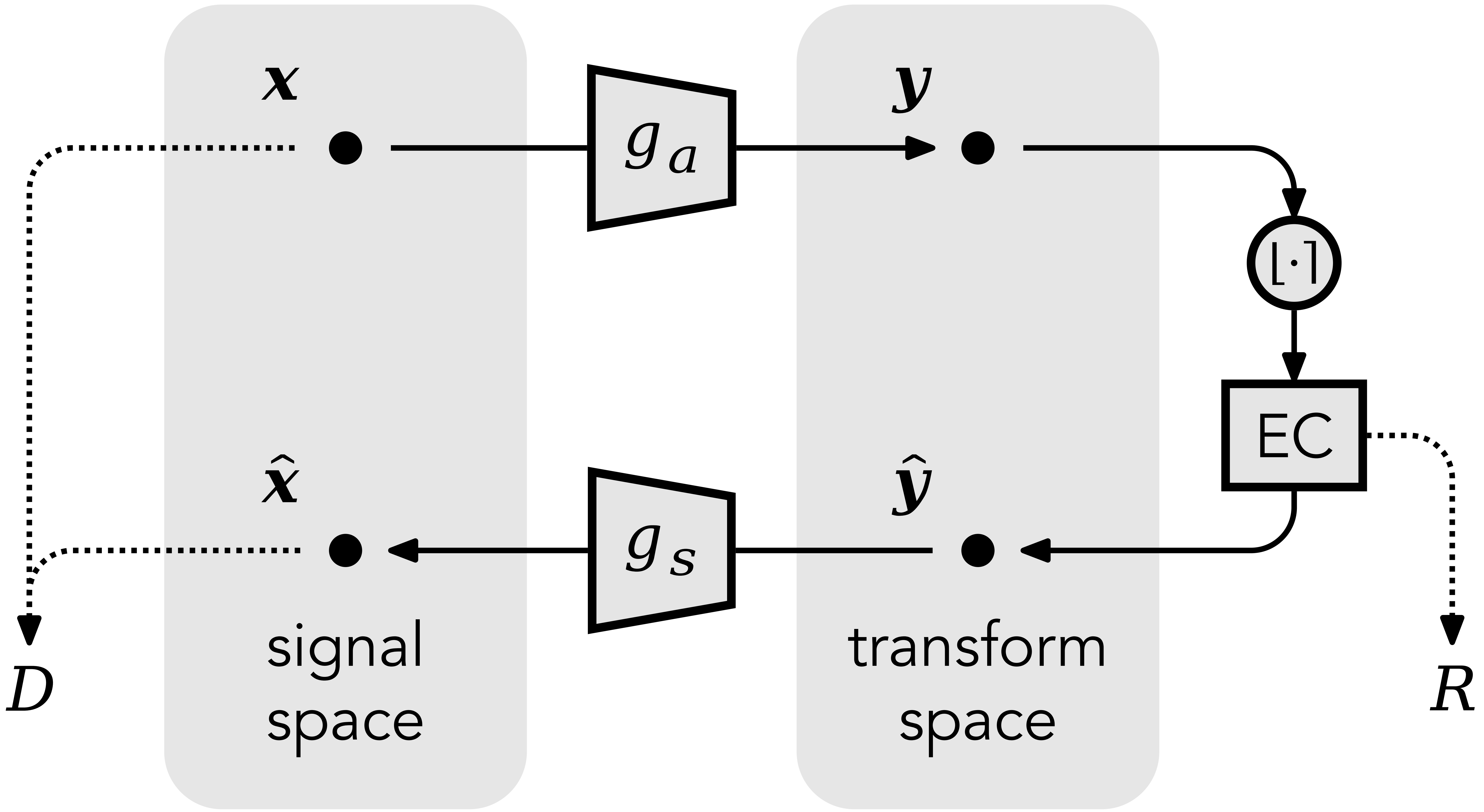
2. New solutions to old problems in information theory:
   a) distributed data compression: Wyner–Ziv and extensions
   b) "compress-and-forward" for the relay channel

Part I

Review:

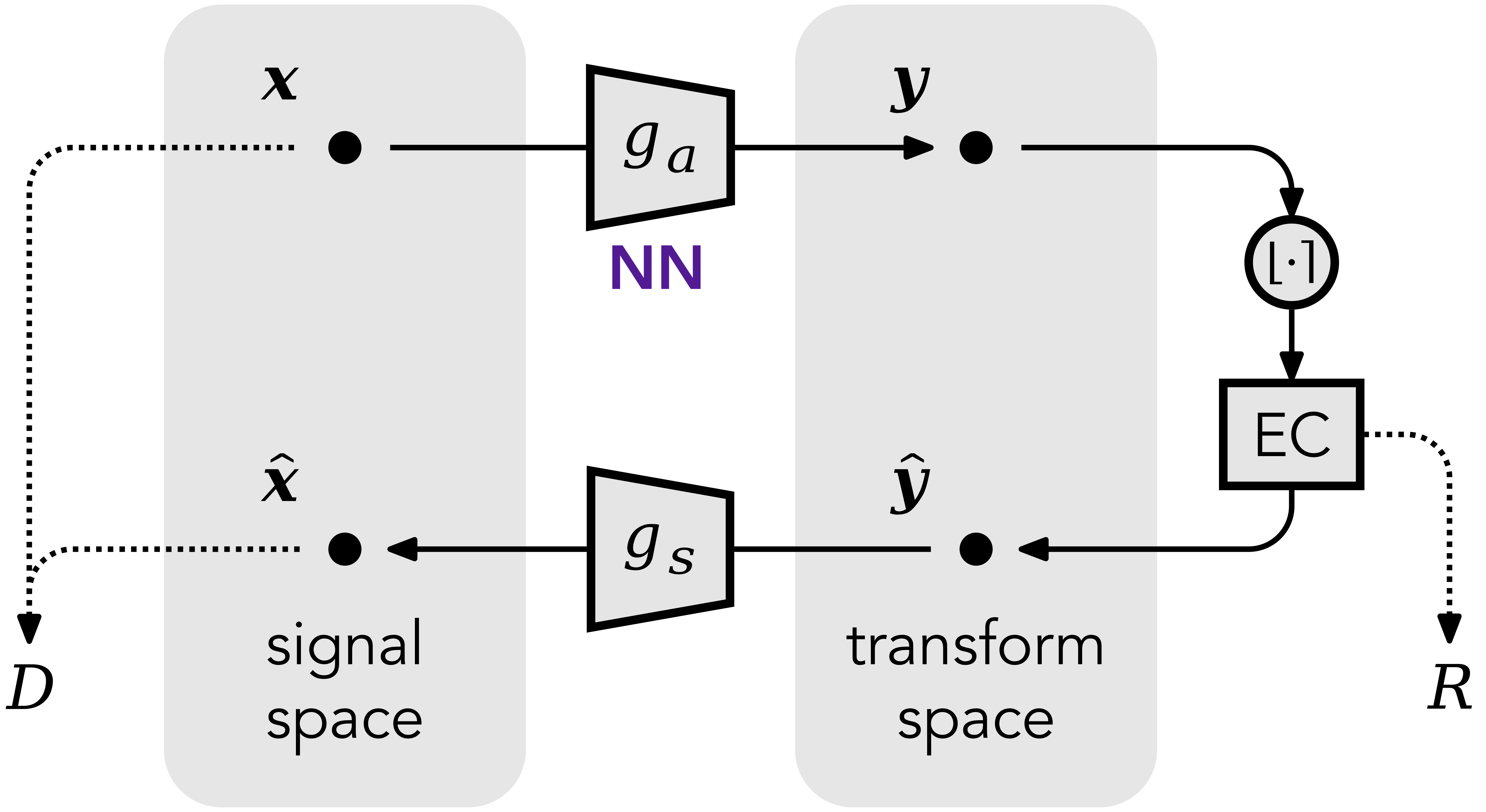
“Learning” Data Compression via *Nonlinear Transform Coding*

Nonlinear Transform Coding (NTC)



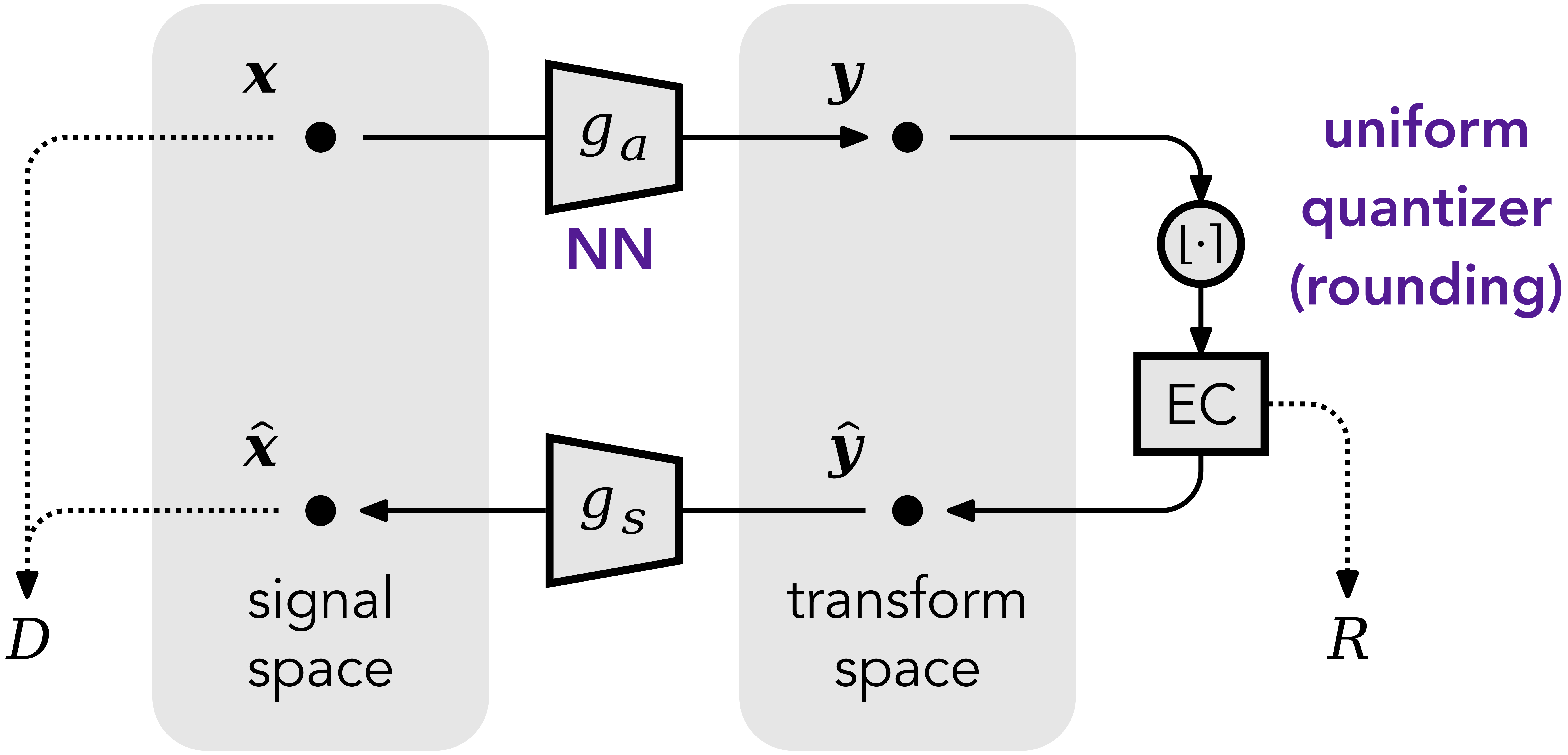
Ballé, Laparra & Simoncelli (ICLR, 2017)

Nonlinear Transform Coding (NTC)



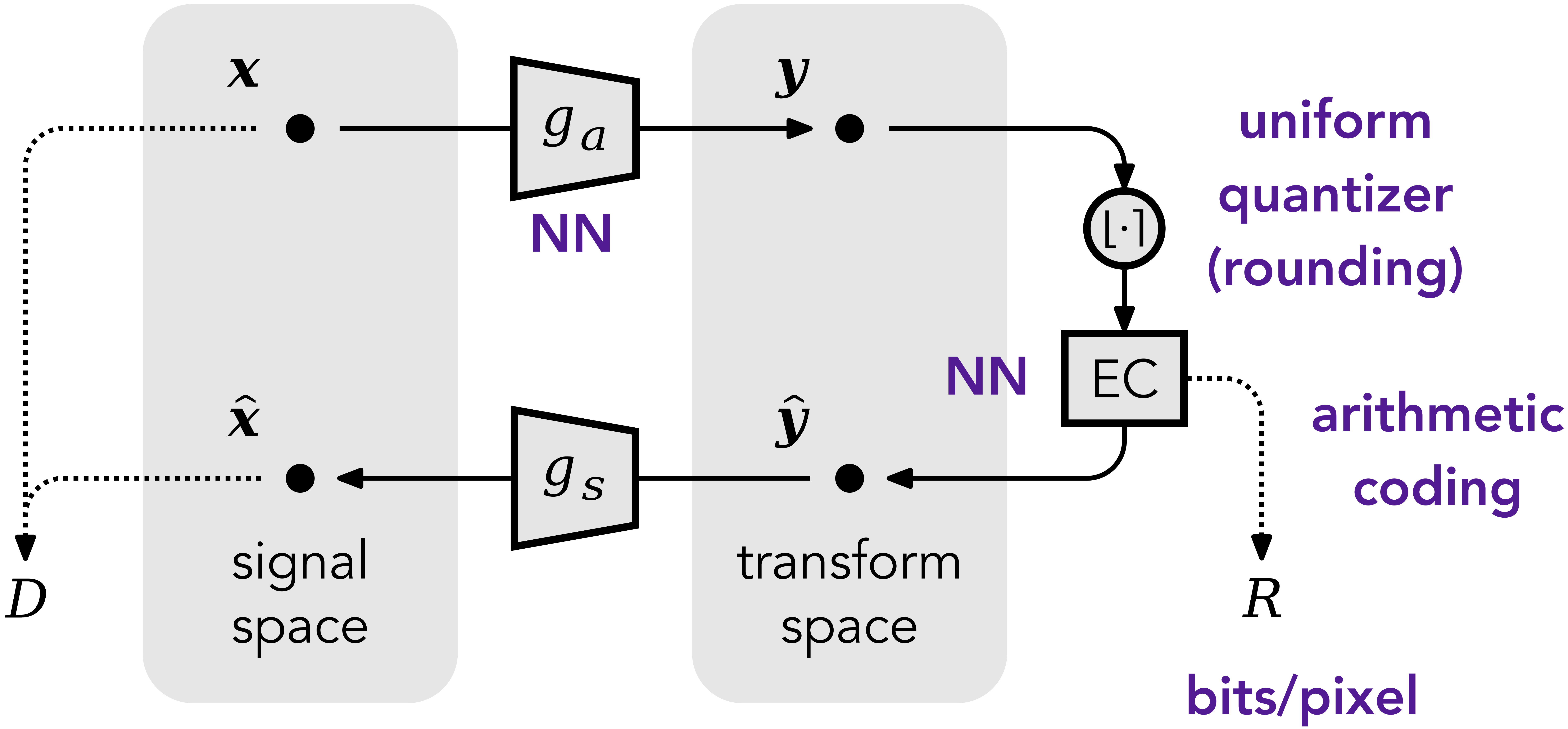
Ballé, Laparra & Simoncelli (ICLR, 2017)

Nonlinear Transform Coding (NTC)



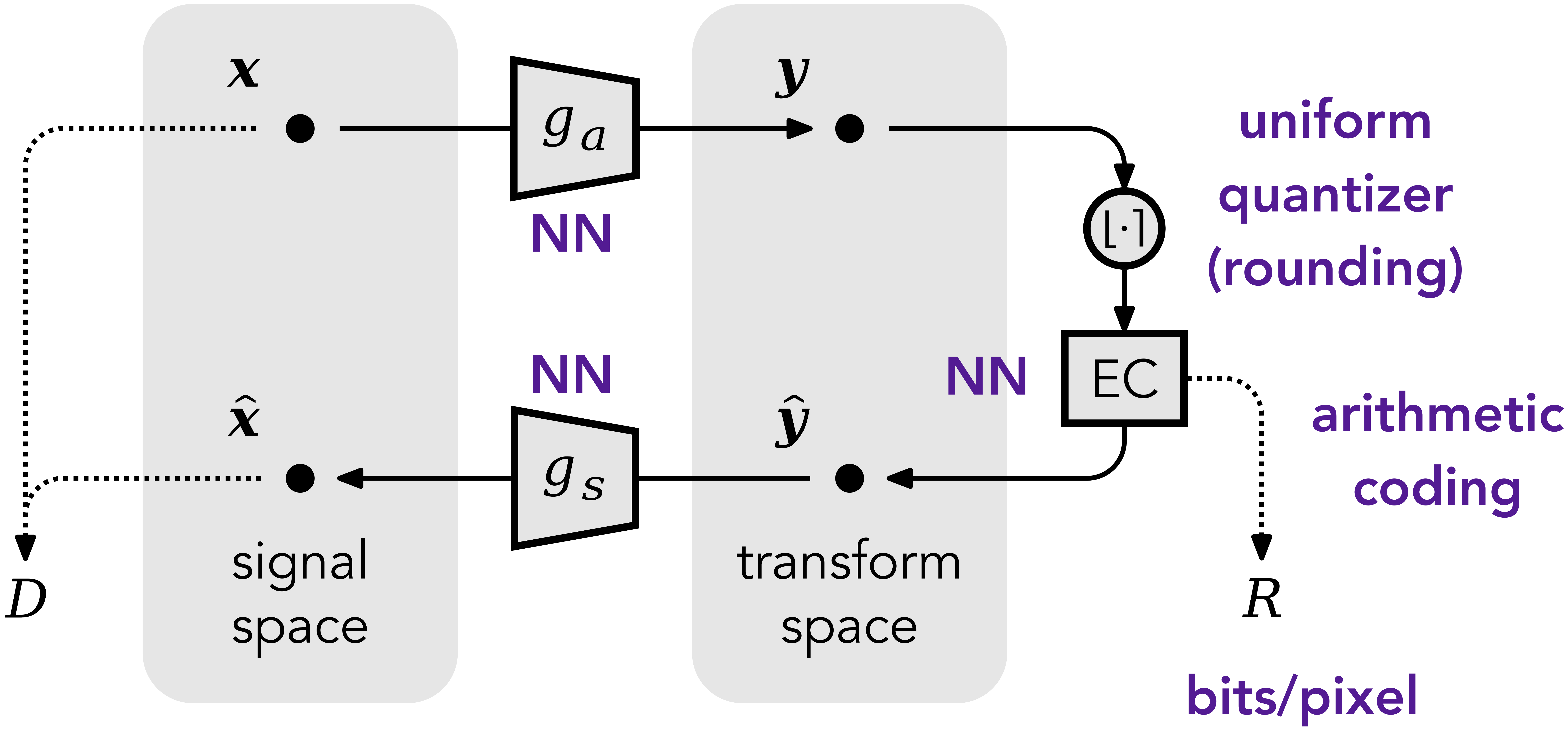
Ballé, Laparra & Simoncelli (ICLR, 2017)

Nonlinear Transform Coding (NTC)



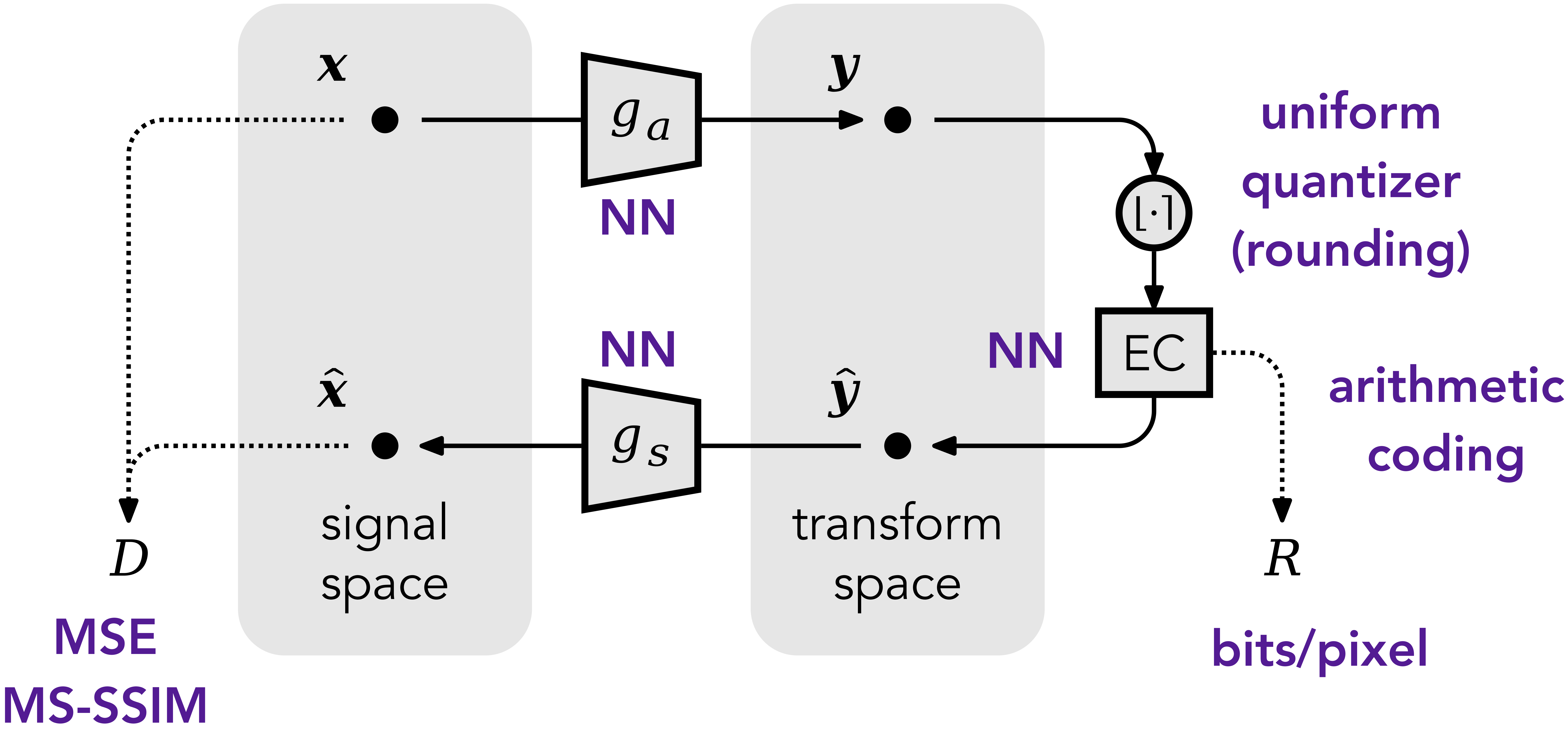
Ballé, Laparra & Simoncelli (ICLR, 2017)

Nonlinear Transform Coding (NTC)



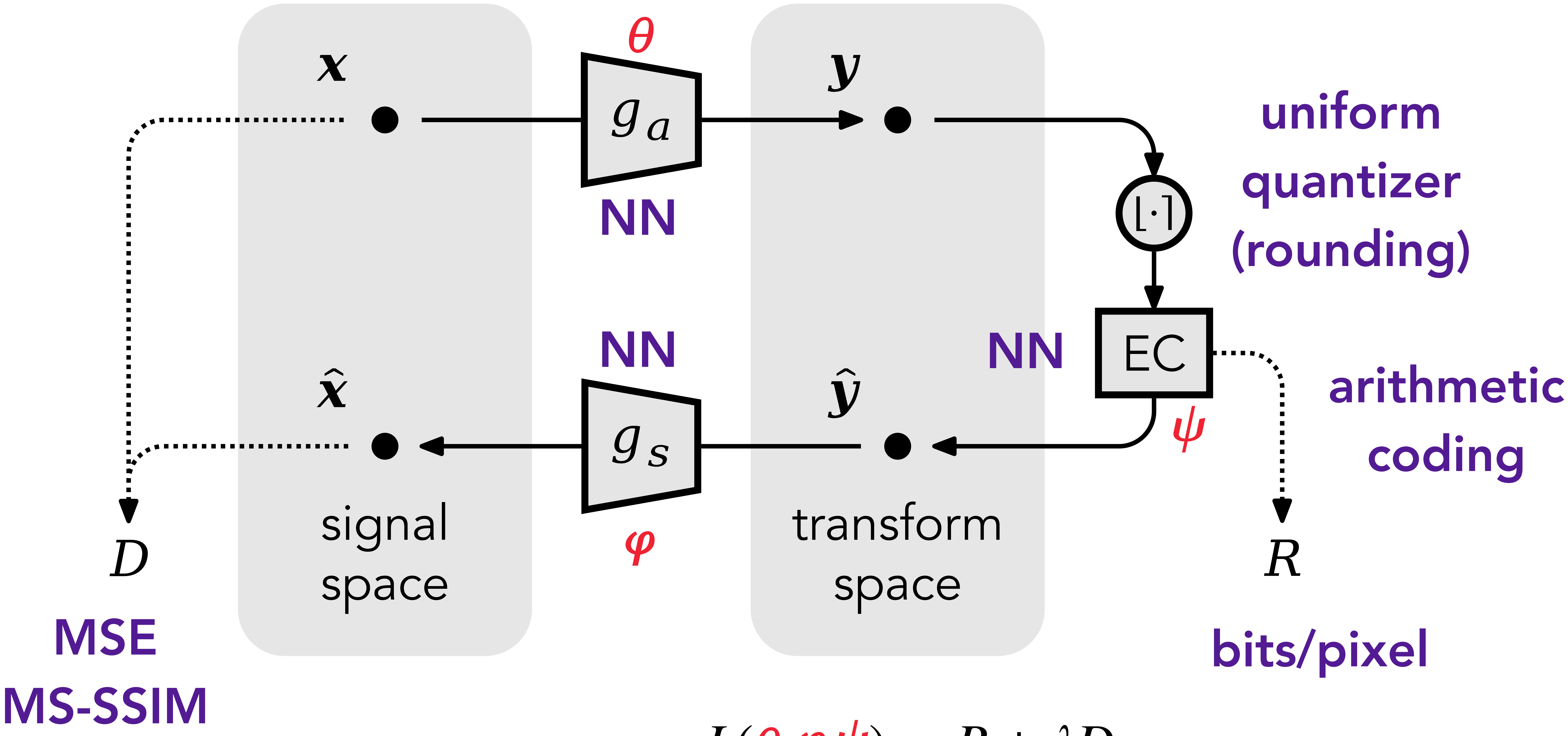
Ballé, Laparra & Simoncelli (ICLR, 2017)

Nonlinear Transform Coding (NTC)



Ballé, Laparra & Simoncelli (ICLR, 2017)

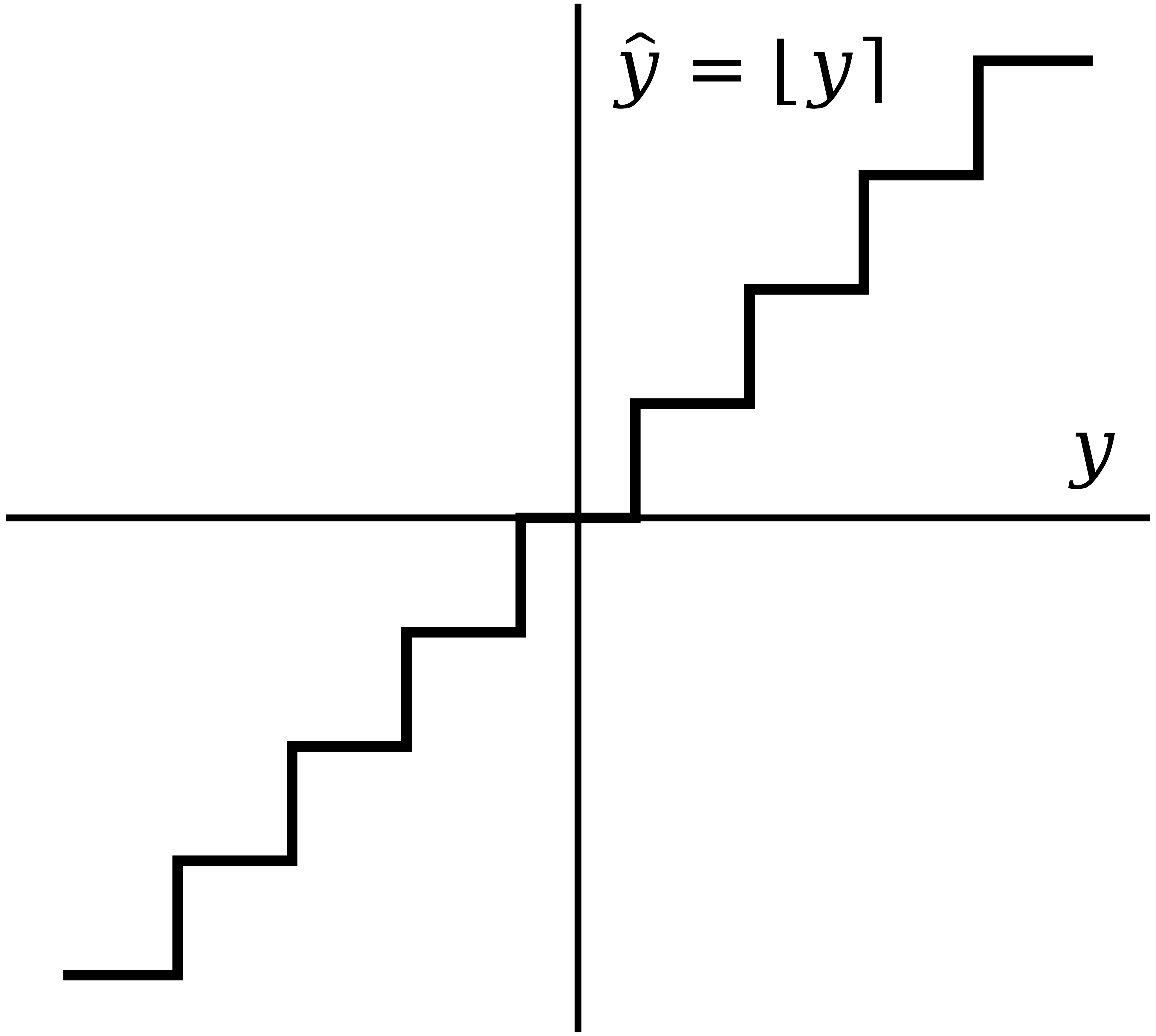
Nonlinear Transform Coding (NTC)



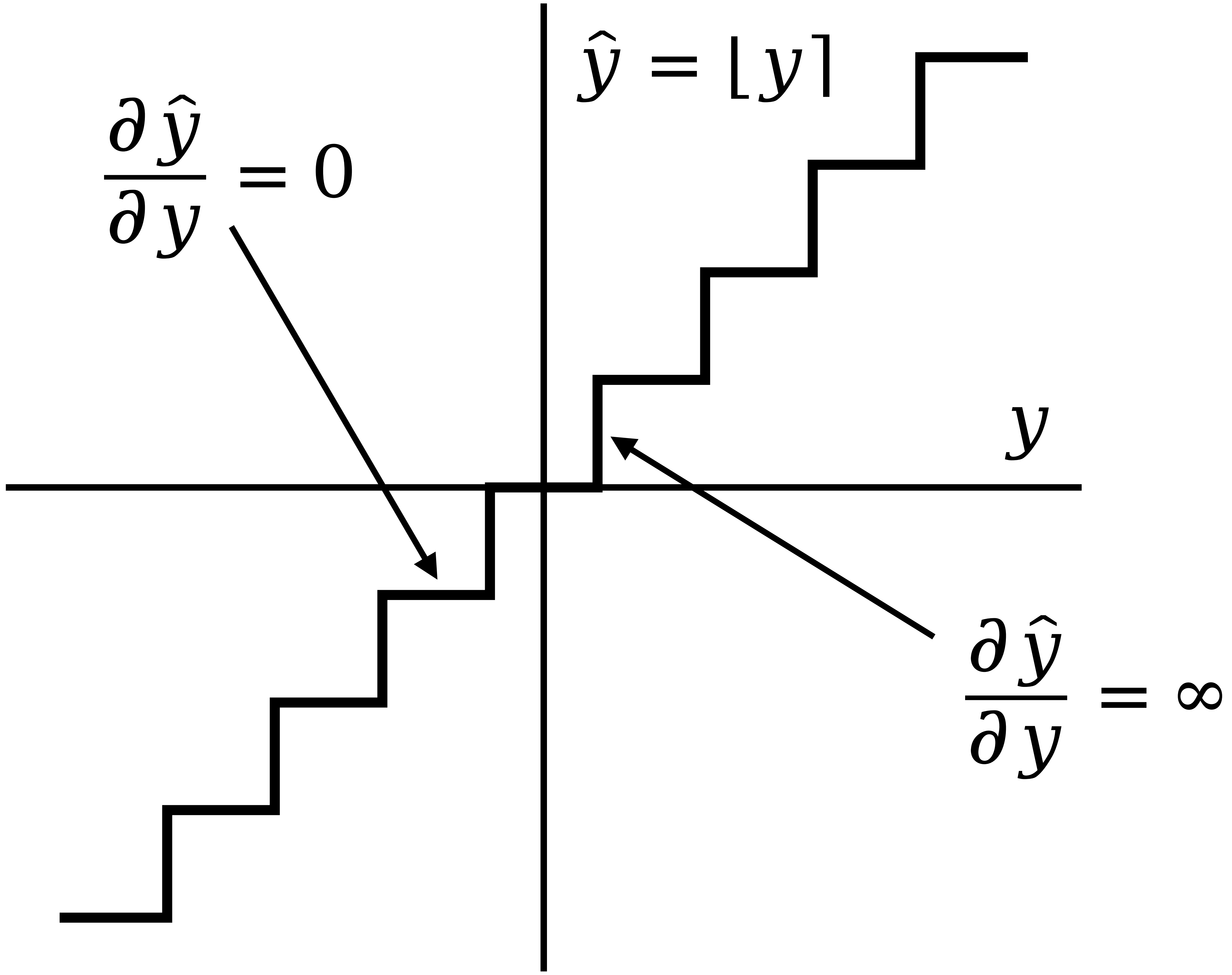
$$L(\theta, \varphi, \psi) = R + \lambda D$$

Ballé, Laparra & Simoncelli (ICLR, 2017)

Gradient is **zero** *almost everywhere*



Gradient is **zero** *almost everywhere*



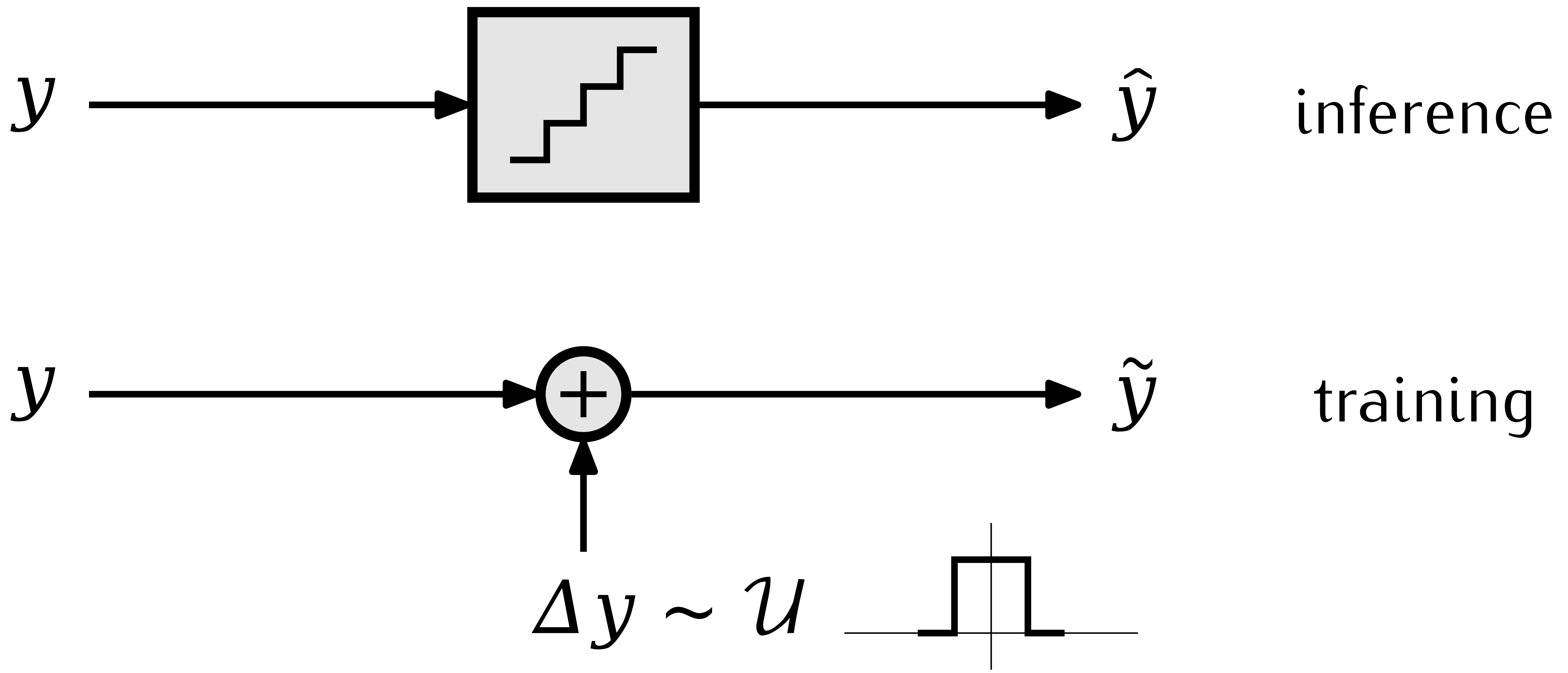
Proxy rate-distortion loss

Replace **rounding** with additive uniform noise.

Ballé, Chou, Minnen et al. (IEEE STSP, 2021)

Proxy rate-distortion loss

Replace **rounding** with additive uniform noise.



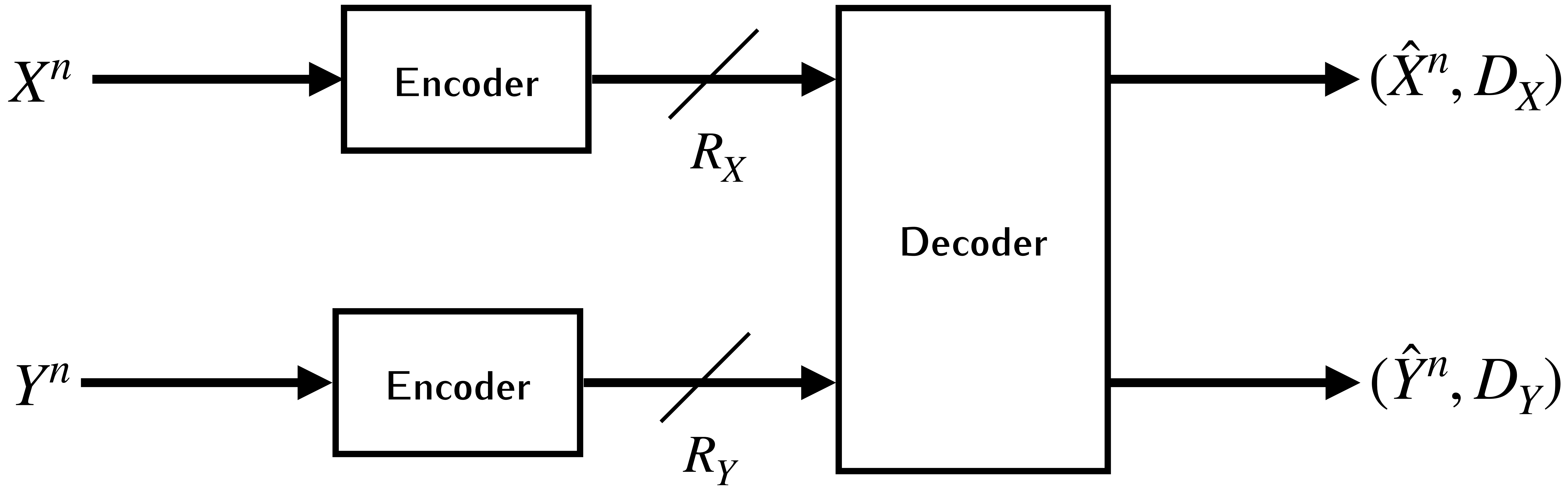
Ballé, Chou, Minnen et al. (IEEE STSP, 2021)

Part II.A

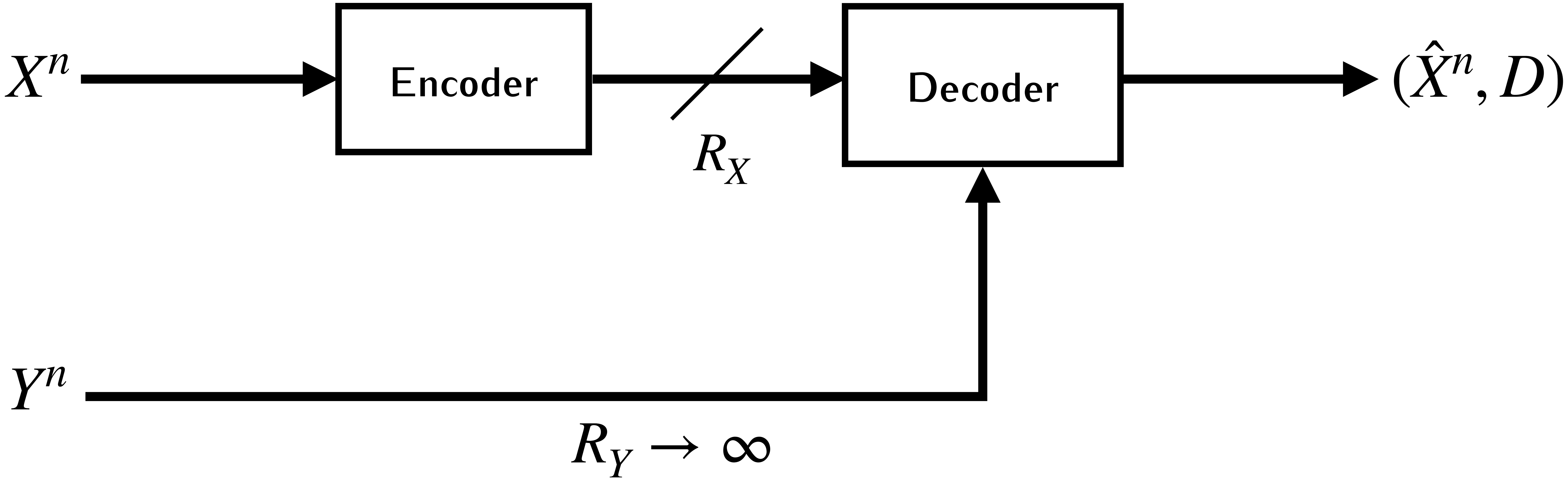
Learning-Based Distributed Data Compression

Distributed compression with 2 sources

Distributed compression with 2 sources

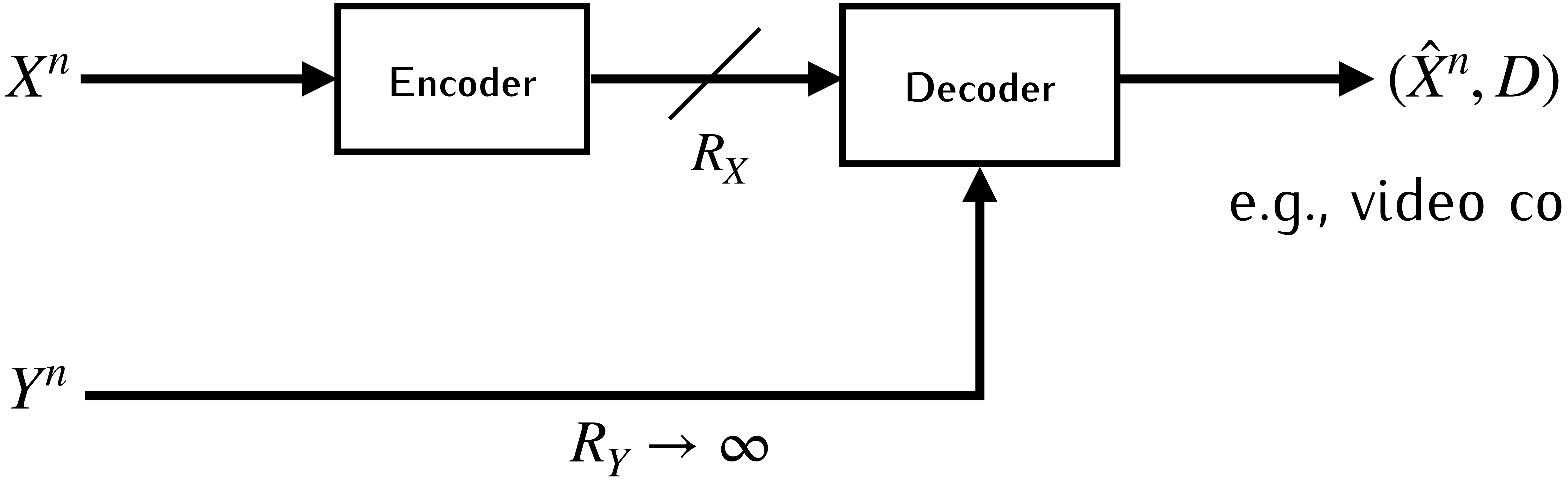


Special case of distributed compression: rate-distortion with (decoder-only) side information



Special case of distributed compression: rate-distortion with (decoder-only) side information

Also known as Wyner-Ziv setup in information theory.

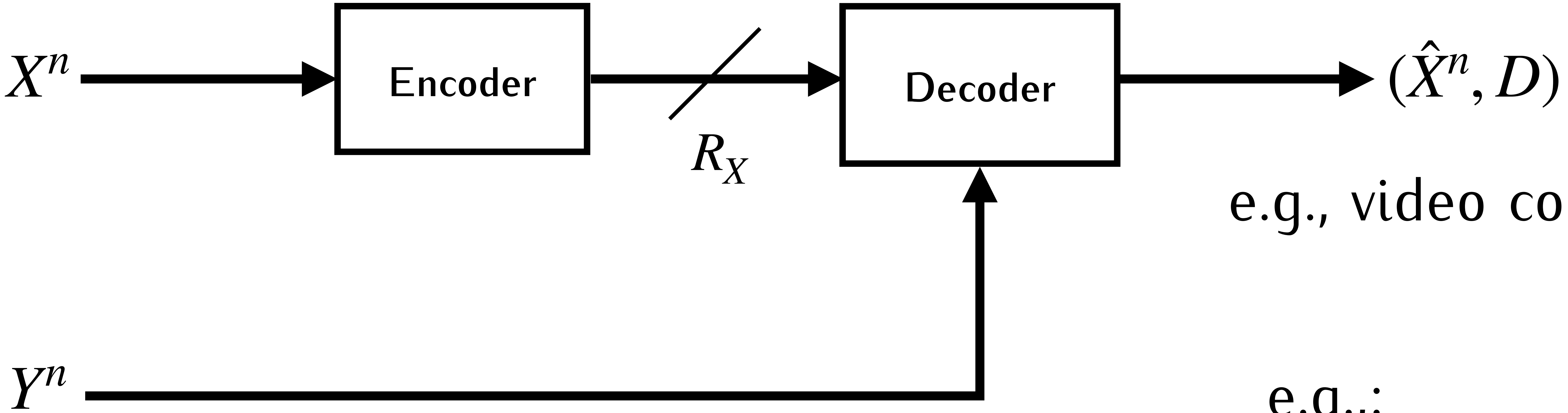


e.g., video coding

Wyner & Ziv (IEEE Trans. on Information Theory, 1976)

Special case of distributed compression: rate-distortion with (decoder-only) side information

Also known as Wyner-Ziv setup in information theory.



e.g., video coding

e.g.,:

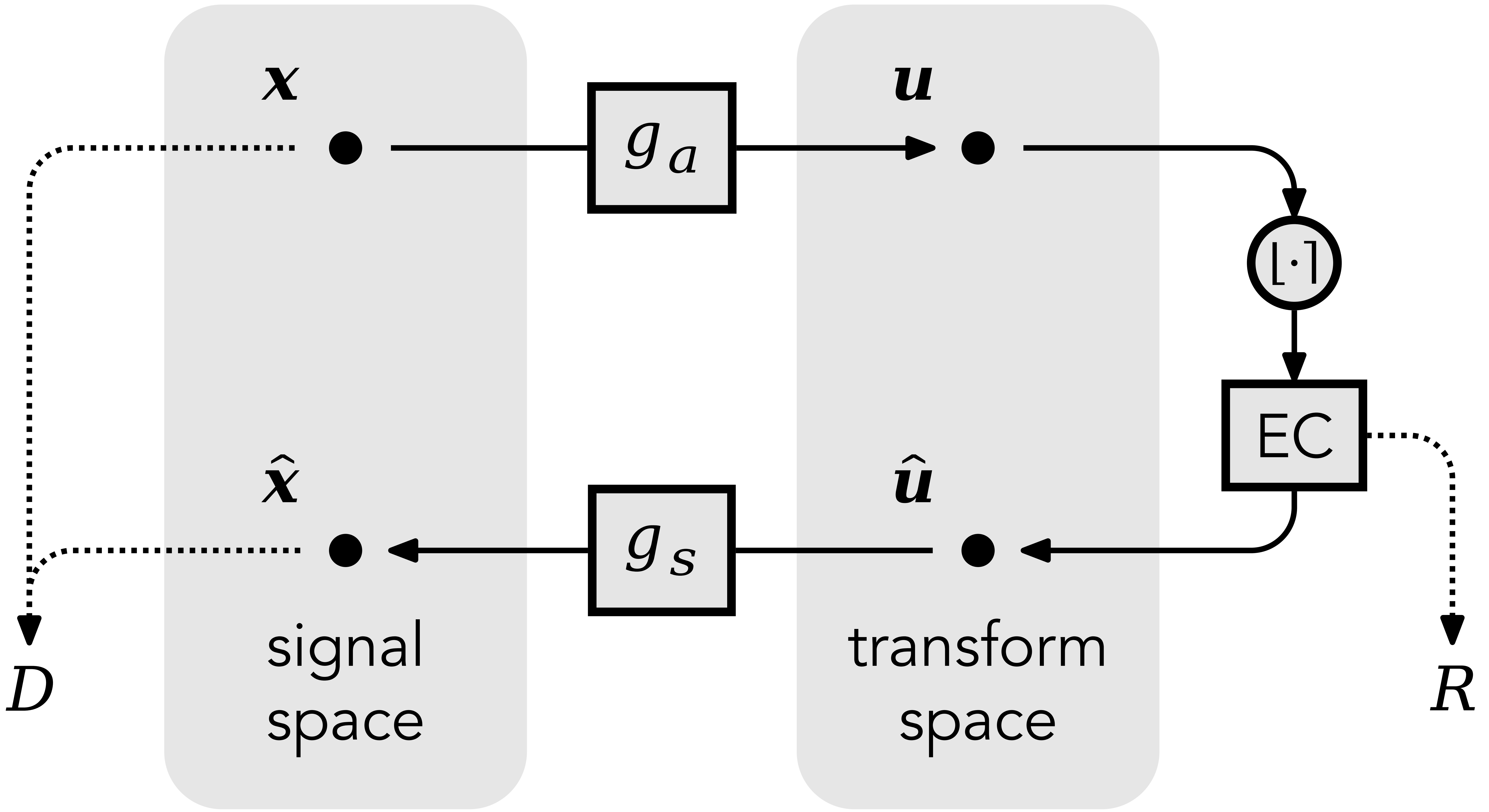
$X = Y + N$

$Y \sim \mathcal{N}(0,1)$

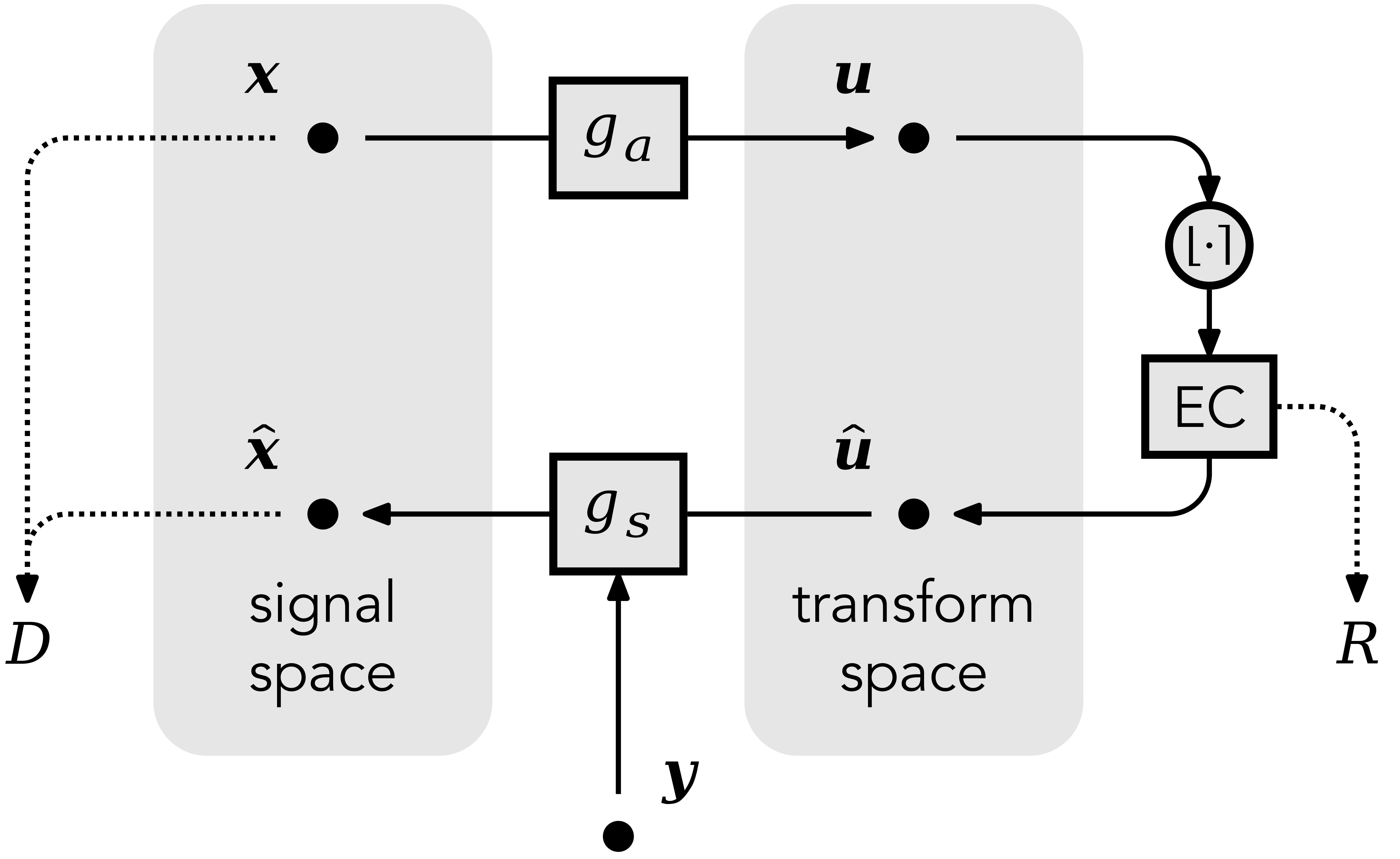
$N \sim \mathcal{N}(0,10^{-1})$

Wyner & Ziv (IEEE Trans. on Information Theory, 1976)

With learned compression, just give y to decoder?

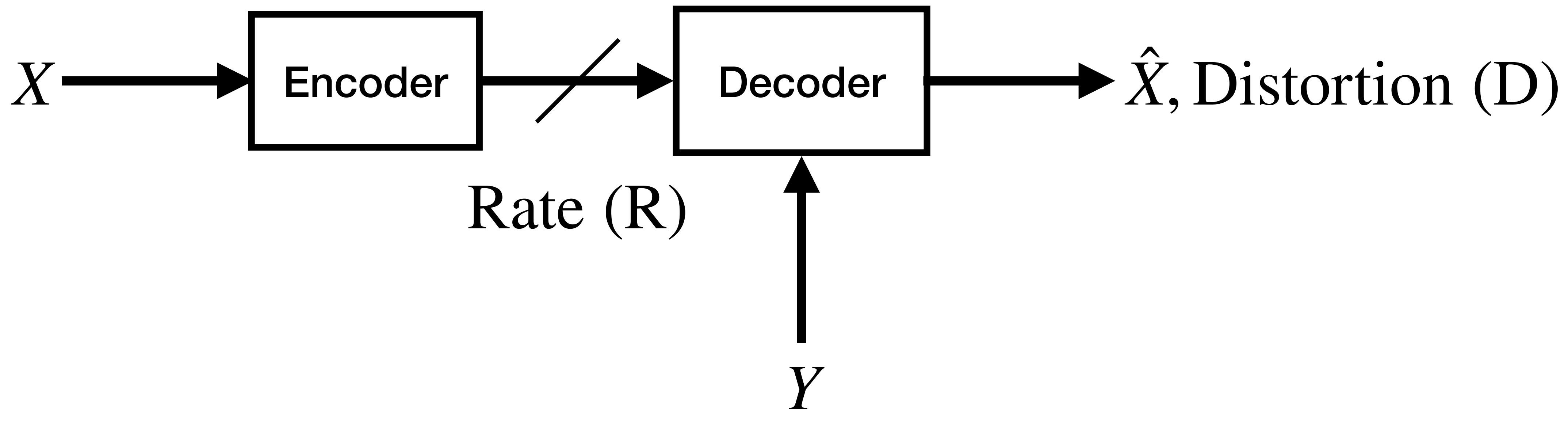


With learned compression, just give y to decoder?



Nope! NTC doesn't work.

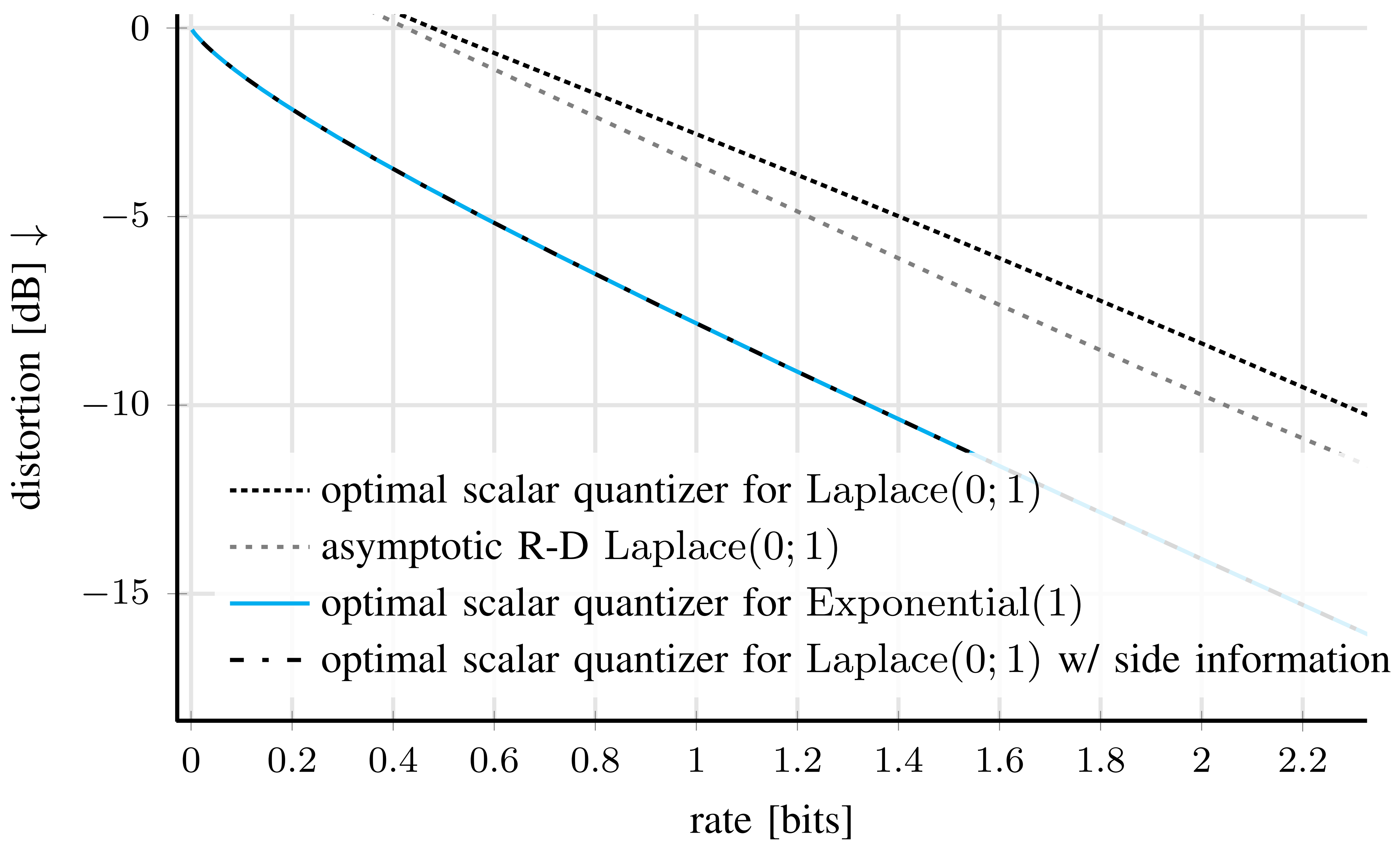
Nope! NTC doesn't work.



$X \sim \mathrm{Laplace}(0,1)$

$Y = \mathrm{sgn}(X)$

Nope! NTC doesn't work.

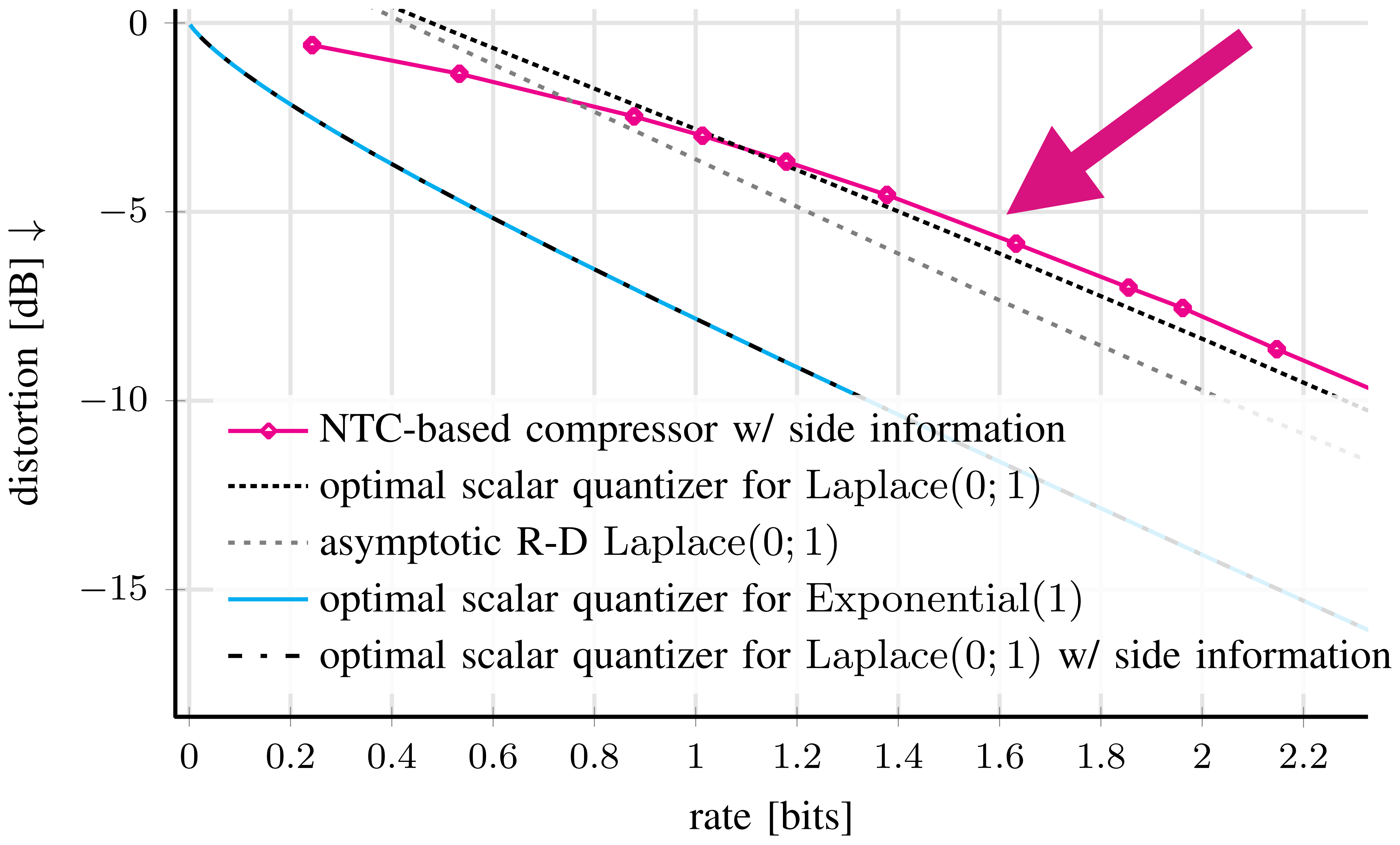


$X \sim \text{Laplace}(0,1)$

$Y = \text{sgn}(\ X\ )$

Nope! NTC doesn't work.

- NTC compressor just barely performs better than point-to-point (when y is not present).



$X \sim \text{Laplace}(0,1)$

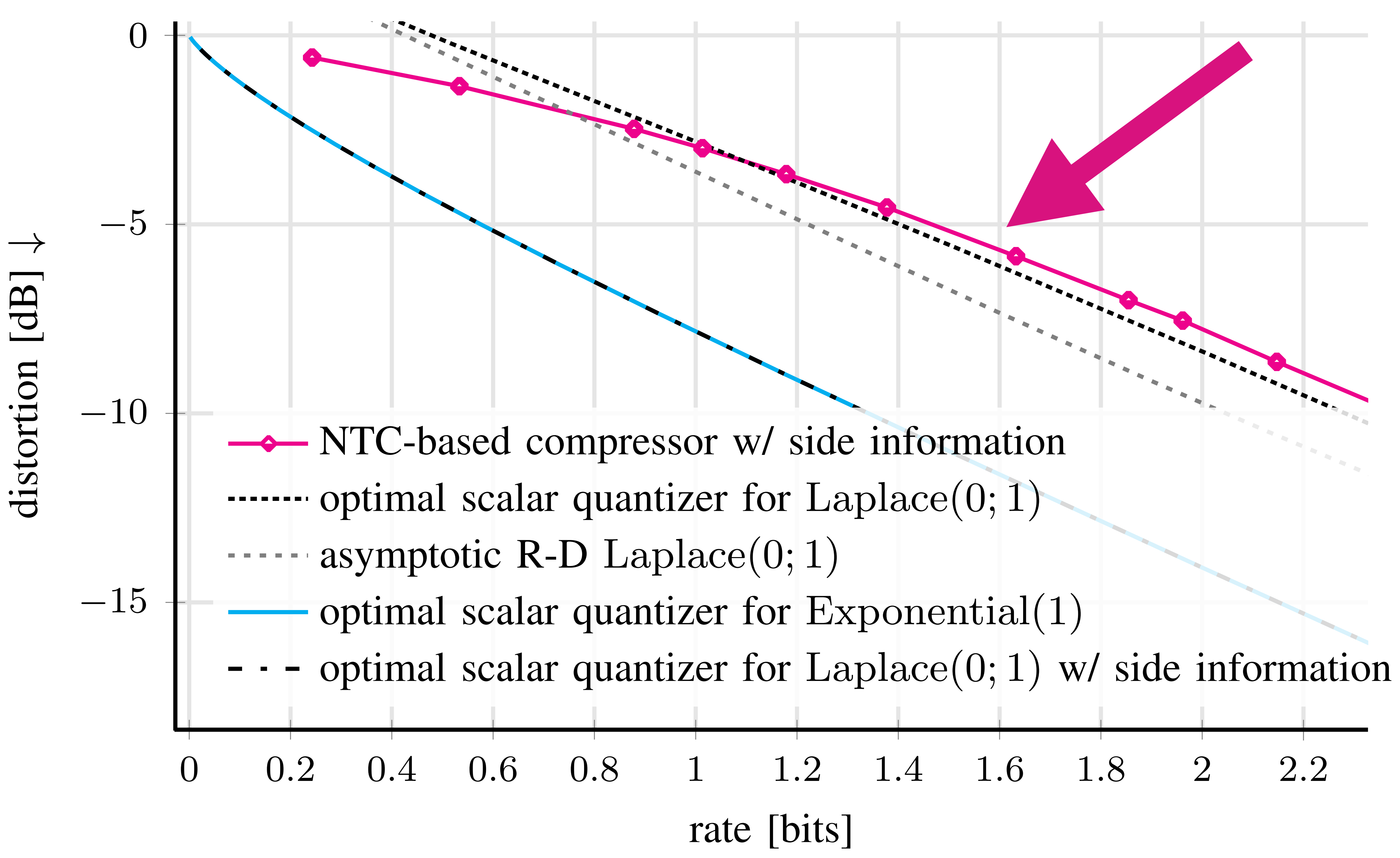
$Y = \text{sgn}(X)$

Ozyilkan, Ballé & Erkip (IEEE J. Sel. Areas Information Theory, 2024)

Nope! NTC doesn't work.

- NTC compressor just barely performs better than point-to-point (when $\boldsymbol{y}$ is not present).

- Some dumb mistake 🤡 ?
  - NTC-based compressors have a **learning bias towards *smooth functions***.



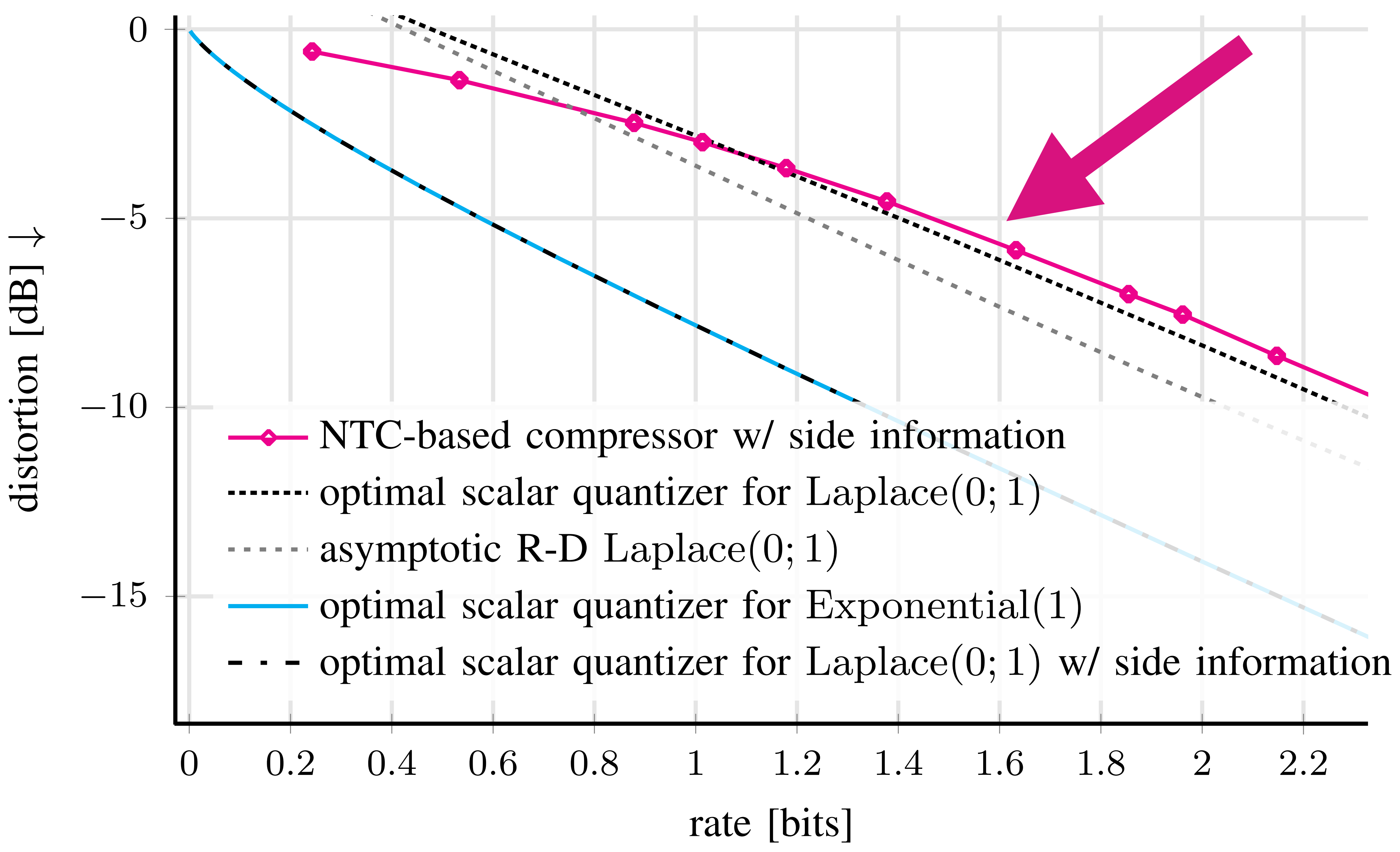
$X \sim \text{Laplace}(0,1)$

$Y = \text{sgn}(X)$

Ozyilkan, Ballé & Erkip (IEEE J. Sel. Areas Information Theory, 2024)
Ozyilkan, Ballé, Bhadane, Wagner & Erkip (Compression Workshop @ NeurIPS, 2024)

Nope! NTC doesn't work.

- NTC compressor just barely performs better than point-to-point (when y is not present).
- Some dumb mistake 🤡 ?
  - NTC-based compressors have a **learning bias towards *smooth functions***.
  - $\Longrightarrow$ cannot recover **many-to-one maps at the encoder**.



$X \sim \text{Laplace}(0,1)$

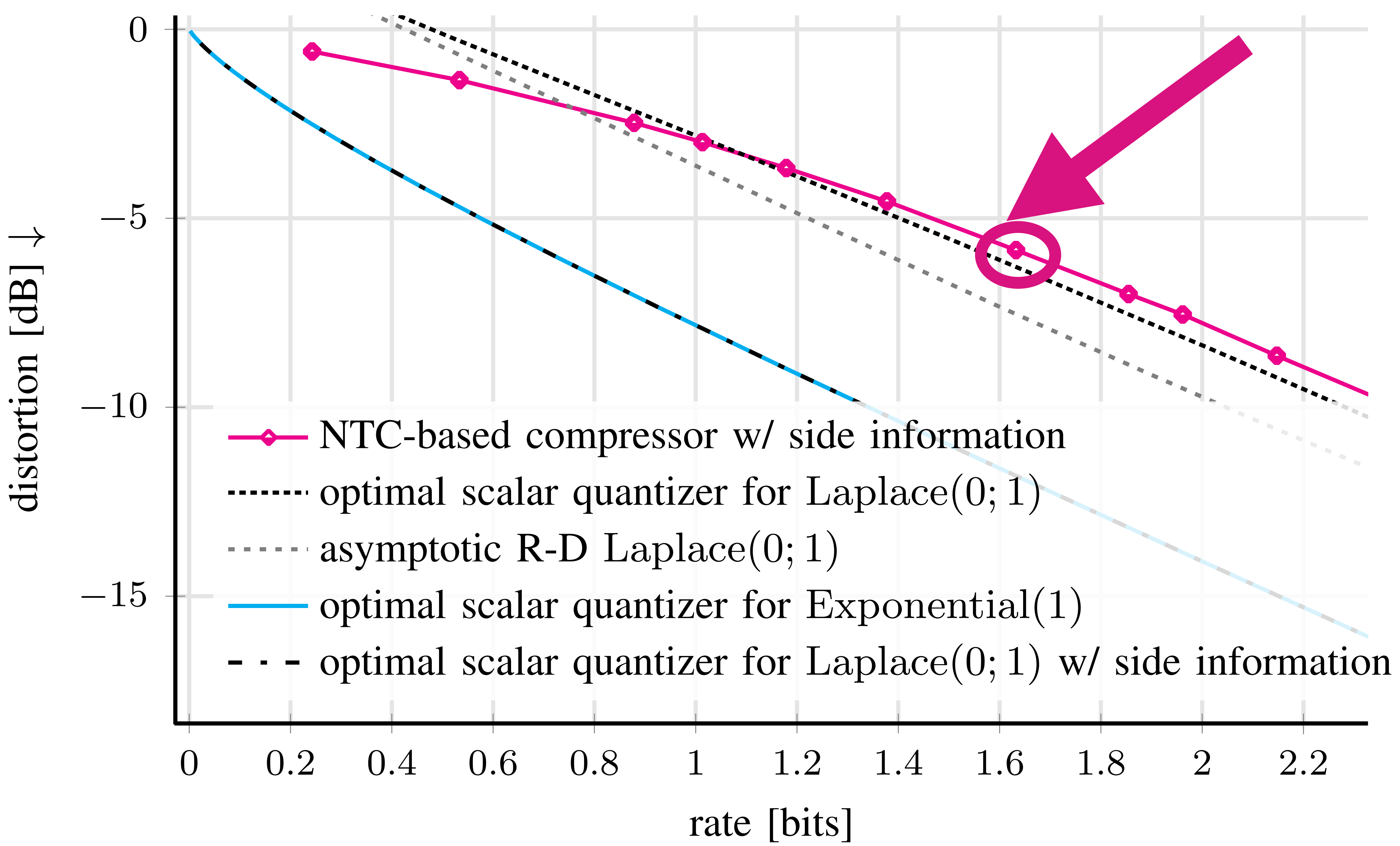
$Y = \text{sgn}(\ X\ )$

Ozyilkan, Ballé & Erkip (IEEE J. Sel. Areas Information Theory, 2024)

Ozyilkan, Ballé, Bhadane, Wagner & Erkip (Compression Workshop @ NeurIPS, 2024)

Nope! NTC doesn't work.

- NTC compressor just barely performs better than point-to-point (when ***y*** is not present).
- Some dumb mistake 🤡 ?
    - NTC-based compressors have a **learning bias towards *smooth functions***.
    - $\Longrightarrow$ cannot recover **many-to-one maps at the encoder**.



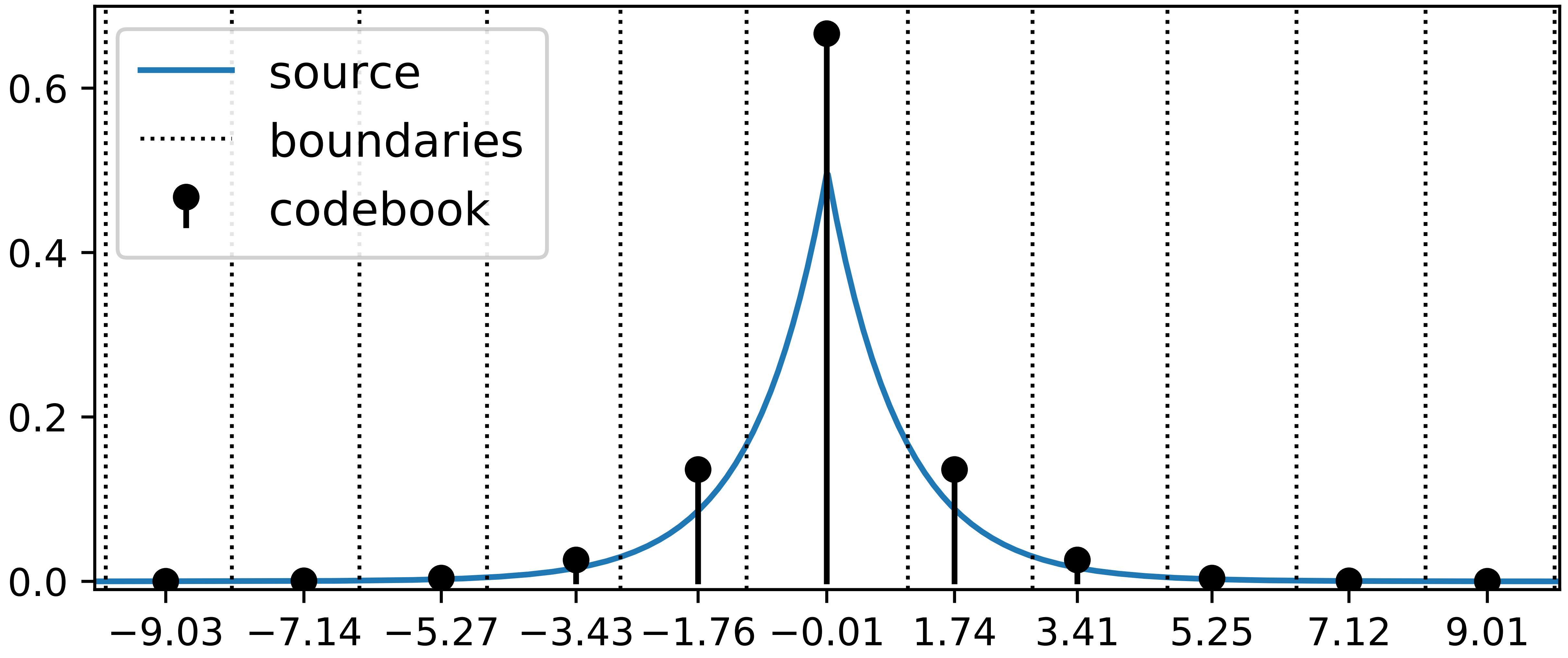
$X \sim \text{Laplace}(0{,}1)$

$Y = \text{sgn}(\ X\ )$

Ozyilkan, Ballé & Erkip (IEEE J. Sel. Areas Information Theory, 2024)
Ozyilkan, Ballé, Bhadane, Wagner & Erkip (Compression Workshop @ NeurIPS, 2024)

What does NTC learn?

What does NTC learn?



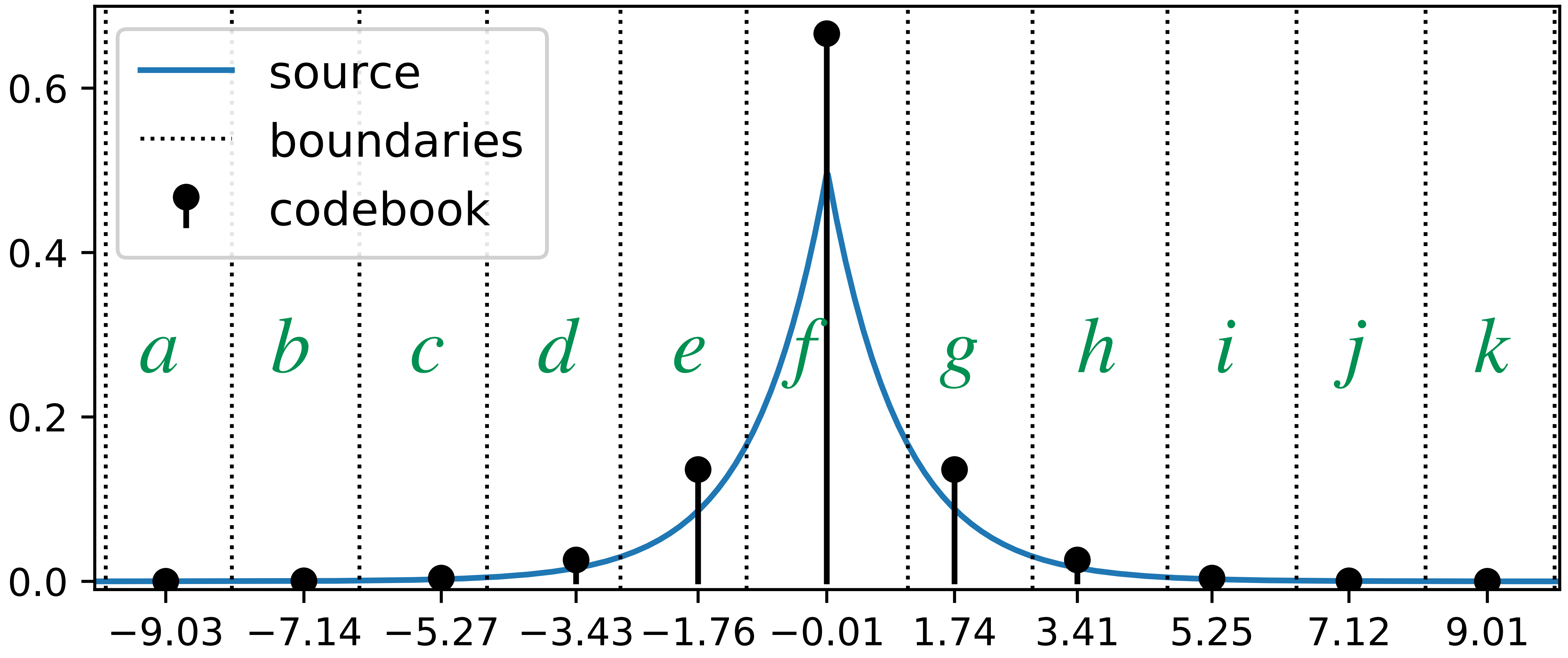
$X \sim \text{Laplace}(0,1)$

$Y = \text{sgn}(\ X\ )$

Sullivan (IEEE Trans. on Information Theory, 1996)

What does NTC learn?

codebook indices:



$X \sim \text{Laplace}(0,1)$

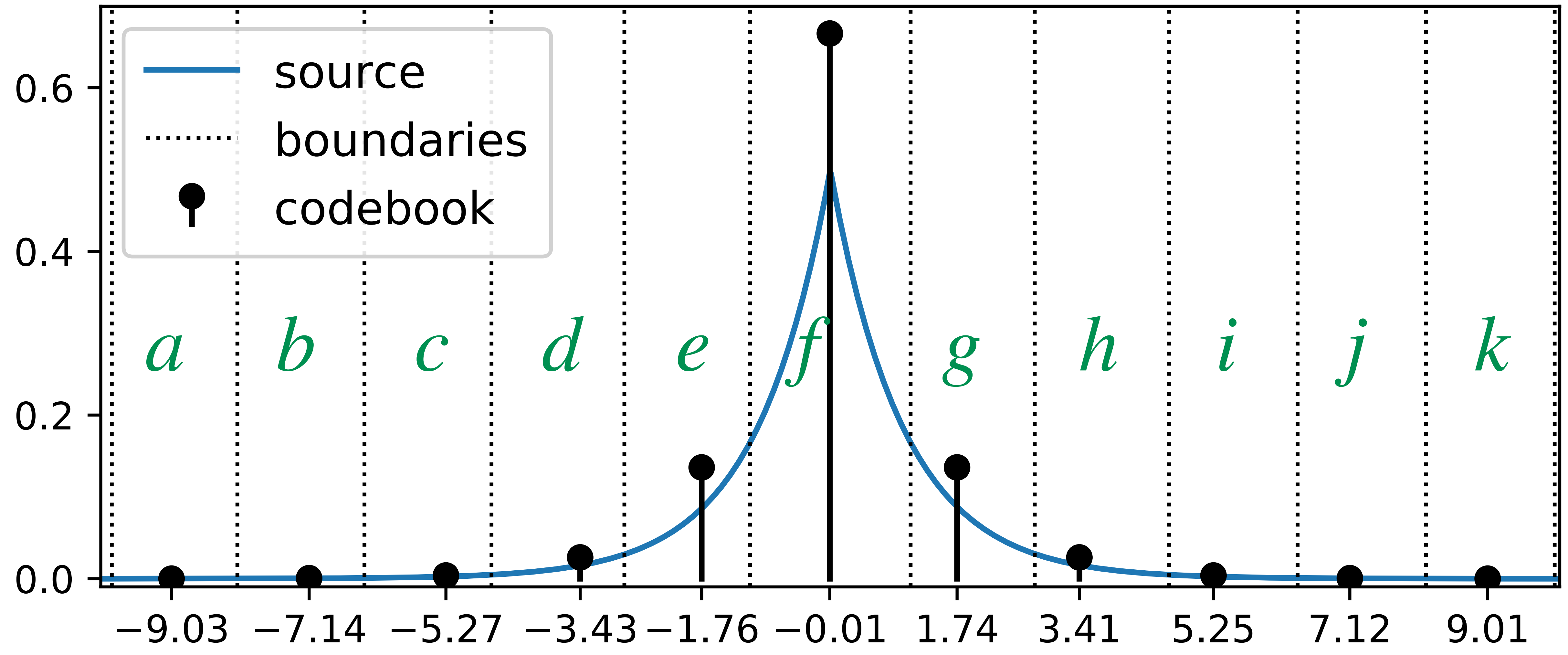
$Y = \text{sgn}(\ X\ )$

NTC assigns ***unique index*** for each interval.

Sullivan (IEEE Trans. on Information Theory, 1996)

What does NTC learn?

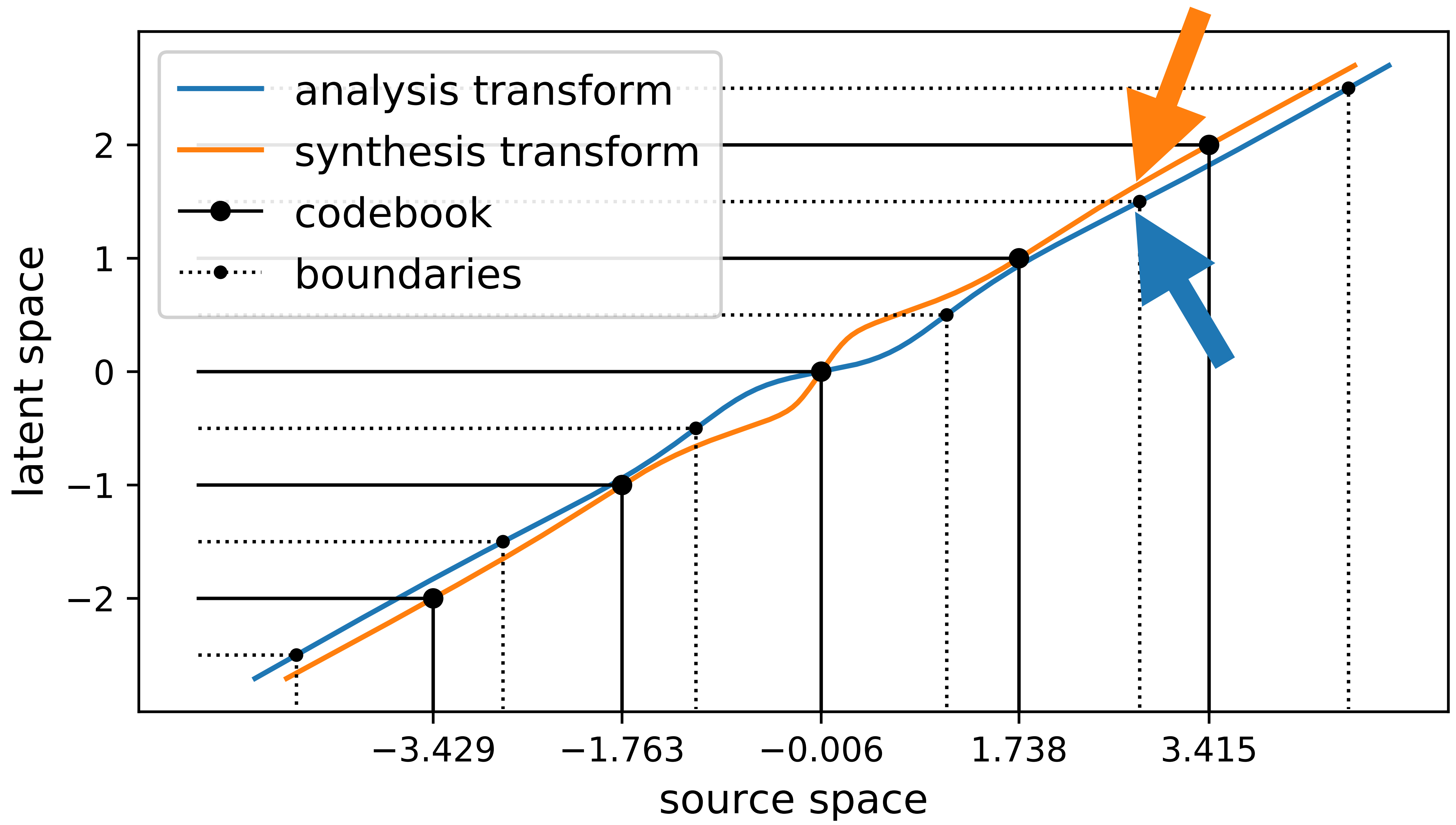
codebook indices:



$X \sim \text{Laplace}(0,1)$

$Y = \text{sgn}(\ X\ )$

NTC assigns *unique index* for each interval.



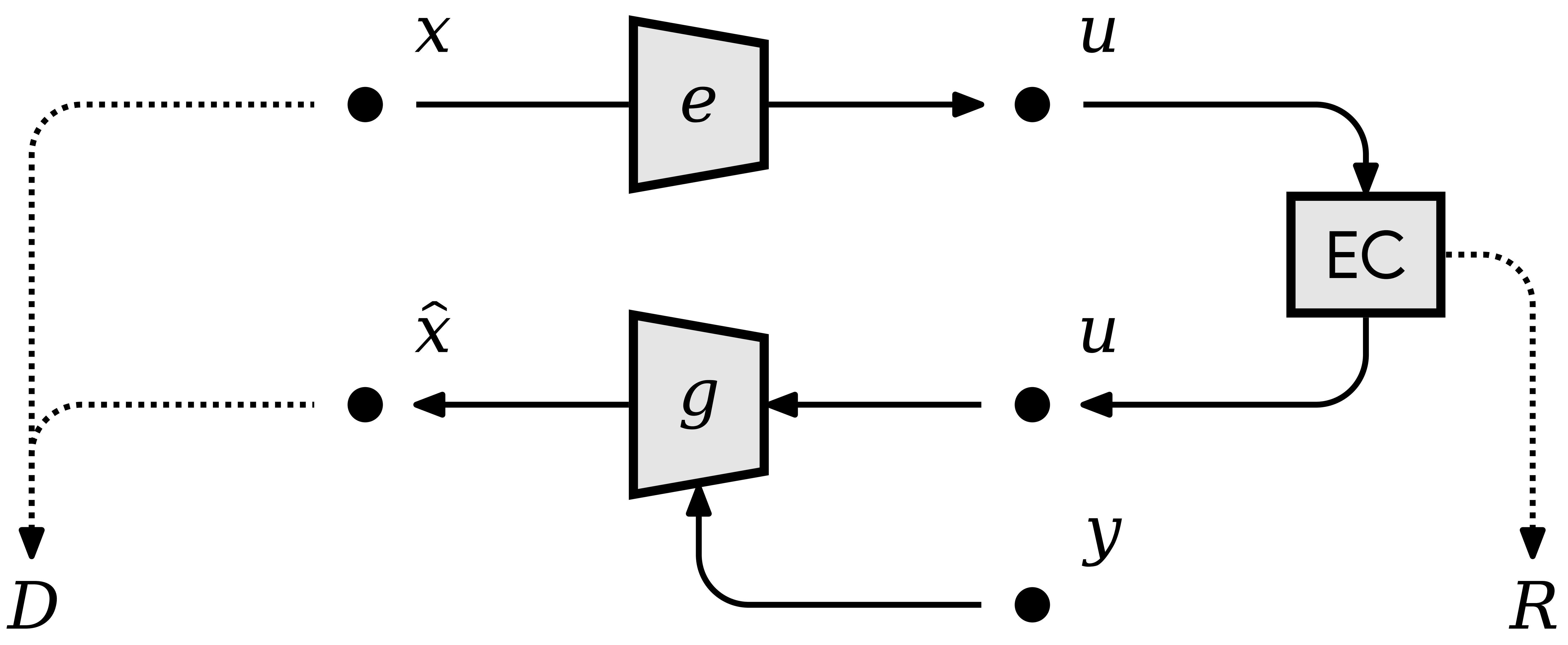
encoder: $\mathbb{R} \mapsto \mathbb{Z}$

decoder: $\mathbb{Z} \mapsto \mathbb{R}$

NTC recovers *smooth* nonlinear transforms.

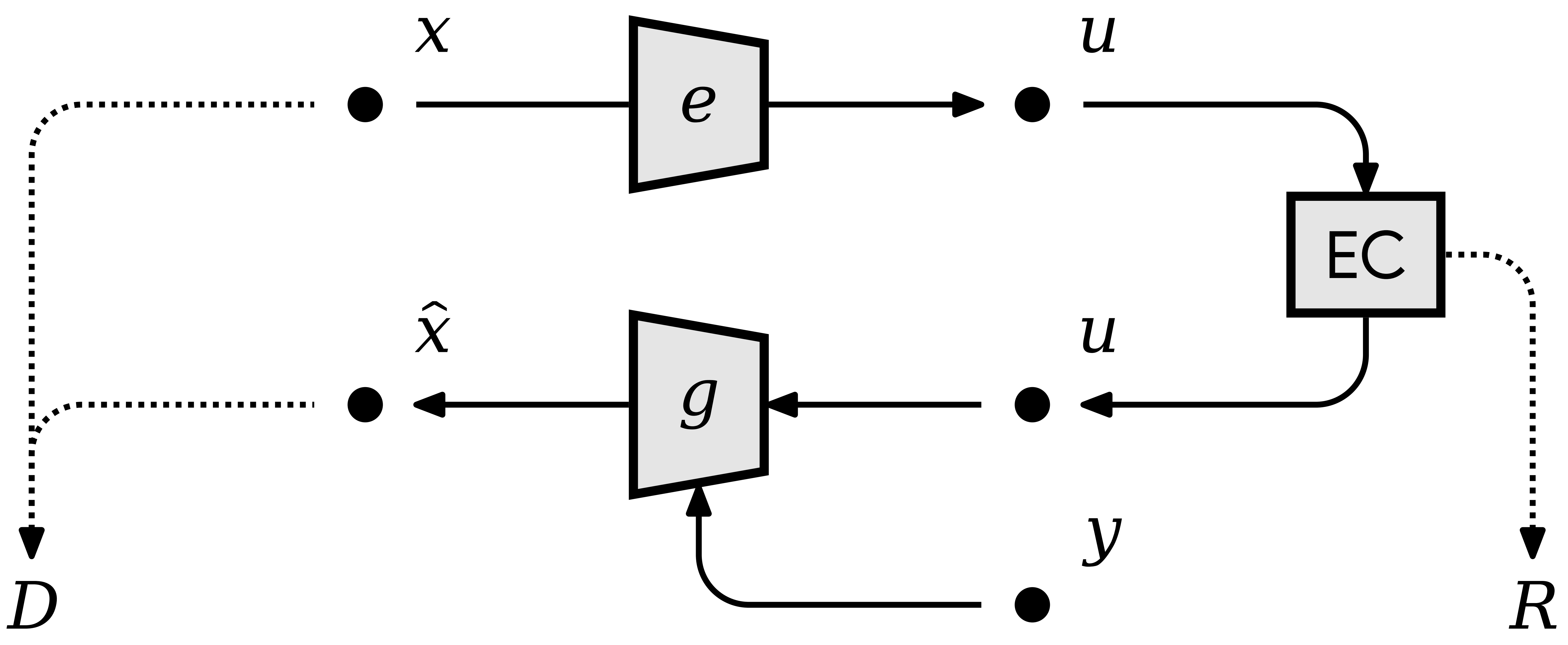
Sullivan (IEEE Trans. on Information Theory, 1996)

How to overcome the smoothness learning bias?



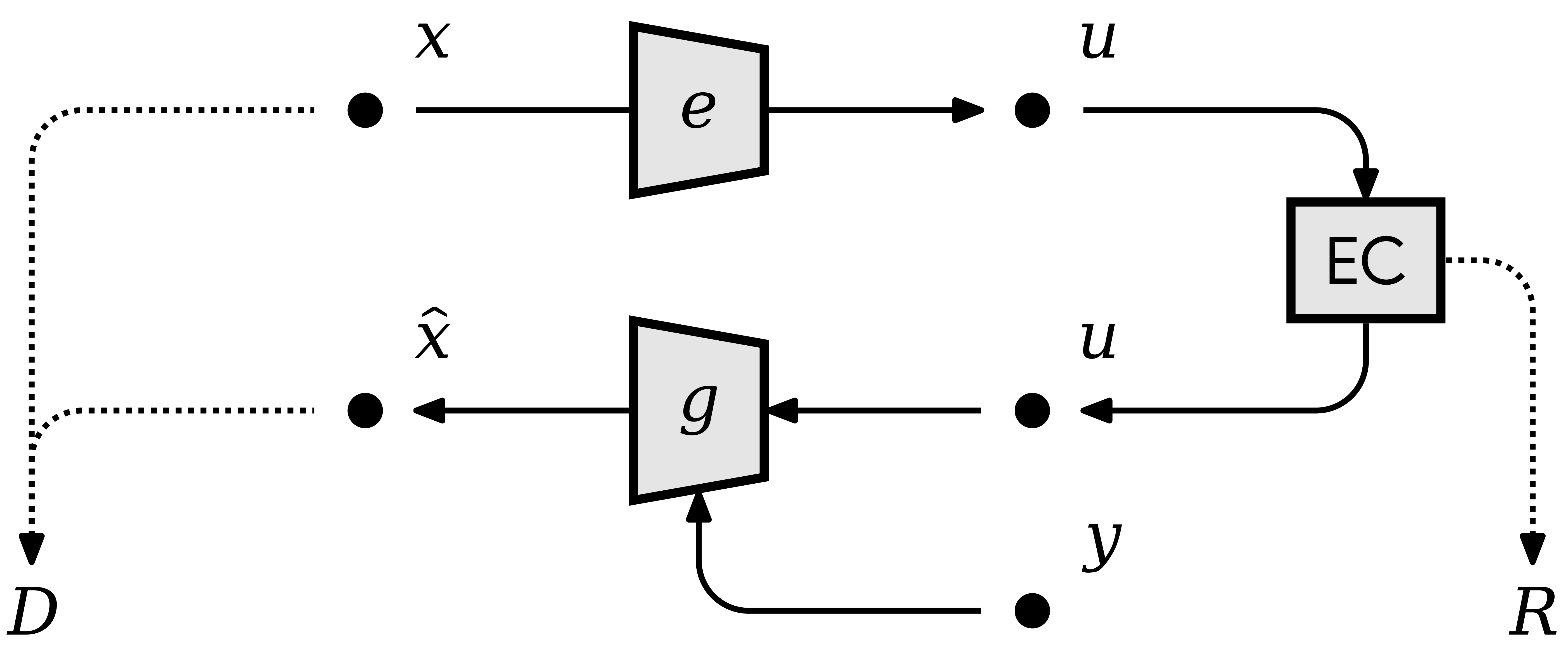
- Motivation: encoder can implement arbitrary maps?

How to overcome the smoothness learning bias?



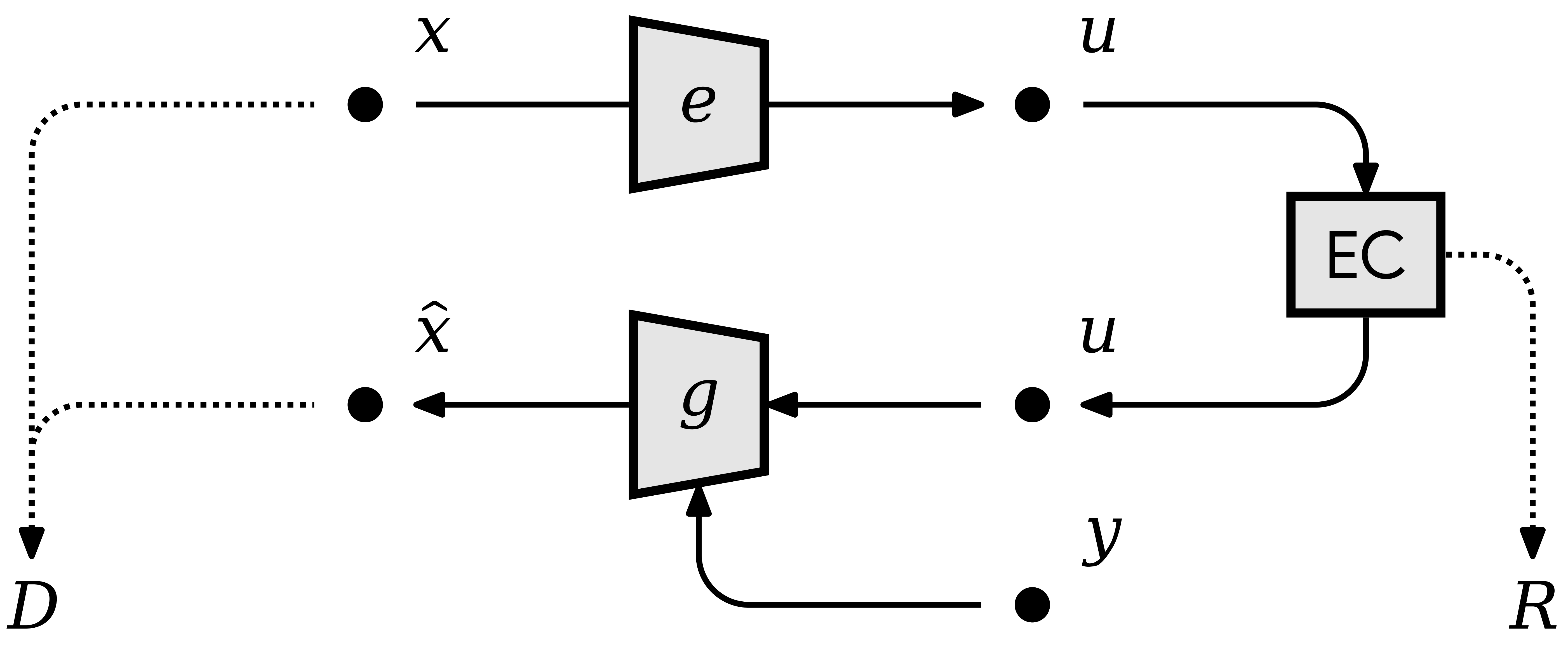
- Motivation: encoder can implement arbitrary maps?
- **Let the encoder output indicator** (one-hot) **functions**, rather than vectors rounded to integers.

How to overcome the smoothness learning bias?



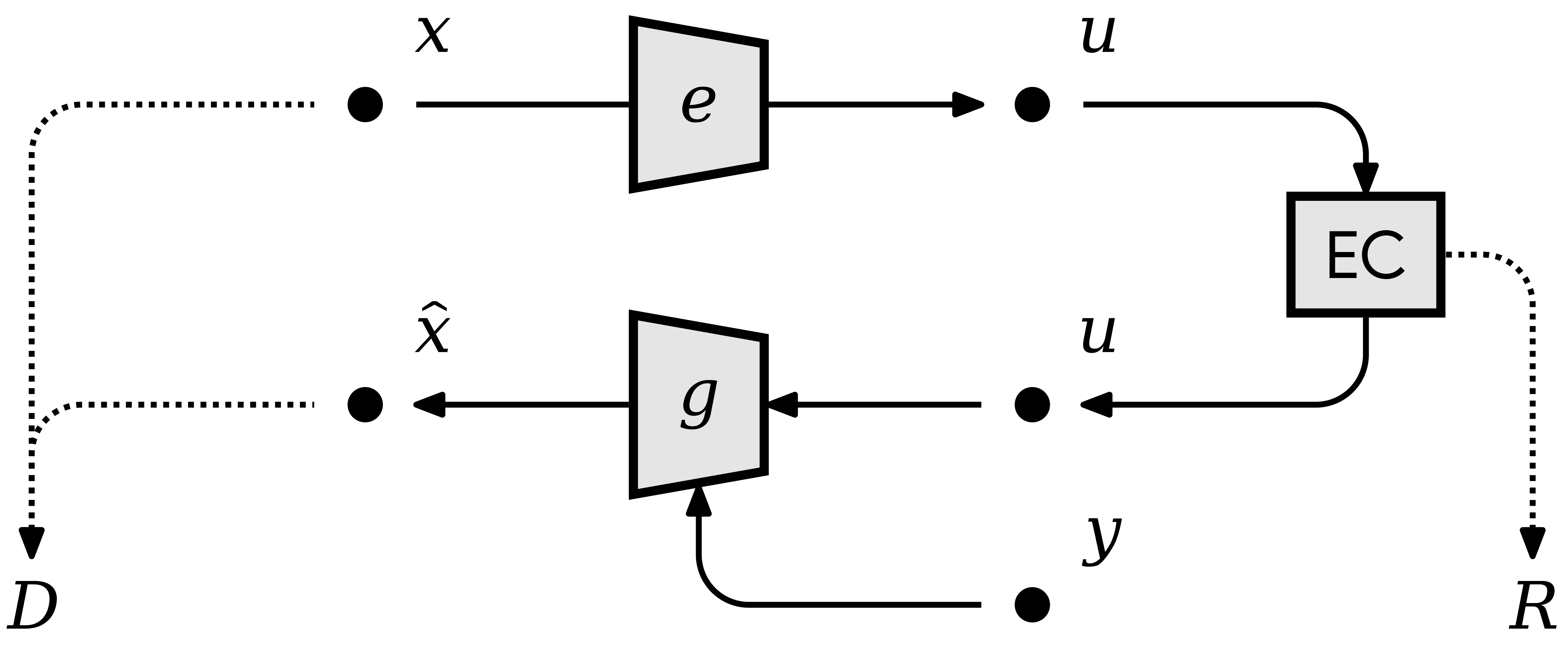
- Motivation: encoder can implement arbitrary maps?
- **Let the encoder output indicator** (one-hot) **functions**, rather than vectors rounded to integers.
  - This gives the encoder the same structure as a **classification network**.
  - Encoder be $e_\theta : \mathbb{R} \mapsto \{a, b, c, d, \cdots\}$, instead of $e_\theta : \mathbb{R} \mapsto \mathbb{Z}$ as in NTC.

Replacing NTC with something less constrained



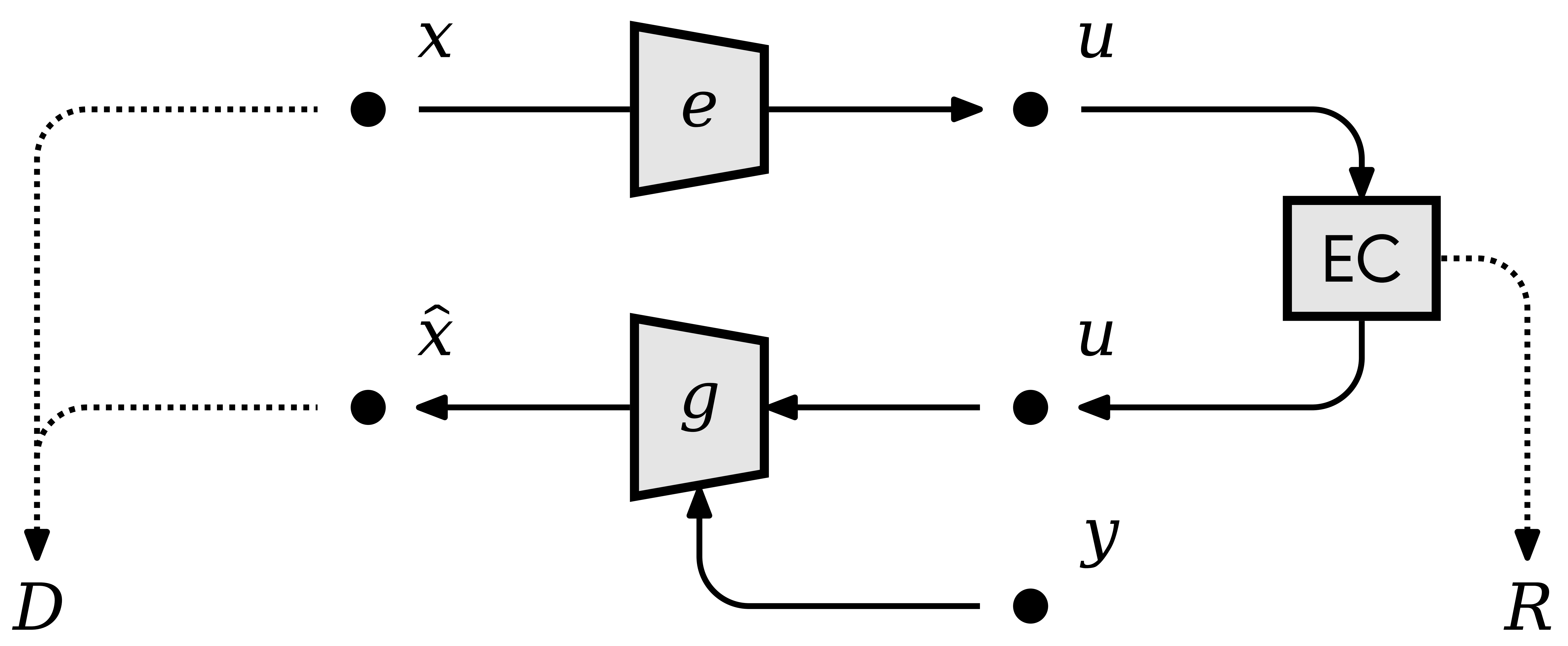
- Let encoder $e_{\boldsymbol{\theta}}(x)$ output "logits" $(\alpha_1, \alpha_2, \alpha_3, \cdots)$.

Replacing NTC with something less constrained



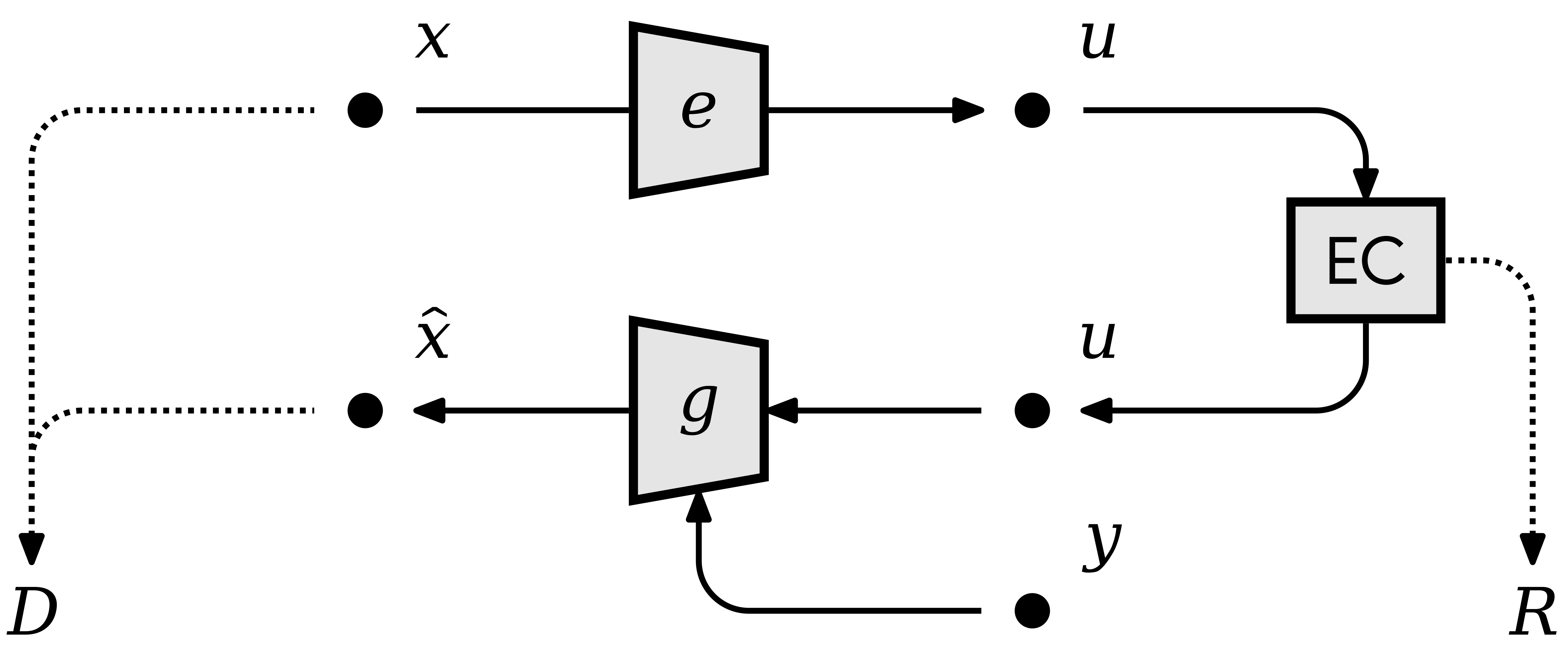
- Let encoder $e_{\boldsymbol{\theta}}(x)$ output "logits" $(\alpha_1, \alpha_2, \alpha_3, \cdots)$.
- Choose quantization index as $u = \operatorname{argmax}_i \alpha_i$.

Replacing NTC with something less constrained



- Let encoder $e_{\boldsymbol{\theta}}(x)$ output "logits" $(\alpha_1, \alpha_2, \alpha_3, \cdots)$.
- Choose quantization index as $u = \operatorname{argmax}_i \alpha_i$.
- $x \mapsto \alpha_k$ can still be a smooth function although $x \mapsto u$ *may not be*!

Replacing NTC with something less constrained



- Let encoder $e_{\boldsymbol{\theta}}(x)$ output "logits" $(\alpha_1, \alpha_2, \alpha_3, \cdots)$.

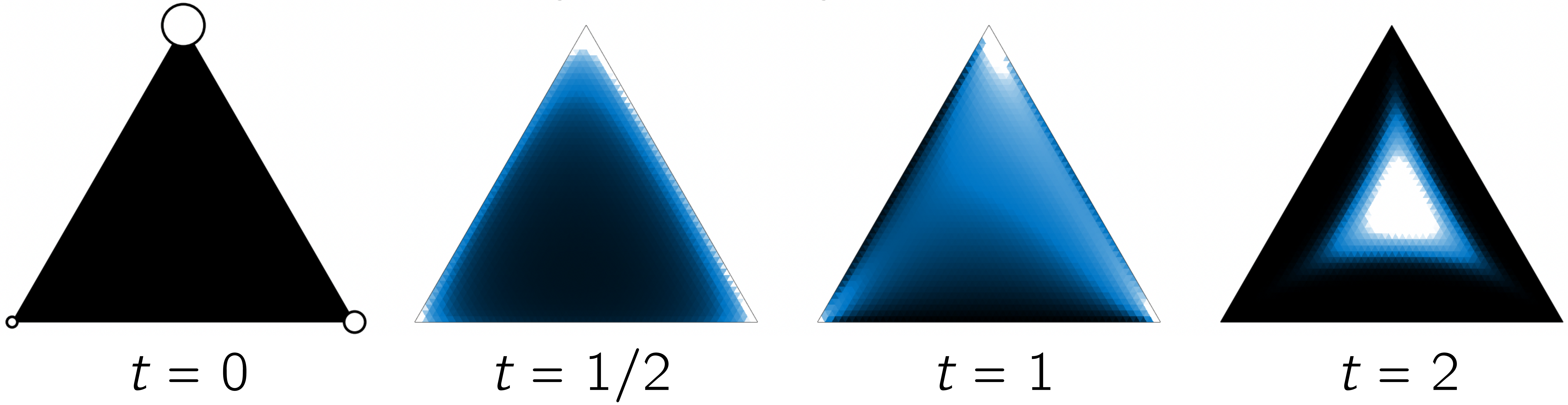
not differentiable 😿

- Choose quantization index as $u = \operatorname{argmax}_i \alpha_i$.

- $x \mapsto \alpha_k$ can still be a smooth function although $x \mapsto u$ *may not be*!

Gumbel-Softmax trick

Allows **differentiable sampling** from categorical-like distribution.



Rather than sampling an index u, we sample a vector $\boldsymbol{u}$:

$$u_k = \frac{\exp((\alpha_k + G_k)/t)}{\sum_i \exp_i((\alpha_i + G_i)/t)}$$

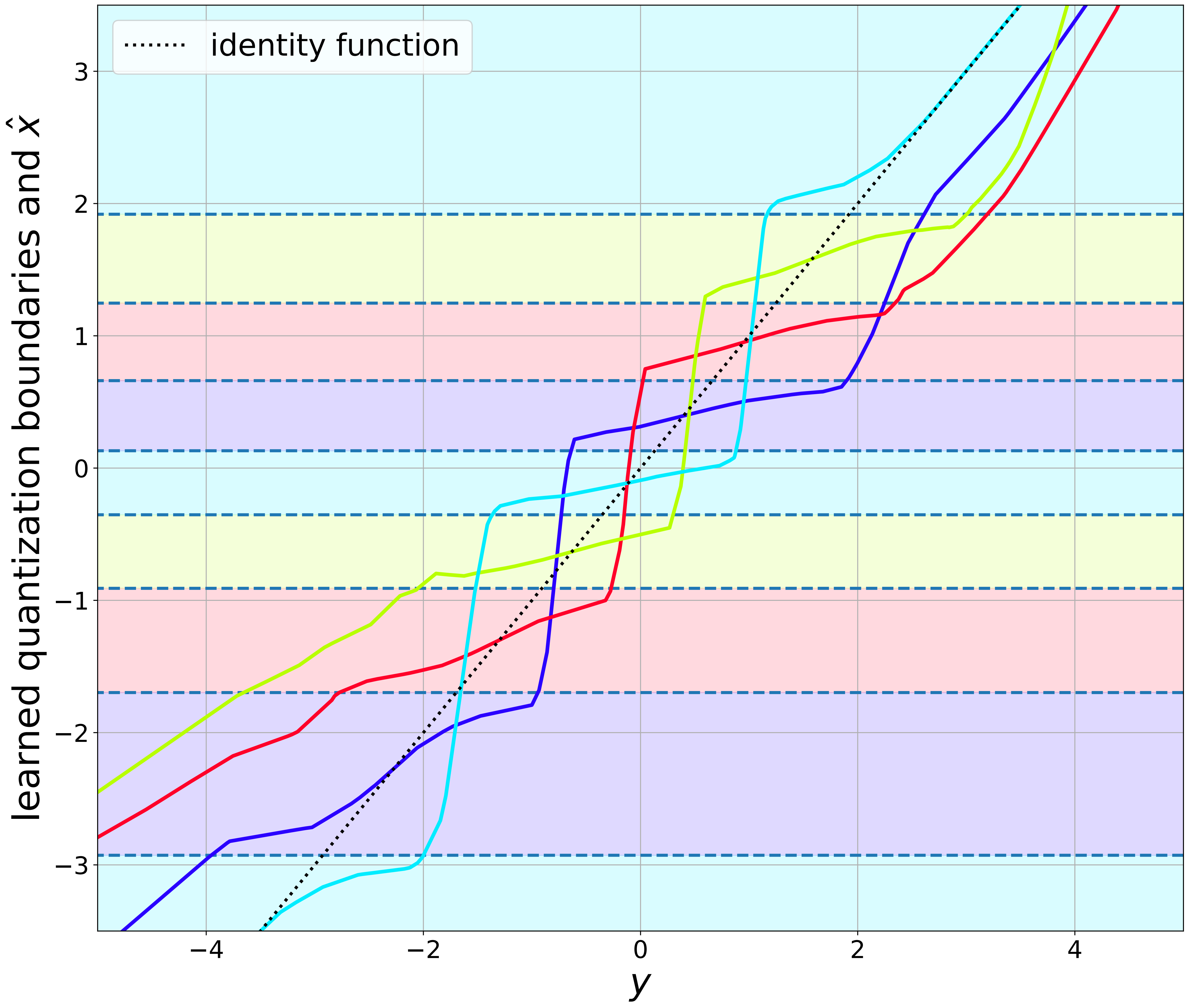
As $t \to 0^+$, we approach $\arg\max$.

softmax is differentiable!!

Maddison et al. (ICLR, 2017)

Learned compressor recovers “binning” (grouping)

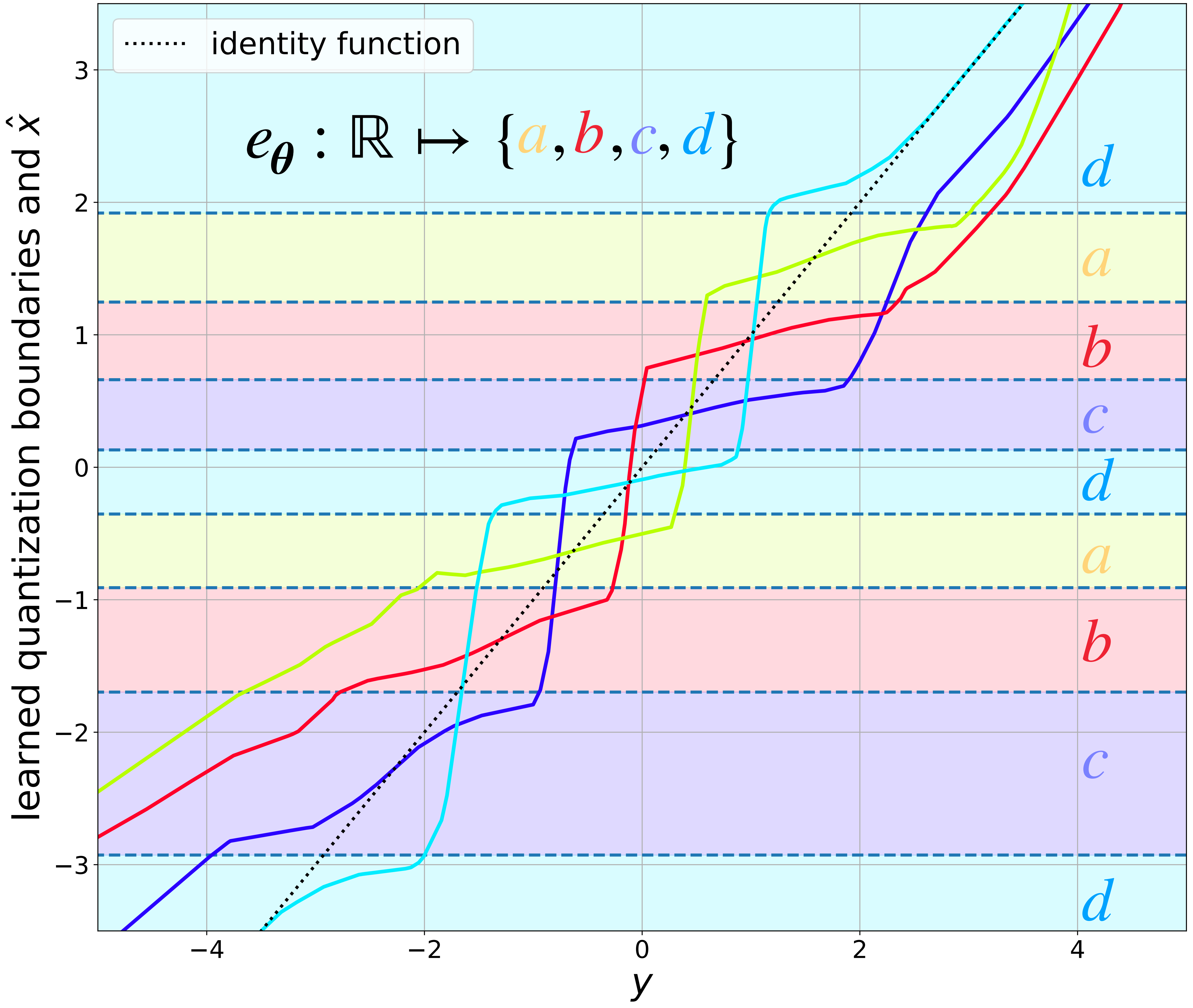
$$X = Y + N$$
$$Y \sim \mathcal{N}(0,1)$$
$$N \sim \mathcal{N}(0,10^{-1})$$



Ozyilkan, Ballé & Erkip (ISIT, 2023 & Neural Compression Workshop @ ICML’23 [oral])

Learned compressor recovers “binning” (grouping)

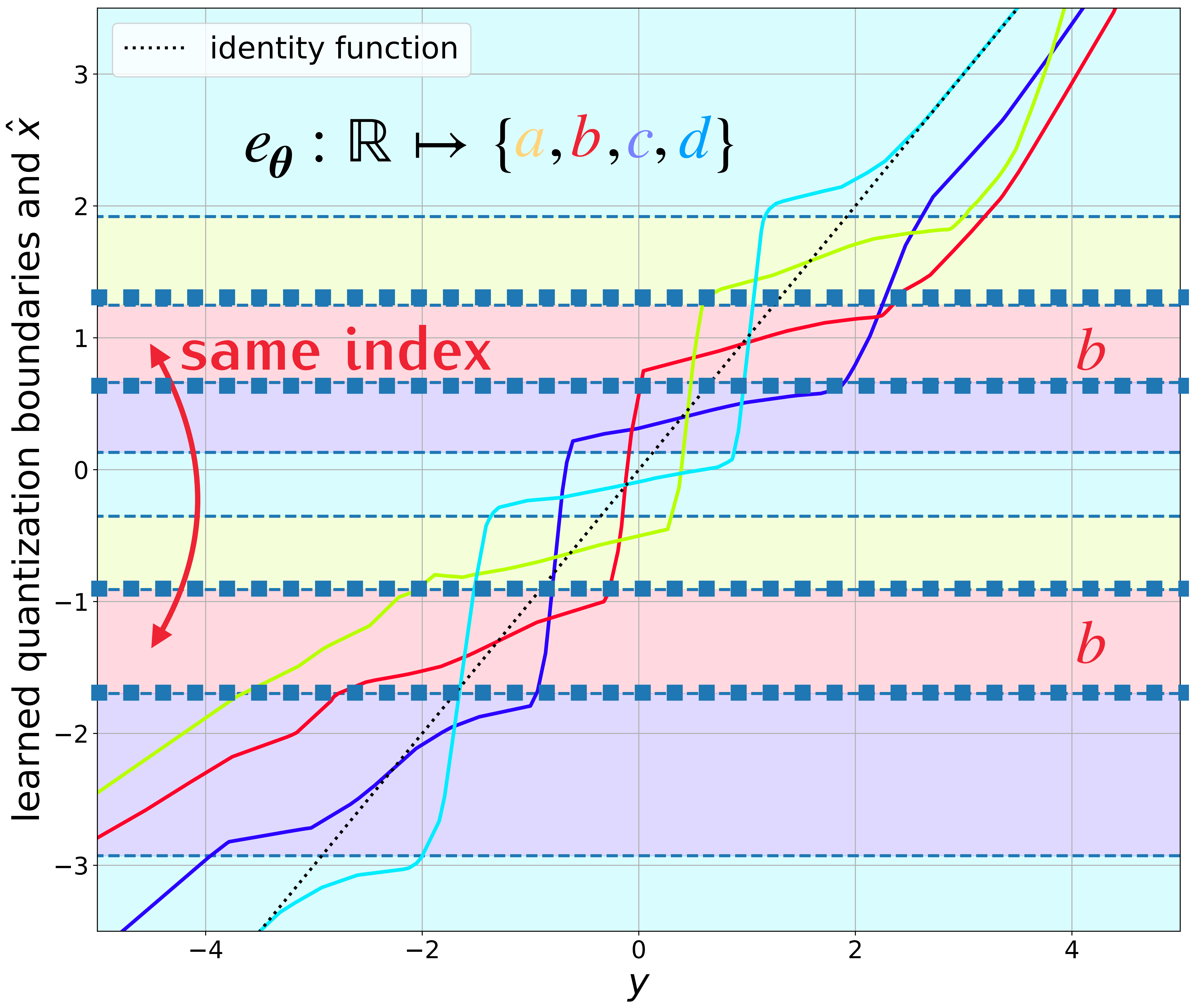
$$X = Y + N$$
$$Y \sim \mathcal{N}(0,1)$$
$$N \sim \mathcal{N}(0,10^{-1})$$



Ozyilkan, Ballé & Erkip (ISIT, 2023 & Neural Compression Workshop @ ICML’23 [oral])

Learned compressor recovers “binning” (grouping)

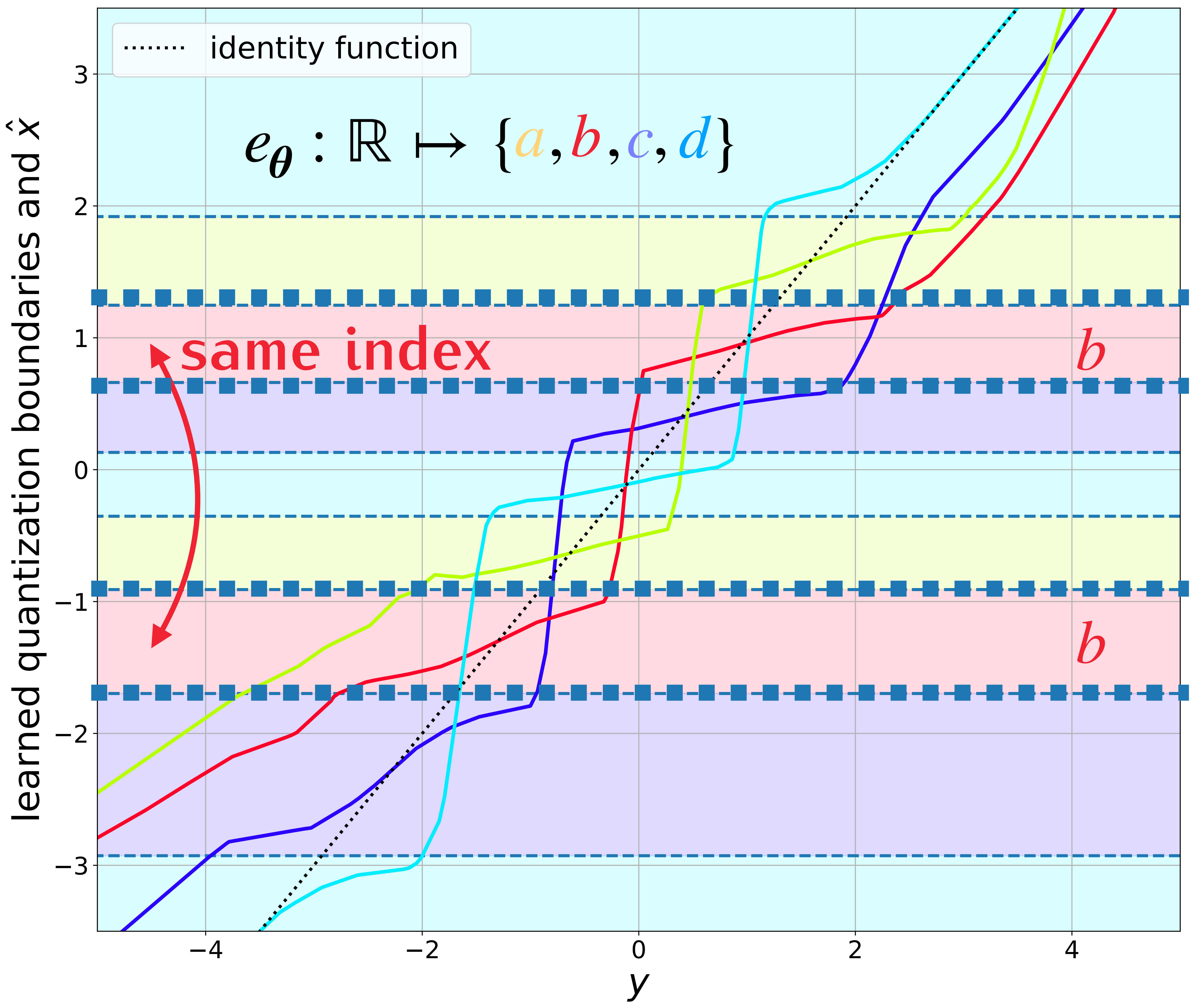
$$X = Y + N$$
$$Y \sim \mathcal{N}(0,1)$$
$$N \sim \mathcal{N}(0,10^{-1})$$



Ozyilkan, Ballé & Erkip (ISIT, 2023 & Neural Compression Workshop @ ICML’23 [oral])

Learned compressor recovers “binning” (grouping)

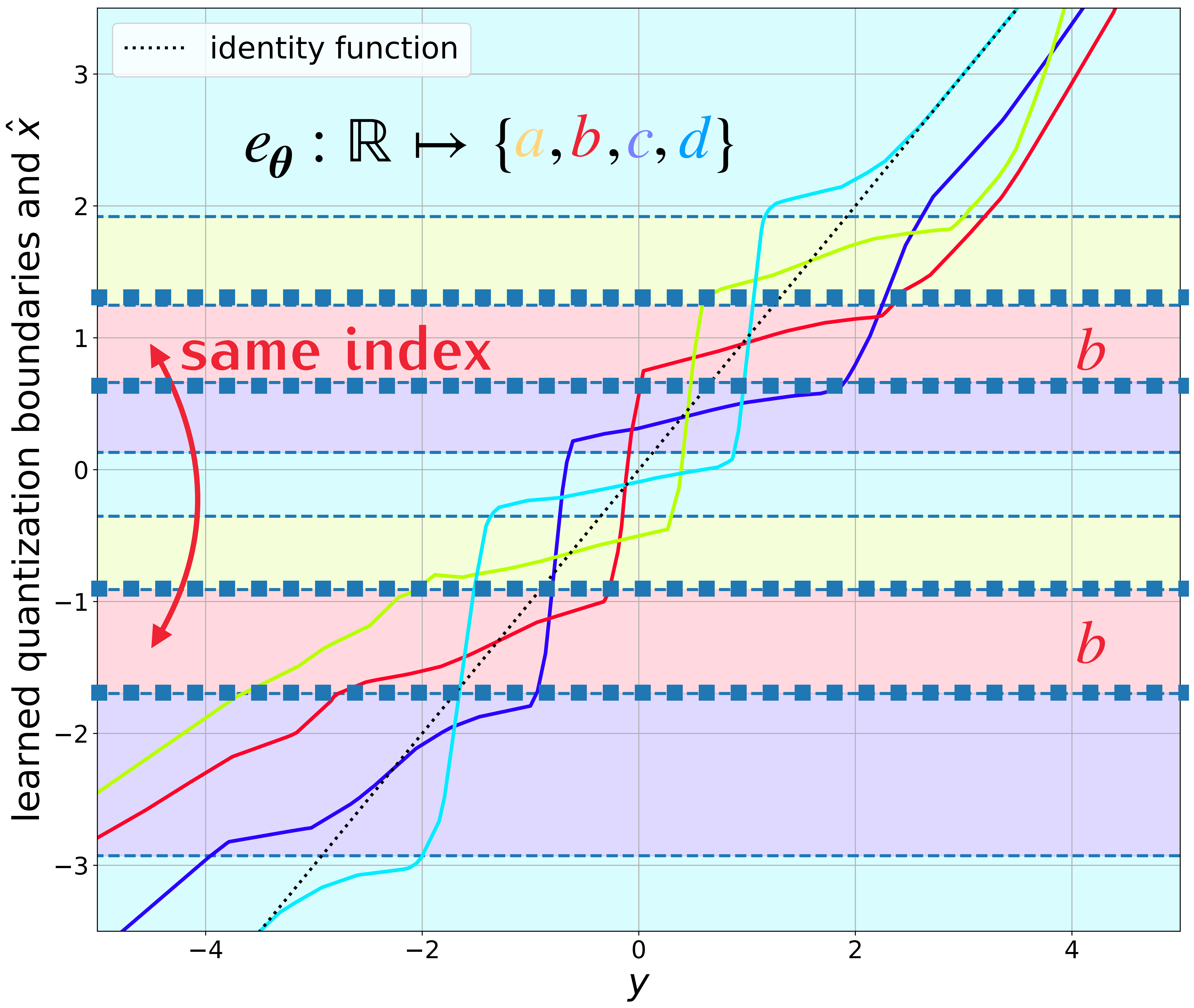
“**Binning**” is widely used to prove bounds in distributed compression,



Ozyilkan, Ballé & Erkip (ISIT, 2023 & Neural Compression Workshop @ ICML'23 [oral])

Learned compressor recovers “binning” (grouping)

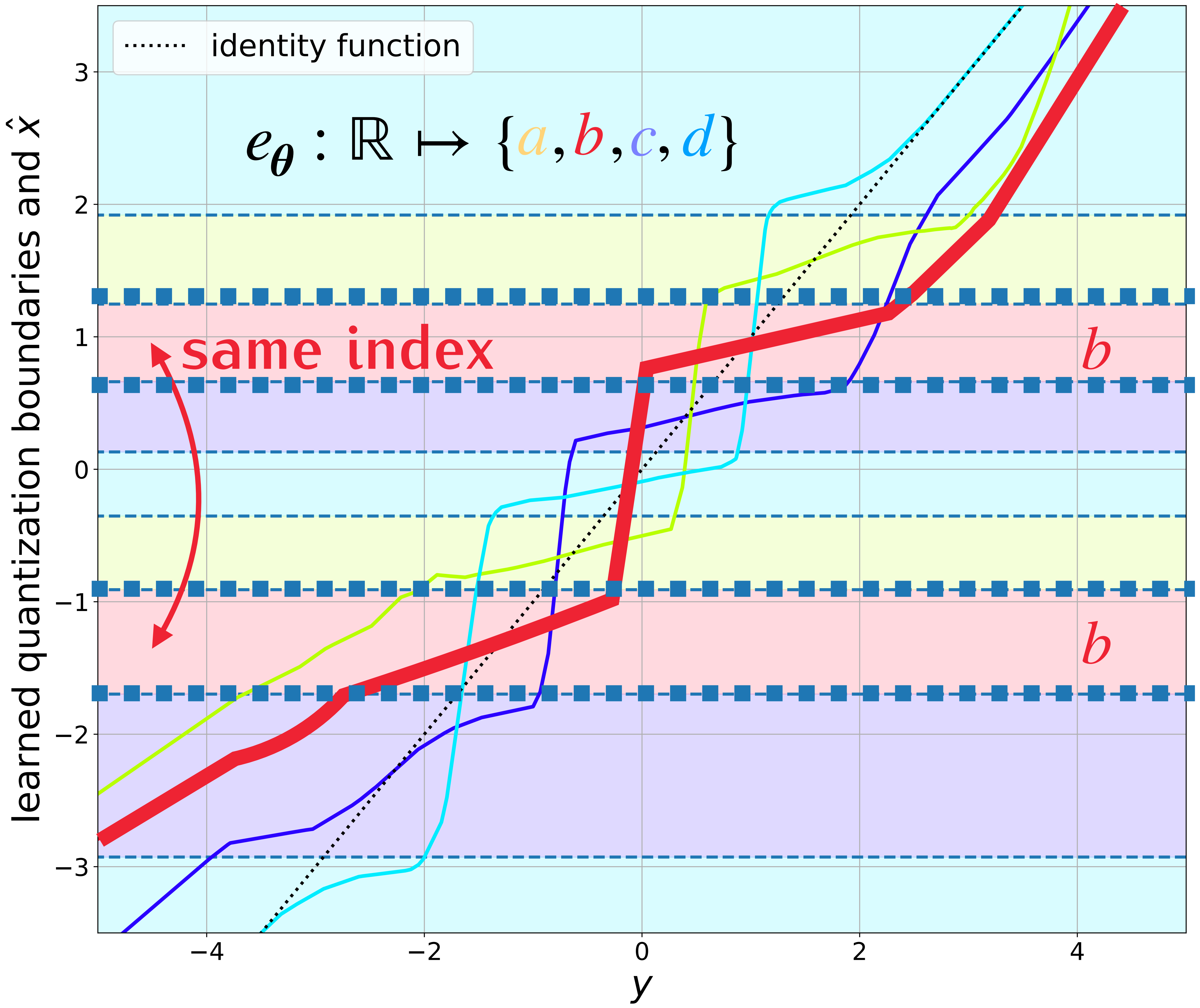
“**Binning**” is widely used to prove bounds in distributed compression, with some constructive attempts at implementation.



Ozyilkan, Ballé & Erkip (ISIT, 2023 & Neural Compression Workshop @ ICML’23 [oral])
Pradhan & Ramchandran (IEEE Trans. on Information Theory, 2003)

Learned compressor recovers "binning" (grouping)

"**Binning**" is widely used to prove bounds in distributed compression, with some constructive attempts at implementation.



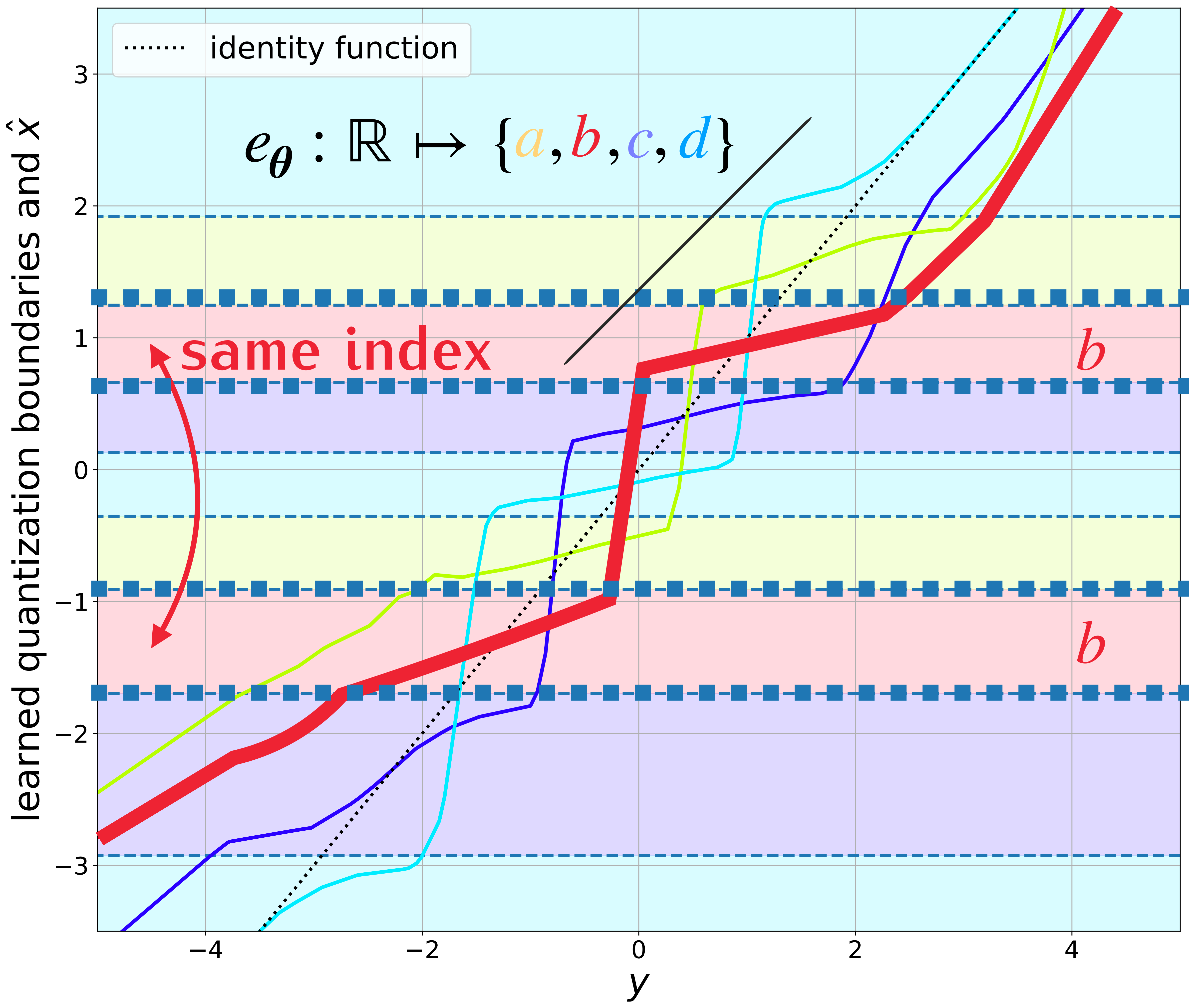
Ozyilkan, Ballé & Erkip (ISIT, 2023 & Neural Compression Workshop @ ICML'23 [oral])

Pradhan & Ramchandran (IEEE Trans. on Information Theory, 2003)

Learned compressor recovers “binning” (grouping)

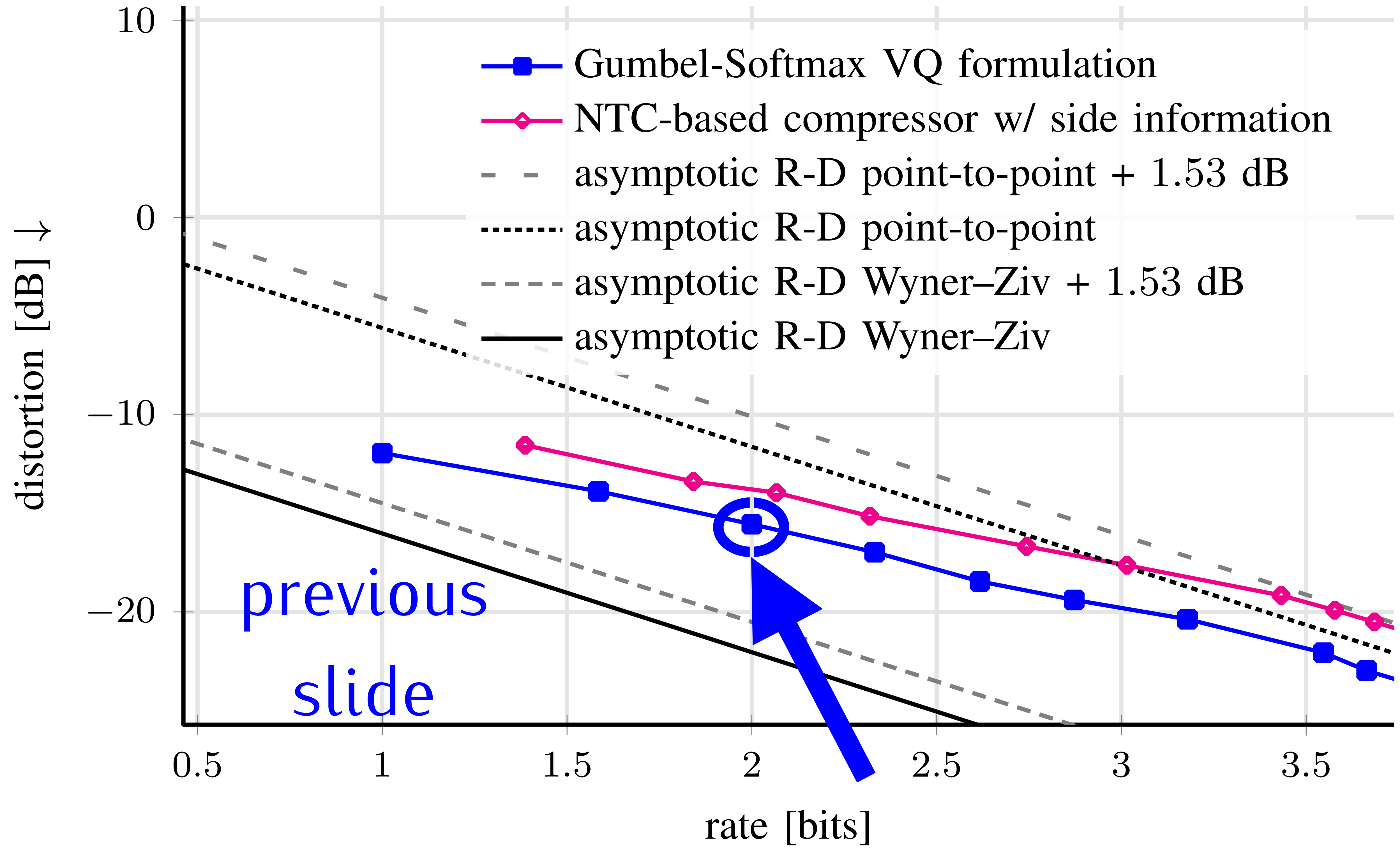
“**Binning**” is widely used to prove bounds in distributed compression, with some constructive attempts at implementation.

For X, Y Gaussian;
the optimal decoder is **linear** given the quantization index.



Ozyilkan, Ballé & Erkip (ISIT, 2023 & Neural Compression Workshop @ ICML’23 [oral])
Pradhan & Ramchandran (IEEE Trans. on Information Theory, 2003)

Empirical results of Gumbel-Softmax



$$X = Y + N$$

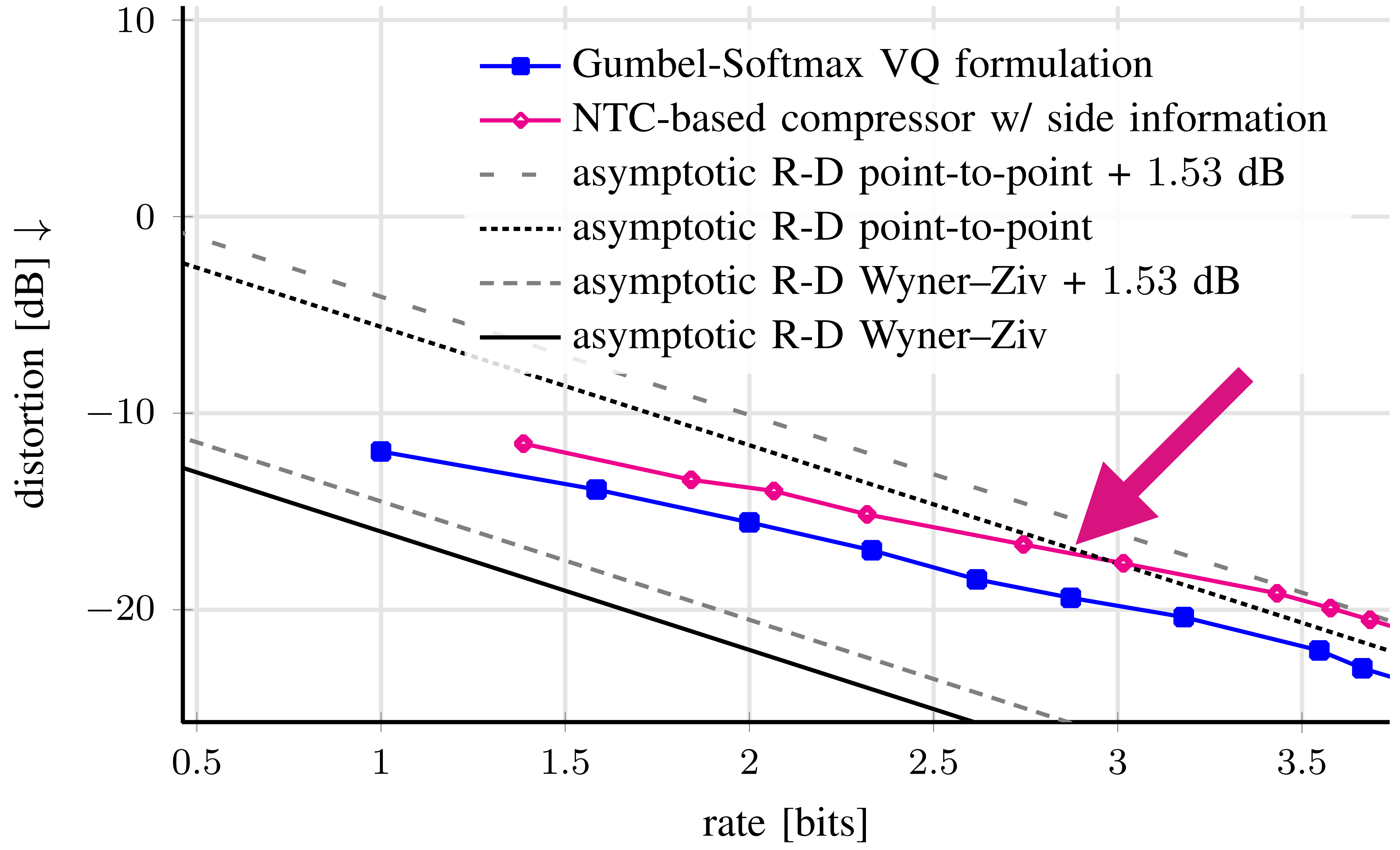
$$Y \sim \mathcal{N}(0,1)$$

$$N \sim \mathcal{N}(0,10^{-1})$$

Ozyilkan, Ballé & Erkip (ISIT, 2023 & Neural Compression Workshop @ ICML'23 [oral])

Empirical results of Gumbel-Softmax

The **inability to effectively perform "binning"** (grouping) hurts the performance of NTC.



$X = Y + N$

$Y \sim \mathcal{N}(0,1)$

$N \sim \mathcal{N}(0,10^{-1})$

Training with Gumbel–Softmax objective

- Optimization requires hyperparameter search (e.g., temperature scheduling).
- The loss function is not an upper bound on true objective (the rate-distortion).

Training with Gumbel–Softmax objective

- Optimization requires hyperparameter search (e.g., temperature scheduling).
- The loss function is not an upper bound on true objective (the rate-distortion).
- **Another solution** for replacing $e_{\boldsymbol{\theta}} : \mathbb{R} \mapsto \mathbb{Z}$ with $e_{\boldsymbol{\theta}} : \mathbb{R} \mapsto \{a, b, c, d, \cdots\}$:

Training with Gumbel–Softmax objective

- Optimization requires hyperparameter search (e.g., temperature scheduling).

- The loss function is not an upper bound on true objective (the rate-distortion).

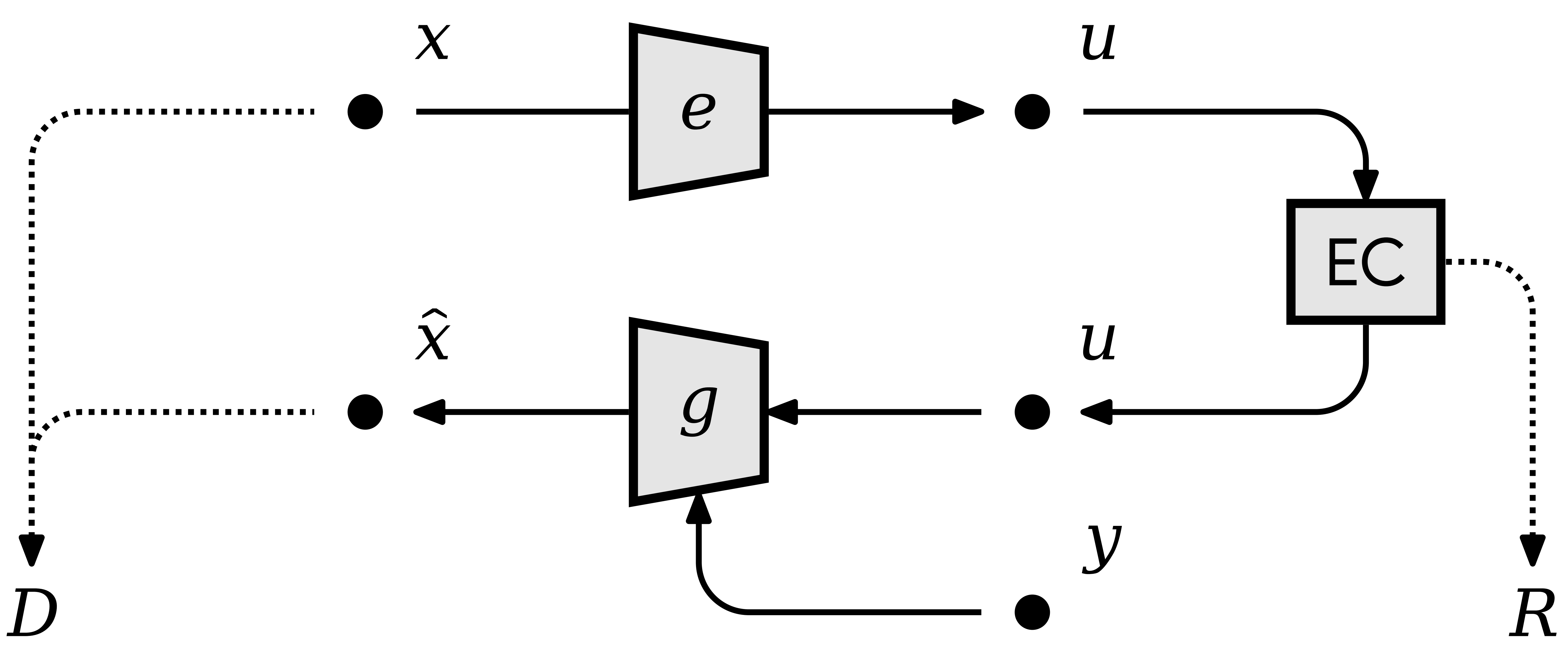
- **Another solution** for replacing $e_{\boldsymbol{\theta}} : \mathbb{R} \mapsto \mathbb{Z}$ with $e_{\boldsymbol{\theta}} : \mathbb{R} \mapsto \{a, b, c, d, \cdots\}$:

  - Encoder follows **classical entropy-coded vector quantizer** (ECVQ).

  - Encoder is completely *unstructured* in this case:
    → can assign any quantization index to input realization.

Ozyilkan, Ballé & Erkip (ISIT, 2023 & Neural Compression Workshop @ ICML'23 [oral])

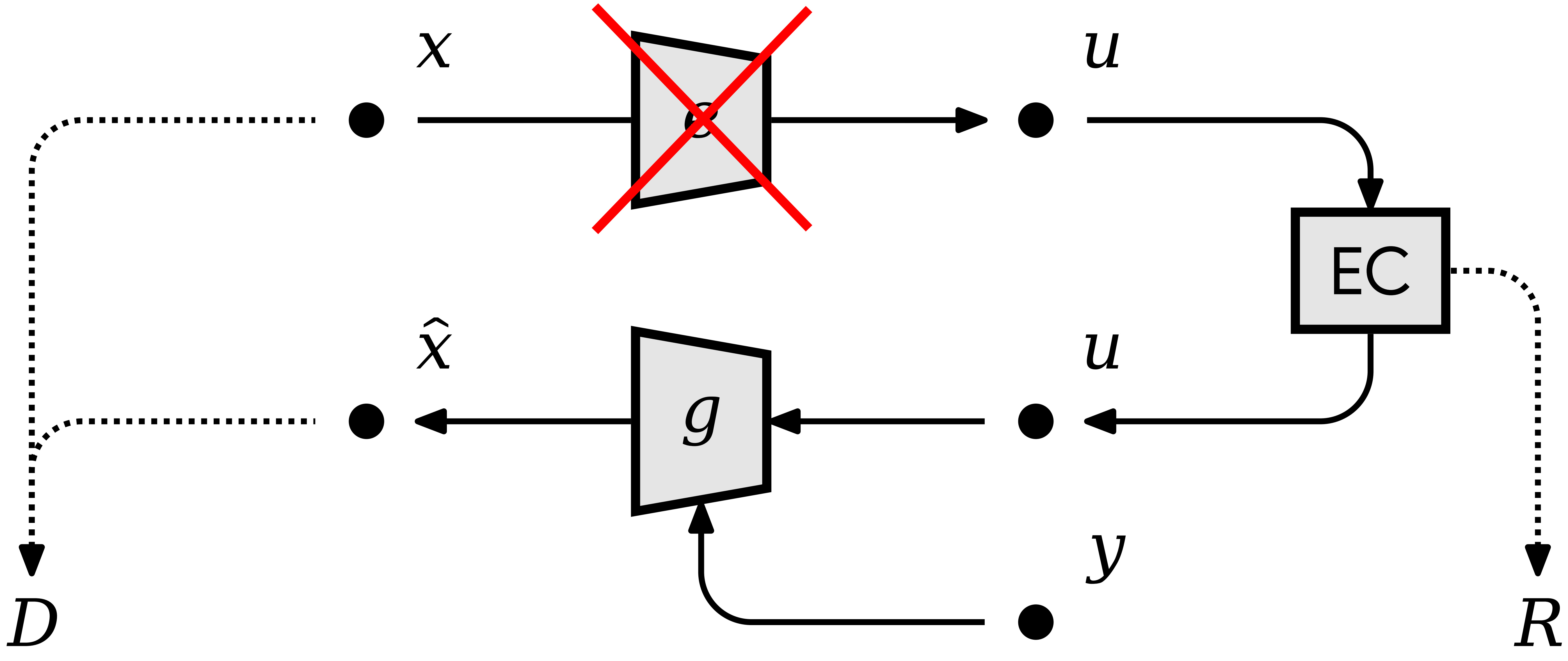
ECVQ version of Wyner–Ziv model



$$u = \underset{k \in K}{\arg\min}\, \mathbb{E}_{p(y|x)}\Big[\underbrace{-\log p_\psi(k)}_{\text{rate}} + \lambda \cdot \underbrace{d(x, g_\phi(k, y))}_{\text{distortion}}\Big]$$

Ozyilkan, Ballé & Erkip (IEEE J. Sel. Areas Information Theory, 2024)

ECVQ version of Wyner–Ziv model



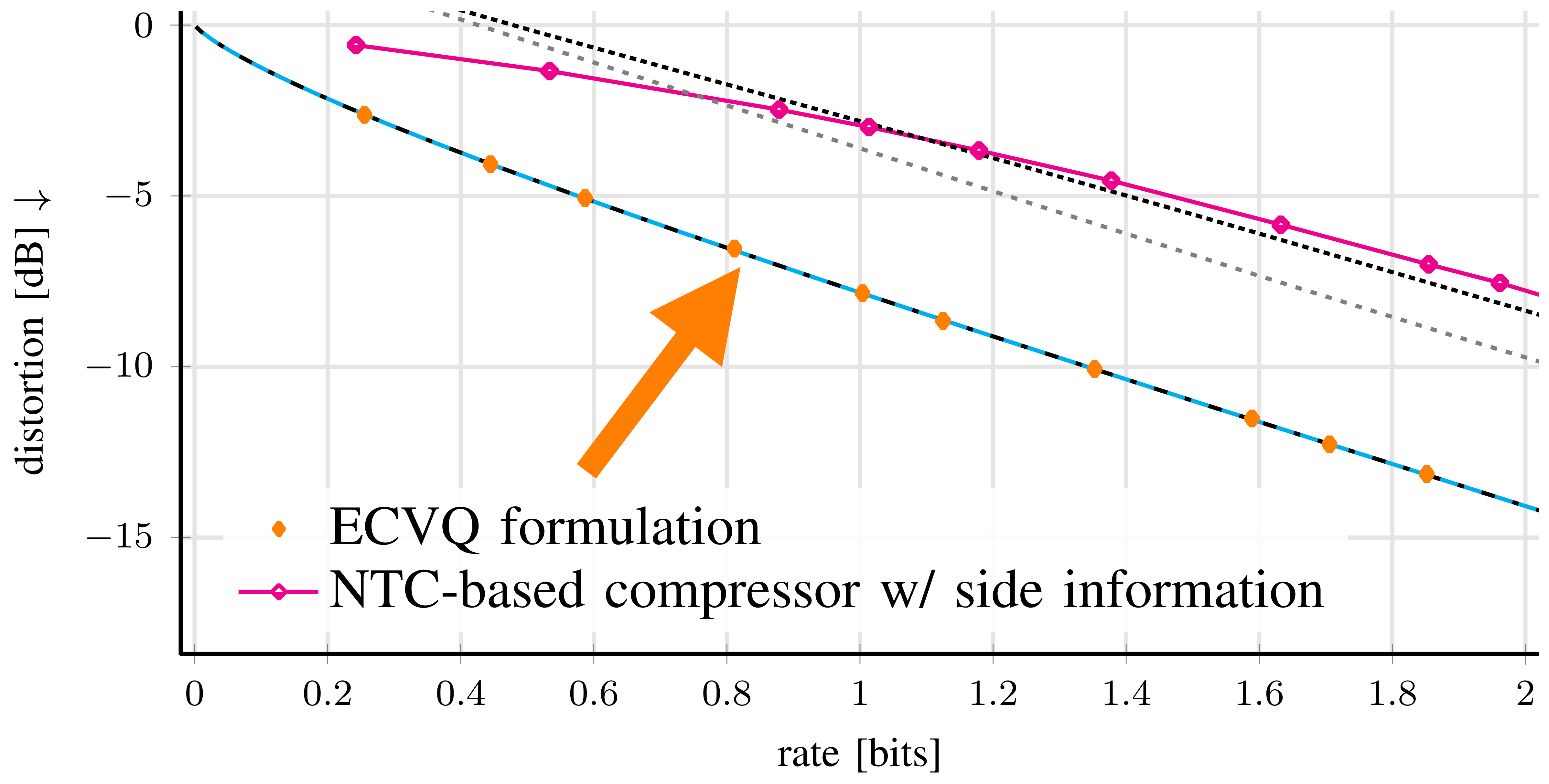
$$u = \arg\min_{k \in K} \mathbb{E}_{p(y|x)} \Big[\underbrace{-\log p_\psi(k)}_{\text{rate}} + \lambda \cdot \underbrace{d(x, g_\phi(k, y))}_{\text{distortion}} \Big]$$

Ozyilkan, Ballé & Erkip (IEEE J. Sel. Areas Information Theory, 2024)

ECVQ vs. NTC

ECVQ vs. NTC

- ECVQ objective is exactly what we want to optimize and requires *no relaxation* (e.g., temperature scheduling).



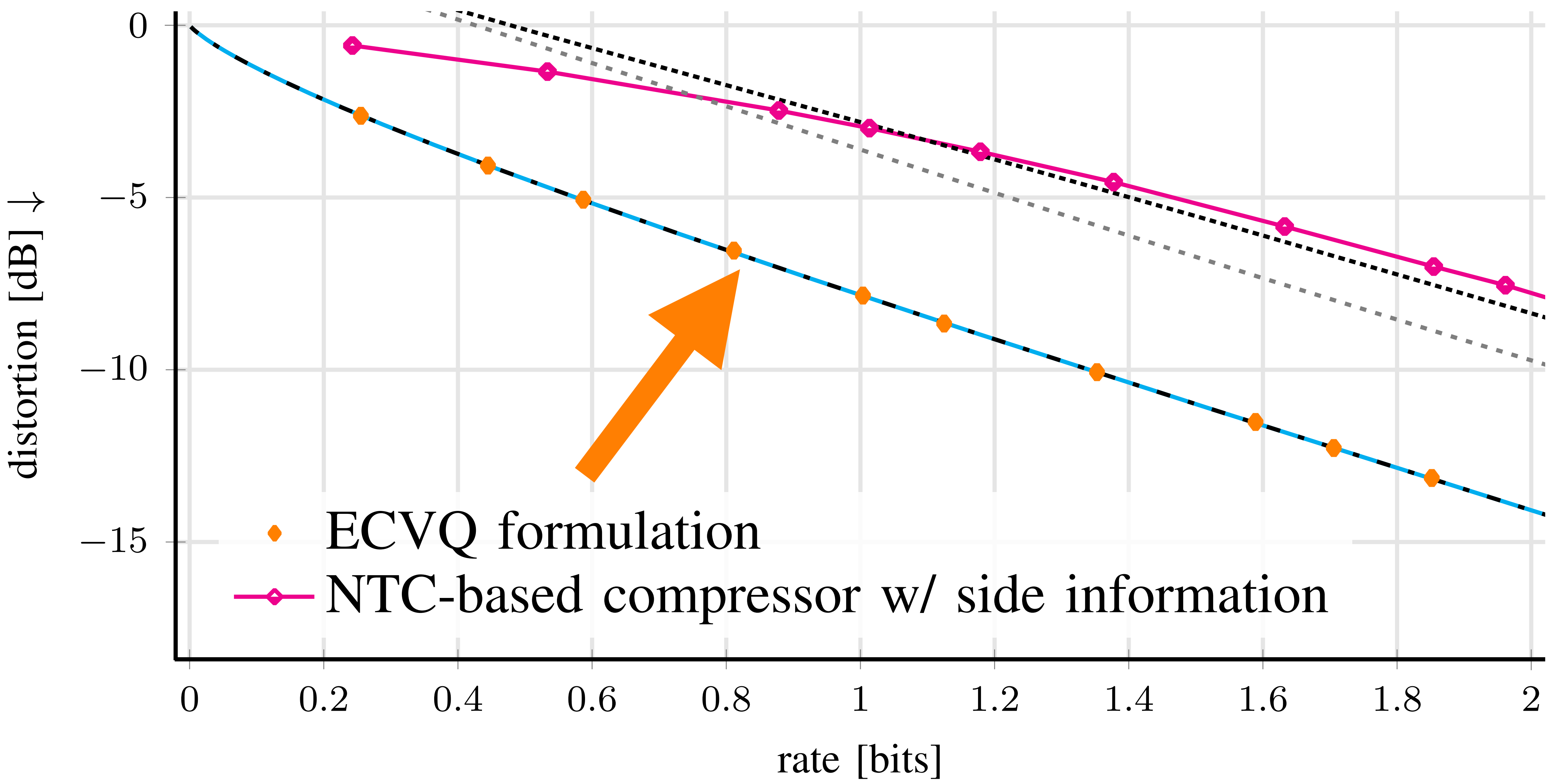
$$X \sim \text{Laplace}(0,1)$$

$$Y = \text{sgn}(\ X\ )$$

Ozyilkan, Ballé & Erkip (IEEE J. Sel. Areas Information Theory, 2024)

ECVQ vs. NTC

- ECVQ objective is exactly what we want to optimize and requires *no relaxation* (e.g., temperature scheduling).
- But, the encoder doesn't scale to higher dimensions (as in "traditional" VQ models)!



$X \sim \text{Laplace}(0,1)$

$Y = \text{sgn}(X)$

Ozyilkan, Ballé & Erkip (IEEE J. Sel. Areas Information Theory, 2024)

Summary on Neural Distributed Compression

- NTC-based solutions showed promise for low rate distributed image compression.
  - → Mital*, Ozyilkan*, Gargani & Gündüz (IEEE/CVF WACV, 2023)
  - → Mital*, Ozyilkan*, Gargani & Gündüz (IEEE DCC, 2022)

Summary on Neural Distributed Compression

- NTC-based solutions showed promise for low rate distributed image compression.
  - → Mital*, Ozyilkan*, Gargani & Gündüz (IEEE/CVF WACV, 2023)
  - → Mital*, Ozyilkan*, Gargani & Gündüz (IEEE DCC, 2022)

- But, using state-of-the-art NTC *naïvely* does not get close to the optimum.
  - The *smoothness learning bias* prevents it from recovering arbitrary maps at the encoder.
    - → Ozyilkan, Ballé & Erkip (ISIT, 2023 & Neural Compression Workshop @ ICML'23)
    - → Ozyilkan, Ballé & Erkip (IEEE J. Sel. Areas Information Theory, 2024)

Summary on Neural Distributed Compression

- NTC-based solutions showed promise for low rate distributed image compression.
  → Mital*, Ozyilkan*, Gargani & Gündüz (IEEE/CVF WACV, 2023)
  → Mital*, Ozyilkan*, Gargani & Gündüz (IEEE DCC, 2022)

- But, using state-of-the-art NTC *naïvely* does not get close to the optimum.
  - The *smoothness learning bias* prevents it from recovering arbitrary maps at the encoder.
    → Ozyilkan, Ballé & Erkip (ISIT, 2023 & Neural Compression Workshop @ ICML'23)
    → Ozyilkan, Ballé & Erkip (IEEE J. Sel. Areas Information Theory, 2024)

- We need *less structured* methods to recover high-frequency mappings (e.g., "binning").
  → Ozyilkan, Ballé, Bhadane, Wagner & Erkip (Compression Workshop @ NeurIPS, 2024)
  → Ozyilkan*, Sriramu*, Wagner, Erkip & Ballé (manuscript in submission)

Extensions and related projects

A. Robust distributed compression (Heegard–Berger)
   Tasci, Ozyilkan, Ulger, Erkip (IEEE ISIT-W, 2024)

B. Distributed Deep Joint Source-Channel Coding
   Yilmaz, Ozyilkan, Gündüz, Erkip (IEEE ICMLCN, 2024)

Extensions and related projects

A. Robust distributed compression (Heegard–Berger)
   Tasci, Ozyilkan, Ulger, Erkip (IEEE ISIT-W, 2024)

B. Distributed Deep Joint Source-Channel Coding
   Yilmaz, Ozyilkan, Gündüz, Erkip (IEEE ICMLCN, 2024)

C. **Developing scalable, beyond one-shot, learning-based ECVQ-like methods (ongoing)**

D. **"Dual" problem of Wyner–Ziv in channel coding: *dirty paper coding***
   Ozyilkan, Ulger, Erkip (IEEE ITW, 2025)

Extensions and related projects

A. Robust distributed compression (Heegard–Berger)
Tasci, Ozyilkan, Ulger, Erkip (IEEE ISIT-W, 2024)

B. Distributed Deep Joint Source-Channel Coding
Yilmaz, Ozyilkan, Gündüz, Erkip (IEEE ICMLCN, 2024)

C. **Developing scalable, beyond one-shot, learning-based ECVQ-like methods (ongoing)**

D. **“Dual” problem of Wyner–Ziv in channel coding: *dirty paper coding***
Ozyilkan, Ulger, Erkip (IEEE ITW, 2025)

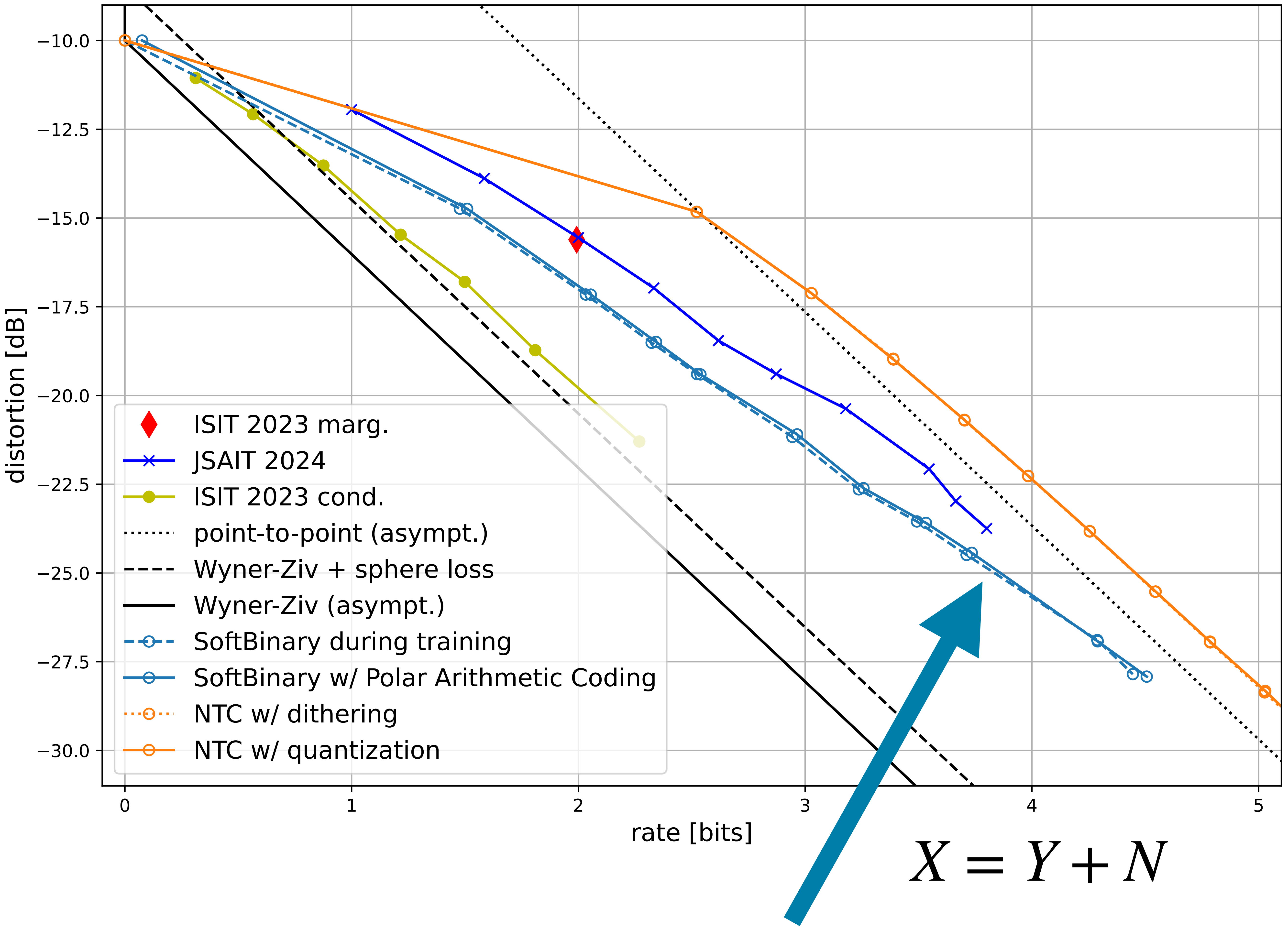
E. **Neural compress-and-forward for the relay → Part II.B**

C. Developing scalable learning-based ECVQ

C. Developing scalable learning-based ECVQ

$$X = Y + N$$
$$Y \sim \mathcal{N}(0,1)$$
$$N \sim \mathcal{N}(0,10^{-1})$$

C. Developing scalable learning-based ECVQ



- *SoftBinary* model: $e_{\boldsymbol{\theta}} : \mathbb{R} \mapsto \{0,1\}^d$, where encoder is now ***stochastic.***

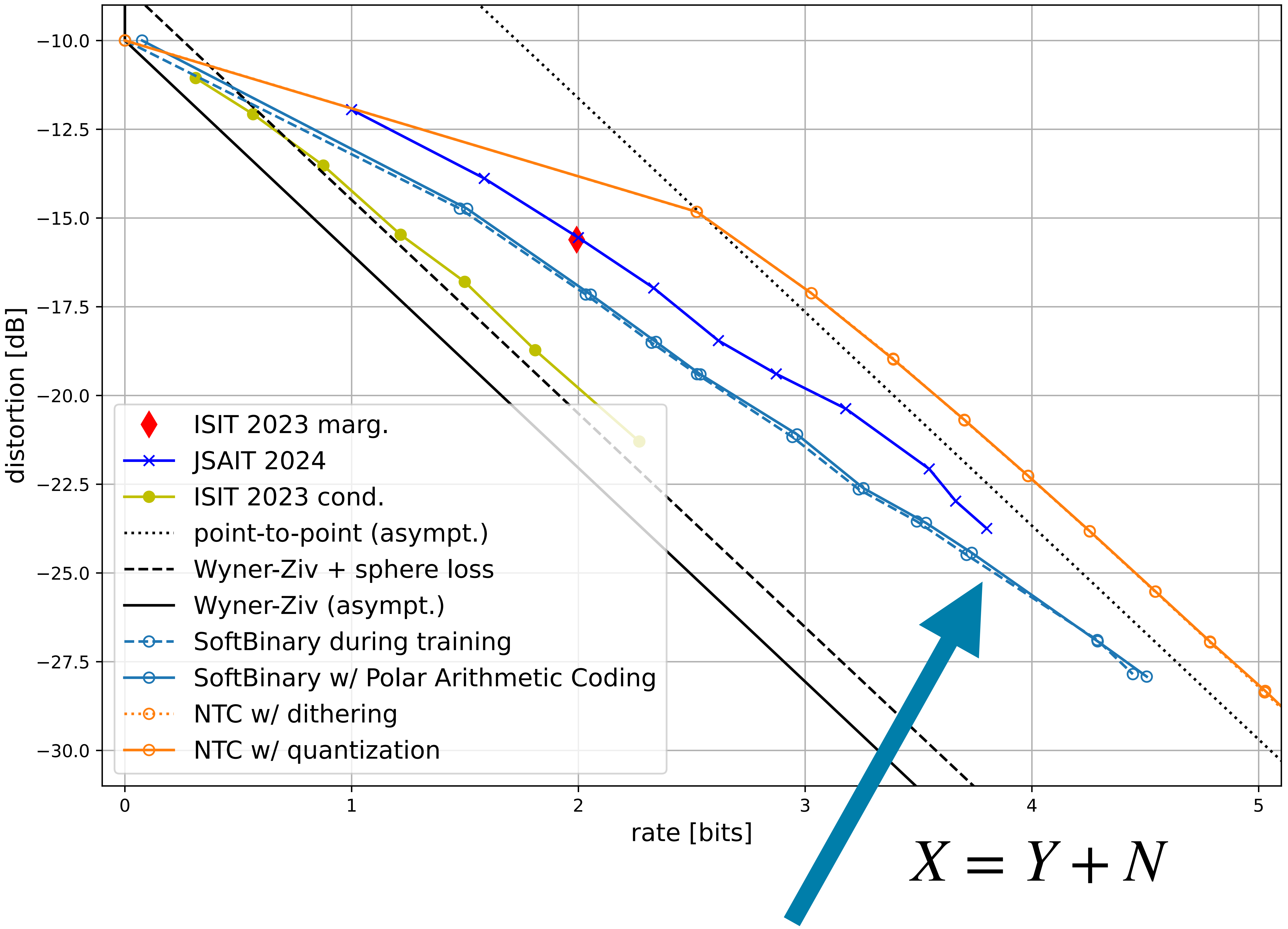
$$X = Y + N$$

$$Y \sim \mathcal{N}(0,1)$$

$$N \sim \mathcal{N}(0,10^{-1})$$

Manuscript in submission.

C. Developing scalable learning-based ECVQ



$X = Y + N$

$Y \sim \mathcal{N}(0,1)$

$N \sim \mathcal{N}(0,10^{-1})$

- *SoftBinary* model: $e_\theta : \mathbb{R} \mapsto \{0,1\}^d$, where encoder is now ***stochastic.***
- Assumes a "channel simulation" instead of (traditional) entropy coding: $R = \mathbb{E}_x[\mathrm{D_{KL}}(q_{\text{enc}} \parallel p_{\text{prior}})]$ instead of $R = \mathbb{E}_{x,q}[-\log p_{\text{prior}}]$.

Manuscript in submission.

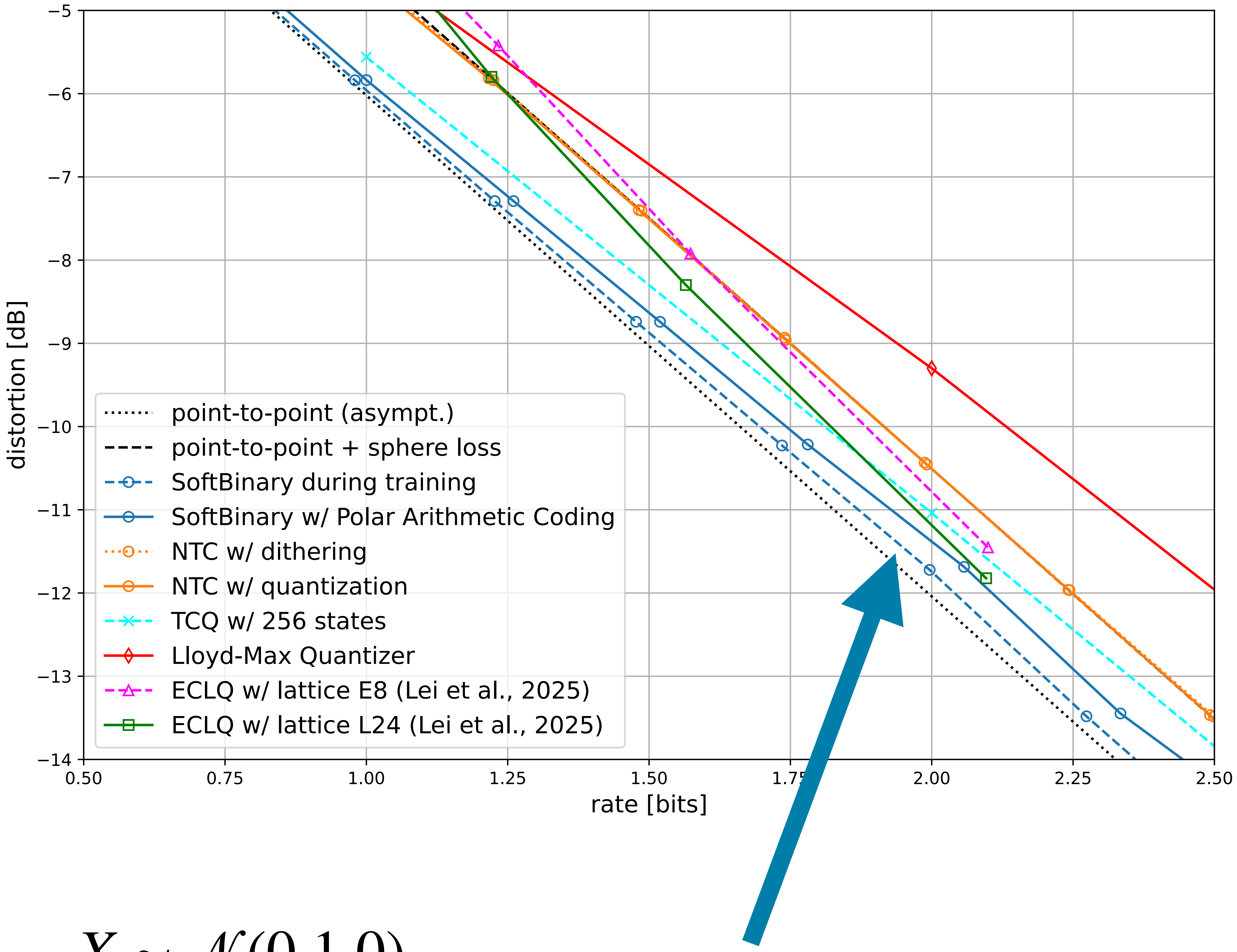
C. Developing scalable learning-based ECVQ

- *SoftBinary* model: $e_{\boldsymbol{\theta}} : \mathbb{R} \mapsto \{0,1\}^d$, where encoder is now ***stochastic.***
- Assumes a "channel simulation" instead of (traditional) entropy coding: $R = \mathbb{E}_x[\mathrm{D}_{\mathrm{KL}}(q_{\mathrm{enc}} \parallel p_{\mathrm{prior}})]$ instead of $R = \mathbb{E}_{x,q}[-\log p_{\mathrm{prior}}]$.

$X \sim \mathcal{N}(0,1.0)$

Manuscript in submission.

C. Developing scalable learning-based ECVQ

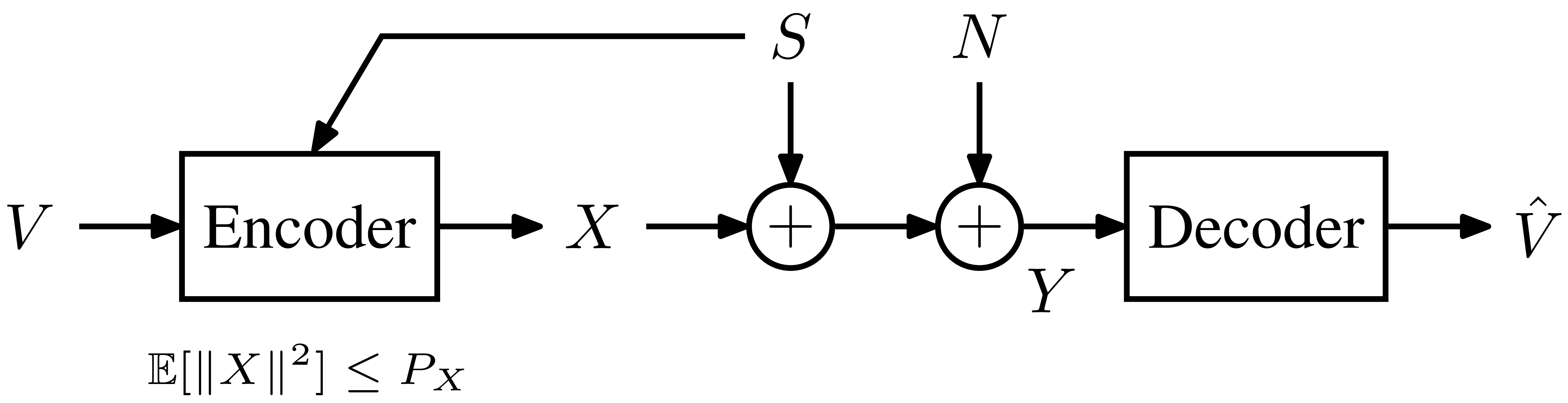


- *SoftBinary* model: $e_{\boldsymbol{\theta}} : \mathbb{R} \mapsto \{0,1\}^d$, where encoder is now ***stochastic.***
- Assumes a "channel simulation" instead of (traditional) entropy coding: $R = \mathbb{E}_x[\mathrm{D}_{\mathrm{KL}}(q_{\text{enc}} \parallel p_{\text{prior}})]$ instead of $R = \mathbb{E}_{x,q}[-\log p_{\text{prior}}]$.

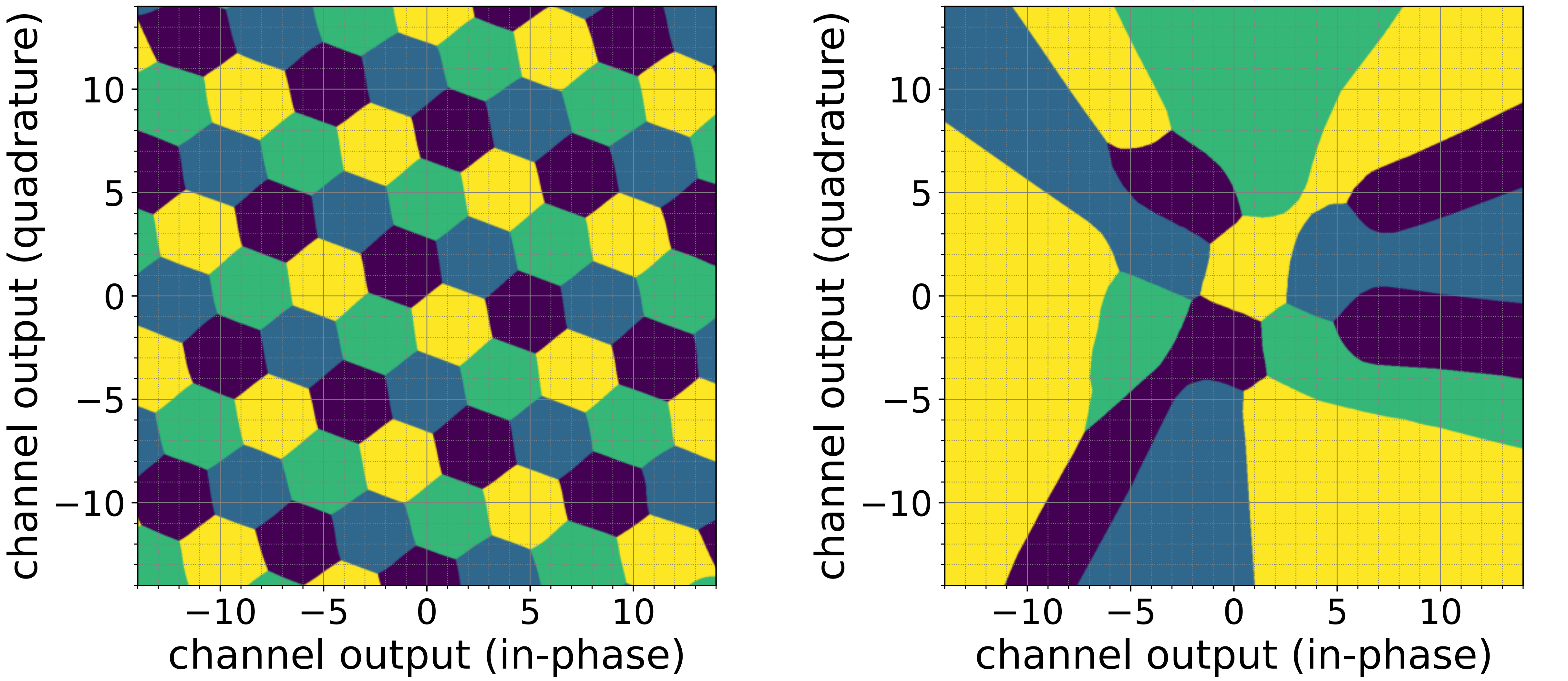
$X \sim \mathcal{N}(0,1.0)$

Manuscript in submission.

D. “Dual” problem of Wyner-Ziv: *Dirty Paper Coding*



The encoder maps V and known interference S to an input X, subject to an average power constraint.



(a) w/sinusoidal activations, scoring SNR : 7.03 dB, SER : -1.10 dB.

(b) w/leaky ReLU activations, scoring SNR : 8.83 dB, SER : -1.11 dB.

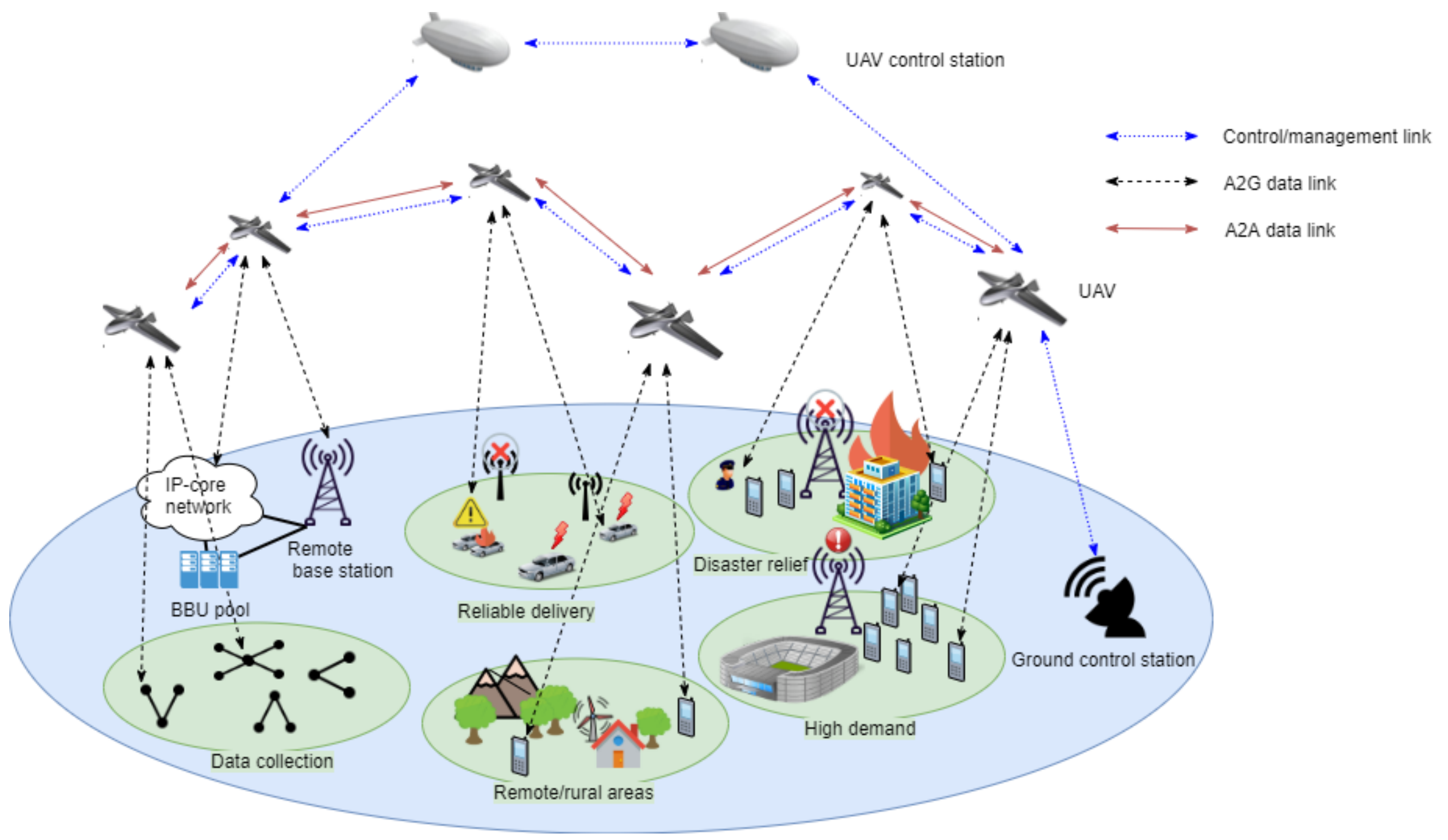
Ozyilkan, Ulger & Erkip (IEEE ITW, 2025)

Part II.B

Neural Compress-and-Forward for the Relay Channel

Introduction

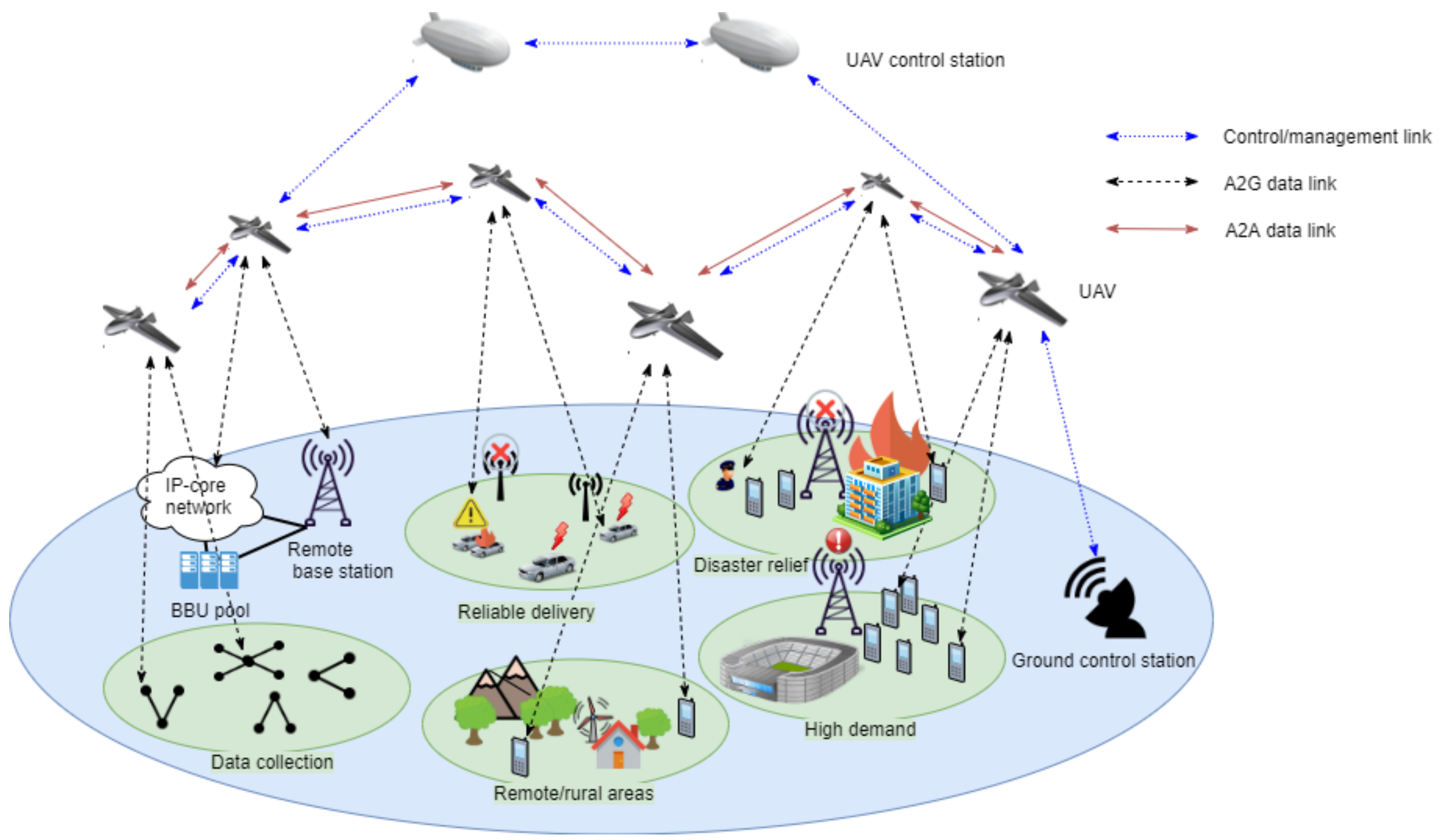
- ***Relay channel***: fundamental building block of "cooperative communications".
  - Applications: relay to improve throughput/coverage, e.g., RIS, drones.



Gholami et. al. "Joint Mobility-Aware UAV Placement and Routing in Multi-Hop UAV Relaying Systems"

Introduction

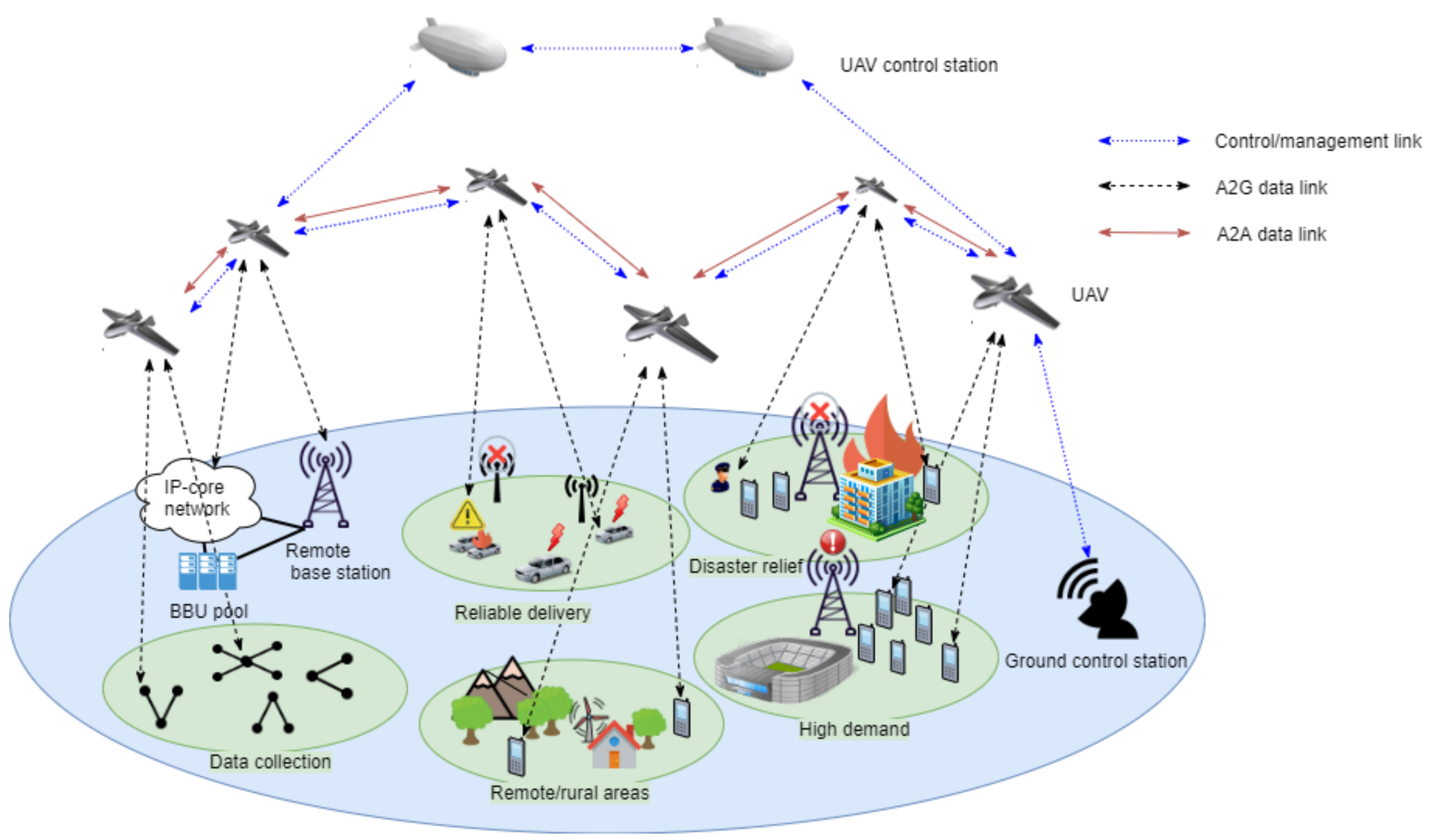
- ***Relay channel***: fundamental building block of "cooperative communications".
  - Applications: relay to improve throughput/coverage, e.g., RIS, drones.



Gholami et. al. "Joint Mobility-Aware UAV Placement and Routing in Multi-Hop UAV Relaying Systems"

Introduction

- ***Relay channel***: fundamental building block of "cooperative communications".
  - Applications: relay to improve throughput/coverage, e.g., RIS, drones.

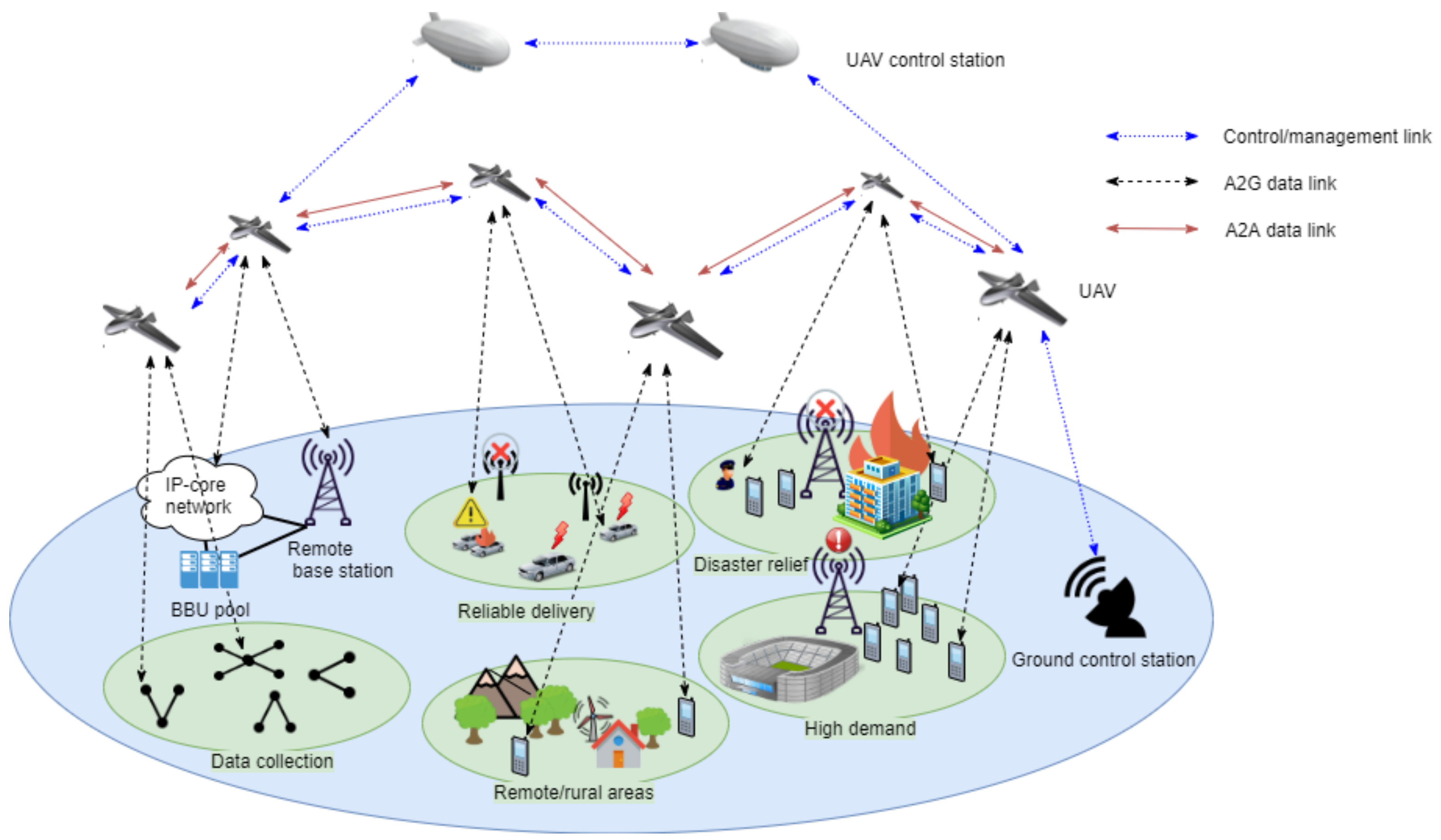


Gholami et. al. "Joint Mobility-Aware UAV Placement and Routing in Multi-Hop UAV Relaying Systems"

  - Capacity for the general relay channel is <u>unknown</u>, but several relaying strategies have been proposed.

Introduction

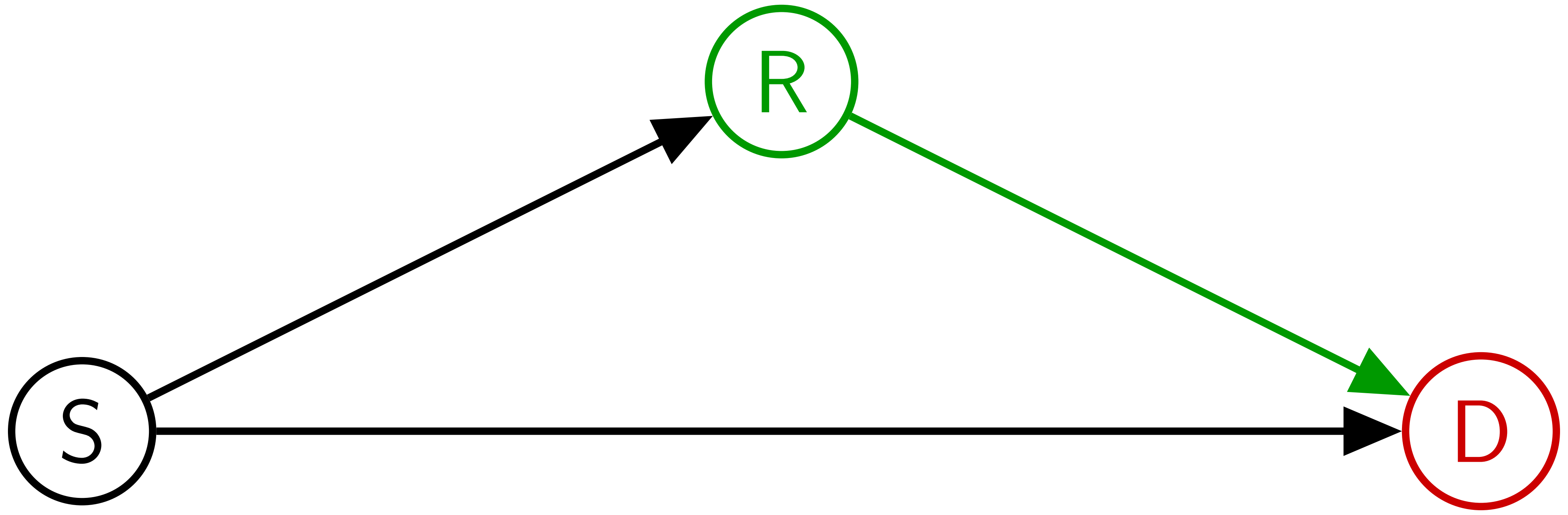
- ***Relay channel***: fundamental building block of "cooperative communications".
  - Applications: relay to improve throughput/coverage, e.g., RIS, drones.



Gholami et. al. "Joint Mobility-Aware UAV Placement and Routing in Multi-Hop UAV Relaying Systems"

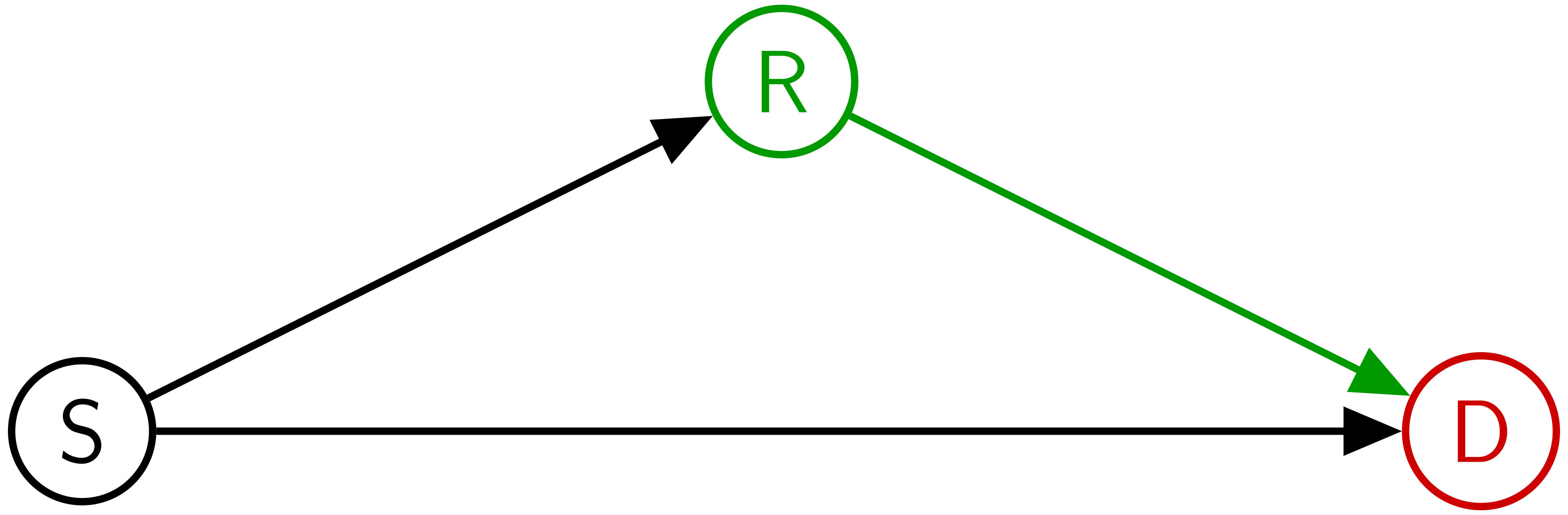
- Capacity for the general relay channel is <u>unknown</u>, but several relaying strategies have been proposed.
  - Amplify-and-forward, decode-and-forward....
  - **Compress-and-forward (CF)**: the relay sends a quantized version of its signal.

Motivation for Distributed Compression



- Relay and destination signals are correlated:
  Distributed compression techniques, like Wyner-Ziv, can be useful!

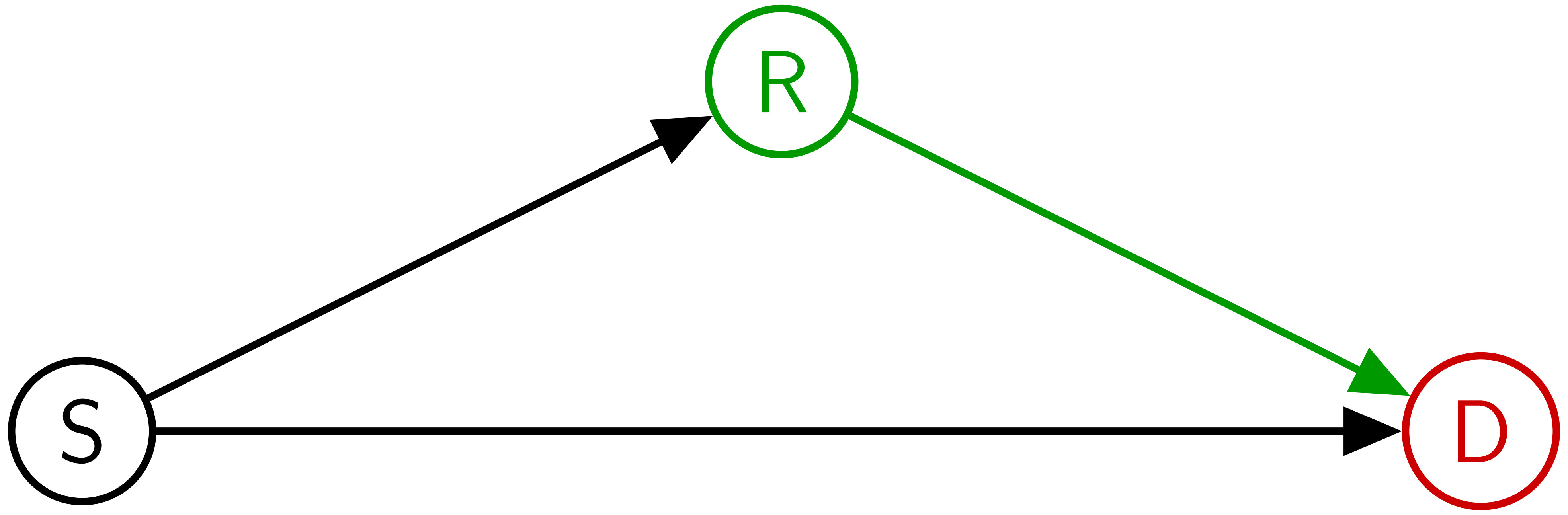
Motivation for Distributed Compression



- Relay and destination signals are correlated:
  Distributed compression techniques, like Wyner-Ziv, can be useful!

- ...but practical relaying schemes have not been fully developed.

Motivation for Distributed Compression



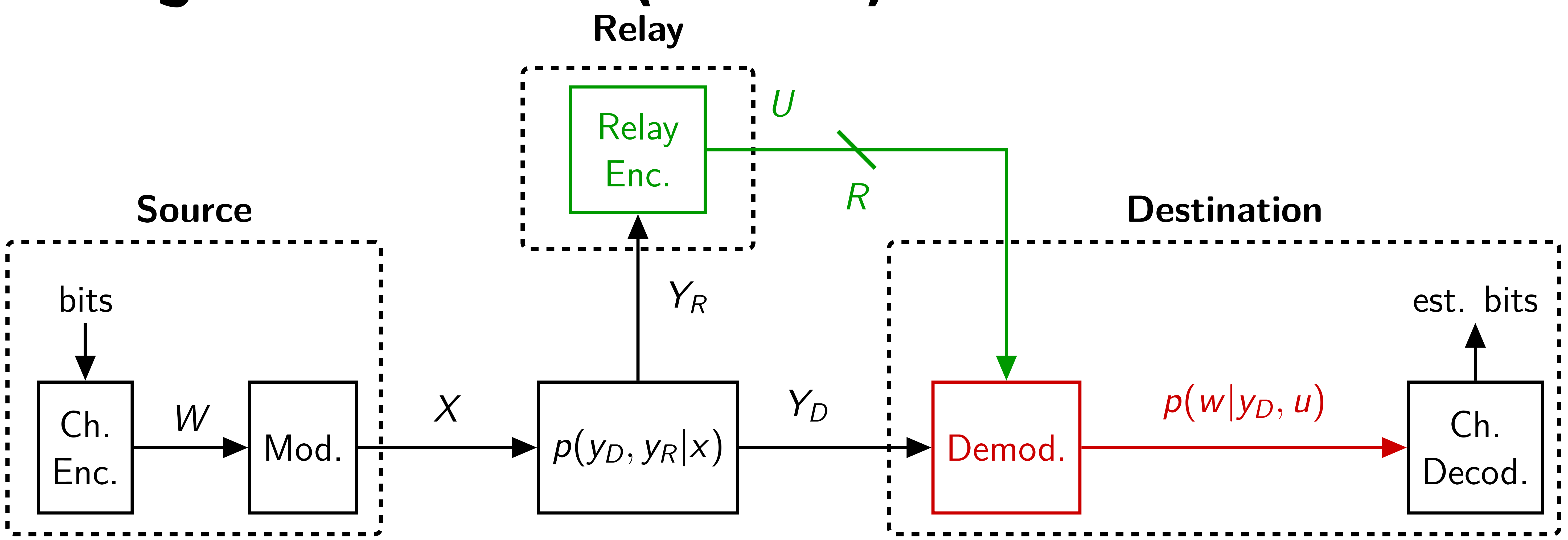
- Relay and destination signals are correlated:
  Distributed compression techniques, like Wyner-Ziv, can be useful!

- ...but practical relaying schemes have not been fully developed.

- We model **relays as learned distributed compressors**
  →learned compress-and-forward strategy.

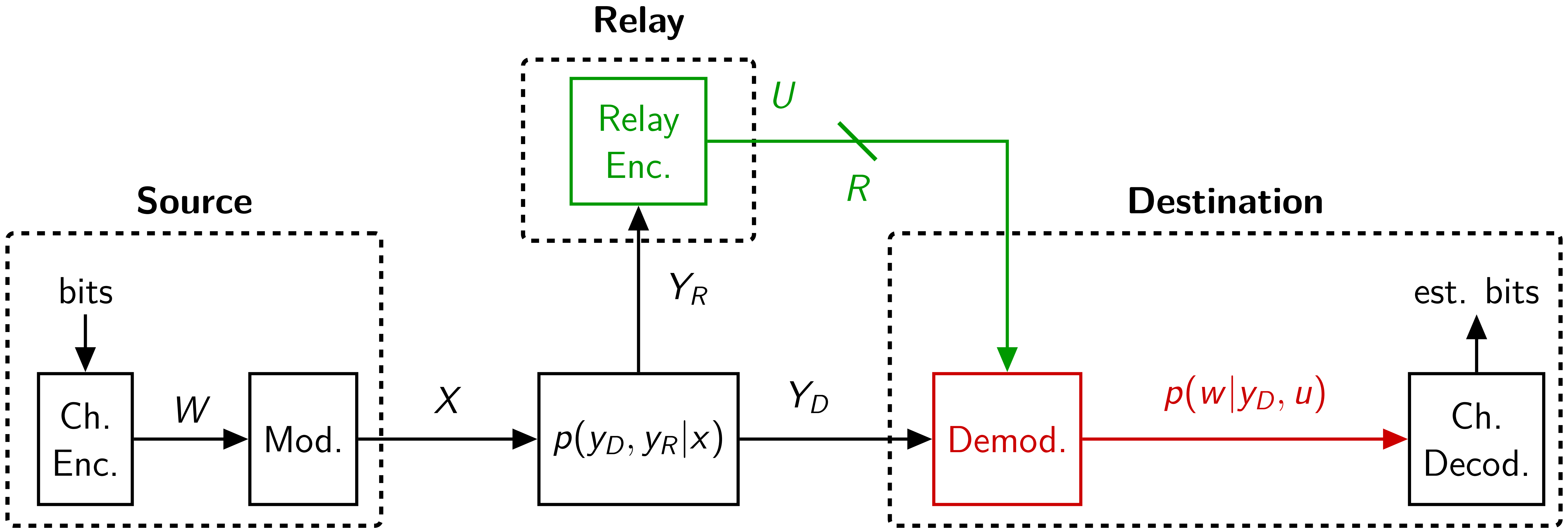
Ozyilkan*, Carpi*, Garg & Erkip (IEEE SPAWC, 2024)
Ozyilkan*, Carpi*, Garg & Erkip (IEEE J. Sel. Areas in Communications, 2025)

Primitive Relay Channel (PRC) — *out-of-band relay*



- **Relay's POV:** Simplest compression setup w/ channel coding.
  - Compress Y_R to help destination decode W via orthogonal noiseless link.

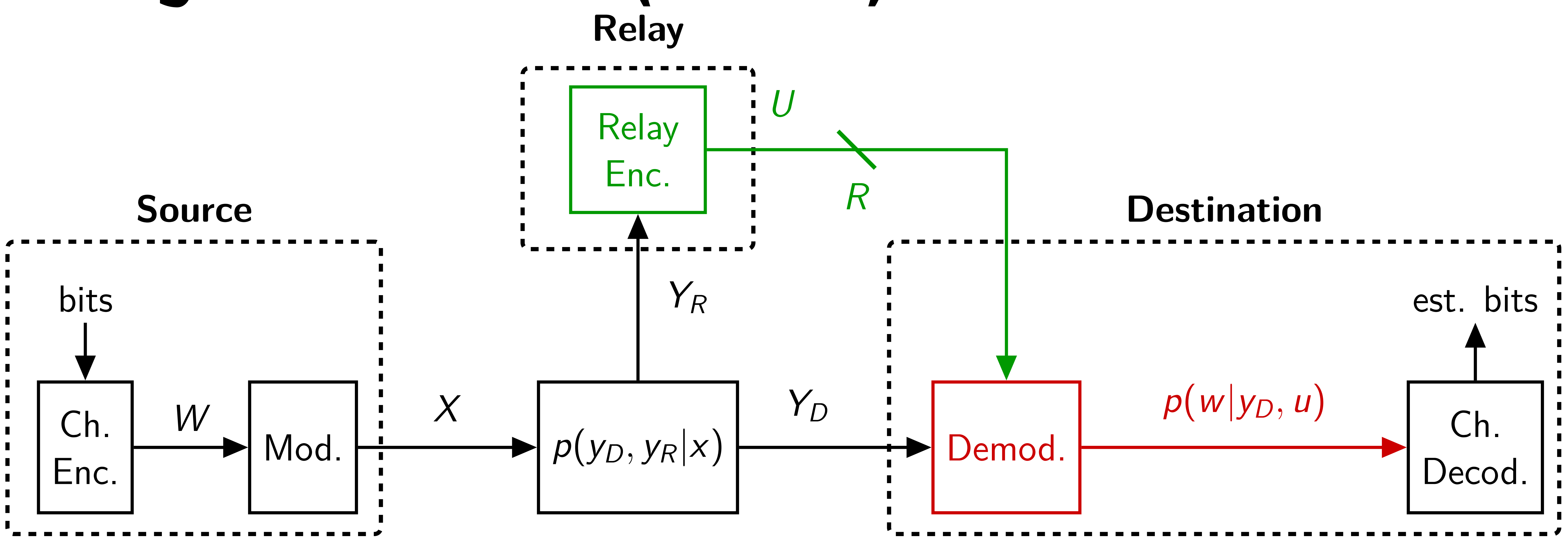
Primitive Relay Channel (PRC) — *out-of-band relay*



- **Relay's POV:** Simplest compression setup w/ channel coding.
  - Compress Y_R to help destination decode W via orthogonal noiseless link.
- Compress-and-forward (CF) is optimal for *oblivious* relaying in PRC.

Simeone, Erkip & Shamai (IEEE Trans. On Information Theory, 2011)

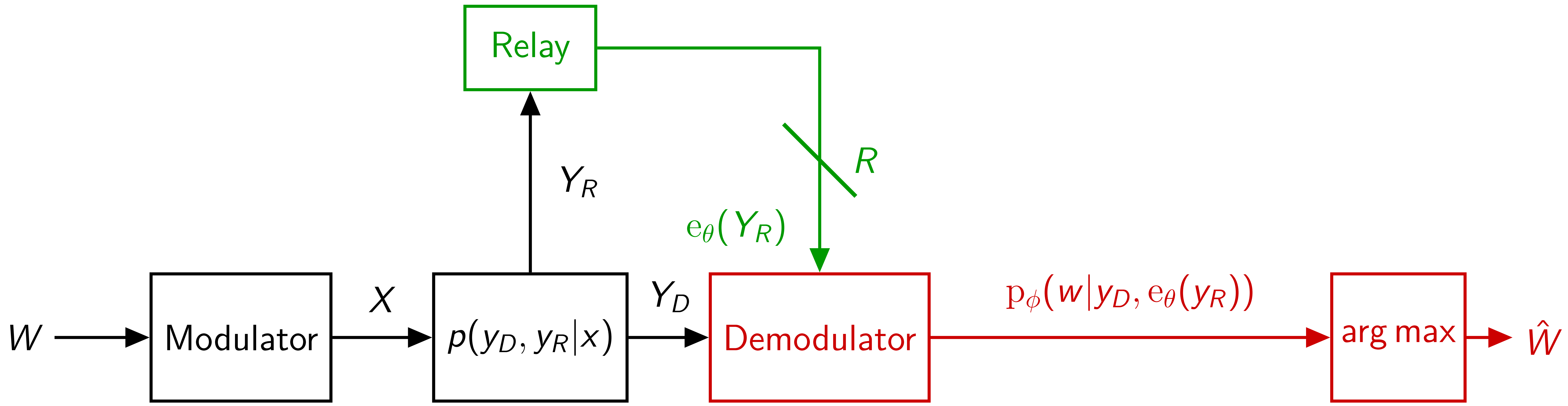
Primitive Relay Channel (PRC) — *out-of-band relay*



- **Relay's POV:** Simplest compression setup w/ channel coding.
  - Compress Y_R to help destination decode W via orthogonal noiseless link.
- Compress-and-forward (CF) is optimal for *oblivious* relaying in PRC.
- Goal: maximize communication rate $I(X; Y_D, U)$ subject to rate constraint R
  - **Task-aware**/semantic **compression**

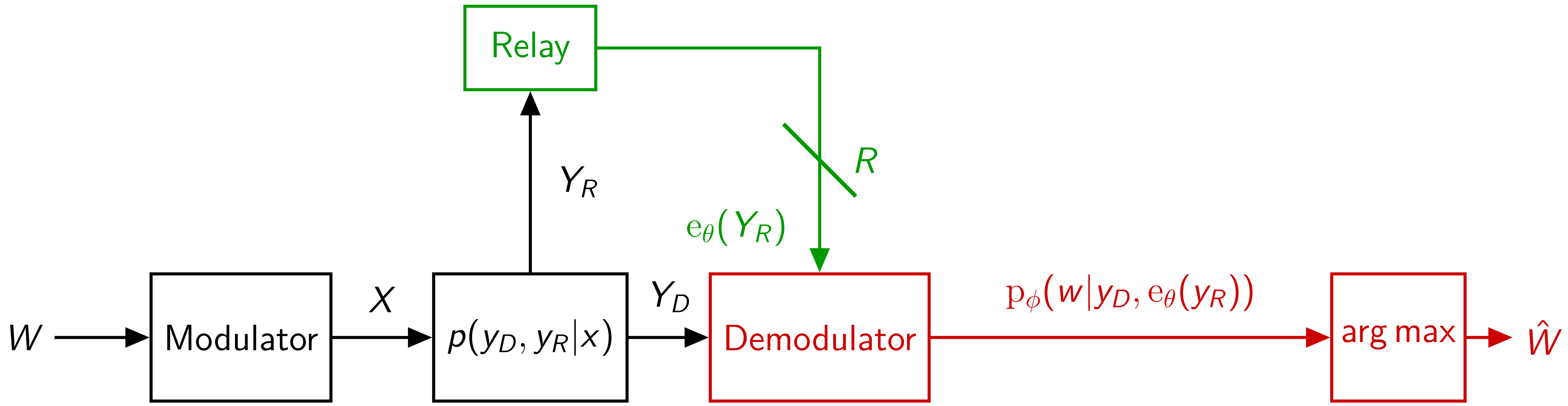
Simeone, Erkip & Shamai (IEEE Trans. On Information Theory, 2011)

Simulation Scenario



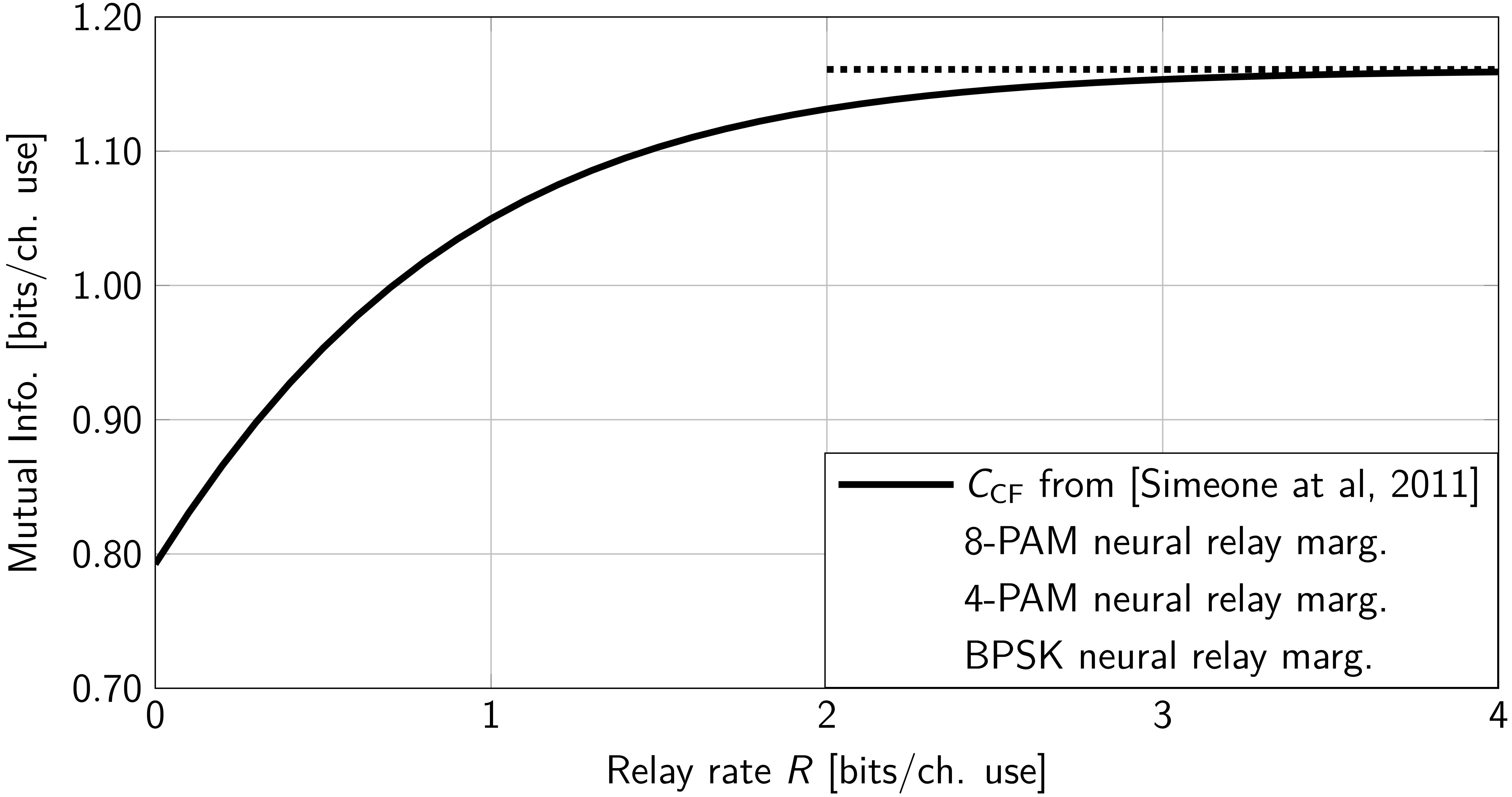
- Source: equally likely symbols, power constraint $P = \mathbb{E}[\|X\|^2]$
  - Real channel: BPSK, 4-PAM, 8-PAM
  - Complex channel: QAM, 16-QAM

Simulation Scenario



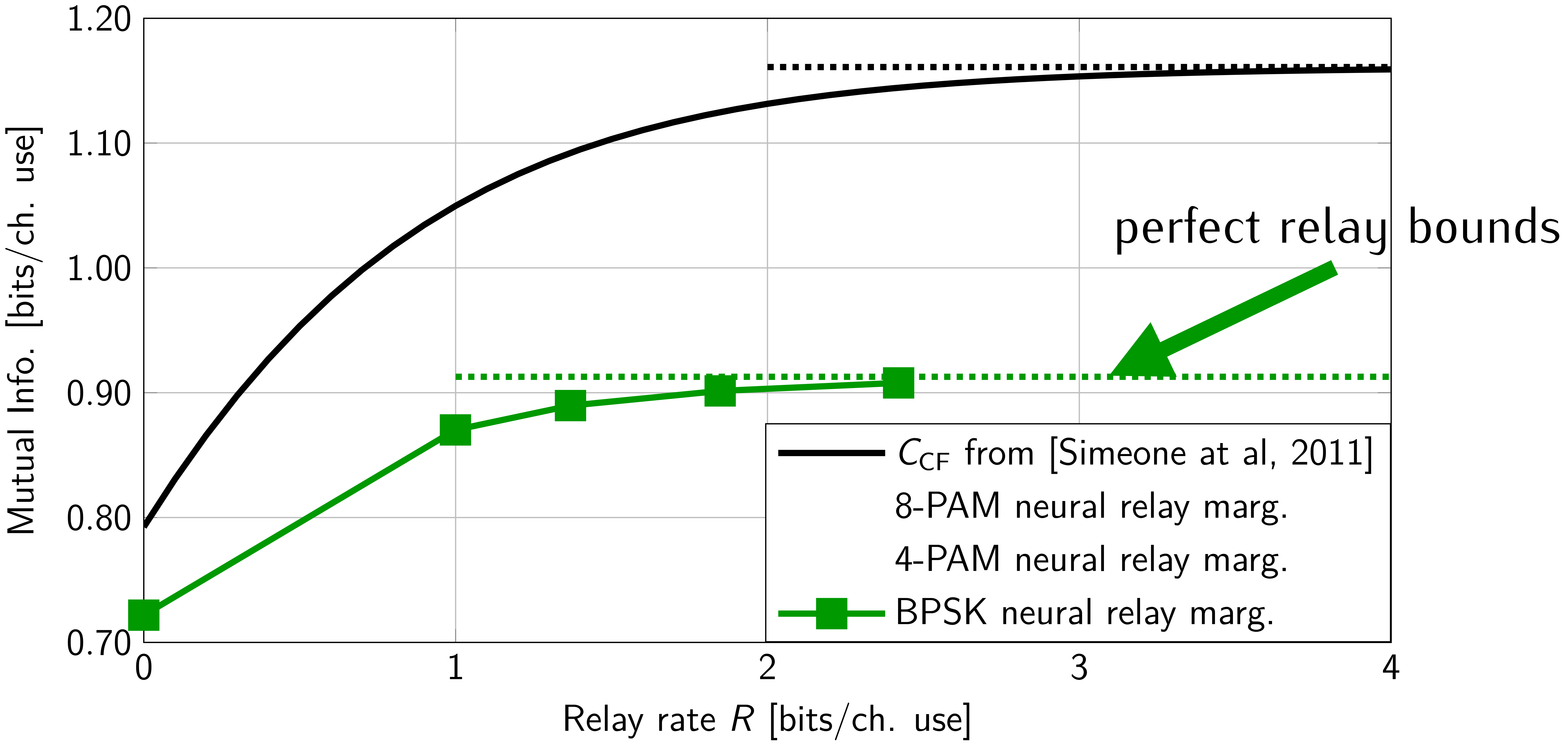
- Source: equally likely symbols, power constraint $P = \mathbb{E}[\|X\|^2]$
  - Real channel: BPSK, 4-PAM, 8-PAM
  - Complex channel: QAM, 16-QAM
- Channel: $Y_D = X + N_D$ and $Y_R = X + N_R$, with $N_D \perp N_R$
  - (N_D, N_R) (complex) Gaussian noise with variance (σ_D^2, σ_R^2).
- SNR: $\gamma_D = P/\sigma_D^2$, $\gamma_R = P/\sigma_R^2$.

Mutual Information for Learned CF at $\gamma_D = \gamma_R = 3$ dB



CF achievable rate [Simeone et al, 2011]: $C_{\text{CF}} = \frac{1}{2}\log_2\left(1 + \gamma_D + \frac{\gamma_R}{1 + \frac{1+\gamma_D+\gamma_R}{(2^{2R}-1)(\gamma_D+1)}}\right)$

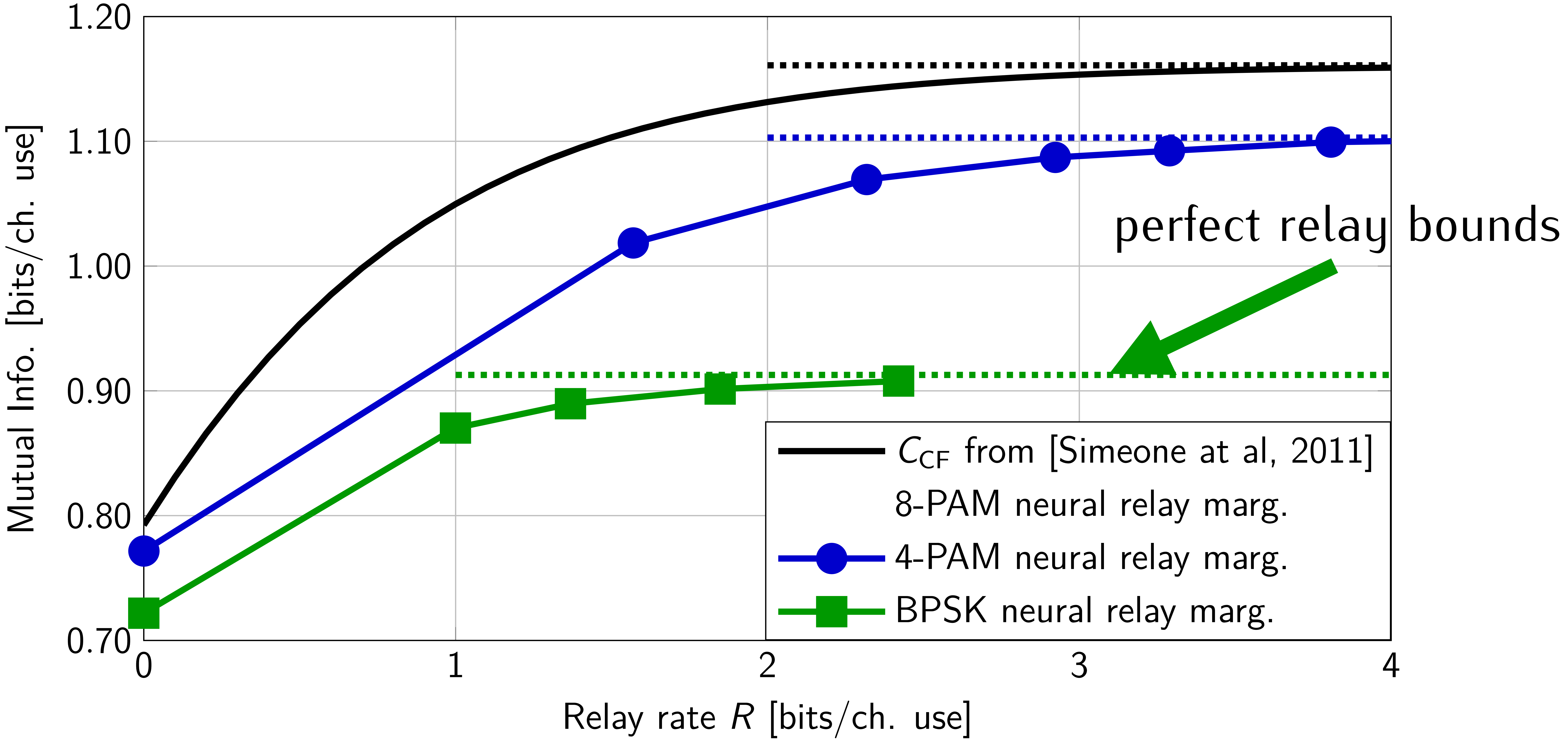
Mutual Information for Learned CF at $\gamma_D = \gamma_R = 3$ dB



CF achievable rate [Simeone et al, 2011]: $C_{\text{CF}} = \frac{1}{2}\log_2\left(1 + \gamma_D + \frac{\gamma_R}{1 + \frac{1+\gamma_D+\gamma_R}{(2^{2R}-1)(\gamma_D+1)}}\right)$

Ozyilkan*, Carpi*, Garg & Erkip (IEEE SPAWC, 2024)

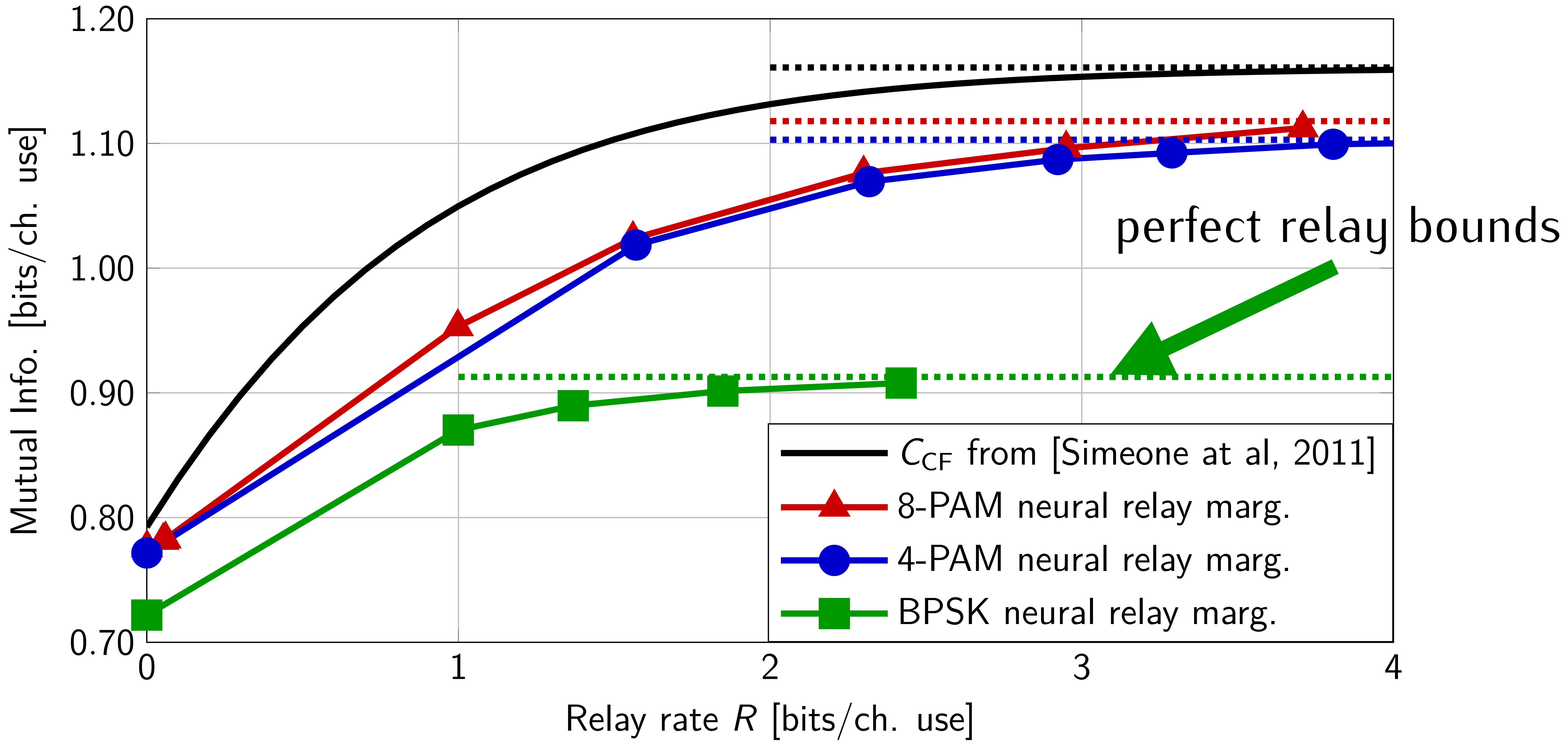
Mutual Information for Learned CF at $\gamma_D = \gamma_R = 3$ dB



CF achievable rate [Simeone et al, 2011]: $C_{\mathrm{CF}} = \frac{1}{2}\log_2\left(1 + \gamma_D + \frac{\gamma_R}{1 + \frac{1+\gamma_D+\gamma_R}{(2^{2R}-1)(\gamma_D+1)}}\right)$

Ozyilkan*, Carpi*, Garg & Erkip (IEEE SPAWC, 2024)

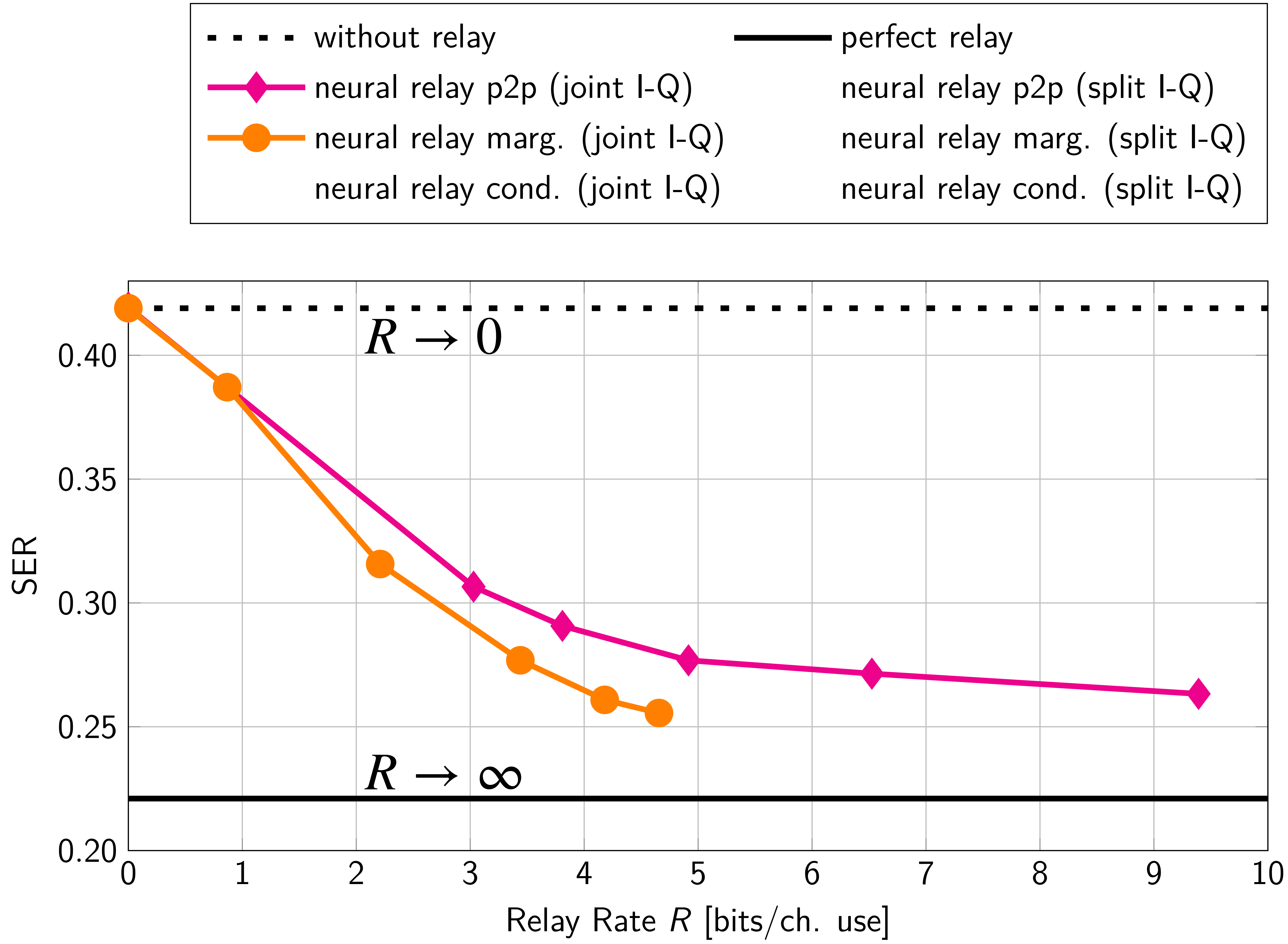
Mutual Information for Learned CF at $\gamma_D = \gamma_R = 3$ dB



CF achievable rate [Simeone et al, 2011]: $C_{\mathrm{CF}} = \frac{1}{2}\log_2\left(1 + \gamma_D + \frac{\gamma_R}{1 + \frac{1+\gamma_D+\gamma_R}{(2^{2R}-1)(\gamma_D+1)}}\right)$

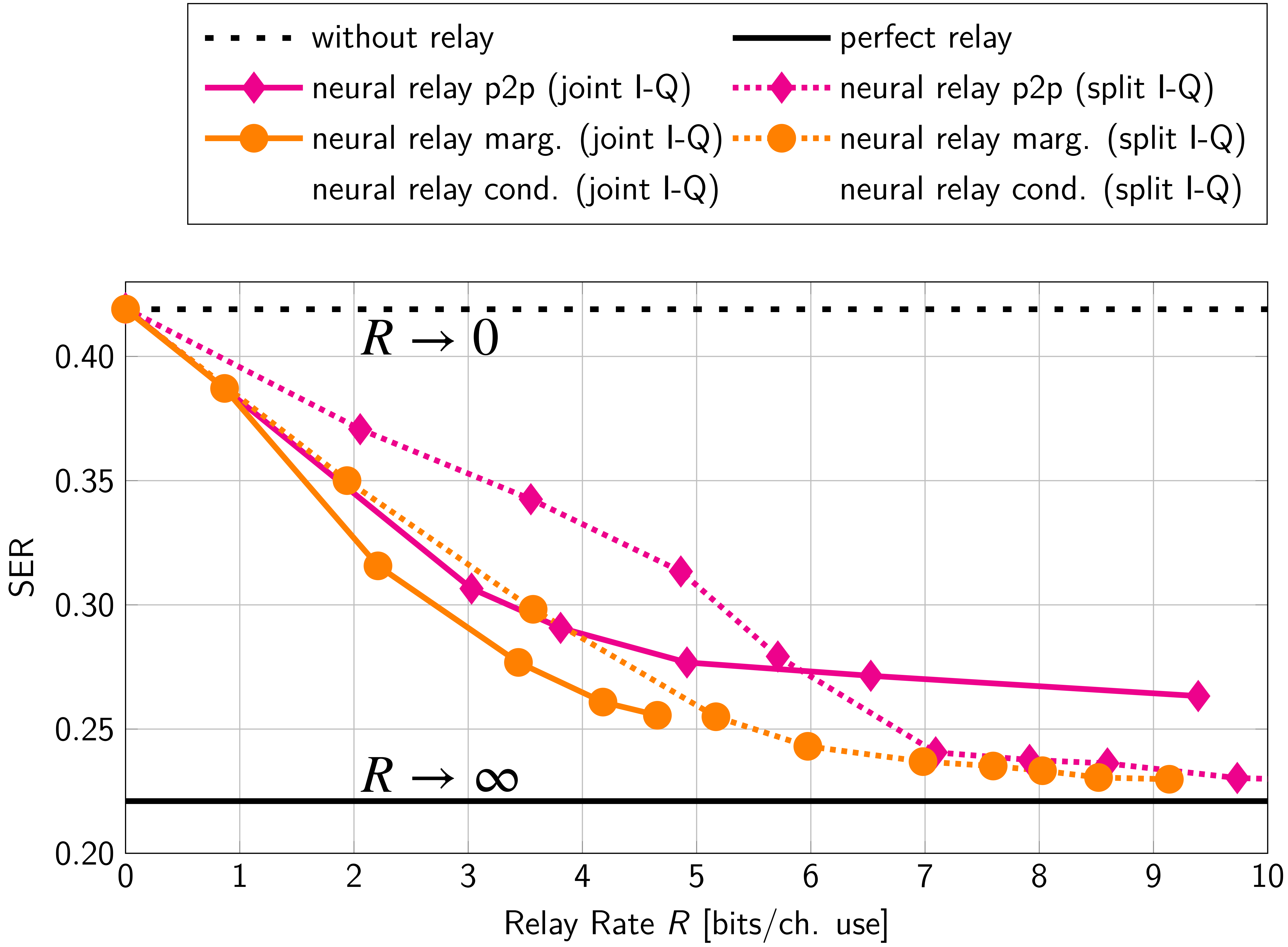
Ozyilkan*, Carpi*, Garg & Erkip (IEEE SPAWC, 2024)

Symbol Error Rate for 16-QAM at $\gamma_D = \gamma_R = 7$ dB



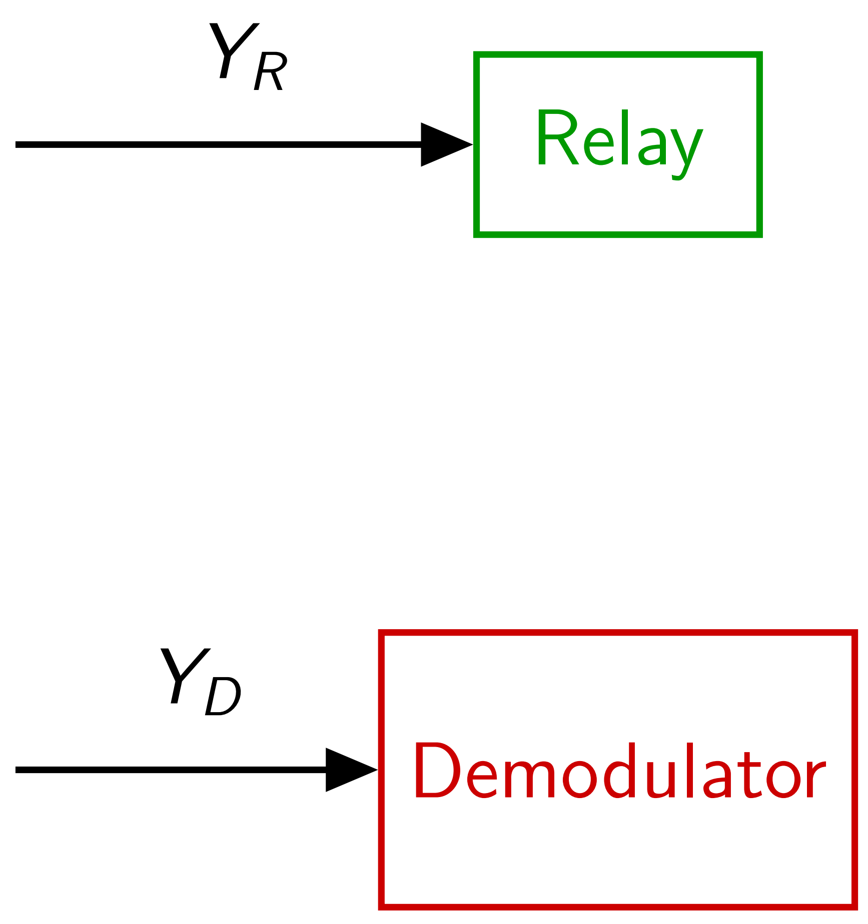
Ozyilkan*, Carpi*, Garg & Erkip (IEEE J. Sel. Areas in Communications, 2025)

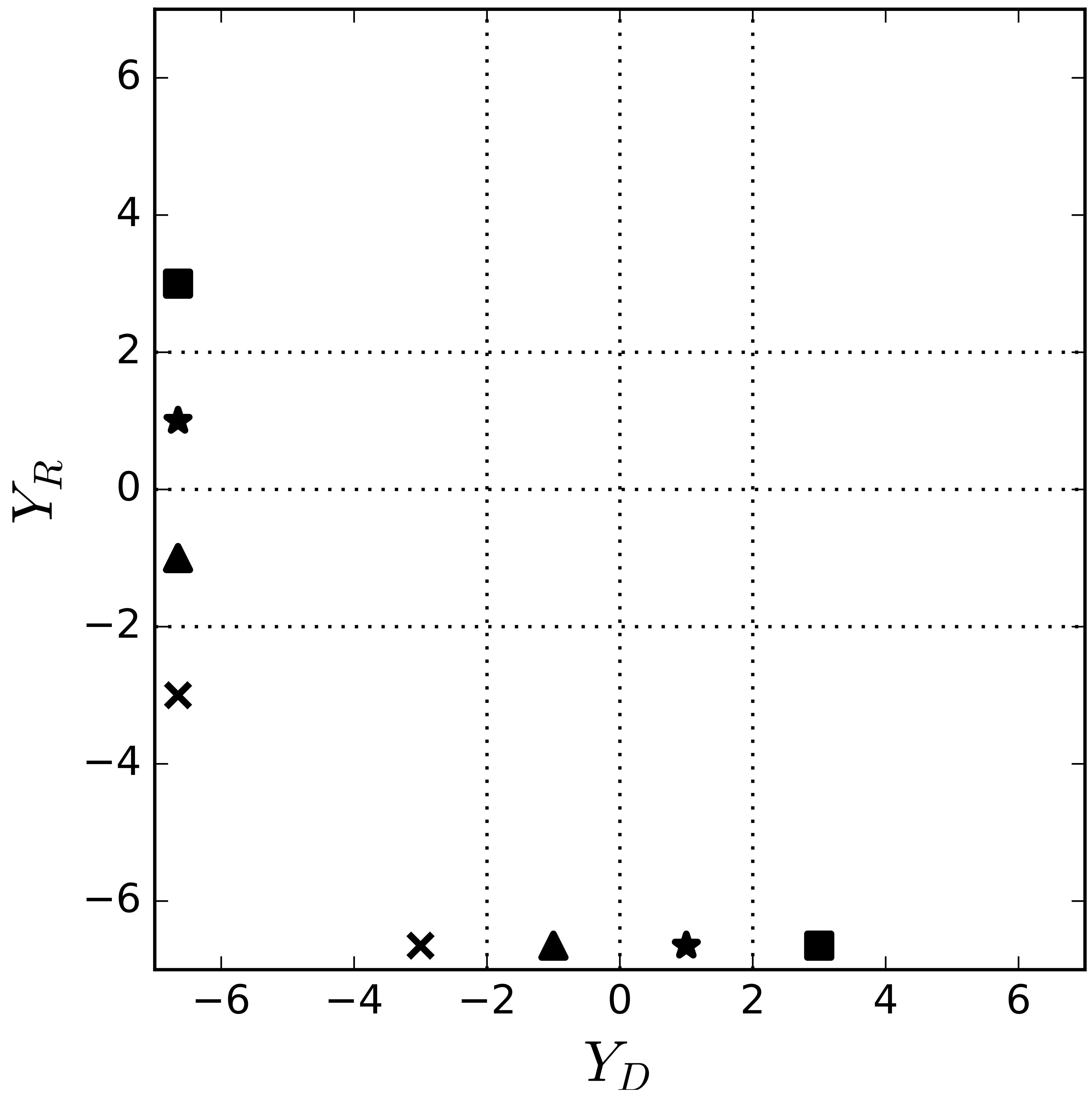
Symbol Error Rate for 16-QAM at $\gamma_D = \gamma_R = 7$ dB



Ozyilkan*, Carpi*, Garg & Erkip (IEEE J. Sel. Areas in Communications, 2025)

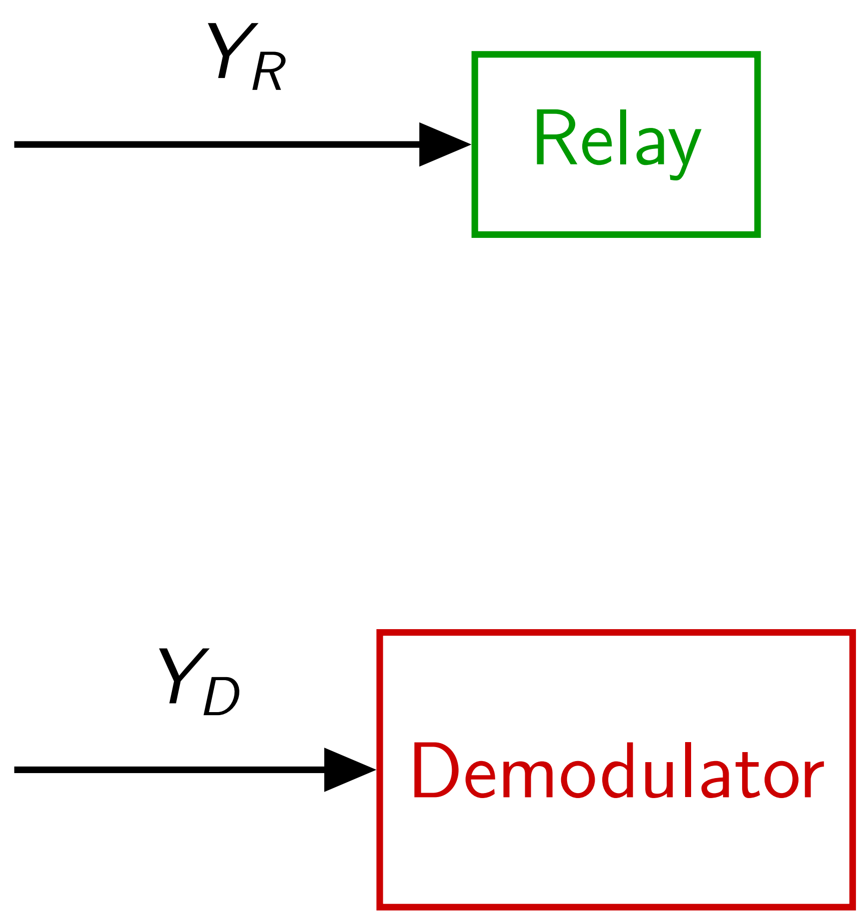
Quantization & Decisions for 4-PAM, $\gamma_D = \gamma_R = 10$ dB, $R \approx 1$

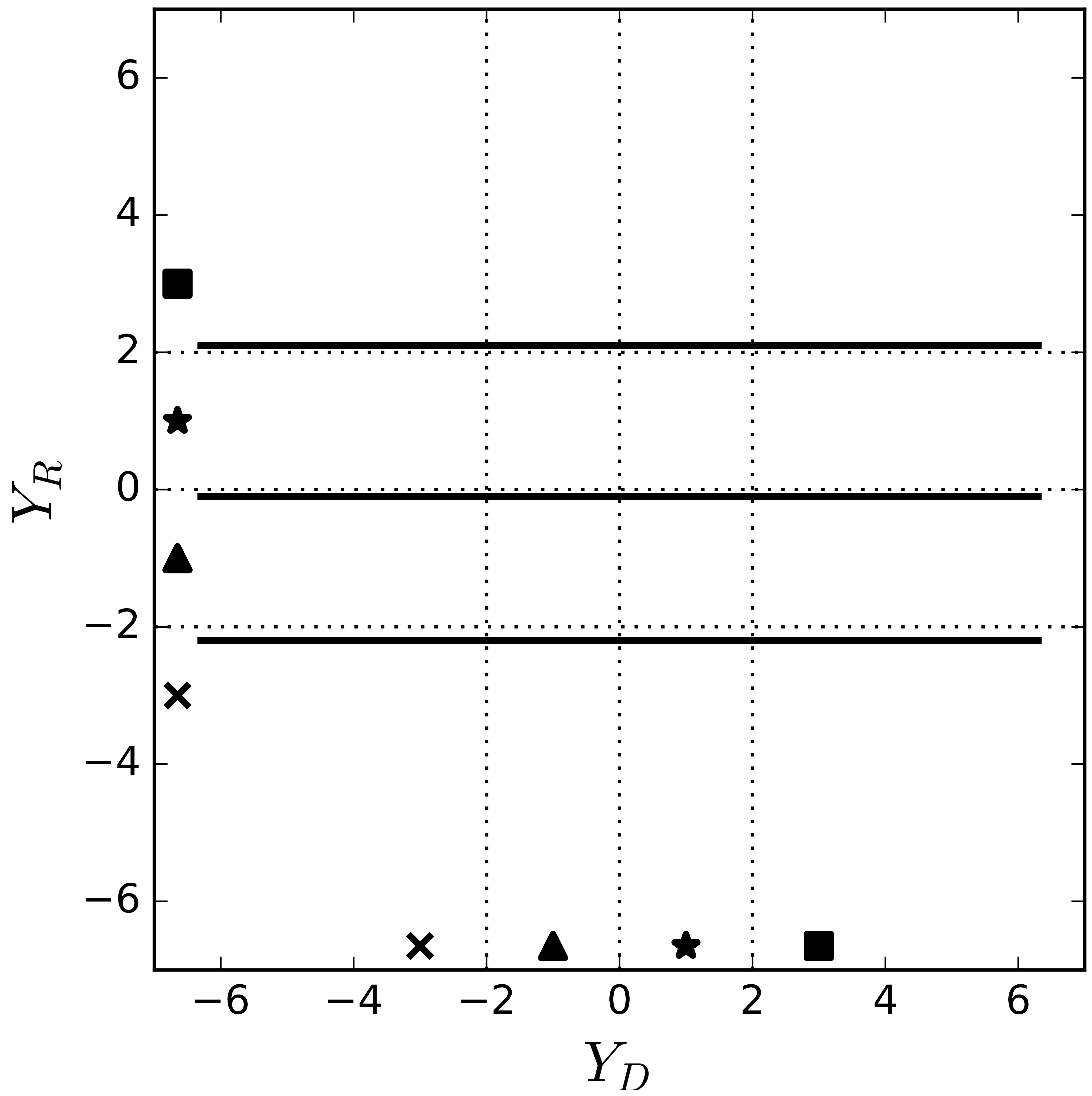




Ozyilkan*, Carpi*, Garg & Erkip (IEEE SPAWC, 2024)

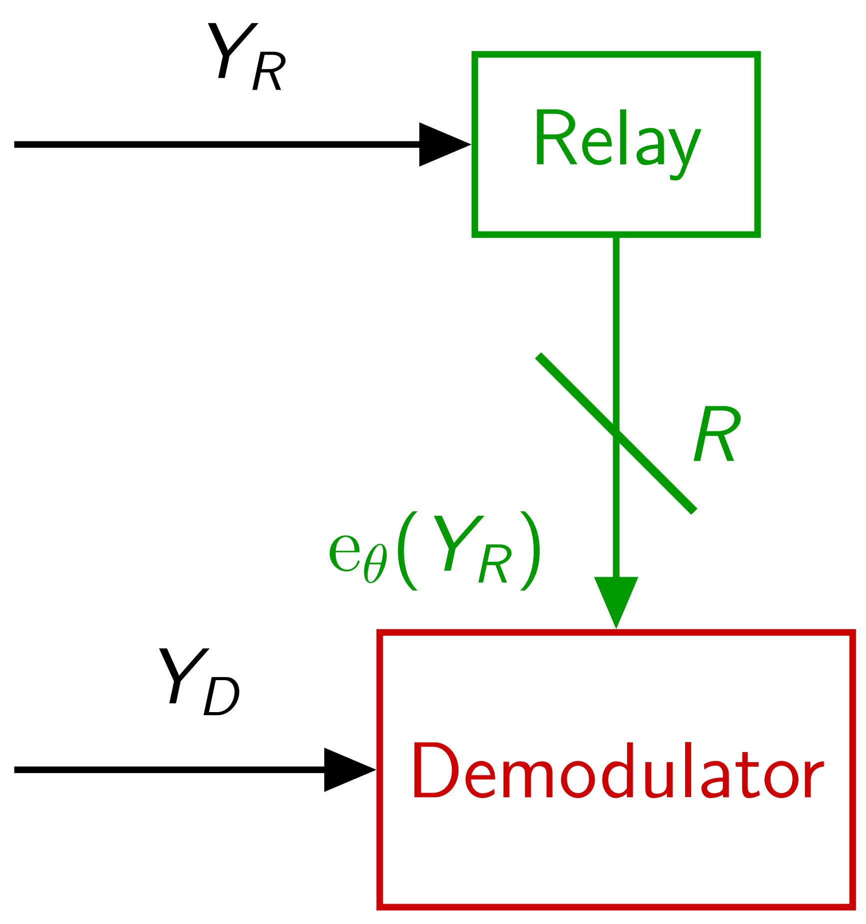
Quantization & Decisions for 4-PAM, $\gamma_D = \gamma_R = 10$ dB, $R \approx 1$

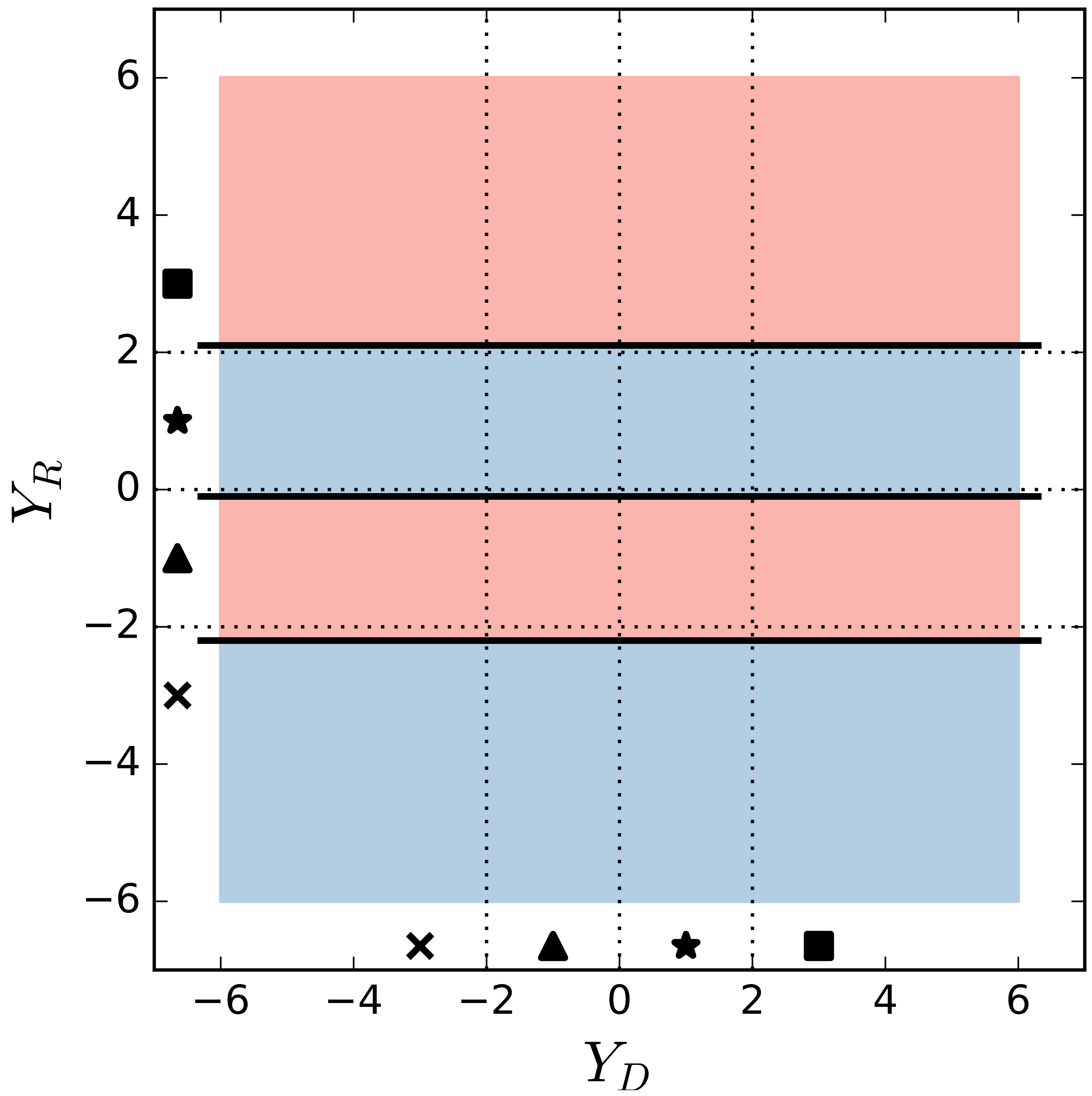




Ozyilkan*, Carpi*, Garg & Erkip (IEEE SPAWC, 2024)

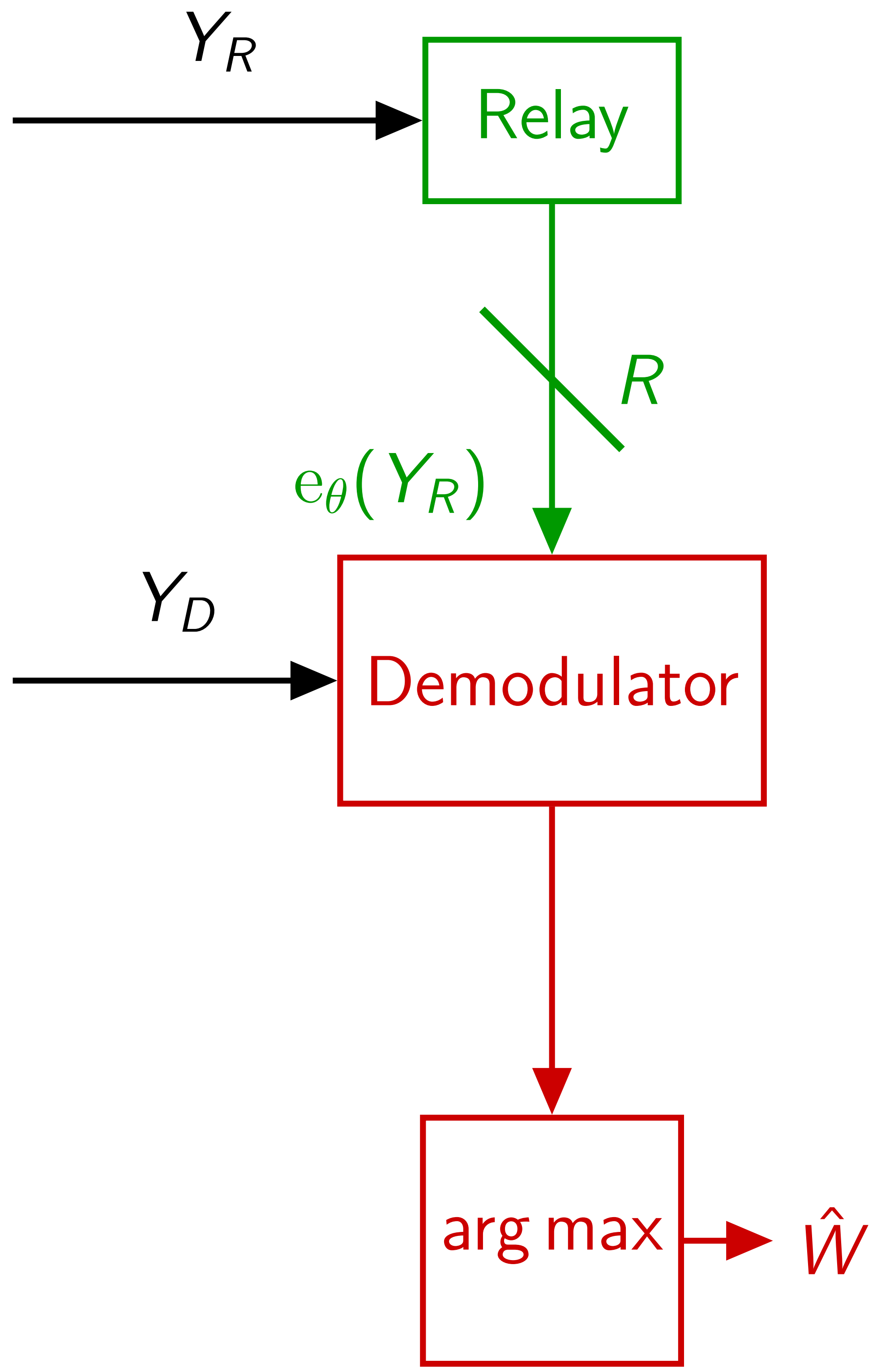
Quantization & Decisions for 4-PAM, $\gamma_D = \gamma_R = 10$ dB, $R \approx 1$

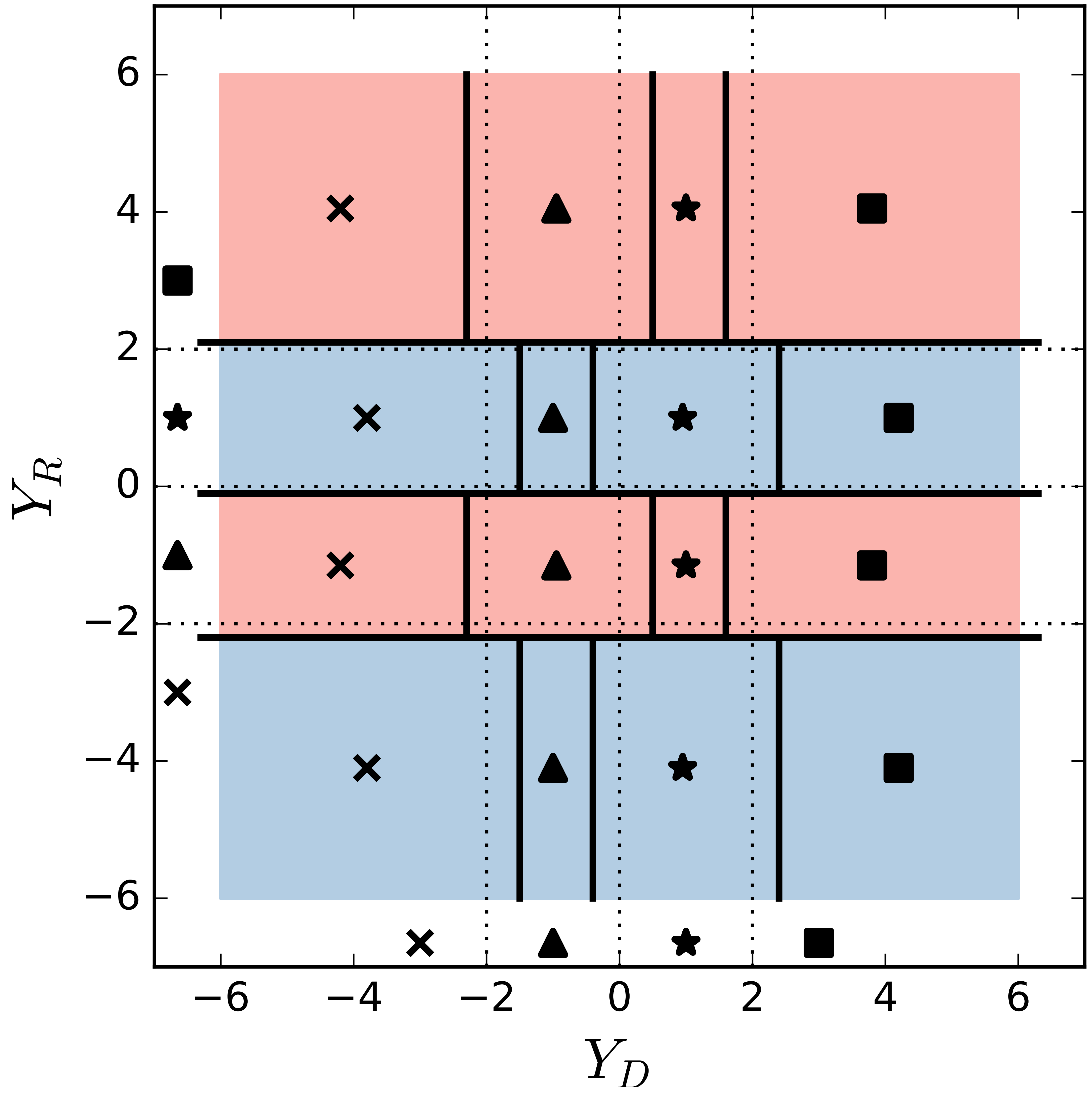




Ozyilkan*, Carpi*, Garg & Erkip (IEEE SPAWC, 2024)

Quantization & Decisions for 4-PAM, $\gamma_D = \gamma_R = 10$ dB, $R \approx 1$

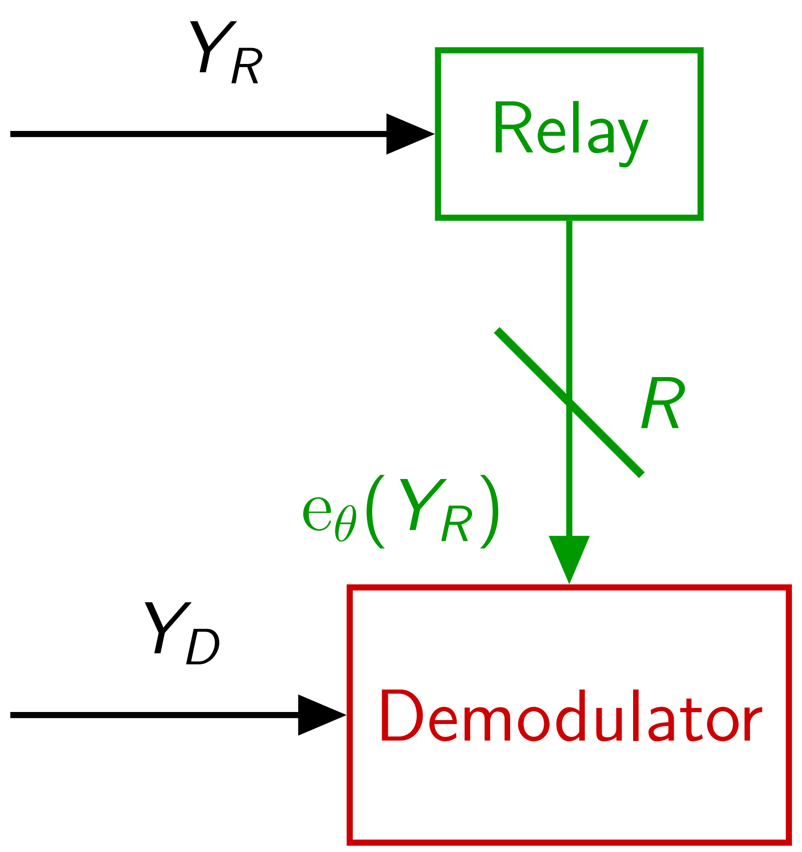


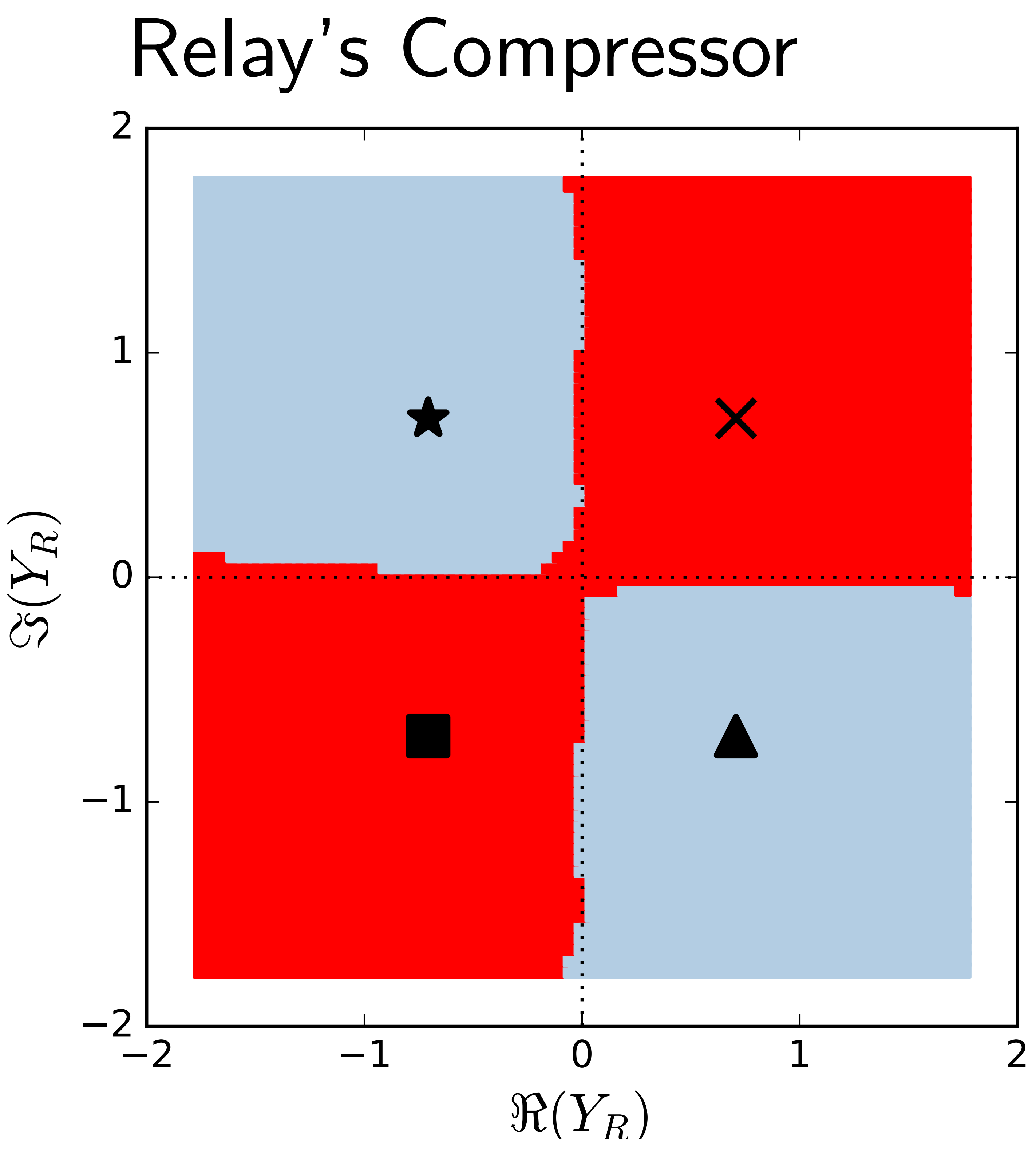


Ozyilkan*, Carpi*, Garg & Erkip (IEEE SPAWC, 2024)

Quantization & Decisions for QAM, $\gamma_D = \gamma_R = 7$ dB, $R \approx 1$

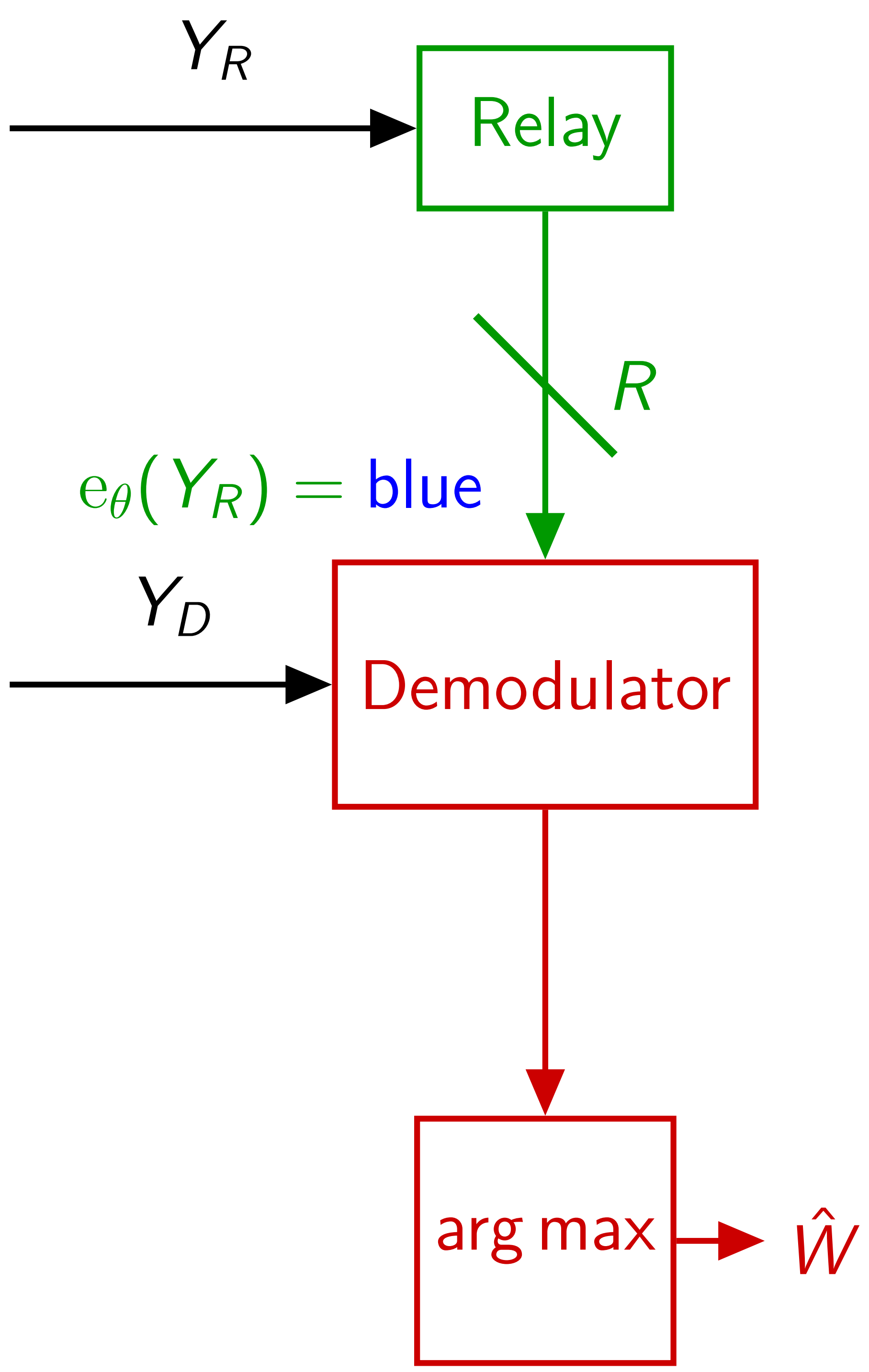
Destination's Demodulator

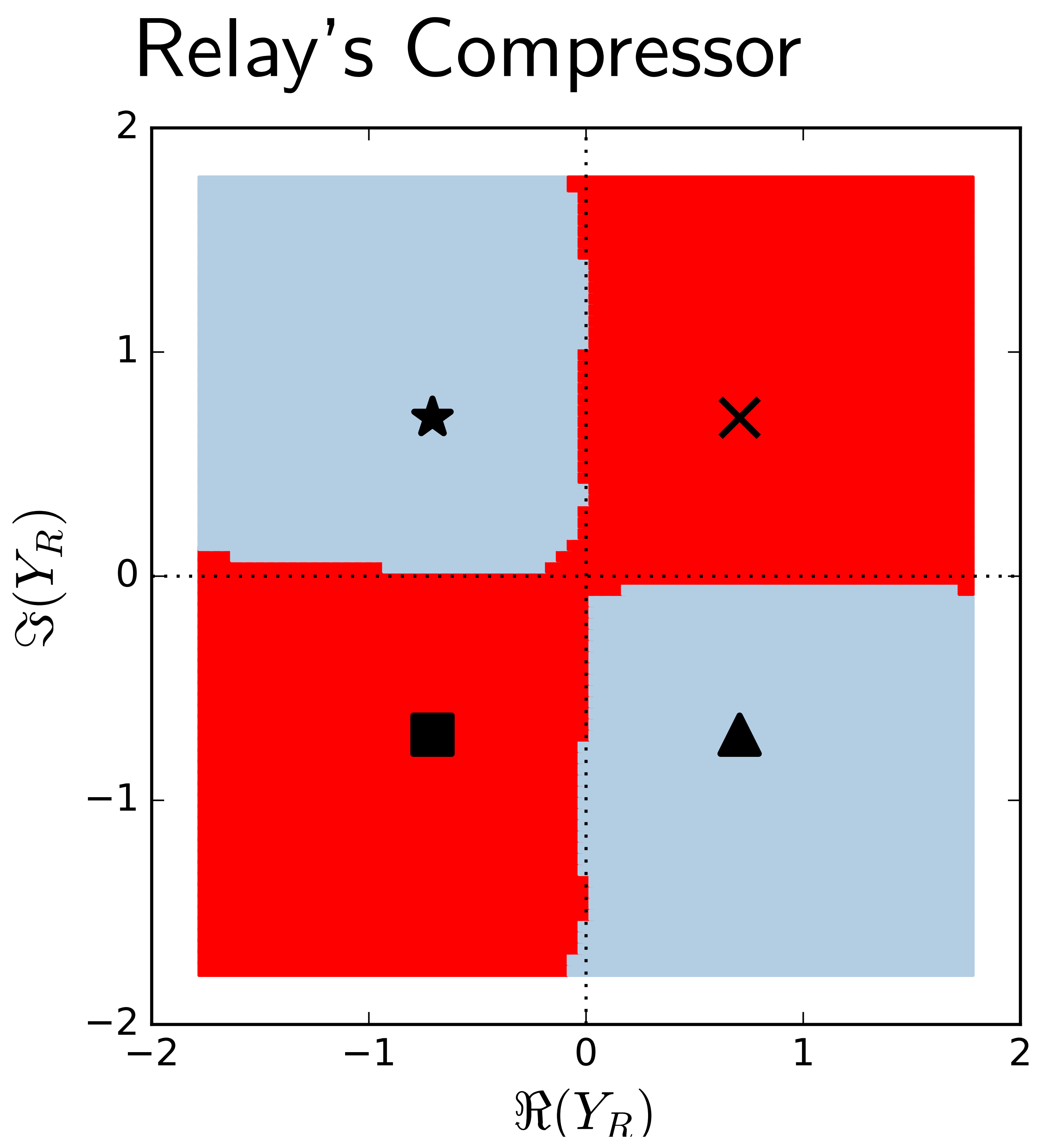


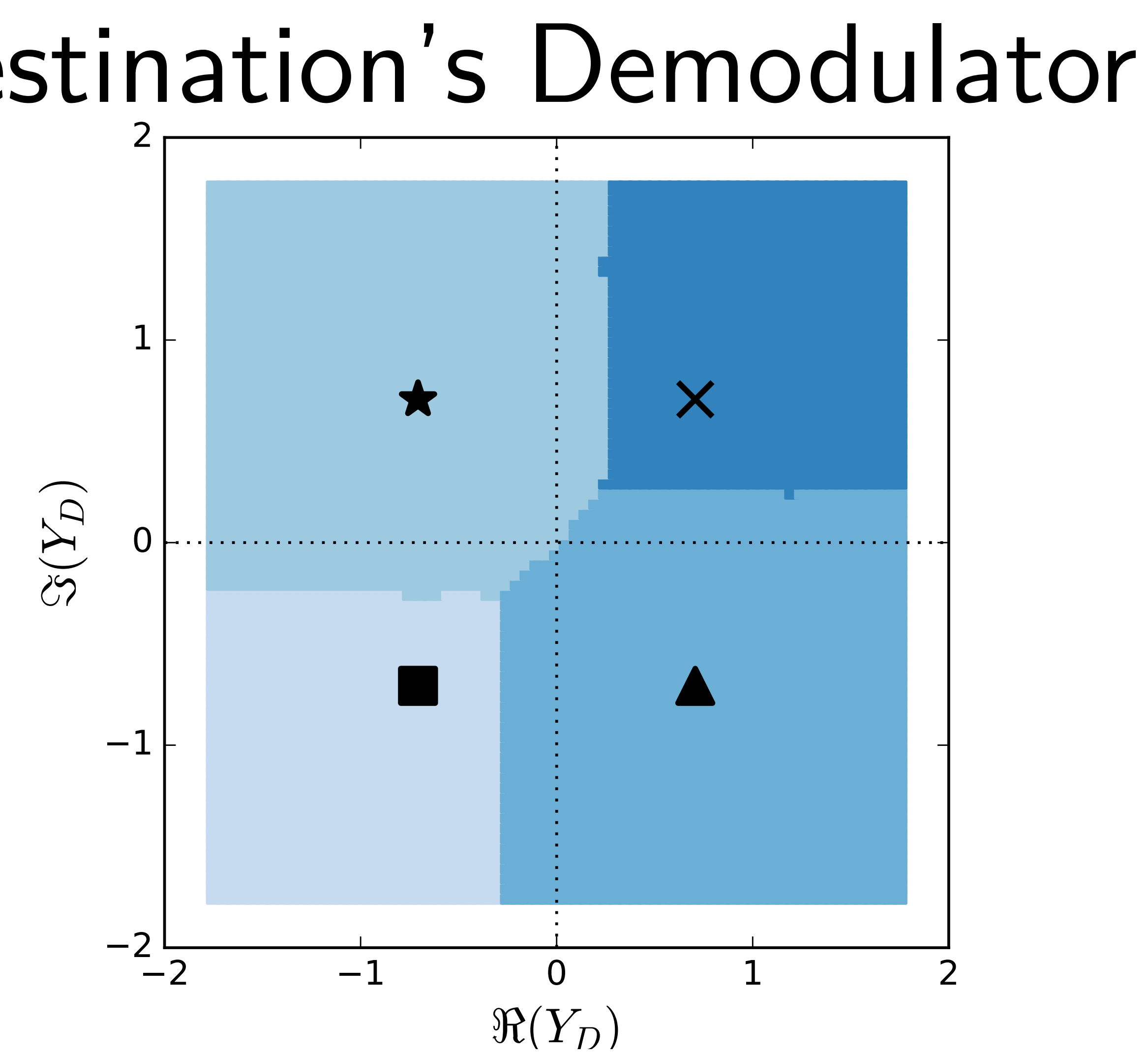


Ozyilkan*, Carpi*, Garg & Erkip (IEEE J. Sel. Areas in Communications, 2025)

Quantization & Decisions for QAM, $\gamma_D = \gamma_R = 7$ dB, $R \approx 1$

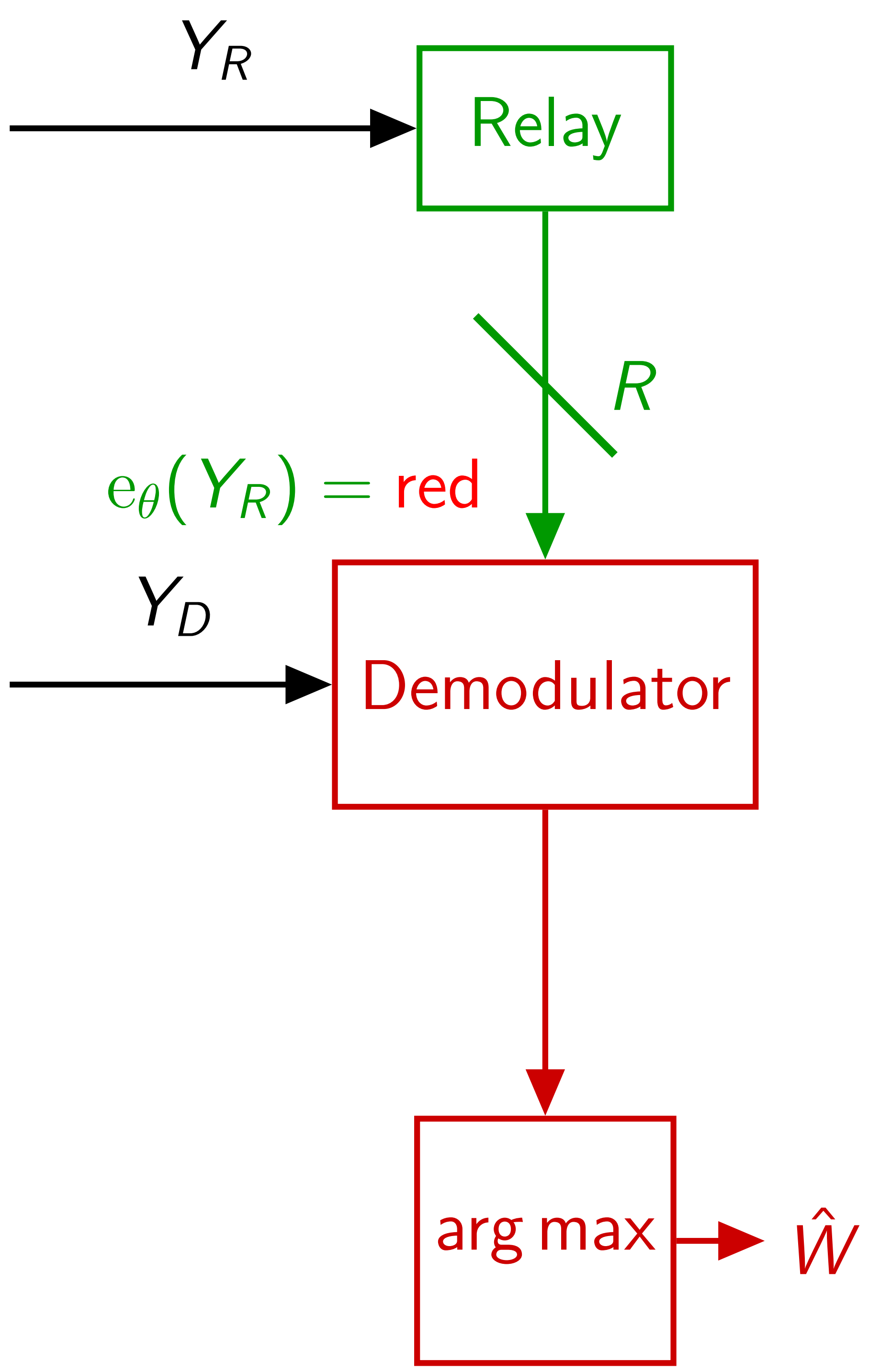


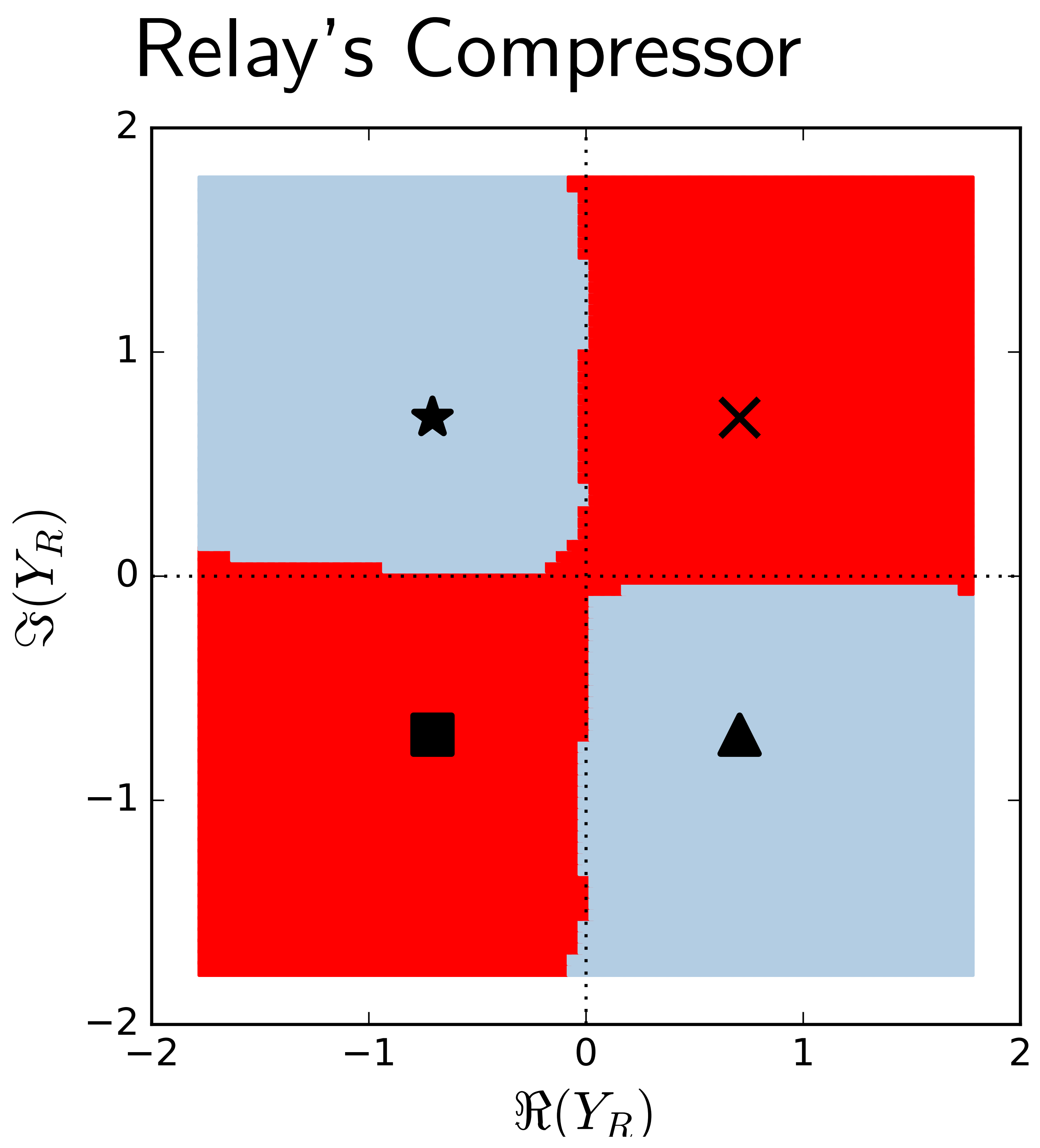


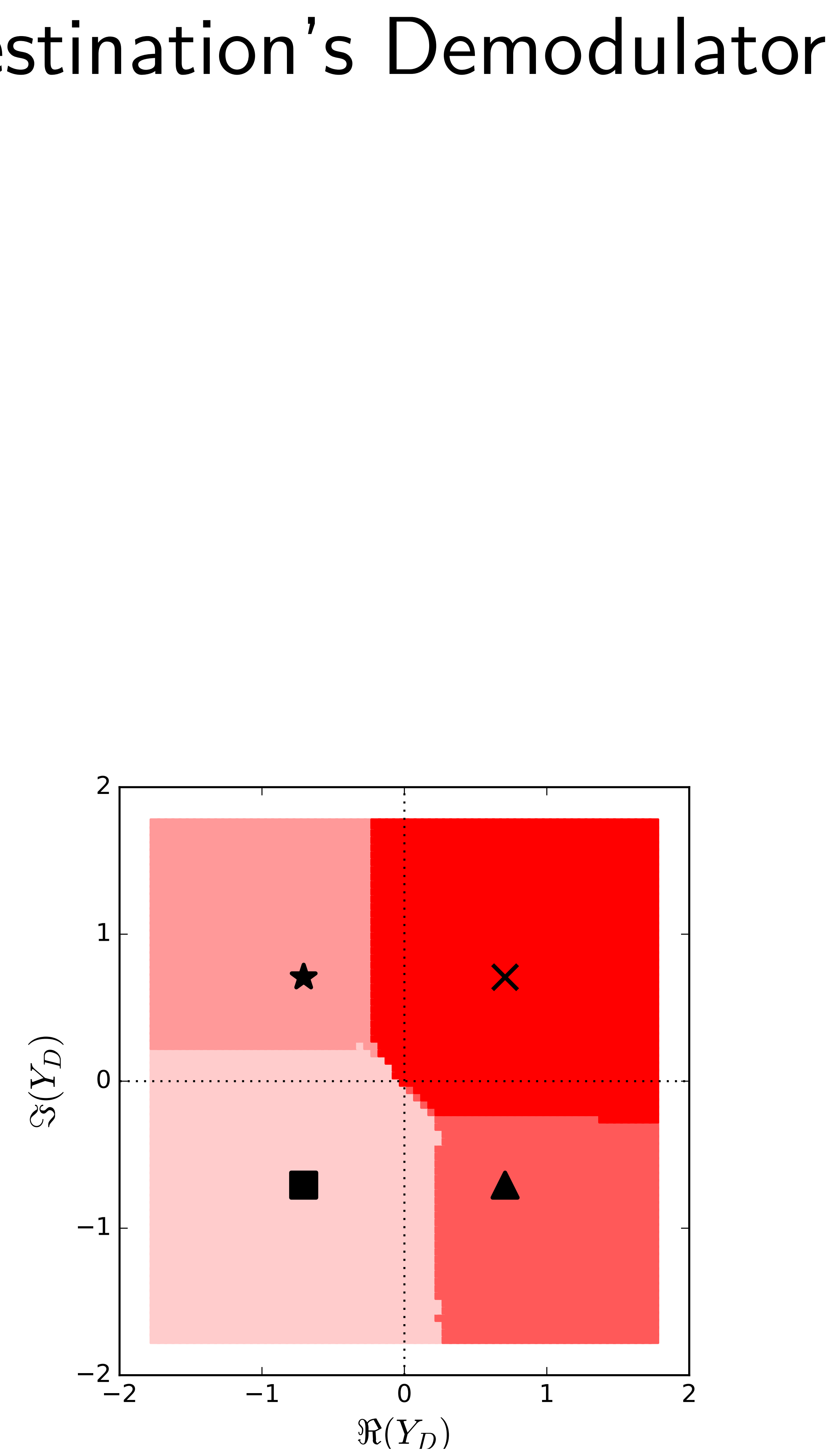


Ozyilkan*, Carpi*, Garg & Erkip (IEEE J. Sel. Areas in Communications, 2025)

Quantization & Decisions for QAM, $\gamma_D = \gamma_R = 7$ dB, $R \approx 1$

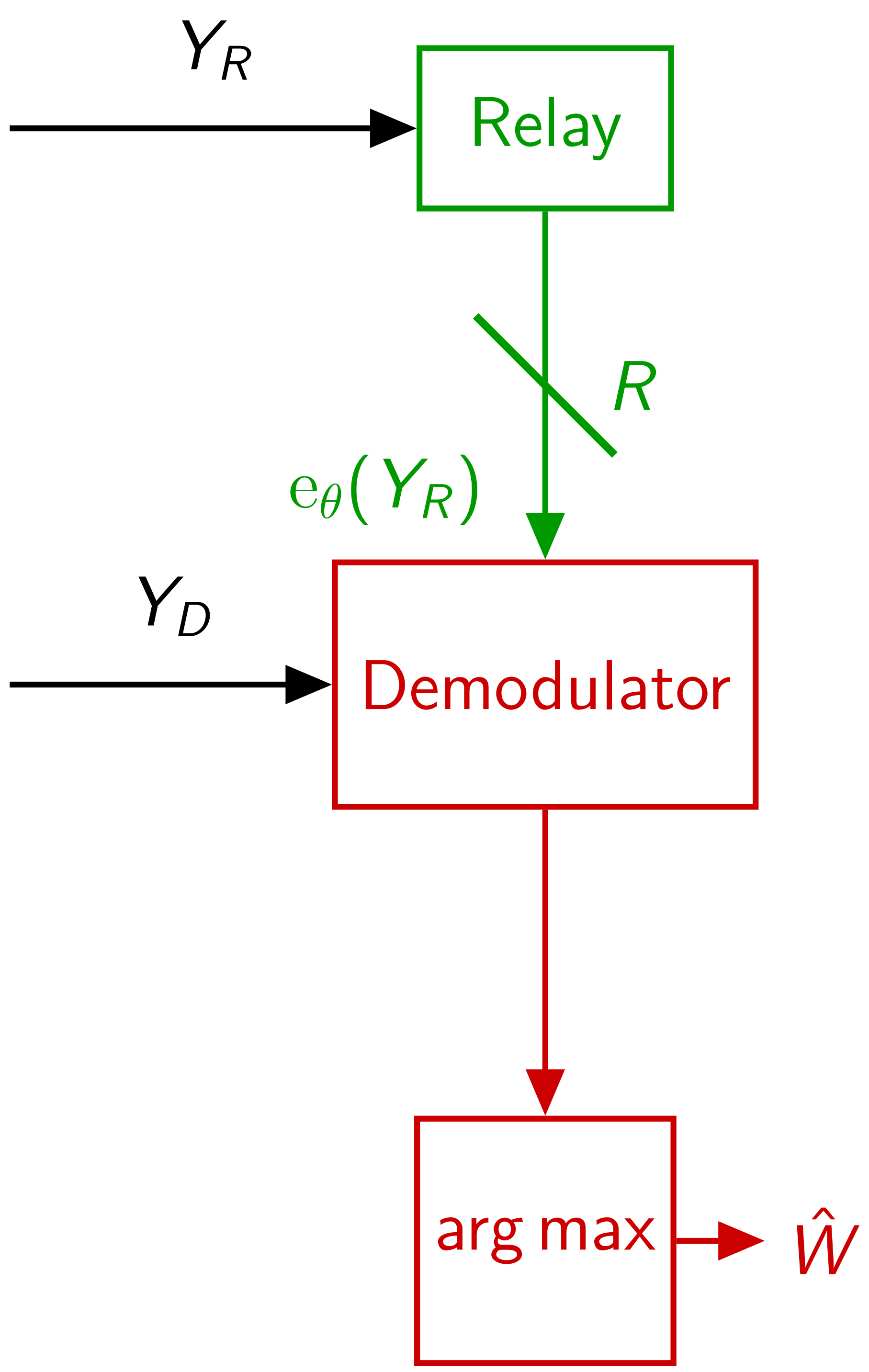




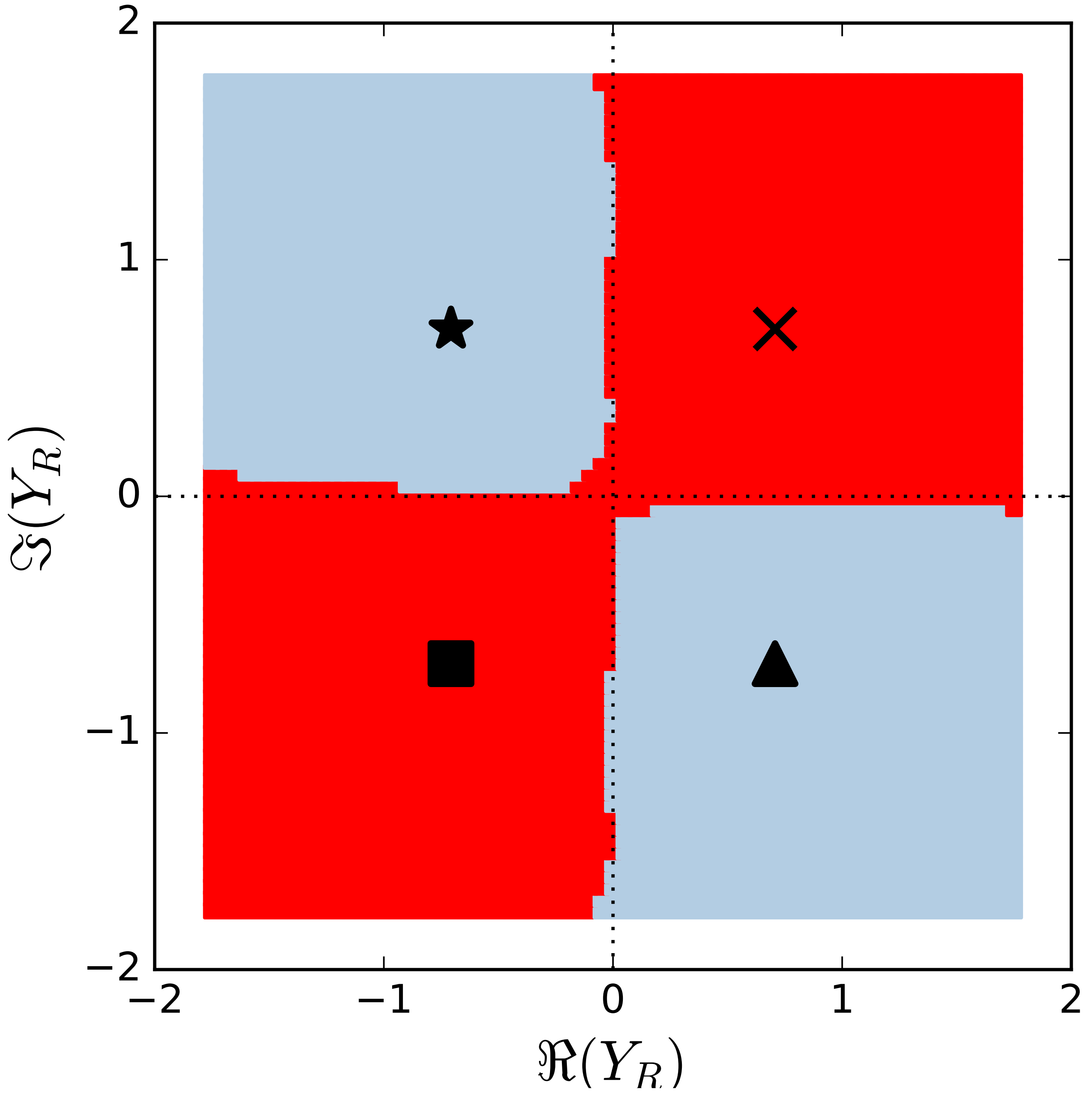


Ozyilkan*, Carpi*, Garg & Erkip (IEEE J. Sel. Areas in Communications, 2025)

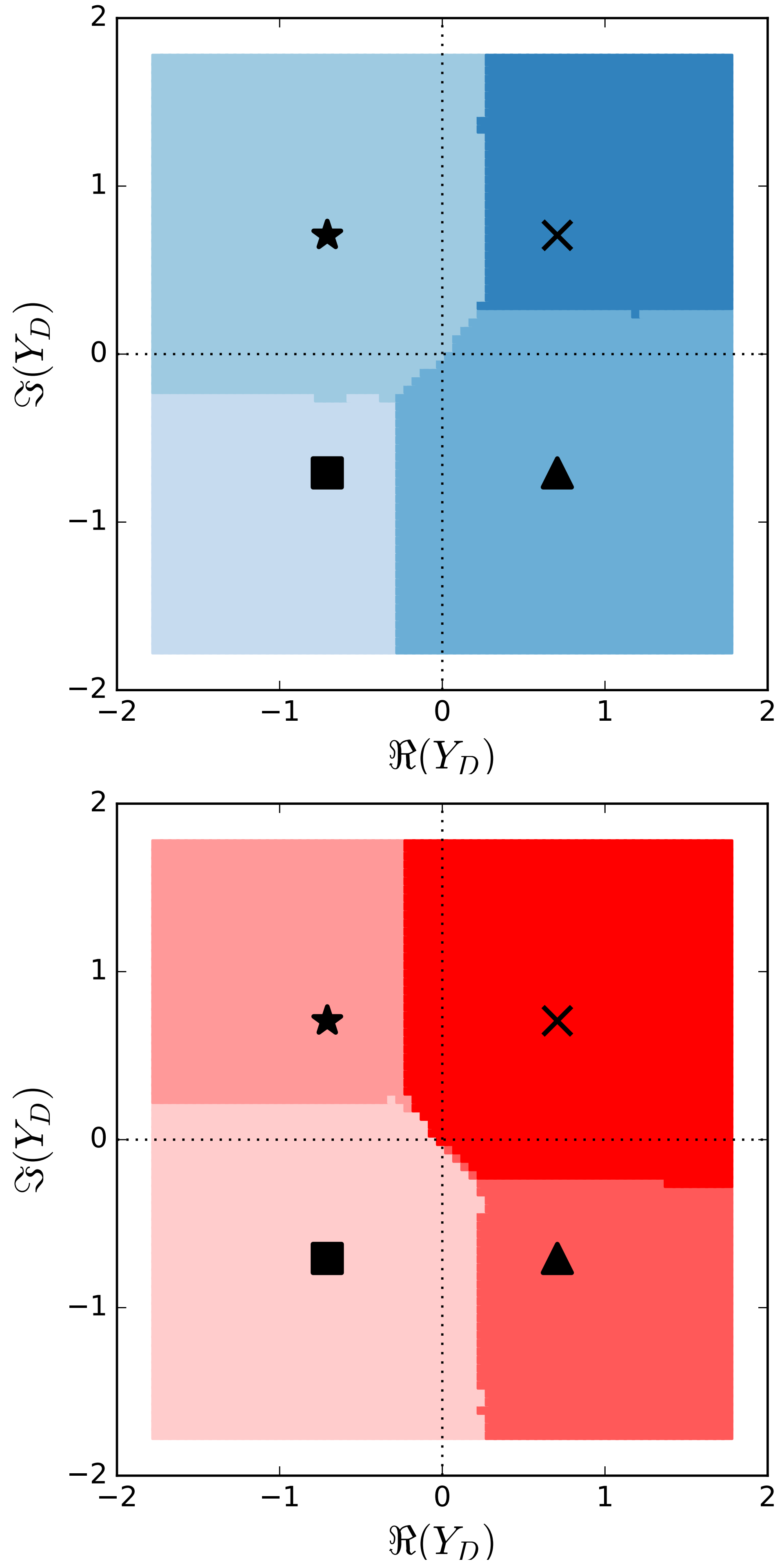
Quantization & Decisions for QAM, $\gamma_D = \gamma_R = 7$ dB, $R \approx 1$



Relay's Compressor



Destination's Demodulator



Ozyilkan*, Carpi*, Garg & Erkip (IEEE J. Sel. Areas in Communications, 2025)

Summary on Neural Compress-and-Forward (CF)

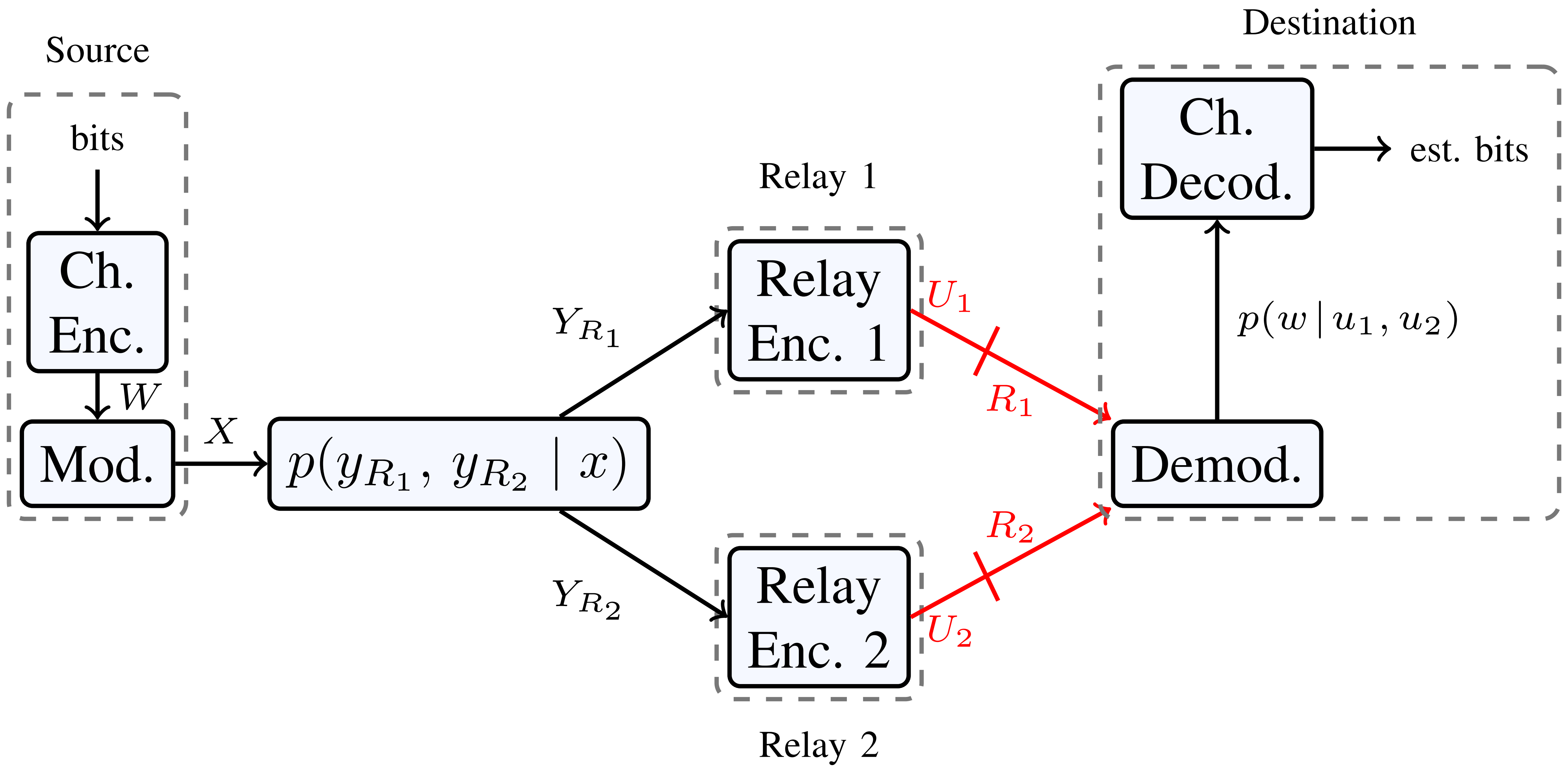
- First proof-of-concept towards practical neural CF relaying scheme.
- Distributed compression helps in exploiting correlation at the destination.

Summary on Neural Compress-and-Forward (CF)

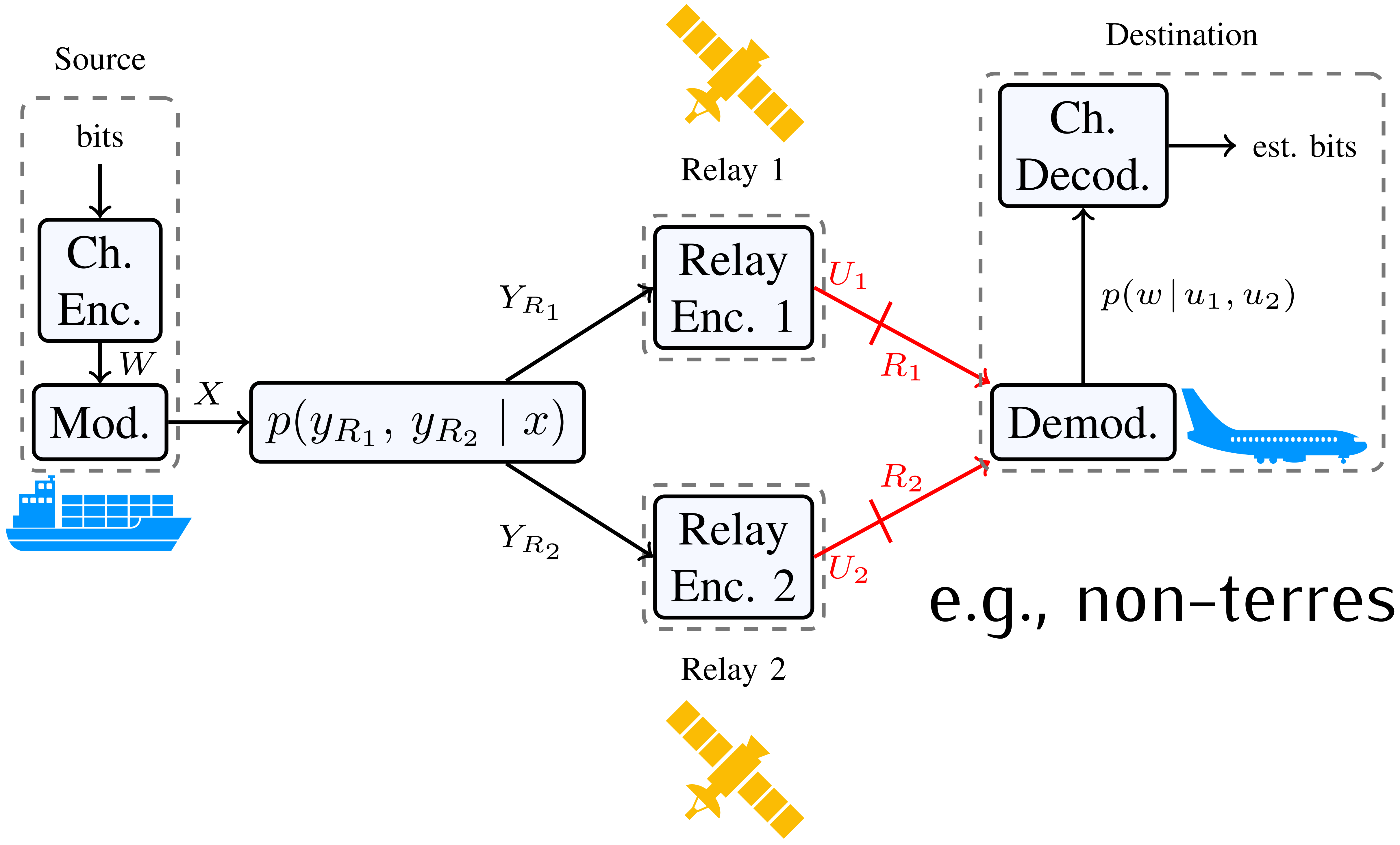
- First proof-of-concept towards practical neural CF relaying scheme.
- Distributed compression helps in exploiting correlation at the destination.

- **Ongoing project:**
  - **Extending neural CF to *diamond relay channel***
    (i.e., w/ two relays connected to the destination via two separate links)

Neural CF for the Diamond Relay Channel

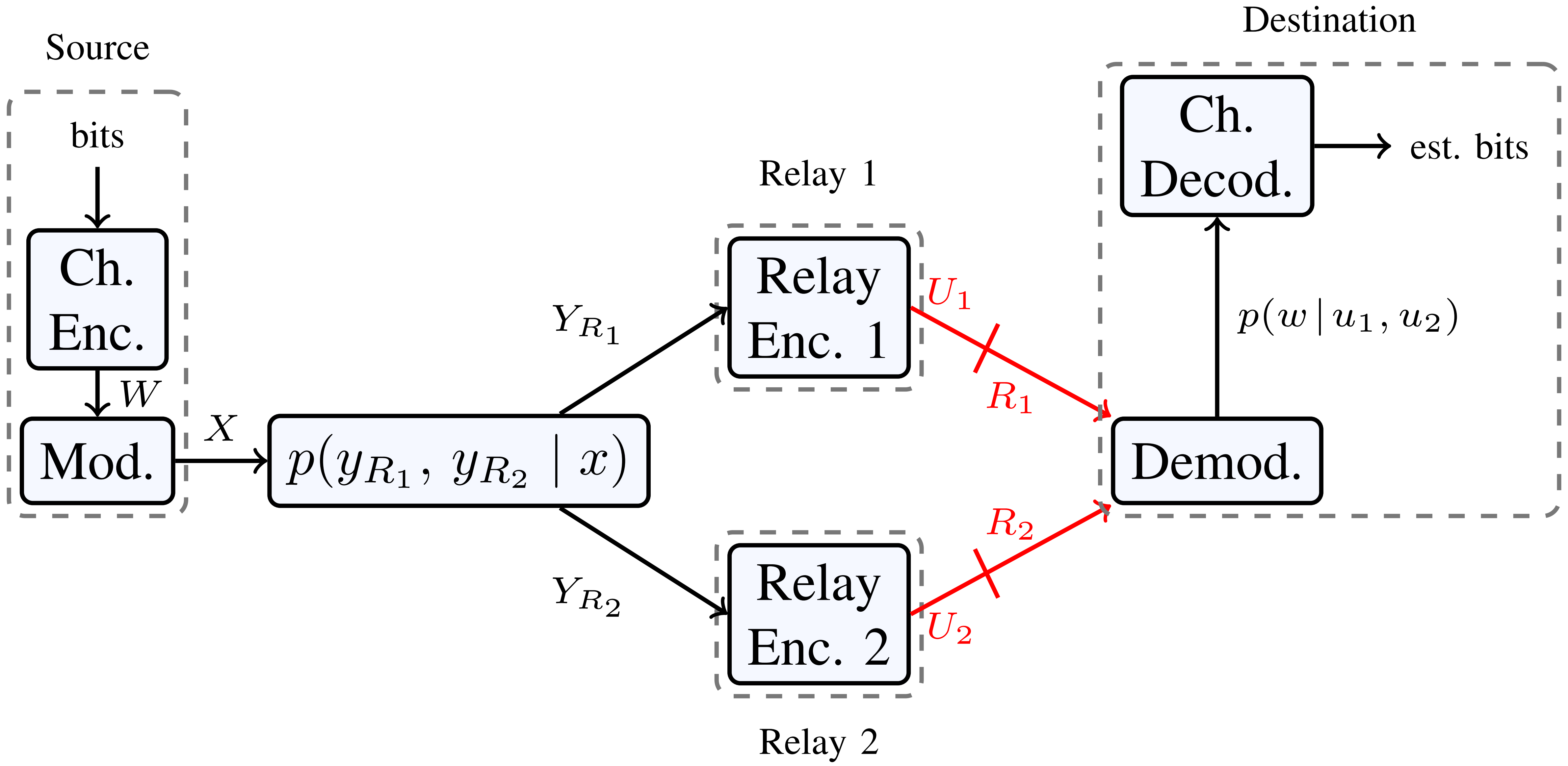


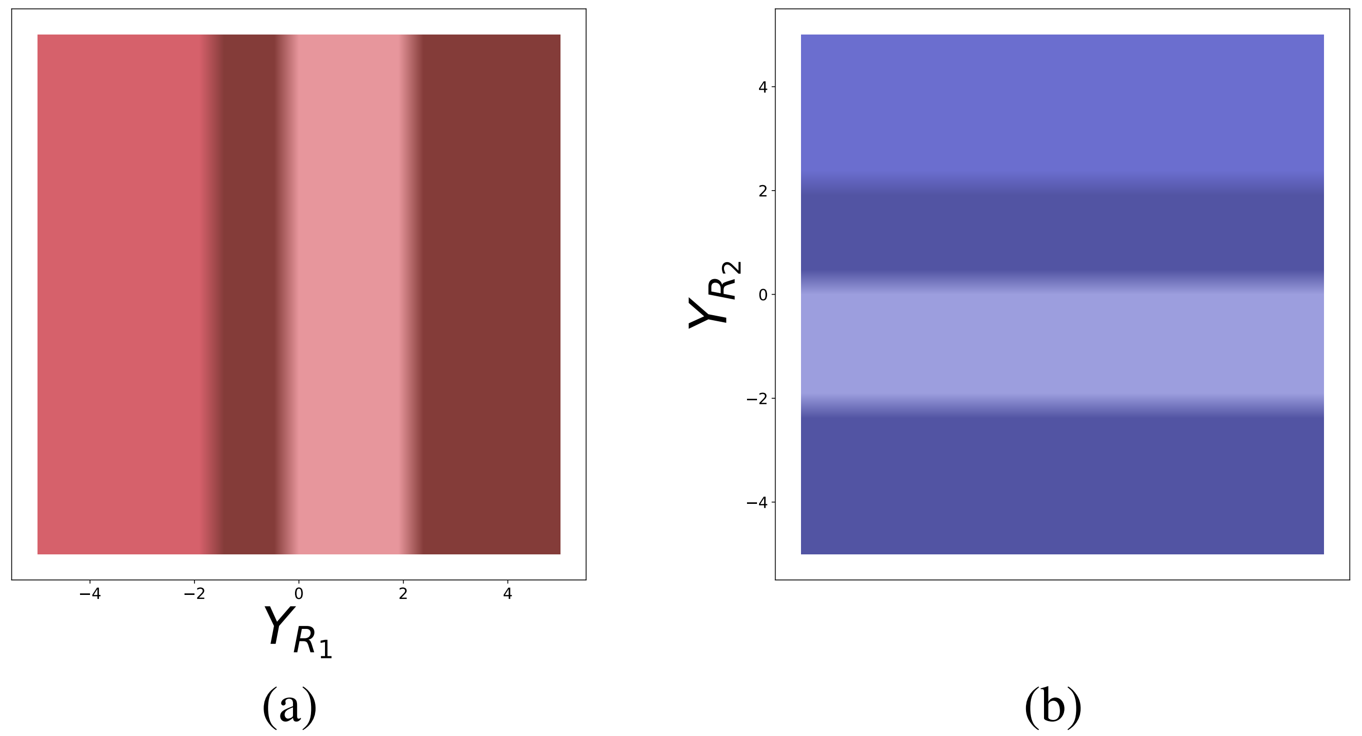
Neural CF for the Diamond Relay Channel

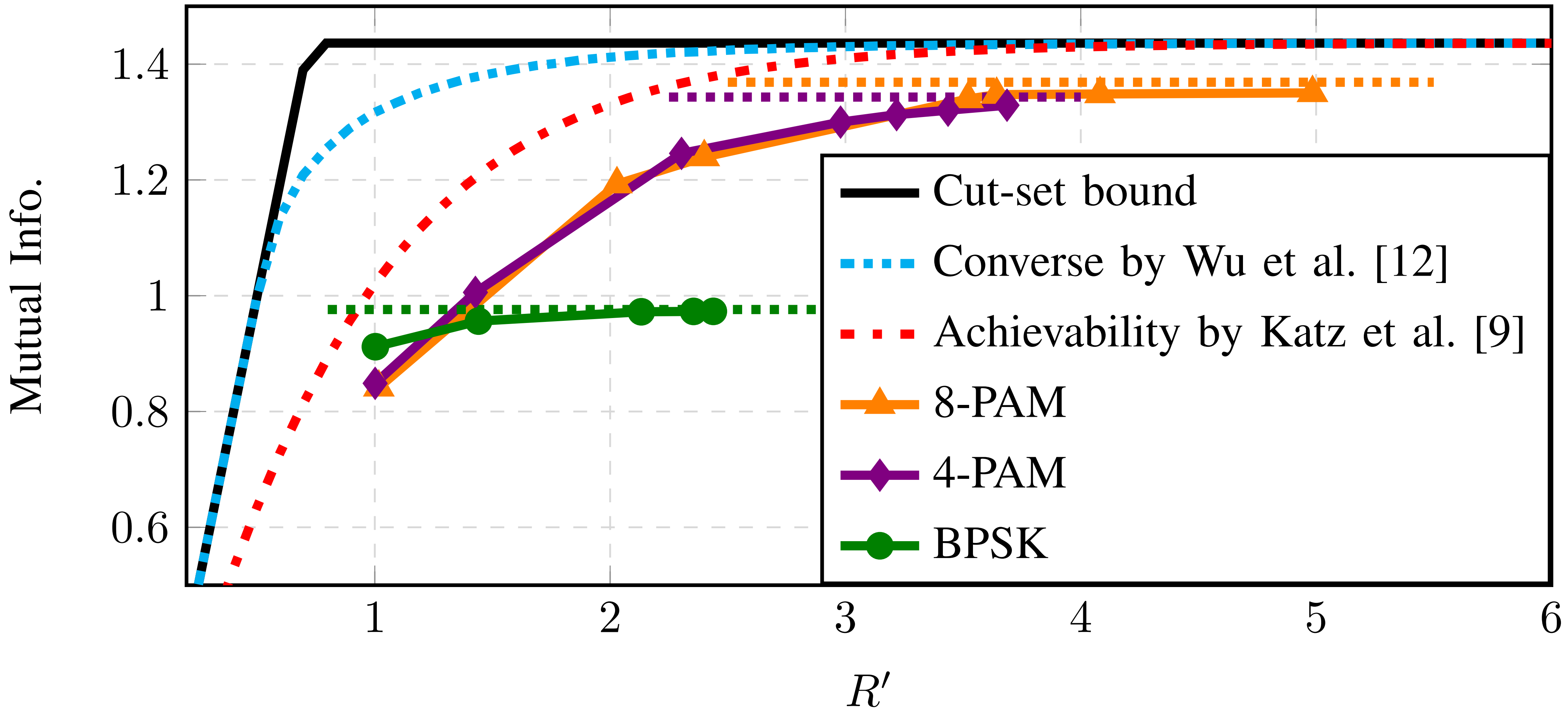


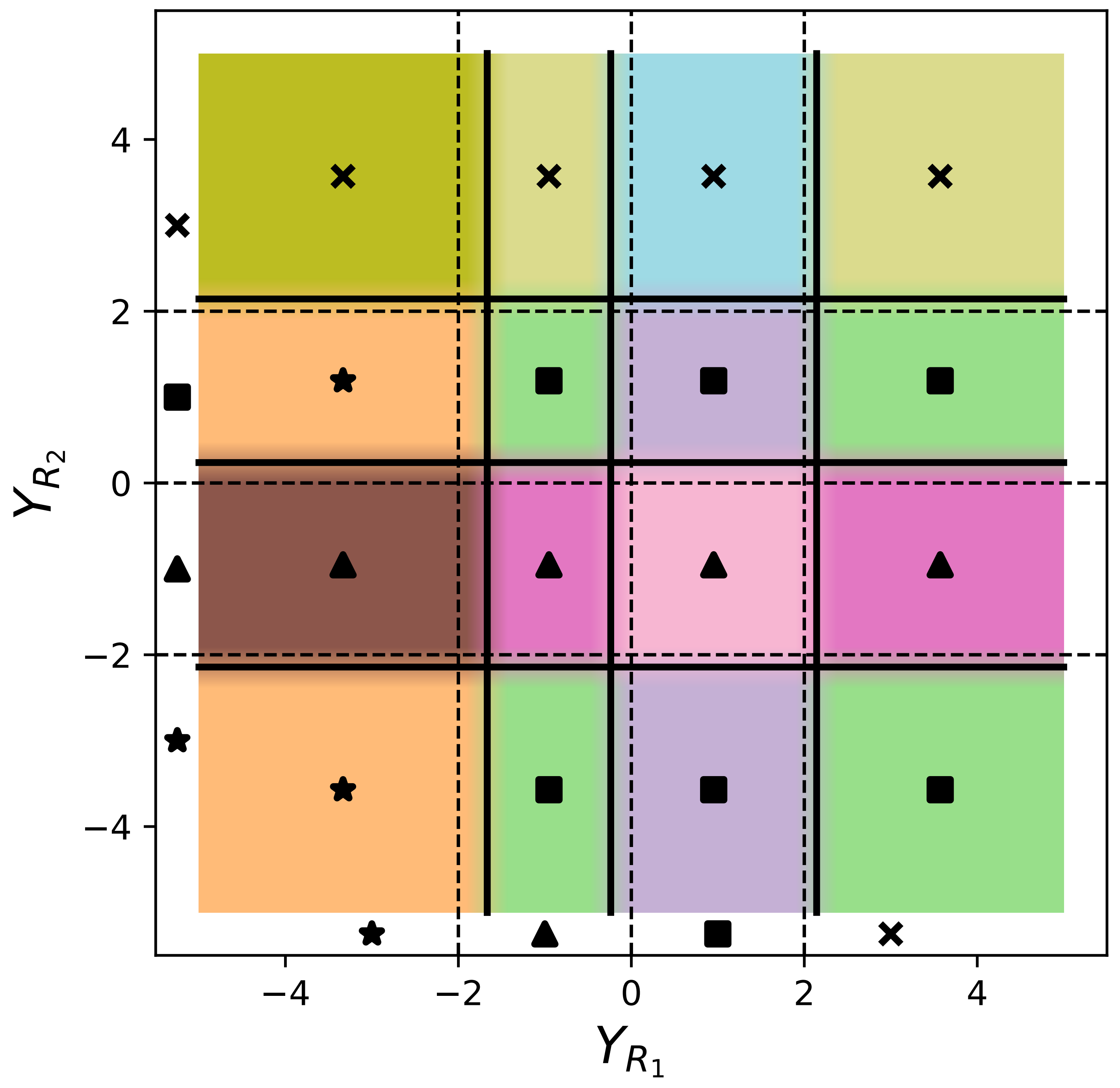
e.g., non-terrestrial networks

Neural CF for the Diamond Relay Channel









Aygun, Ozyilkan & Erkip (IEEE Asilomar, 2025)

Summary on Neural Compress-and-Forward

- Learning-based distributed compressors are useful for task-aware/semantic communication problems while being interpretable!

Summary on Neural Compress-and-Forward

- Learning-based distributed compressors are useful for task-aware/semantic communication problems while being interpretable!
  - Operating close to the capacity in primitive relay channel:

    → Ozyilkan*, Carpi*, Garg & Erkip (IEEE SPAWC, 2024)

    → Ozyilkan*, Carpi*, Garg & Erkip (IEEE J. Sel. Areas in Communications, 2025)

    ✓ Patent application in preparation.

Summary on Neural Compress-and-Forward

- Learning-based distributed compressors are useful for task-aware/semantic communication problems while being interpretable!
  - Operating close to the capacity in primitive relay channel:

    → Ozyilkan*, Carpi*, Garg & Erkip (IEEE SPAWC, 2024)

    → Ozyilkan*, Carpi*, Garg & Erkip (IEEE J. Sel. Areas in Communications, 2025)

    ✓ Patent application in preparation.
  - Observations generalize to multi-relay setups as well:

    → Aygun, Ozyilkan & Erkip (IEEE Asilomar, 2025)

Summary on Neural Compress-and-Forward

- Learning-based distributed compressors are useful for task-aware/semantic communication problems while being interpretable!
  - Operating close to the capacity in primitive relay channel:

    → Ozyilkan*, Carpi*, Garg & Erkip (IEEE SPAWC, 2024)

    → Ozyilkan*, Carpi*, Garg & Erkip (IEEE J. Sel. Areas in Communications, 2025)

    ✓ Patent application in preparation.
  - Observations generalize to multi-relay setups as well:

    → Aygun, Ozyilkan & Erkip (IEEE Asilomar, 2025)
- Revised objective: Distortion ↔ Classification.
- MPEG activity, "Video Coding for Machines"

  → Ozyilkan*, Ulhaq*, Choi & Racapé (IEEE DCC, 2023)

If you found this talk interesting ...

Publications: **https://ezgi.space**

Reach out to me at **ezgi.ozyilkan@nyu.edu**!

If you found this talk interesting ...

Publications: **https://ezgi.space**

Reach out to me at **ezgi.ozyilkan@nyu.edu**!

Moving forward, will be working on:
perceptual optimization
+ **compression** + **3D vision**

NYU

Apple, ML + Video Research

Thank you! Q&A?

Neural Distributed Data Compression and Communication

Ezgi Ozyilkan

https://ezgi.space

ezgi.ozyilkan@nyu.edu