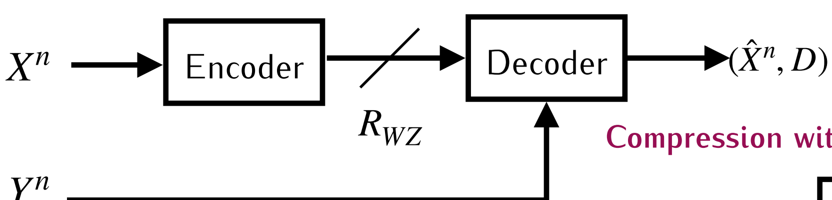
Neural Distributed Compressor Does Binning

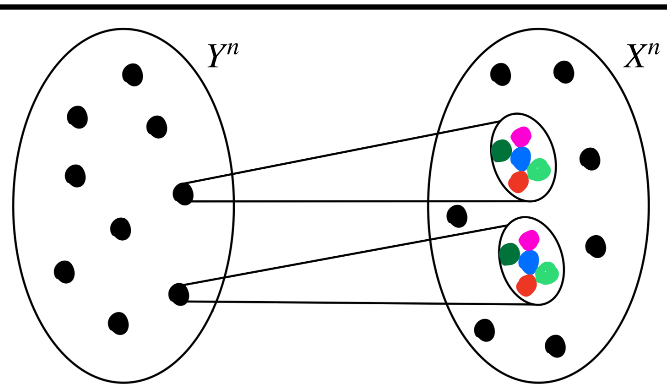
Ezgi Özyılkan*, Johannes Ballé†, Elza Erkip*
*NYU, †Google Research
{ezgi.ozyilkan,elza}@nyu.edu, jballe@google.com

Overview

Summary: Demonstrate that neural distributed compressor mimics the Wyner-Ziv theorem and does binning, although **no particular structure** was imposed onto the model.

Although Wyner-Ziv setup is heavily studied in network information theory and has real-life applications (e.g., federated learning and sensor networks), constructing a practical framework for arbitrary sources is still **an open problem**.





For X^n, the optimal compressor sends the color within the "fan" $\Longrightarrow$ **Binning**.

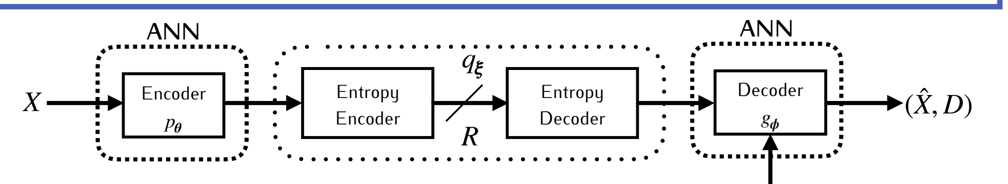
Rate-Distortion (R-D) with Side Information
(Wyner & Ziv, 1976)
Let (X, Y) be correlated and drawn i.i.d. $\sim p(x, y)$. The R-D function for X when Y available at the decoder is:

$$R_{WZ}(D) = \min(I(X; U) - I(Y; U)),$$

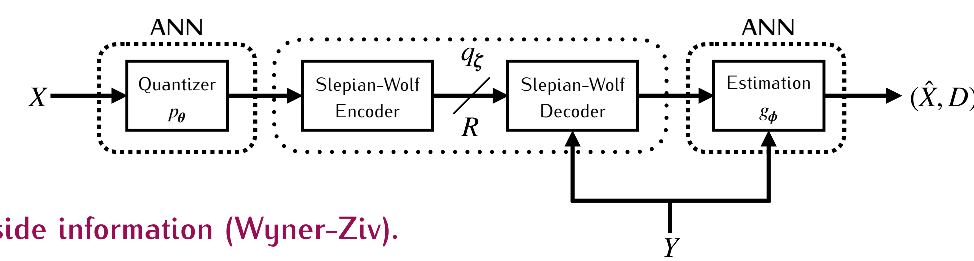
where the minimization is over all $p(u\,|\,x)$ and all $g(u, y)$ satisfying the average distortion criterion.

Framework

Main idea: Leverage universal function approximation capability of neural networks to find **constructive solutions** for one-shot **Wyner-Ziv compression**.



Marginal (up) and **conditional (down)** system models.



Compression with side information (Wyner-Ziv).

Assume that the encoder is represented by a probability model $p_\theta(u\,|\,x)$,

$$I(X; U) - I(Y; U) = I(X; U\,|\,Y) = \mathbb{E}\left[\log \frac{p_\theta(u\,|\,x)}{p(u\,|\,y)}\right].$$

Set encoder output as $u = \operatorname{argmax}_v p_\theta(v\,|\,x)$. Have U as discrete.
Choose one of two variational upper bounds:

$$I(X; U\,|\,Y) \leq \mathbb{E} \log \frac{p_\theta(u\,|\,x)}{q_\xi(u)},$$

$$I(X; U\,|\,Y) \leq \mathbb{E} \log \frac{p_\theta(u\,|\,x)}{q_\zeta(u\,|\,y)}.$$

Relax the constrained formulation of Wyner-Ziv theorem using Lagrange multipliers:

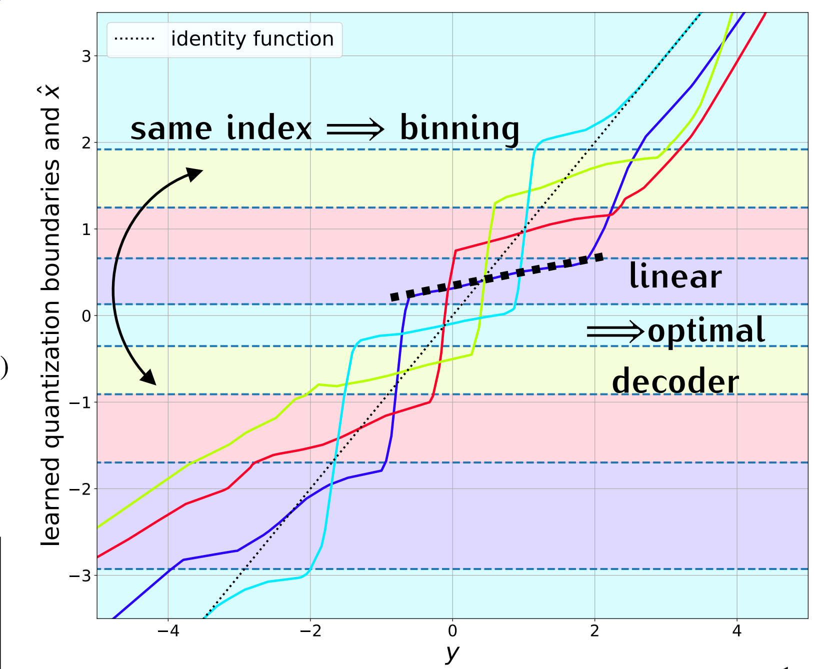
$$L_{\mathrm{m}}(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\xi}) = \mathbb{E}\left[\log \frac{p_\theta(u\,|\,x)}{q_\xi(u)} + \lambda \cdot d(x, g_\phi(u, y))\right],$$

$$L_{\mathrm{c}}(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\zeta}) = \mathbb{E}\left[\log \frac{p_\theta(u\,|\,x)}{q_\zeta(u\,|\,y)} + \lambda \cdot d(x, g_\phi(u, y))\right].$$

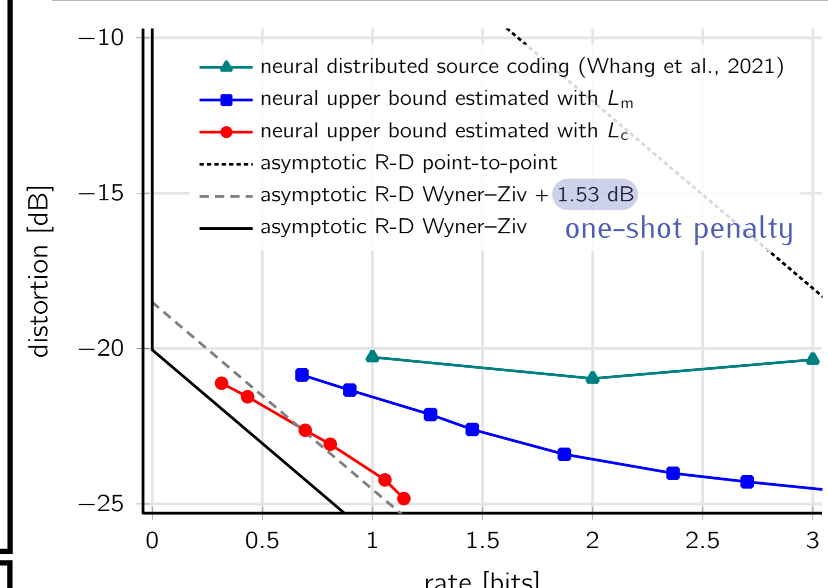
Define all probabilistic models as **discrete distributions**.
Use Gumbel-max (Gumbel, 1954) to draw samples, and Concrete distributions (Maddison et al., 2016) to facilitate optimization.

Results

To evaluate **how close we get to the R-D bound**, choose X and Y as i.i.d. Gaussian memoryless sources and $d(\cdot)$ as mean-squared error.



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



$Y = X + N$ with $X \sim N(0,1)$ and N $\sim N(0,10^{-2})$.

Paper available at https://arxiv.org/abs/2305.04380
To appear in ISIT, June 2023