Neural Distributed Compressor Does Binning

Ezgi Özyılkan [1] **Johannes Ballé** [2] **Elza Erkip** [1]

Abstract

We consider lossy compression of an information source when the decoder has lossless access to a correlated one. This setup, also known as the *Wyner–Ziv* problem in information theory, is a special case of distributed source coding. To this day, real-world applications of this problem have neither been fully developed nor heavily investigated. We find that our neural network-based compression scheme re-discovers some principles of the optimum theoretical solution of the Wyner–Ziv setup, such as *binning* in the source space as well as linear decoder behavior within each quantization index, for the quadratic-Gaussian case. Binning is a widely used tool in information theoretic proofs and methods, and to our knowledge, this is the first time it has been explicitly observed to emerge from data-driven learning.

1. Introduction

Consider a distributed sensor network consisting of individual cameras that independently capture images at different locations across the same city. Suppose that each sensor node compresses and transmits its highly correlated image to a joint central processing unit that reproduces a unified visual map of the city, by fusing the information collected by all of the nodes. If the sensors could directly communicate with each other in a cooperative manner, they could avoid some degree of redundancy by transmitting less correlated information. However, direct communication between nodes is often infeasible.

Given that, what is the best strategy to exploit the correlation between sensor data? Slepian & Wolf (1973) (SW) proved a remarkable and well-known information theoretic result that the distributed compression is asymptotically as efficient as the joint one, if the joint distribution statistics are known and compression is lossless. Their proof invokes *random binning* arguments and is non-constructive. Establishing a practical framework building onto these concepts is a challenging open problem to this day.

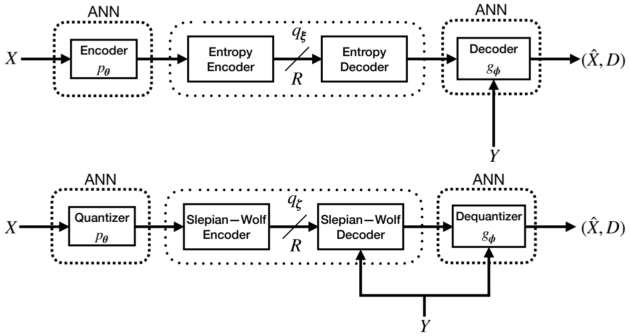

Figure 1. The two lossy compression systems that we consider: learned compressor using a classic entropy coder (**top**) and learned quantizer and dequantizer, using an ideal Slepian–Wolf coder (**bottom**).

Here, we investigate the setup characterized by Wyner & Ziv (1976) (WZ), which is both more general than SW as it encompasses lossy compression, and a simpler special case, as it assumes the decoder has access to a correlated source, the *side information*, losslessly. For WZ coding, there has been vast prior work considering synthetic setups and specific correlation patterns. Zamir et al. (2002) outlined the asymptotically optimal constructive mechanisms using nested linear and lattice codes for binary and Gaussian sources, respectively. Since then, the constructive and non-asymptotic research effort has been spearheaded by distributed source coding using syndromes (DISCUS) (Pradhan & Ramchandran, 2003), which formulated the WZ setup as a dual quantizer-channel coding problem. The complex interaction between the quantization, channel coding and estimation parts was also highlighted in competitive practical code design frameworks proposed in Liu et al. (2004); Yang et al. (2003). These methods achieve performances close to the theoretical bound, but are only applicable for Gaussian sources.

We propose to leverage the universal function approximation capability of artificial neural networks (ANNs) (Leshno et al., 1993; Hornik et al., 1989) to find constructive solutions for the non-asymptotic regime. More specifically, we consider the *one-shot* case, i.e., compressing each source

[1] Dept. of Electrical and Computer Engineering, New York University, NY 11201, USA. [2] Google Research New York, NY 10011, USA. Correspondence to: Ezgi Özyılkan <ezgi.ozyilkan@nyu.edu>, Johannes Ballé <jballe@google.com>.

realization one at a time, similarly to popular ANN-based compressors (e.g., Ballé et al., 2017). We provide two distinct solutions to the WZ problem, where we either handle the quantization and binning parts jointly or take a two-step approach by having a learned quantizer that is coupled with an ideal SW coder (see Fig. 1). We defer discussion of the entropy coders to later sections.

In order to establish the training objectives for these solutions, aiming for optimality, we minimize upper bounds on mutual information. These are expressed through one of the two probabilistic models utilizing ANNs (Section 2). Next, we explain how each probabilistic model is interpretable as one of the operational schemes shown in Fig. 1 (Sections 3 and 4, respectively). We discuss empirical results and connections to related work in Section 5.

2. Estimating Neural Upper Bounds on Wyner–Ziv

Since our choice of objective functions is inspired by the rate–distortion function of the case where side information is only available at the decoder, we briefly recap the WZ theorem and the accompanying information theoretic concepts. For the complete proof, refer to the original paper (i.e., Wyner & Ziv, 1976) and to Gamal & Kim (2012).

Theorem. *(Wyner–Ziv Theorem [1976]) Let (X, Y) be correlated sources, drawn i.i.d. $\sim p(x, y)$, and let $d(x, \hat{x})$ be a single-letter distortion measure. The rate–distortion function for X with side information Y available at the decoder side is as follows:*

$$R_{\mathrm{WZ}}(D) = \min(I(X;U) - I(Y;U)), \tag{1}$$

where the minimization operation is over all conditional probability distribution functions $p(u|x)$ and all functions $g(u, y)$ such that $\mathbb{E}_{p(x,y)p(u|x)} d(x, g(u, y)) \leq D$.

The achievability part of the WZ theorem invokes the covering lemma, resulting in the rate of $I(X;U)$, followed by a random binning argument based on joint typicality, which yields the rate discount of $I(Y;U)$ in Eq. (1) (Cover & Thomas, 2006). This achievability, which is shown to be tight, assumes a Markov chain constraint $U - X - Y$.

Assuming further that the encoder in the achievability proof is represented by a probability model $p_{\boldsymbol{\theta}}(u|x)$ with parameters $\boldsymbol{\theta}$, the difference of mutual informations in Eq. (1) can be written as:

$$I(X;U|Y) = \mathbb{E}_{\substack{p(x,y)\\ p_{\boldsymbol{\theta}}(u|x)}} \log \frac{p_{\boldsymbol{\theta}}(u|x)}{\cancel{p(u)}} \cdot \frac{\cancel{p(u)}}{p(u|y)}. \tag{2}$$

We will use the probabilistic model $p_{\boldsymbol{\theta}}(u|x)$ to facilitate the learning procedure of an encoder, and we set our encoder output as $u = \arg\max_v p_{\boldsymbol{\theta}}(v|x)$. To consider a practical compression setting, we also have U as discrete. For our objective functions, we choose one of two variational upper bounds:

$$I(X;U|Y) \leq \mathbb{E}_{\substack{p(x,y)\\ p_{\boldsymbol{\theta}}(u|x)}} \log \frac{p_{\boldsymbol{\theta}}(u|x)}{q_{\boldsymbol{\xi}}(u)}, \tag{3}$$

$$I(X;U|Y) \leq \mathbb{E}_{\substack{p(x,y)\\ p_{\boldsymbol{\theta}}(u|x)}} \log \frac{p_{\boldsymbol{\theta}}(u|x)}{q_{\boldsymbol{\zeta}}(u|y)}. \tag{4}$$

Here, $q_{\boldsymbol{\xi}}(u)$ and $q_{\boldsymbol{\zeta}}(u|y)$ (with parameters $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$, respectively), are two different models of the distribution $p(u|y)$, which is generally not known in closed form. We will discuss the operational meaning of these two variants in Sections 3 and 4. The upper bounds in Eqs. (3) and (4) follow from cross-entropy being larger or equal to entropy (Cover & Thomas, 2006).

We define all probabilistic models $p_{\boldsymbol{\theta}}(u|x)$, $q_{\boldsymbol{\xi}}(u)$ and $q_{\boldsymbol{\zeta}}(u|y)$, as discrete distributions with probabilities $P_k = \frac{\exp \alpha_k}{\sum_{i=1}^{K} \exp \alpha_i}$ for $k \in \{1, \ldots, K\}$, where K is a model parameter. The unnormalized log-probabilities (*logits*) α_k are computed by ANNs as functions of the conditioning variable (i.e., x for $p_{\boldsymbol{\theta}}(u|x)$ and y for $q_{\boldsymbol{\zeta}}(u|y)$), where the parameters represent the ANN weights, or treated as learnable parameters directly (for $q_{\boldsymbol{\xi}}(u)$). This choice keeps the parametric families as general as possible and does not unnecessarily impose any structure. Specifically, this allows the model $p_{\boldsymbol{\theta}}(u|x)$ to learn, if needed, quantization schemes that involve discontiguous bins, akin to the *random binning* operation in the achievability part of the WZ theorem, and resembling the systematic partitioning of the quantized source space with *cosets* in DISCUS, according to the virtual channel arising between the side information and the quantized source.

Next, we relax the constrained formulation of the WZ theorem to an unconstrained one using a Lagrange multiplier. This yields either a marginal or a conditional loss function:

$$L_{\mathrm{m}}(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\xi}) = \mathbb{E}\Big[\log \frac{p_{\boldsymbol{\theta}}(u|x)}{q_{\boldsymbol{\xi}}(u)} + \lambda d(x, g_{\boldsymbol{\phi}}(u, y))\Big], \tag{5}$$

$$L_{\mathrm{c}}(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\zeta}) = \mathbb{E}\Big[\log \frac{p_{\boldsymbol{\theta}}(u|x)}{q_{\boldsymbol{\zeta}}(u|y)} + \lambda d(x, g_{\boldsymbol{\phi}}(u, y))\Big], \tag{6}$$

where $\{\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\xi}, \boldsymbol{\zeta}\}$ are optimization parameters, and $g_{\boldsymbol{\phi}}(u, y)$ is the decoding function, also represented by an ANN with parameters $\boldsymbol{\phi}$, which outputs the reconstruction $\hat{x} = g_{\boldsymbol{\phi}}(u, y)$. The optimized $p_{\boldsymbol{\theta}}(u|x)$ and $g_{\boldsymbol{\phi}}(u, y)$ models yield the ANN-based encoder–decoder and quantizer–dequantizer components, respectively, depicted in Fig. 1.

We use a well-known technique, that is Gumbel-max (Gumbel, 1954), to draw samples from the discrete distributions. Moreover, we use Concrete distributions (Maddison et al., 2016) to facilitate stochastic optimization. To match the distribution of u samples, we also choose $q_{\boldsymbol{\xi}}(u)$ and $q_{\boldsymbol{\zeta}}(u|y)$ as Concrete during training.

2.1. Evaluation and Experimental Setup

The WZ formula in Eq. (1) has a closed-form expression only in a few special cases. To evaluate how close our neural bounds get to the known rate–distortion function, we consider the following correlation model: let X and Y be correlated, zero mean and stationary Gaussian memoryless sources, and let the distortion metric be mean-squared error. Then, the WZ rate–distortion function is

$$R_{\mathrm{WZ}}(D) = \frac{1}{2}\log\left(\frac{\sigma^2_{x|y}}{D}\right), \quad 0 \leq D \leq \sigma^2_{x|y}, \tag{7}$$

where $\sigma^2_{x|y}$ denotes the conditional variance of X given Y. For $X = Y + N$, where $N \sim \mathrm{N}(0, \sigma_n^2)$, which is considered throughout the paper except Fig. 3, we have $\sigma^2_{x|y} = \sigma^2_n$. The rate–distortion function for $Y = X + N$, considered in Fig. 3, can also be derived similarly. Note that in spite of considering Gaussian sources, we do *not* make any assumptions on the distribution of information sources in our formulations of the models.

For the conditional probabilistic models and the decoding function, we employ ANNs of three dense layers, with 100 units each, and leaky rectified linear units as activation functions for each of the layers. We use Adam (Kingma & Ba, 2014) and conduct our experiments using the JAX (Bradbury et al., 2018) framework. For evaluation, we switch from Concrete distributions back to their discrete counterparts, and use a deterministic encoding function that is equal to the mode of $p_{\boldsymbol{\theta}}(u|x)$, rather than sampling from it.

3. Operational Meaning and Evaluation of L_{m}

We first consider the system model at the top of the Fig. 1. Note that the upper bound in Eq. (3) corresponds to the rate of a system employing a one-shot encoder and an entropy code which asymptotically achieves a rate equal to the cross-entropy $\mathbb{E}_x\left[\mathbb{E}_{u\sim p_{\boldsymbol{\theta}}(u|x)}[-\log q_{\boldsymbol{\xi}}(u)]\right]$.

In Figs. 2 and 4 (in the appendix), we visualize the learned compressors obtained with this formulation. We remark that the learned compressors exhibit periodic grouping, binning-like behavior with respect to the source space, although no explicit structure was imposed onto the model architecture. Color coding of the bin indices reveals discontiguous quantization bins. This demonstrates that ANN-based methods are indeed capable of recovering very similar solutions to some of the handcrafted frameworks proposed for the WZ problem, such as DISCUS. Note that this behavior is also analogous to the random binning procedure in the achievability part of the WZ theorem.

These figures also show that the learned compressors exhibit optimal decoder behavior within each quantization index. In the given setup, the optimal decoder disambiguates the quantization index from the received bin index u, and reconstructs the source as (Zamir et al., 2014),

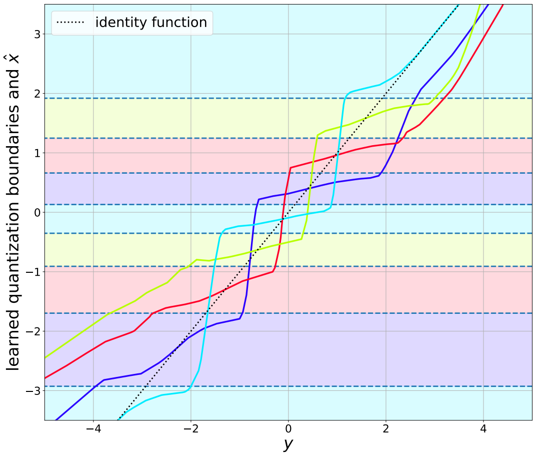

Figure 2. Visualization (best viewed in color) of the learned deterministic encoder $u = \arg\max_v p_{\boldsymbol{\theta}}(v|x)$ and decoder $\hat{x} = g_{\boldsymbol{\phi}}(u, y)$ of the marginal formulation (Eq. (5)), for the quadratic–Gaussian WZ setup. Here, we consider the correlation structure of $X = Y + N$ with $Y \sim N(0, 1)$ and N $\sim N(0, 10^{-1})$. The dashed horizontal lines are quantization boundaries, and the colors between boundaries represent unique values of u. We depict the decoding function as separate plots for each value of u, using the same color assignment.

$$\hat{x} = (1-\beta)\cdot y + \beta \cdot M(u), \text{ where } \beta \propto \sigma^2_n\,, \tag{8}$$

where $M(\cdot)$ denotes the disambiguation procedure. The slopes of the learned curves are also sensitive to σ^2_n, as is evident from comparing both Figs. 2 and 4 (in the appendix).

We explain the behavior of the learned encoder and decoder as follows. The encoder quantizes the source and subsequently bins the quantization index using the learned joint statistics of $Q(X)$ and Y, where $Q(\cdot)$ refers to the quantization, yielding u. Note that the encoder does *not* explicitly have access to the realization $Y = y$. The decoder then disambiguates the received bin index and deduces the quantization index, with the help of the side information. It subsequently estimates the source as $\hat{x}$, yielding the linear decoding functions within each quantization index with respect to the matching curve shown in Fig. 2.

As seen in Fig. 3 and in both panels of Fig. 5 (in the appendix), our learned compressors yield a better performance compared to the point-to-point rate–distortion functions. We argue that this is mainly due to the learned binning behavior, resulting in rate reduction. However, the compressors do not reach the asymptotic WZ rate–distortion bound provided in Eq. (7). As the ANN model compresses and consecutively bins each scalar input one by one, it is subjected both to the space-filling loss (Lookabaugh & Gray, 1989) during the quantization step, as well as to the loss coming from binning non-uniformly distributed quantization indices.

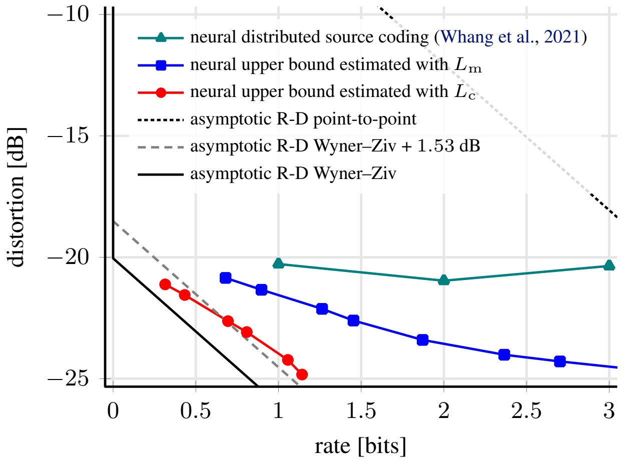

Figure 3. Rate–distortion (R-D) performances obtained with the marginal and conditional formulation, L_m and L_c, as in Eqs. (5) and (6) respectively. We consider quadratic-Gaussian WZ setup with the correlation structure of $Y = X + N$, having $X \sim \mathrm{N}(0, 1)$ and $N \sim \mathrm{N}(0, 10^{-2})$. The 1.53 dB distortion offset refers to the space-filling loss that the entropy-constrained one-shot lattice quantizer is subjected to in a high-rate regime (Gray & Neuhoff, 1998).

The achievability part of the WZ theorem, by comparison, considers binning of long sequences. This type of *compress–bin* (Gamal & Kim, 2012) is much more efficient than the one-shot case we consider, as it exploits the correlated side information in a better way.

4. Operational Meaning and Evaluation of L_{c}

We next consider the system model at the bottom of Fig. 1. The upper bound in Eq. (4) corresponds to the rate of a system employing a one-shot quantizer and an ideal SW entropy coder which asymptotically achieves the cross-entropy $\mathbb{E}_x\left[\mathbb{E}_{u\sim p_{\boldsymbol{\theta}}(u|x)}[-\log q_{\boldsymbol{\zeta}}(u|y)]\right]$.

The experimental results are provided in Fig. 3 and in both panels of Fig. 5 (in the appendix). We observe that unlike the previous case, this model's performance is closer to the asymptotic WZ rate–distortion bound. We find no evidence of binning occurring in these quantizers (not depicted). We explain the improved rate–distortion performance of this model as follows. When binning is left to the ideal SW code, which may make use of a high dimensional channel code (e.g., as in DISCUS), the performance loss of such a learned Wyner–Ziv compressor only comes from the quantization part alone. This line of reasoning was also followed by the practical code design in (Yang et al., 2003). The authors make use of a combination of a classic quantizer (without binning) and a powerful SW coding scheme, implemented with irregular low-density parity-check (LDPC) codes, in order to achieve the theoretical limit of $H(Q(X)|Y)$, where $Q(X)$ refers to the quantized source. Hence, minimizing L_{c} corresponds to learning one-shot quantizer and dequantizer components, reducing the WZ problem to a SW problem in a data-driven fashion.

5. Discussion

Our experiments yield interesting data-driven insights about the nature of a classical source coding problem with side information. Figs. 2 and 4 (in the appendix) provide the first explicit evidence of ANN-based learned compressors recovering some elements of the optimal theoretical WZ solution, both through binning with respect to the source space, and piecewise linear behavior of the decoding function.

We linked two neural upper bounds (Section 2) with two corresponding operational schemes (Sections 3 and 4) by picking a suitable entropy coding technique for each one. In the case of the marginal formulation in Eq. (5), it is attainable with high-order classic entropy coding, operating on discrete values (Rissanen & Langdon, 1981). Considering the conditional formulation in Eq. (6), we make use of an ideal SW coding scheme, which compresses sufficiently large blocks of quantized source elements to the rate of $H(Q(X)|Y)$. The role of SW coding is to additionally exploit the correlation between $Q(X)$ and Y. This explains our empirical finding that in this case, there is no binning observed in the quantization (as SW coding takes care of this). State-of-the-art channel coding schemes such as LDPC (e.g., Liu et al., 2004) and turbo codes (e.g., Aaron & Girod, 2002) have been demonstrated to yield results coming close to the theoretical SW bound. To be fair, in order to achieve optimality, these schemes make certain assumptions about the virtual channel, which might not be met in our case.

Previous work (Whang et al., 2021) investigated the construction of neural WZ schemes using a loss function based on VQ-VAE (van den Oord et al., 2017). We note that both of our methods outperform this scheme (see Fig. 3). We attribute the suboptimal performance of the scheme to the lack of explicit accounting for entropy in the learning objective. Notable prior work on the machine learning side include Alemi et al. (2017); Fischer (2020), which are related to the *information bottleneck* problem (Tishby et al., 2000). The learning objectives are comparable to our marginal and conditional formulation, respectively. However, both of these are strictly concerned with probabilistic model fitting, not with operational compression schemes.

Going forward, by actually implementing the two aforementioned entropy coding techniques, and reporting the actual bit rates, we hope to demonstrate the feasibility of our neural schemes as a complete constructive end-to-end solution to the WZ problem. In the case of SW coding, learned channel coding techniques (Kim et al., 2020; 2018) could be investigated, to relax the assumptions about the virtual channel arising between the quantized source and side information.

References

Aaron, A. and Girod, B. Compression with side information using turbo codes. In *Proceedings DCC 2002. Data Compression Conference*, pp. 252–261, 2002. doi: 10.1109/DCC.2002.999963.

Alemi, A. A., Fischer, I., Dillon, J. V., and Murphy, K. Deep variational information bottleneck. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=HyxQzBceg.

Ballé, J., Laparra, V., and Simoncelli, E. P. End-to-end optimized image compression. In *International Conference on Learning Representations (ICLR)*, 2017.

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., and Zhang, Q. JAX: Composable transformations of Python+NumPy programs, 2018. URL http://github.com/google/jax.

Cover, T. M. and Thomas, J. A. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, USA, 2006. ISBN 0471241954.

Fischer, I. The conditional entropy bottleneck, 2020. URL https://arxiv.org/abs/2002.05379.

Gamal, A. E. and Kim, Y.-H. *Network Information Theory*. Cambridge University Press, USA, 2012. ISBN 1107008735.

Gray, R. and Neuhoff, D. Quantization. *IEEE Transactions on Information Theory*, 44(6):2325–2383, 1998. doi: 10.1109/18.720541.

Gumbel, E. J. Statistical theory of extreme values and some practical applications : A series of lectures, 1954.

Hornik, K., Stinchcombe, M., and White, H. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, jul 1989. ISSN 0893-6080.

Kim, H., Jiang, Y., Rana, R., Kannan, S., Oh, S., and Viswanath, P. Communication algorithms via deep learning. In *6th Int. Conf. on Learning Representations (ICLR)*, 2018.

Kim, H., Oh, S., and Viswanath, P. Physical layer communication via deep learning. *IEEE Journal on Selected Areas in Information Theory*, 1(1):5–18, 2020. doi: 10.1109/JSAIT.2020.2991562.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization, 2014. URL https://arxiv.org/abs/1412.6980.

Leshno, M., Lin, V. Y., Pinkus, A., and Schocken, S. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6):861–867, January 1993. doi: 10.1016/s0893-6080(05)80131-5. URL https://doi.org/10.1016/s0893-6080(05)80131-5.

Liu, Z., Cheng, S., Liveris, A., and Xiong, Z. Slepian-Wolf coded nested quantization (SWC-NQ) for Wyner-Ziv coding: performance analysis and code design. In *Data Compression Conference, 2004. Proceedings. DCC 2004*, pp. 322–331, 2004. doi: 10.1109/DCC.2004.1281477.

Lookabaugh, T. and Gray, R. High-resolution quantization theory and the vector quantizer advantage. *IEEE Transactions on Information Theory*, 35(5):1020–1033, 1989. doi: 10.1109/18.42217.

Maddison, C. J., Mnih, A., and Teh, Y. W. The concrete distribution: A continuous relaxation of discrete random variables, 2016. URL https://arxiv.org/abs/1611.00712.

van den Oord, A., Vinyals, O., and Kavukcuoglu, K. Neural discrete representation learning, 2017. URL https://arxiv.org/abs/1711.00937.

Pradhan, S. and Ramchandran, K. Distributed source coding using syndromes (DISCUS): design and construction. *IEEE Transactions on Information Theory*, 49(3):626–643, 2003. doi: 10.1109/TIT.2002.808103.

Rissanen, J. and Langdon, G. Universal modeling and coding. *IEEE Transactions on Information Theory*, 27 (1):12–23, 1981.

Slepian, D. and Wolf, J. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19(4):471 – 480, 1973.

Tishby, N., Pereira, F. C., and Bialek, W. The information bottleneck method, 2000. URL https://arxiv.org/abs/physics/0004057.

Whang, J., Acharya, A., Kim, H., and Dimakis, A. G. Neural distributed source coding, 2021. URL https://arxiv.org/abs/2106.02797.

Wyner, A. and Ziv, J. The rate–distortion function for source coding with side information at the decoder. *IEEE Transactions on Information Theory*, 22(1):1 – 10, 1976.

Yang, Y., Cheng, S., Xiong, Z., and Zhao, W. Wyner-Ziv coding based on TCQ and LDPC codes. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 1, pp. 825–829 Vol.1, 2003. doi: 10.1109/ACSSC.2003.1292028.

Zamir, R., Shamai, S., and Erez, U. Nested linear/lattice codes for structured multiterminal binning. *IEEE Transactions on Information Theory*, 48(6):1250–1276, 2002. doi: 10.1109/TIT.2002.1003821.

Zamir, R., Nazer, B., Kochman, Y., and Bistritz, I. *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation and Multiuser Information Theory*. Cambridge University Press, 2014. doi: 10.1017/CBO9781139045520.

A. Appendix.

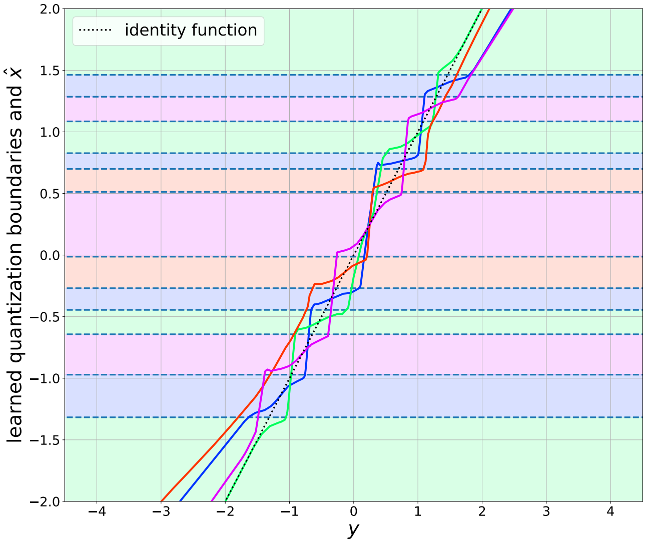

Figure 4. Visualization (best viewed in color) of the learned deterministic encoder $u = \arg\max_v p_{\boldsymbol{\theta}}(v|x)$ and decoder $\hat{x} = g_{\boldsymbol{\phi}}(u, y)$ of the marginal formulation (Eq. (5)), for the quadratic–Gaussian WZ setup. Here, we consider the correlation structure of $X = Y + N$ with $Y \sim N(0, 1)$ and N $\sim N(0, 10^{-2})$.

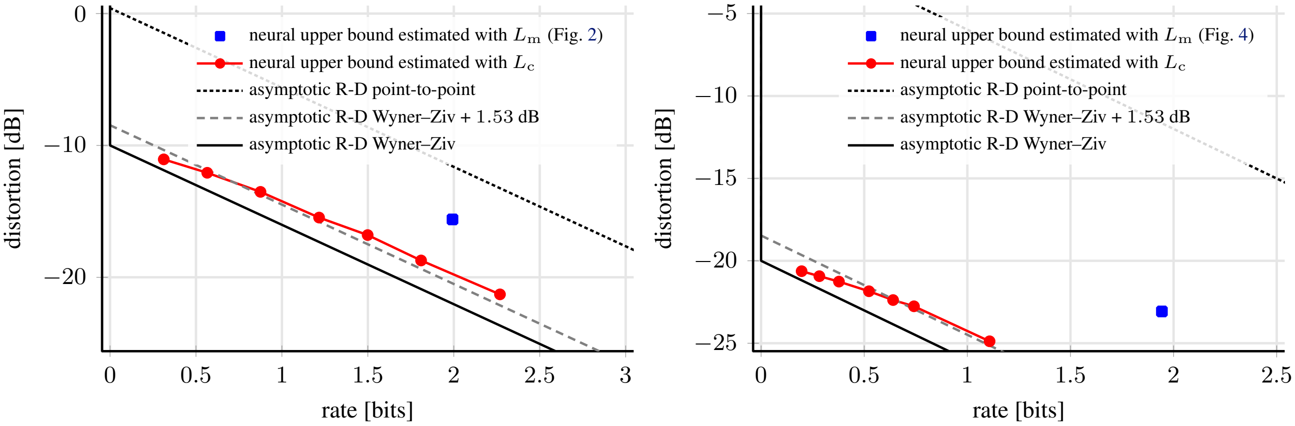

Figure 5. Rate–distortion (R-D) performances obtained with marginal and conditional formulations, as in Eqs. (5) and (6), respectively. We consider the quadratic-Gaussian WZ setup, having correlation structure of $X = Y + N$, and plot the empirical results versus the asymptotic bounds. On the **left** panel, we have $Y \sim \mathrm{N}(0, 1)$ and $N \sim \mathrm{N}(0, 10^{-1})$. On the **right** panel, we have $Y \sim \mathrm{N}(0, 1)$ and $N \sim \mathrm{N}(0, 10^{-2})$. The 1.53 dB distortion offset refers to the space-filling loss that the entropy-constrained one-shot lattice quantizer is subjected to in a high-rate regime (Gray & Neuhoff, 1998).