Low-Delay Analog Distributed Joint Source-Channel Coding using SIRENs

Ziwei Xuan and Krishna Narayanan
Deptartment of Electrical and Computer Engineering
Texas A&M University, College Station, USA
Email: {xuan64, krn}@tamu.edu

***Abstract*—We consider the design of joint source-channel coding (JSCC) schemes for two multiterminal source-channel coding problems - namely, the transmission of a Gaussian source with side information at the receiver and the transmission of bivariate Gaussian sources over two independent Gaussian channels. Our focus is on low-delay transmission. We formulate the design problem as optimization of an autoencoder (AE) model, and show that sinusoidal representation networks (SIRENs) are a good choice due to their inherent periodicity and the ability of stretched sinusiods to cover the source space. We show that SIRENs outperform parametric ReLU based networks. The complexity of the proposed method scales better with source dimension than the best traditional schemes known in the literature while their performance is comparable to or better than that of the best traditional schemes. We demonstrate that the spontaneously learned encoder mappings share resemblance to the classical Wyner-Ziv mappings for JSCC with side information, and exhibits structured patterns in the case of distributed coding that are interpretable.**

***Index Terms*—Distributed source-channel coding, joint source-channel coding, SIRENs, deep learning**

I. Introduction

We consider the problem of designing source-channel codes for two problems in multiterminal information theory - (1) transmission of a Gaussian source over a Gaussian channel with side information at the receiver (joint source-channel coding with side information or the JSCCSI problem); (2) transmitting correlated sources over orthogonal channels to be recovered by a central decoder (distributed joint source-channel coding or the DJSCC problem). Unlike in the single user case, separate source and channel coding is not optimal for JSCCSI and DJSCC in general, and more intricate joint source-channel coding (JSCC) schemes need to be constructed. In this paper, we consider the design of such JSCC schemes for the low-delay case (i.e. sources have small dimensions).

Construction of codes for these problems has been considered in many papers in the literature. For the zero-delay JSCCSI problem, a hybrid digital-analog (HDA) coding based method has been proposed in [1], and several reconstruction approaches have been discussed. For the zero-delay DJSCC problem, a Shannon-Kotel'nikov (S-K) mapping based encoder has been designed in [2], and comparison with other schemes including sawtooth mappings and digital quantization has been presented. Parametric mappings, including piece-wise linear mappings, spirals, sinusoids, and nested hexagons have been designed for both problems [3]–[6]. Optimization of non-parametric mappings based on noisy channel relaxation (NCR) has been implemented for both problems by [7], and it has been improved using deterministic annealing (DA) in [8].

Recently, there has been a lot of interest in solving communication problems based on deep learning methods, especially in the design of JSCC for single-user systems (e.g. [9]–[11]). Inspired by these successes, we revisit the design of codes for the JSCCSI and DJSCC problems through the perspective of deep learning. To the best of our knowledge, deep learning based design of codes for the two aforementioned problems have not been considered in the literature in the past.

We model the entire system with an autoencoder (AE), and we propose the use of encoder and decoder that are represented with multiple layers of sinusoidal representation networks (SIRENs). For the case of JSCCSI with bandwidth reduction, we construct a neural nested lattice structure. The highlights of our proposed scheme are - (i) for some parameters, they outperform the best known traditional schemes in terms of mean squared error, (ii) they have lower decoding complexity than the best performing traditional baselines, (iii) we show that SIRENs result in structured mappings that are well suited for the JSCC problems considered here and that they outperform networks without such structure, e.g. parametric ReLU networks, (iv) our training procedure spontaneously learns structured encoder mappings that are interpretable,

II. Problem Formulation

A. Case I: JSCCSI

The block diagram of the JSCCSI problem is presented in Fig. 1(a). We wish to transmit a source $\mathbf{u_1} \in \mathbb{R}^{k_1}$, after encoding, through an additive white Gaussian noise (AWGN) channel to a receiver. The receiver also has side information $\mathbf{u_2} \in \mathbb{R}^{k_2}$ that is correlated with $\mathbf{u_1}$. We assume that $\{(\mathbf{u_1}, \mathbf{u_2})\}$ are independently and identically generated along discrete time. $\mathbf{u_1}$ is encoded by an encoder function $f_\phi(\cdot) : \mathbb{R}^{k_1} \to \mathbb{R}^{n_1}$, parameterized by ϕ and we denote the transmitted channel codeword as $\mathbf{x_1} = f_\phi(\mathbf{u_1})$. The channel codeword satisfies a power constraint given by $P_\phi = \frac{1}{n_1}\mathbb{E}[||\mathbf{x_1}||]^2] \leq P_T$, and without loss generality, it is assumed $P_T = 1$. The noisy signal observed at the receiver is denoted as $\mathbf{y_1} = \mathbf{x_1} + \mathbf{n_1}$, where $\mathbf{n_1} \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathbf{n}}^2\mathbf{I})$. The decoder function, $g_\psi(\cdot) : \mathbb{R}^{n_1} \times \mathbb{R}^{k_2} \to \mathbb{R}^{k_1}$, transforms the

This work was funded by the National Science Foundation under grant CCF-1718886.

received signal into the estimate $\hat{\mathbf{u}}_1 = g_\psi(\mathbf{y}_1, \mathbf{u}_2)$ with the help of side information.

The distortion between $\mathbf{u}_1$ and $\hat{\mathbf{u}}_1$ is measured by mean-squared-error (MSE) and denoted by D_1, i.e. $D_1 = \frac{1}{k_1}\mathbb{E}[||\mathbf{u}_1 - \hat{\mathbf{u}}_1||_2^2]$.

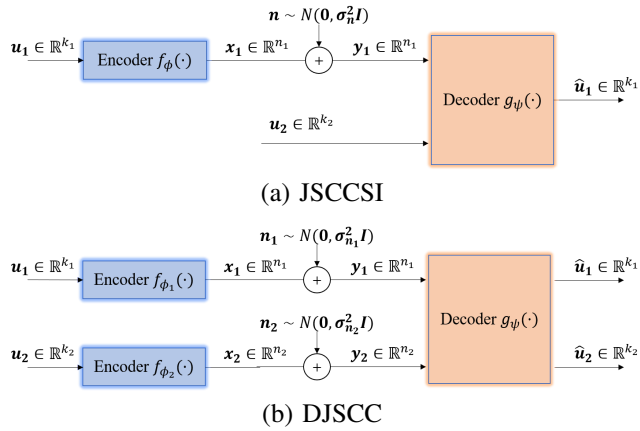


Fig. 1: System models for the JSCCSI and DJSCC problems.

B. Case II: DJSCC

We show the settings for the DJSCC problem in Fig. 1(b). We assume a pair of physically separated but correlated sources, denoted as $\mathbf{u_i} \in \mathbb{R}^{k_i}$, for $i = 1, 2$, which are generated independently and identically along discrete time. The sources are processed by distinct encoders $f_{\phi_i}(\cdot) : \mathbb{R}^{k_i} \to \mathbb{R}^{n_i}$, which are parameterized by ϕ_i. The corresponding channel codewords are denoted as $\mathbf{x_i} = f_{\phi_i}(\mathbf{u_i})$. We denote the power of each encoder as $P_{\phi_i} = \frac{1}{n_i}\mathbb{E}[||\mathbf{x_i}||]^2]$. Two types of power constraints are considered here - total power constraint, for which $P_{\phi_1} + P_{\phi_2} \leq 2P_T$, and individual power constraints, for which $P_{\phi_1} \leq P_T$ and $P_{\phi_2} \leq P_T$. The codewords are transmitted over two orthogonal AWGN channels and the receiver observes $\mathbf{y_i} = \mathbf{x_i} + \mathbf{n_i}$, where $\mathbf{n_i}$ is a zero-mean Gaussian random variable with variance $\sigma^2_{\mathbf{n_i}}$. The common decoder, $g_\psi(\cdot) : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}^{k_1} \times \mathbb{R}^{k_2}$, has access to both received signals, and forms estimates $\hat{\mathbf{u}}_\mathbf{i}$ for $i = 1, 2$, respectively. We define the distortion between estimates and sources with equal weights as MSE and denoted it by D. Hence, $D = D_1 + D_2 = \frac{1}{k_1}\mathbb{E}[||\mathbf{u_1} - \hat{\mathbf{u}}_1||_2^2] + \frac{1}{k_2}\mathbb{E}[||\mathbf{u_2} - \hat{\mathbf{u}}_2||_2^2]$.

C. Optimum Performance Theoretically Attainable

Optimum performance theoretically attainable (OPTA) depicts the minimum distortion achievable at infinite delay, and we include the OPTA in our baseline. From here, we specifically focus on using bivariate Gaussian sources as the source information pair, for which the OPTA is well-known. We consider the case that $(\mathbf{u_1}, \mathbf{u_2}) \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathbf{u_1},\mathbf{u_2}})$, with $\mathbf{\Sigma}_{\mathbf{u_1},\mathbf{u_2}} = \sigma^2_\mathbf{u}\begin{bmatrix} \mathbf{I} & \rho\mathbf{I} \\ \rho\mathbf{I} & \mathbf{I} \end{bmatrix}$, where $\sigma^2_\mathbf{u}$ is the source variance and ρ is the correlation coefficient.

For JSCCSI, OPTA can be computed using the Wyner-Ziv rate distortion function [12]. We describe channel condition using the channel signal-to-noise ratio (CSNR) as $\text{CSNR} := 10\log_{10}\left(\frac{P_\phi}{\sigma^2_\mathbf{n}}\right)$ (in dB), and distortion level using signal-to-distortion ratio (SDR) as $\text{SDR} := 10\log_{10}\left(\frac{\sigma^2_\mathbf{u}}{D_1}\right)$ (in dB).

The OPTA for distributed coding can be numerically calculated according to results in [13]. In this case, we define $\text{CSNR} := 10\log_{10}\left(\frac{P_{\phi,1}+P_{\phi,2}}{2\sigma^2_\mathbf{n}}\right)$ (in dB), and $\text{SDR} := 10\log_{10}\left(\frac{2\sigma^2_\mathbf{u}}{D_1+D_2}\right)$ (in dB), respectively.

III. Proposed Method

From this section, we focus on the case of $k_1 = k_2$ and $n_1 = n_2$, and denote $k_1 = n_1$ as equal BW, $k_1 < n_1$ as BW expansion, and $k_1 > n_1$ as BW reduction.

A. Autoencoder

We adopt the AE model in our method, represent the encoder and decoder with deep neural networks, and set AWGN channel as non-trainable layer in between.

B. Loss

For the problem of JSCCSI, we set the loss function as the corresponding Lagrangian cost associated with power constraint, $J_{\phi,\psi} = D_1 + \lambda P_\phi$. For the case of DJSCC, we consider two kinds of power constraints described previously. To enforce a total power constraint, we set the optimization objective as the Lagrangian cost, $J_{\phi_1,\phi_2,\psi} = D + \lambda(P_{\phi_1} + P_{\phi_2})$. To realize individual power constraints, we set normalization layers in between the encoder networks and the channels. In this case, the MSE serves as the loss function that we optimize in an end-to-end manner. Note in the Lagrangian cost, λ is the slope of the optimal distortion (denoted as D^*)-power curve $\lambda = -\frac{\mathrm{dD}^*}{\mathrm{dP_T}}$ [14], but nevertheless, we don't know the exact curve for optimal distortion versus power at finite delay in general. In experiments, we empirically select λ by taking as reference the distortion curves of best traditional schemes as we know, and this is effective for all scenarios we considered.

C. SIREN

We design the network structure for AE based on the following observations.

First, periodicity is a universal characteristic of the encoder mappings in literature for both JSCCSI and DJSCC problems (e.g., [1], [4], [6], [8]. In the Wyner-Ziv structure, it is usually postulated that the presence of the side information assists at locating the interval where the source is encoded into by a periodic transformation. This acts as the channel decoding to neutralize the corruption from the channel noise and distortion from the source encoder. For the DJSCC problem, the pair of received signals acts as side information to each other in a similar way to the JSCCSI problem. Therefore, periodicity is a desirable property of our network.

Secondly, in [6], a Cramer-Rao lower bound on the distortion is derived as given by, $D_1 \geq \frac{1}{\frac{1}{\sigma^2_\mathbf{n}}\mathbb{E}[||f'_\phi(\mathbf{u_1})||^2] + \frac{1}{(1-\rho^2)\sigma^2_{\mathbf{u_1}}}}$. This bound reveals that the effect of the channel can be suppressed by designing the encoder mapping with higher stretching factor $||f'_\phi(\mathbf{u_1})||^2$. A similar observation of this relationship was made in [2] for the DJSCC problem. On the other hand, too small distance between different folds of the encoder mapping, which is stretched by twisting and bending, hinders decoding the received signal into the correct fold. A proper encoding

mapping should strike the balance between the two concerns. Specifically for JSCCSI problem with BW expansion, [6] uses parametric sinusoids as the encoder mapping. This fact inspires us to design neural network with sinusoidal activation functions, which has high flexibility to twist and bend, and let it spontaneously learn this balance through training.

In addition, recently the work in [15] showed that network with sine as activation function can be trained in a stable manner, and the learned smooth latent representations are reminiscent of spiral-based encoding mappings adopted in [2], [4]. Meanwhile, sine functions with very low frequency can approximate piece-wise linear mappings, which were studied in [2] or have resemblance to the learned mapping in [8]. This guarantees that networks with sine activation functions can learn diverse mappings.

Therefore, we represent encoder and decoder networks in the AE with SIRENs as in [15].

$$\mathbf{\Psi}(\mathbf{x}) = \mathbf{W}_l(\psi_{l-1} \circ \psi_{l-2} \circ \cdots \circ \psi_0)(\mathbf{x}) + \mathbf{b}_n,$$
$$\mathbf{z_i} \mapsto \psi_i(\mathbf{z_i}) = \sin(\mathbf{W}_i \mathbf{x}_i + \mathbf{b_i})$$

where $\psi_i : \mathbb{R}^{M_i} \mapsto \mathbb{R}^{N_i}$ is the i-th layer of the network, with M_i and N_i as the corresponding input and the output dimensions. $\mathbf{W}_i \in \mathbb{R}^{N_i \times M_i}$ and $\mathbf{b}_i$ are the learnable weight matrix and biases of the i-th layer network.

For the case of JSCCSI with equal BW and BW expansion, we build the encoder (decoder) using 3 (4) layers of SIRENs, and each layer has 256 hidden units; we concatenate $\mathbf{y_1}$ and $\mathbf{u_2}$ and decode them together. For the case of DJSCC, we construct each encoder using 3 layers of SIRENs with 200 hidden units per layer at the transmitter; at the receiver, we concatenate $\mathbf{y_1}$ and $\mathbf{y_2}$ and decode them together using 4 layers of SIRENs with 300 hidden units per layer. The model is trained end-to-end.

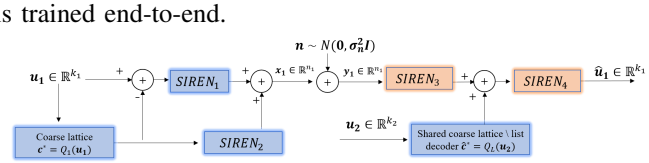


Fig. 2: Model for the JSCCSI with BW reduction.

For the case of JSCCSI with BW reduction, we consider a nested framework [16] as shown in Fig. 2. The coarse lattice $Q(\cdot)$, which acts as the channel code, is learned based on k-nearest neighbors rule at the transmitter, and trained separately from other parts. Two candidate methods are considered: Voronoi cell based method learns the codewords as a set of general Voronoi cell centers in the source space; or hexagon based method learns the codewords as centers of regular hexagons, and the side length is the learnable variable. The coarse lattice center $\mathbf{c}^*$ is picked as the one with minimal Euclidean distance to $\mathbf{u_1}$, and in practice we approximate it using $\mathbf{c}^* = \sum_j \frac{\exp(-||\mathbf{u_1}-\mathbf{c}_j||/t)}{\sum_l \exp(-||\mathbf{u_1}-\mathbf{c}_l||/t)} \mathbf{c}_j$ as in [17], with t being the temperature parameter controlling how close this $softmax$ is to a Dirac Delta function. Since the residual $\mathbf{r} = \mathbf{u_1} - \mathbf{c}^*$ will have a different distribution than $\mathbf{u_1}$, we integrate information from both the residual and the selected center and set $\mathbf{x_1} = \text{SIREN}_1(\mathbf{r}) + \text{SIREN}_2(\mathbf{c}^*)$. The received signal is first decoded into estimates of residuals as $\hat{\mathbf{r}}$ by SIREN_3. The coarse lattice is shared with the receiver, and we use a list decoder to select $L > 1$ candidate codewords $\hat{\mathbf{c}_\mathbf{i}^*}$ that are closest to the given side information $\mathbf{u_2}$ in terms of Euclidean distance. The list decoder provides robustness against the channel noise. For each candidate codeword in the list $\mathbf{c}_\mathbf{i}^*$, we produce an estimate of $\mathbf{u_1}$ according to $\hat{\mathbf{c}_\mathbf{i}^*} + \hat{\mathbf{r}}$. The L estimates are concatenated and input to SIREN_4 which produces the final estimate of $\mathbf{u_1}$.

We intialize the SIRENs network according to suggestions in [15]. At each iteration, we generate sources of batch size 1.024×10^5, and train the model with Adam optimizer [18]. We decay the learning rate when the loss performance doesn't improve, and continue training for up to 26000 epochs for high CSNRs. For the model using an individual power constraint, we keep the weights associated with best validation performance. For the model using a total power constraint, we calculate the true CSNR in validation, evaluate the gap to the OPTA under that CSNR, and store the weights of the model that minimizes the gap to the OPTA.

D. Complexity

For our proposed schemes, the model is trained offline and hence, the training complexity is not a major concern. During testing time, for each pair of received signal and side information, the computational complexity of the proposed method grows linearly with (k, n) - more specifically, it is $\mathcal{O}(Ld_h^2 + kd_h + nd_h)$ where d_h and L are the number of hidden units and the number of hidden layers, respectively. On the contrary, all the traditional baselines (except the one from [2]) that we consider adopt an MMSE decoder for best decoding performance. The MMSE decoder requires the computation of multi-dimensional integrals which have to be numerically approximated in practice. For the JSCCSI problem, the receiver can either a) perform Monte Carlo simulation for each pair of received signal and side information, or b) build a table of decoder mapping offline and decode using the lookup table. Both approaches require quantization of $\mathbf{u_1}$ to a grid of values and the grid precision has to scale proportional to $\sqrt{D^*}$. Thus, for the former method, the computation complexity is roughly dominated by $\mathcal{O}(c^k(1 + \text{CSNR})^{\frac{n}{2}} n)$ with constant c, which grows exponentially with (k, n); for the latter method, the memory requirement will also increase exponentially with (k, n), which is not affordable. Note that while [2] uses an decoder whose complexity is lower than that of the MMSE decoder, its performance is worse. Therefore, the use of SIRENs results in simple structures that are easy to train, have lower complexity than the baselines and are flexible to extend to different source and channel dimensions (for e.g. for BW $k_1 : n_1 = 2 : 3$, it is not clear how to design the corresponding encoder mapping using traditional methods).

IV. Experimental Results

We first present results for JSCCSI with BW 1:1. We train our model using jointly Gaussian sources with unit variance and correlation coefficient $\rho = 0.9$ and $\rho = 0.99$ for different

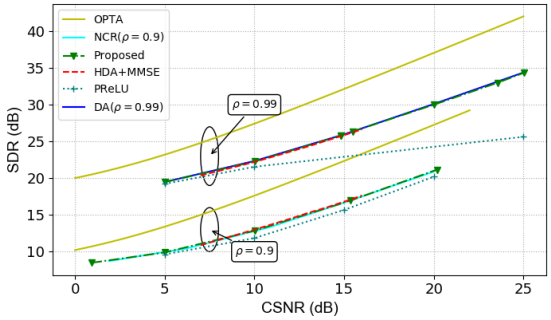


Fig. 3: Plot of SDR versus CSNR for JSCCSI, BW 1:1.

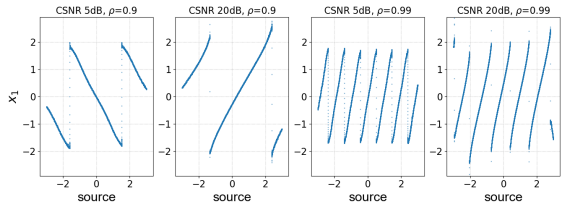
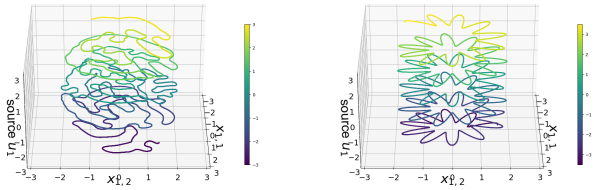


Fig. 4: Learned encoder mappings for JSCCSI, BW 1:1.

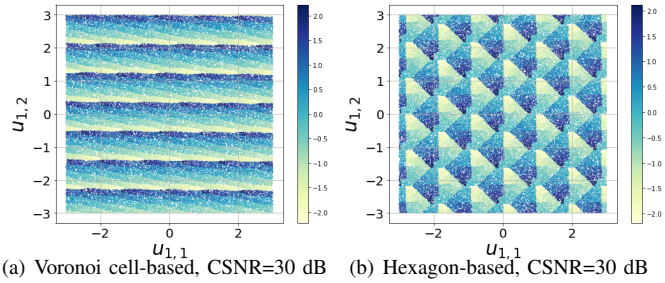
CSNRs. Several baselines are considered based on the best results available in the literature for different parameters. For both $\rho = 0.9$ and $\rho = 0.99$, the first baseline is the HDA-based scheme by [1] with an MMSE estimator at the receiver. We provide comparison with the NCR technique-based encoder and decoder in [7] for $\rho = 0.9$. For $\rho = 0.99$, we compare our results to those of the DA method in [8]. A plot of SDR versus CSNR is shown in Fig. 3 for these schemes and our proposed scheme. It can be seen that the proposed method performs on par with other three methods over the entire CSNR range. Note that our baselines use the computationally complex MMSE decoders, and sub-optimal decoders with lower complexity will result in worse performances. We also consider a deep learning based baseline which replaces the activation function as parametric ReLU (PReLU), and keeps rest of the structure the same. The network with PReLU failed to learn a feasible scheme, and the SDR did not improve with CSNR.



(a) Proposed scheme, CSNR=20 dB (b) Encoder mapping in [6]

Fig. 5: Encoder mappings for JSCCSI with BW 1:2.

We also plot the learned encoder mappings for our proposed scheme and examine the impact of ρ and CSNR in Fig. 4. Firstly, the transmitted signal as a function of the source symbol varies periodically in a many-to-one modulo pattern analogous to an effective digital Wyner-Ziv mapping. We further observe that for the same CSNR, increasing the correlation coefficient ρ leads to a denser partition of the source-



(a) Voronoi cell-based, CSNR=30 dB (b) Hexagon-based, CSNR=30 dB

Fig. 6: Encoder mappings for JSCCSI with BW 2:1.

space. Both these properties are intuitively pleasing. A larger correlation coefficient ρ leads to more common information shared between the source and the side information, and thus the decoder can rely on the side information to a larger extent so as to neutralize the impairment from the channel noise. This is how the 'period' of the mapping is primarily determined. When the correlation coefficient is fixed, we observe that the learned mapping is approximately linear in each period. As the CSNR increases, the corresponding period is slightly larger. Also, the mappings are amplified and attempt to fill the space in a twisted way. This phenomenon results from the trade-off between the goals of minimizing distortion, for which more amplified mappings are preferred at high CSNRs, and satisfying the power constraint, for which the curve is twisted.

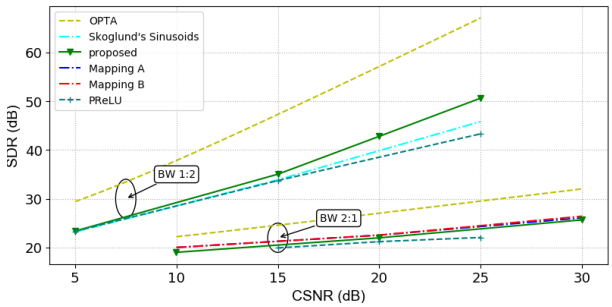
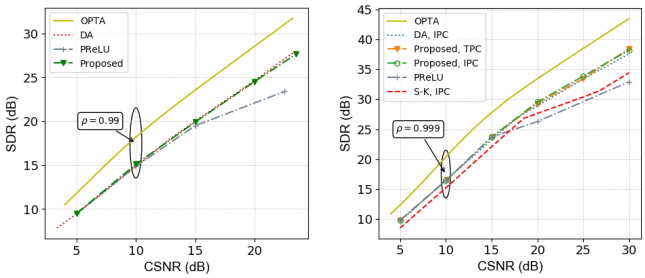


Fig. 7: Plot of SDR vs. CSNR for JSCCSI, BW 1:2 and 2:1.

Next, we consider the case of BW 1:2. Let $x_{1,1}$ and $x_{1,2}$ represent the first and second components of the transmitted signal $\mathbf{x_1}$. The mapping learned by the proposed scheme is shown in Fig. 5(a). In the channel codeword plane, a family of similar space filling curves are learned ; meanwhile, along the axis of the source, such curves are repeated in a modulo manner. The overall mapping is therefore shaped by both effects. Compared with parametric sinusoids designed in [6], the proposed mapping has a higher extent of stretching, and leaves tiny phase difference between the learned folds, which reveals a better balance between source coding and error correction. Fig. 7 shows the learned scheme has a gain over the baseline at high CSNR region.

Fig. 6 shows the encoder mappings for BW 2:1 case. Both the Voronoi cell and hexagon based schemes are capable of learning the structure of a nested lattice, where the coarse lattice is repeated periodically over the source space, and the fine lattices work as source coding within each coarse lattice. Also Fig. 7 reveals that our proposed Voronoi cell based scheme performs closely (slightly worse) to mapping A and B in [5],

which nest hexagons with spiral and discontinuous curve based mappings, and is better than the PReLU based scheme, which uses the similar structure as PReLU based scheme in the BW expansion case.



(a) $\rho = 0.99$ (b) $\rho = 0.999$

Fig. 8: Plot of SDR versus CSNR for DJSCC, BW 1:1.

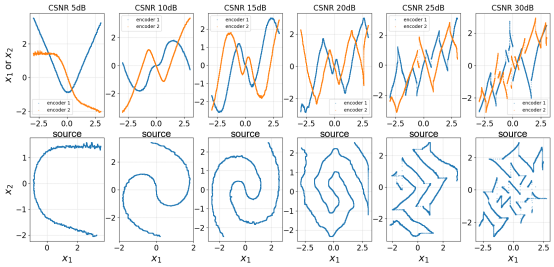


Fig. 9: Learned encoder mappings for DJSCC with $\rho = 0.999$, BW 1:1, under total power constraint.

Next, we consider the case of DJSCC. We first implement the total power constraint and compare the performance of the proposed scheme with that of the DA-based method in [8] in Fig. 8(a) for $\rho = 0.99$. Over the entire CSNR range, the two methods perform comparably. Fig. 8(b) shows a comparison between DA-based method [8], S-K mappings [2], and the proposed method under individual power constraint (IPC) when $\rho = 0.999$. We also present the performance of the proposed method under a total power constraint (TPC). Fig. 8(b) illustrates that the proposed method is at least as good as the traditional schemes. It also shows that the freedom of allocating power to each encoder does not necessarily provide a clear gain over the scheme under individual power constraint.

Meanwhile, the learned encoder mappings are different from the ones in [8]. As shown in Fig. 9, the top row represents the two encoder mappings respectively, and the bottom row exhibits how the channel space is filled when the same source is input to both encoder, i.e. $\rho = 1$. The top row reveals that each source serves as the side information to another (especially one encoding function is likely to exhibit modulo pattern over the intervals where the other encoder function is linear). As the CSNR increases, the learned mappings are more wiggly and closer to being piece-wise linear. The bottom row demonstrates that the channel space is filled in a pattern similar to spirals. When the CSNR increases, the distance between arms decreases and the spirals are approximated by piece-wise linear encoding functions. Further, in experiments, such patterns are also found in the learned mappings under individual power constraints.

V. Conclusion

In this work, we revisited the problem of transmitting correlated sources in two different network settings, using SIRENs based models. The experimental results demonstrated a comparable or better performance than best traditional schemes as we know with lower computational complexity. In the future, we will investigate sources with longer block-length and more practical channels.

References

[1] X. Chen and E. Tuncel, "Zero-delay joint source-channel coding using hybrid digital-analog schemes in the Wyner-Ziv setting," *IEEE Transactions on Communications*, vol. 62, no. 2, pp. 726–735, 2014.

[2] Pål Anders Floor, Anna N Kim, Tor A Ramstad, and Ilangko Balasingham, "Zero delay joint source channel coding for multivariate Gaussian sources over orthogonal Gaussian channels," *Entropy*, vol. 15, no. 6, pp. 2129–2161, 2013.

[3] B. Lu and J. Garcia-Frias, "Analog mappings for flexible rate transmission of Gaussian sources with side information," in *48th Annual Conference on Information Sciences and Systems*, 2014, pp. 1–6.

[4] B. Lu and J. Garcia-Frias, "Analog joint source-channel coding for transmission of correlated senders over separated noisy channels," in *2015 49th Annual Conference on Information Sciences and Systems*, 2015, pp. 1–5.

[5] Iker Alustiza, Aitor Erdozain, Pedro Crespo, and Baltasar Beferull-Lozano, "Bandwidth-reduction analog mappings for awgn channels with side information," in *2012 Proceedings of the 20th European Signal Processing Conference*. IEEE, 2012, pp. 1603–1607.

[6] Johannes Karlsson and Mikael Skoglund, "Analog distributed source-channel coding using sinusoids," in *2009 6th International Symposium on Wireless Communication Systems*. IEEE, 2009, pp. 279–282.

[7] E. Akyol, K. Rose, and T. Ramstad, "Optimized analog mappings for distributed source-channel coding," in *Data Compression Conference*, 2010, pp. 159–168.

[8] M. S. Mehmetoglu, E. Akyol, and K. Rose, "Deterministic annealing-based optimization for zero-delay source-channel coding in networks," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 5089–5100, 2015.

[9] Eirina Bourtsoulatze, David Burth Kurka, and Deniz Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 567–579, 2019.

[10] Yashas Malur Saidutta, Afshin Abdi, and Faramarz Fekri, "Joint source-channel coding of Gaussian sources over AWGN channels via manifold variational autoencoders," in *57th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2019, pp. 514–520.

[11] Z. Xuan and K. Narayanan, "Analog joint source-channel coding for Gaussian sources over AWGN channels with deep learning," in *International Conference on Signal Processing and Communications*, 2020, pp. 1–5.

[12] Aaron D Wyner, "The rate-distortion function for source coding with side information at the decoder\3-ii: General sources," *Information and control*, vol. 38, no. 1, pp. 60–80, 1978.

[13] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," in *IEEE International Symposium on Information Theory*, 2006, pp. 1404–1408.

[14] E. Akyol, K. B. Viswanatha, K. Rose, and T. A. Ramstad, "On zero-delay source-channel coding," *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7473–7489, 2014.

[15] Vincent Sitzmann, Julien N.P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein, "Implicit neural representations with periodic activation functions," in *arXiv*, 2020.

[16] Ram Zamir, Shlomo Shamai, and Uri Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1250–1276, 2002.

[17] Tobias Plötz and Stefan Roth, "Neural nearest neighbors networks," *Advances in Neural Information Processing Systems*, vol. 31, pp. 1087–1098, 2018.

[18] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.