Deep Joint Source-Channel Coding for Transmission of Correlated Sources over AWGN Channels

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

ICC 2021 - IEEE International Conference on Communications | 978-1-7281-7122-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICC42927.2021.9500692

***Abstract*—We revisit the joint source-channel coding (JSCC) problem of transmitting correlated sources over the additive white Gaussian noise channel using deep learning methods. Specifically, we consider the design of JSCC schemes for transmitting multivariate Gaussian sources, and Gauss-Markov processes over noisy channels with bandwidth (BW) compression and low delay. We show that encoding and decoding schemes represented by deep neural networks can be optimized jointly to obtain good JSCC schemes. Specifically, we adopt sinusoidal representation networks (SIRENs) for the transmission of multivariate Gaussian sources. The new architecture not only provides similar performance as the state-of-the-art (SOTA) with higher flexibility, but also results in interpretable encoder mappings. For the transmission of Gauss-Markov sources, recurrent neural networks (RNNs) are implemented to extract temporal information without resorting to explicit decorrelation prior to transmission. Experimental results show improved performance compared with traditional schemes.**

I. Introduction

We consider the joint source-channel coding (JSCC) problem of transmitting a discrete-time analog-valued source of dimension k, over a discrete-time additive white Gaussian noise (AWGN) channel with $n = \gamma k$-uses, subject to power constraint on the encoded symbols. Our focus is on the bandwidth (BW) compression case, i.e. $\gamma < 1$. It is well known that separation-based source and channel coding is generally suboptimal at finite delay, and in the absence of perfect knowledge of the channel signal-to-noise ratio (CSNR) at the transmitter. In such situations, JSCC has shown promising potential (see for example, [1], [2], [3]). A majority of the works have focused on memoryless Gaussian sources, albeit this assumption is often not satisfied in practice. The design of optimal JSCC schemes for transmitting correlated sources over the AWGN channel with small delay remains an interesting open problem. This is the focus of this paper.

One solution is to design a hybrid digital-analog (HDA) source-channel coding scheme. For the Gauss-Markov source, Skoglund *et. al* employ a turbo channel code in the digital part, and linear/nonlinear coding in the analog part, which are followed by superposing the analog and digital signals for transmission [4]. In [5], a simpler vector quantizer is implemented in the digital part and linear coding is considered in the analog part. Pure analog mappings have also been broadly studied. A scheme using non-parametric mappings has been proposed and optimized in [6] and its performance has been shown to be close to optimal. For bivariate Gaussian sources, parametric encoder mappings based on sinusoidal functions which reduce encoding and decoding complexities have been proposed in [7].

On the other hand, recent advancements in the design of deep learning based communication systems (e.g., see [8], [9], [10]), have provided impetus for revisiting JSCC problems. For the problem of transmitting i.i.d. Gaussian sources over AWGN channel with BW compression, a variational auto-encoder based method has been put forward in [11]. Both the BW compression and BW expansion cases have been reinspected in [12], where a recurrent neural network (RNN)-based structure is adopted together with fine tuning techniques. Meanwhile, [13] has designed JSCC schemes for images using convolutional neural networks and generalized divisive normalization. For image transmission over the noisy channel with feedback, [14] has proposed a neural architecture inspired by the Schalkwijk-Kailath scheme.

In this work, we focus on two kinds of correlated sources – multivariate Gaussian source and Gauss-Markov source, and design JSCC schemes for their transmission over the AWGN channel. The main contributions of this article are as follows,

- We formulate the JSCC problem for correlated sources as training an auto-encoder (AE). Similar to [13], we set the AWGN channel to be a non-trainable layer.
- For the multivariate Gaussian source, we employ sinusoidal representation networks (SIRENs) in both the transmitter and receiver, where the sine function is selected as the activation function. This model has a simple structure and is easy to train. Our experimental results show that the performance of this scheme matches that of the non-parametric mappings based method in [7], and closely approaches the asymptotic limit at low CSNRs. It also exhibits robustness to mismatch between the true source correlation and the correlation that the scheme is designed for. Further, the learned encoder transformation reveals a completely different mapping from the classical spiral-like curve for the i.i.d. Gaussian source.
- For the Gauss-Markov source, we leverage the ability of RNNs to exploit temporal correlation in the source and represent the encoder and decoder networks with stacked RNNs. We rely on the network to extract temporal features, without explicitly decorrelating the source. Following a similar fine tuning technique as in our prior work [12], the AE learns robust encoding and decoding mappings to mismatched source correlations and channel conditions.

978-1-7281-7122-7/21/$31.00 ©2021 IEEE

II. Problem Formulation

A. System model

We consider the system model shown in Fig. 1. We assume that the source to be transmitted is a k-dimensional vector of random variables $\mathbf{u} \in \mathbb{R}^k$, where u_i is ith component of this vector, for $i = 1, ..., k$.

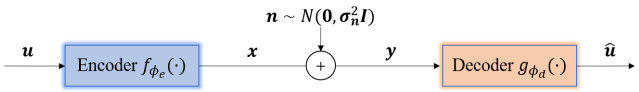


Fig. 1: System model with single transmitter and single receiver for transmitting a correlated source over an AWGN channel.

We consider two types of correlated sources. In the first case, we assume the source is a multivariate Gaussian source. For simplicity, here we consider a 2-dimensional source following a bivariate Gaussian distribution with zero mean and covariance matrix $\Sigma_{\mathbf{u}} = \sigma^2 \left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right]$, where ρ denotes the correlation coefficient. The analysis can be easily extended to higher dimensions within reasonable range. In the second case, the sources follow an order-1 zero-mean stationary Gauss-Markov process, i.e., $u_i = z_i + \rho u_{i-1}$, for $i = 1, ..., k$, with $z_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_z^2)$, $\sigma_z^2 = 1 - \rho^2$, $u_0 = z_0$, and the correlation coefficient $|\rho| < 1$.

We assume the parametrized encoder function as $f_{\phi_e}(\cdot) : \mathbb{R}^k \to \mathbb{R}^n$, and the transmitted channel codeword is denoted as $\mathbf{x} = f_{\phi_e}(\mathbf{u})$. We focus on the case of BW compression, i.e. $k > n$. The channel codeword satisfies a power constraint given by $\frac{1}{n}\mathbb{E}[||\mathbf{x}||]^2] \leq P_T$, and without loss generality it is assumed $P_T = 1$. The transmission is impaired by additive white Gaussian noise, i.e., $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathbf{n}}^2 \mathbf{I})$, and the channel outputs $\mathbf{y} = \mathbf{x} + \mathbf{n}$. The received vector $\mathbf{y}$ is decoded by applying the decoding function $g_{\phi_d}(\cdot) : \mathbb{R}^n \to \mathbb{R}^k$ parameterized by ϕ_d, and the estimate is denoted as $\hat{\mathbf{u}} = g_{\phi_d}(\mathbf{y})$. The quality of the restoration is evaluated by the mean squared error (MSE) distortion between $\mathbf{u}$ and $\hat{\mathbf{u}}$,

$$D(\mathbf{u}, \hat{\mathbf{u}}) = \frac{1}{k}\mathbb{E}[||\mathbf{u} - \hat{\mathbf{u}}||_2^2].$$

B. Optimum performance theoretically attainable (OPTA)

OPTA is the minimum distortion theoretically achievable under the given channel condition, and we include it in our baselines. It should be noted that the OPTA is the asymptotically achievable limit and can only be approached with asymptotically large block length (and hence, delay).

We denote $R(D)$ and $D(R)$ as the rate-distortion function pair for the source, and C as the channel capacity. It is well known from [15] that, sources can be compressed at rate $R(D)$ with distortion D, and also it is feasible to transmit C bits of information with arbitrarily low error rate. The OPTA on distortion can be derived when the source coding and channel coding are optimal, and $kR(D) = nC$.

The channel capacity for the AWGN channel is given by

$$C = \frac{1}{2}\log\left(1 + \frac{P_T}{\sigma_{\mathbf{n}}^2}\right).$$

For the case of bivariate Gaussian sources, we follow the the parametric rate-distortion function given in [7],

$$D(\theta) = \frac{1}{M}\sum_{i=1}^{M} \min\left[\theta, \lambda_i\right]$$
$$R(\theta) = \frac{1}{M}\sum_{i=1}^{M} \max\left[0, \frac{1}{2}\log\left(\frac{\lambda_i}{\theta}\right)\right]$$

where λ_i's are the eigenvalues of the covariance matrix $\Sigma_{\mathbf{u}}$, and M is the number of positive eigenvalues. The corresponding OPTA is given in [7], if $k = 2, n = 1$.

$$\text{OPTA} = \begin{cases} \dfrac{2\sigma_{\mathbf{n}}^2 + 2}{2\sigma_{\mathbf{n}}^2 + (1-\rho)}, & \dfrac{P_T}{\sigma_{\mathbf{n}}^2} < \dfrac{2\rho}{1-\rho} \\ \sqrt{\dfrac{\sigma_{\mathbf{n}}^2 + 1}{\sigma_{\mathbf{n}}^2(1-\rho^2)}}, & \dfrac{P_T}{\sigma_{\mathbf{n}}^2} \geq \dfrac{2\rho}{1-\rho} \end{cases}$$

For the case of Gauss-Markov sources with a variance σ^2 and a correlation coefficient ρ, according to [16], the parametric rate-distortion function can be calculated through spectral density function of source $\Phi_{uu}(\omega)$ using Grenander and Szegö's theorem and is given in parametric form by

$$D(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \min\left(\Phi_{uu}(\omega), \theta\right) \mathrm{d}\omega$$
$$R(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \max\left(0, \frac{1}{2}\log_2\frac{\Phi_{uu}(\omega)}{\theta}\right) \mathrm{d}\omega$$

For small rate, rate-distortion function can be calculated through numerical approximation. The Shannon lower bound matches the rate-distortion function when the rate is moderately large,

$$R(D) = \frac{1}{2}\log_2\frac{\sigma^2(1-\rho^2)}{D}, \qquad D \leq \sigma^2\frac{1-\rho}{1+\rho}$$
$$D(R) = (1-\rho^2)\sigma^2 2^{-2R}, \qquad R \geq \log_2(1+\rho)$$

We define the channel signal-to-noise ratio (CSNR) as $\text{CSNR} := 10\log_{10}\left(\frac{P_T}{\sigma_{\mathbf{n}}^2}\right)$ (in dB), and signal-to-distortion ratio (SDR) as $\text{SDR} := 10\log_{10}\left(\frac{\sigma^2}{D}\right)$ (in dB).

III. Proposed Method

In this section, we elaborate on our proposed deep learning based models for transmitting bivariate Gaussian sources and Gauss-Markov sources over AWGN channels with BW compression, as well as their corresponding training techniques.

A. Autoencoder

AE is not only a classical model for latent representation learning, but also a successful tool in promoting end-to-end wireless communication system performance. The encoder neural network and the decoder neural network replace the traditional manually-designed modules in the transmitter and receiver. An AWGN channel can be represented as a non-trainable layer in between the encoder and decoder networks such that the forward or backward propagation will not be impaired. To satisfy the power constraint, we set $\mathbf{x}_{L,i} = \frac{\mathbf{x}_{L-1,i} - \bar{\mathbf{x}}_{L-1,i}}{\sigma_{x_{L-1,i}}}$, where L, i denote the last layer of the encoder and i-th element,

respectively. And this can be implemented as a normalization layer at the end of the encoder network.

MSE is not only a common distortion measure for restoration, but also a widely-used cost function for regression. Here we set MSE given by

$$\text{MSE} := \frac{1}{k}\mathbb{E}_{\mathbf{u},\mathbf{n}}\left[||\mathbf{u} - g_{\phi_d}(f_{\phi_e}(\mathbf{u}) + \mathbf{n})||_2^2\right]$$

as the end-to-end optimization objective during training. The following sections will employ this AE model with channel and the end-to-end loss function.

B. Case I: Bivariate Gaussian Source

1) SIREN: In [7], it has been shown that both parametric mappings and non-parametric mappings present periodicity and sinusoidal-like contours. Through successive optimization of the encoder-decoder pair at each iteration [6], non-parametric mappings have been shown to work well in low dimensions. Despite that, it needs to implement numerical approximations to integrals involved in its alternating optimization, and store the corresponding table for the encoder and decoder. Hence this scheme heavily relies on the accuracy for discretizing the source and channel space, and the size of the lookup table size grows exponentially with the dimension of the source. In contrast, parametric mappings are defined with respect to sinusoidal functions, and an approximation to minimum-mean-squared-error decoder is utilized. It reduces the storage requirements and lowers the computational complexity, but sacrifices the performance at high CSNRs due to the sub-optimality of the decoder.

Our proposed scheme is based on two facts. On one side, we are inspired by the observation that the encoder mapping for both of the aforementioned schemes resemble sinusoids. On the other side, recent works have shown that [17] SIRENs can be trained in a stable manner and they are very effective in solving a series of challenging representation problems and boundary value problems. Particularly, compared with other activation functions, SIREN are closed on many operations, such as additions and derivatives, and the learned representations are relatively smoother. This renders it possible to gracefully optimize the encoder and decoder networks when their weights are initialized properly. Based on these desirable properties of SIRENs, we propose to build our encoding and decoding schemes based on SIRENs.

According to [17], a simple layer of SIREN is the concatenation of a fully connected layer and a sine function as the activation function,

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

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.

The structure of the model can be found in Table. I. All the encoders in this case consist of 3 layers of SIRENs with 200 hidden units per layer, and all decoders consist of 4 layers of SIRENs with 300 hidden units per layer. For faster convergence, we use a scaled version of a SIREN with scale coefficient 1.2. The correct initialization of weights is the key to keeping the dot product between weights and input of the layer following the same distribution and thus avoiding vanishing or exploding gradients. Here we initialize the weights using a uniform distribution as suggested in [17]. The scalar factor associated with weights of the first layer of encoder and decoder networks is set at 30, and at 1 for the rest of layers, so as to keep frequency components of activation spectrum similar and accelerate the training process.

Layer	Hidden size / Annotation	Output size
Case: Bivariate Gaussian source, transmitter		
Input	Input layer	k
[SIREN] $\times$ 3	200	n
Normalization	Power constraint	n
Case: Bivariate Gaussian source, receiver		
[SIREN] $\times$ 4	300	k
Case: Gauss-Markov source, transmitter		
Input	Reshape	$n \times 2$
BiLSTM	48/direction	$n \times 96$
BiLSTM	48/direction	$n \times 96$
FF+Reshape	1	n
Normalization	power constraint	n
Case: Gauss-Markov source, receiver		
Reshape+BiLSTM	64/direction	$n \times 128$
BiLSTM	64/direction	$n \times 128$
FF+Reshape	2	k

TABLE I: Model parameters of channel AE.

2) Training Techniques: For all experiments corresponding to this case, we generate bivariate Gaussian sources of batch size 1.024×10^5 every iteration, train the model with Adam optimizer with learning rate 0.0004, and use the same batch size for validation data. We continue the optimization until no further improvement on validation data is observed, and keep the model with minimum validation loss for testing.

C. Case II: Gauss-Markov Sources

1) RNN: A Gauss-Markov source is a temporally correlated sequence. RNNs, especially long-short-term-memory (LSTM, in [18]) networks are well-known for achieving significant successes with various tasks involving temporally correlated sequences. Therefore, we adopt bidirectional LSTM (BiLSTM) networks for each layer, and build both the encoder and decoder with stacked layers of BiLSTMs, respectively.

The implementation details are as follows. The source sequence is first reshaped into a $n \times \frac{k}{n}$ tensor, which acts as the input to the 2 stacked layers of BiLSTMs. The output goes through a position-wise fully-connected feedforward (FF) layer, followed by a normalization layer that produces the channel codewords to satisfy the power constraint. The receiver structure is similar. The observed symbol sequence is first reshaped into dimension $n \times 1$, and processed by 2 stacked layers of BiLSTMs subsequently. It is then used in tandem with another fully-connected FF layer, and the output is of dimension $n \times \frac{k}{n}$. The recovered symbols are finally obtained after reshaping.

Let h_e and h_d denote the number of hidden units for LSTM cells employed in the encoder and decoder networks respectively. We set $h_e = 48$ and $h_d = 64$. We find that despite the fact that no specific layers play the role of a decorrelator, the architecture is sufficient to handle the task. The details of structure can be found in Table I with $k/n = 2$.

2) Training Techniques: According to the reported result in [12], empirically we can take advantage of decent fine tuning techniques, such as larger batch size and learning rate schedule, for the faster convergence and better performance.

First, as suggested by [19], a proper pair of larger batch size and higher learning rate can achieve the same performance level with fewer parameter updates without impairing generalization of the model. More specifically, for JSCC, it is important that the training data adequately covers the space of possible inputs. Without large batch sizes, the space of possible inputs may not be covered well rendering our learned encoder mapping inefficient for such corner data points. This problem is accentuated at high CSNRs. This can be alleviated by using a larger batch size which implies more data points available to generalize the model well. Moreover, the encoder also benefits from a more accurate approximation of second moment brought by a larger batch size.

Secondly, we schedule the learning rate with warm restart and cosine annealing as suggested in [20]. For any epoch t that satisfies $\sum_{j=0}^{i-1} T_o^{(j)} \leq t < \sum_{j=0}^{i} T_o^{(j)}$, the learning rate η_t is set according to

$$\eta_t = \eta_{min}^{(i)} + 0.5(\eta_{max}^{(i)} - \eta_{min}^{(i)})(1 + \cos \pi T_{cur}/T_o^{(i)}).$$

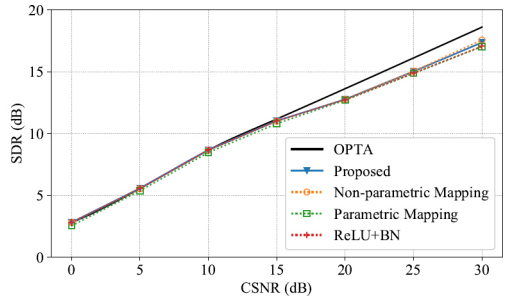
Here, $\eta_{min}^{(i)}$ and $\eta_{max}^{(i)}$ are the minimum and maximum learning rates w.r.t. the i-th restart cycle. $T_o^{(i)}$ and $T_{cur} = t - \sum_{j=0}^{i-1} T_o^{(j)}$ stand for the length of current restart cycle, and the number of epochs that have lapsed since the recent restart cycle, respectively. This schedule has been empirically shown effectiveness of achieving almost better anytime performance in [20]. This was shown to be effective for analog JSCC of i.i.d. Gaussian sources over AWGN channel with BW mismatch in [12].

In the experiments, we let $\eta_{max}^{(i)}$ decay exponentially every restart cycle, and $T_o^{(i)}$ grow exponentially every restart cycle, so that it can explore the minima with finer resolution in the later stages. We set $T_o^{(0)} = 50$ with growing factor 2, $\eta_{max}^{(0)} = 0.01$ with decay factor 0.75, and $\eta_{min}^{(i)} = 0$ for all i. We use Adam optimizer, and train the model until the loss converges. We test the model with the weights corresponding to the best validation performance.

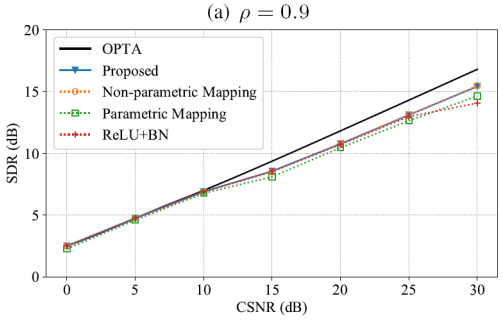
IV. Experimental results

A. Case I: Bivariate Gaussian Source

In this section, we present experimental results for transmitting bivariate Gaussian sources with 2:1 BW compression. A plot of the SDR versus CSNR is shown in Fig. 2(a) for a source correlation of $\rho = 0.9$. We use the parametric mapping scheme in [7] and the non-parametric mapping scheme in [6] as baselines for comparison. In addition to these traditional schemes, we also designed a neural network based JSCC scheme for comparison, which uses ReLU activation and batch normalization. The performance of the proposed scheme is close to that of the non-parametric mapping in [6] over the entire CSNR range, and closely approaches the OPTA at low to middle CSNRs. It should be noted that the OPTA is an asymptotic bound and is achieved at infinite delay in general. Our scheme outperforms parametric mappings in [7] and the ReLU based network at high CSNRs. Similar results are obtained for the case of $\rho = 0.75$ as depicted in Fig. 2(b), where the gap between the performance of our scheme and that of parametric mappings as well as ReLU based network is larger than for the $\rho = 0.9$ case. Although the proposed method and non-parametric mappings based method can both approach the SOTA performance, the complexity of non-parametric method increases exponentially with k, n; on the contrary, that of neural network based methods increases linearly with k, n for small values of k, n. Besides, the proposed scheme is able to learn a smoother and more interpretable encoding mappings than ReLU based networks (see Fig. 4(a-c) and (e)).



(a) $\rho = 0.9$



(b) $\rho = 0.75$

Fig. 2: Performance comparison of JSCC schemes for transmission of bivariate Gaussian source with 2:1 BW compression.

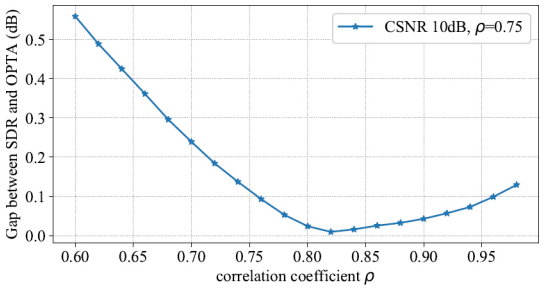
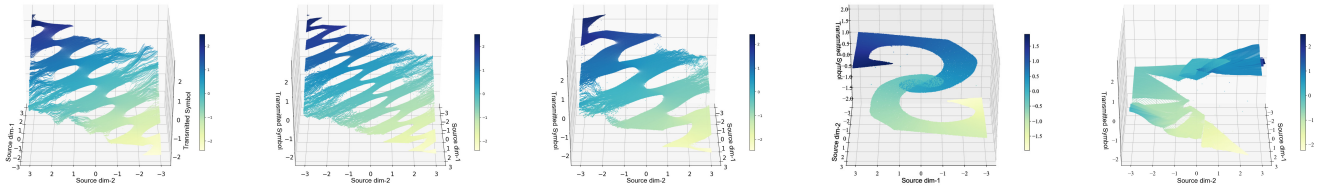


Fig. 3: Simulation results showing model robustness to source correlation.

The robustness of the proposed scheme to the source



(a) SIREN CSNR 20 dB, $\rho = 0.9$ (b) SIREN CSNR 30 dB, $\rho = 0.9$ (c) SIREN CSNR 20 dB, $\rho = 0.75$ (d) Normal source, CSNR 20 dB (e) ReLU CSNR 30 dB, $\rho = 0.75$

Fig. 4: Encoder transformations for 2:1 BW compression of a bivariate Gaussian source.

correlation coefficient is shown in Fig. 3. The model is trained at with $\rho = 0.75$, and CSNR 10dB, but the performance is tested for a range of correlations. It can be seen that the performance smoothly degrades with respect to the mismatched source correlations.

In order to gain interpretability about the encoding scheme, we plot the encoder transformation learned by the proposed scheme in Fig. 4. The x-axis and y-axis represent the first and second element in the source vector, and the z-axis serves as the scalar to be transmitted over the channel. The color is selected according to the value of z-axis, such that deeper blue corresponds to larger values and lighter yellow to smaller values. The plot (d) in Fig. 4 shows the encoder mapping when $\rho = 0$ with 2:1 BW compression that was designed in [12]. It resembles a spiral-like curve filling the entire source-channel space. The first three plots correspond to different source correlations and it can be seen that the encoder functions learned by the proposed scheme for the correlated case are clearly different from that of the uncorrelated case.

Let the eigenvectors and eigenvalues of the covariance matrix of the bivariate Gaussian sources be $\{\mathbf{v}_1, \mathbf{v}_2\}$, and $\{\lambda_1, \lambda_2\}$ respectively, where $\lambda_1 > \lambda_2$. Note any point in the source space can be expressed as $\alpha\mathbf{v}_1 + \beta\mathbf{v}_2$. Implied by the plots, it is consistent that all the learned mappings partition the source space along the direction of $\mathbf{v}_1$. Given fixed value of α and within small vicinity of β along the direction of $\mathbf{v}_2$, all source symbols will be encoded into similar values; but given fixed value of β, two neighbor points in the source space can be encoded into different scalars if they are located on the two sides of the boundary.

For a fixed correlation coefficient, a higher CSNR leads to a denser partition along the direction of $\mathbf{v}_1$, and the contour of the sinusoidal-like wave changes with higher frequency. This is in line with our intuition - at high CSNRs, it suffices to maintain a smaller Euclidean distance between the transmitted symbols corresponding to points that are farther apart. Thus more periods of the sinusoid can fit through the source space.

For a fixed CSNR, a lower correlation coefficient leads to a sparser partition along the direction of $\mathbf{v}_1$, and the contour of the sinusoidal-like wave has lower frequency. This is also reasonable, since lower correlation coefficient implies an increase in the length of λ_2 and a decrease in the length of λ_1. This leaves relatively less space to partition along the direction of $\mathbf{v}_1$ under the same noise level, and thus there are fewer periods for the sinusoidal-like wave.

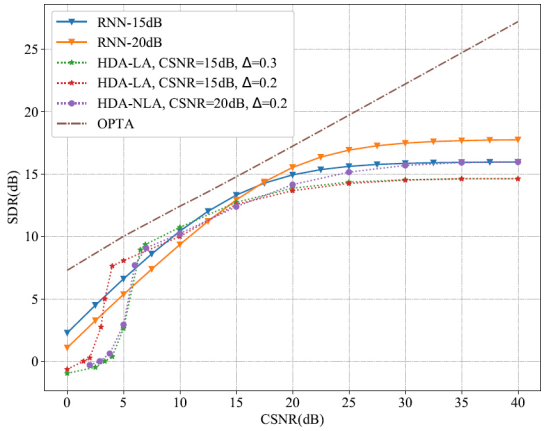


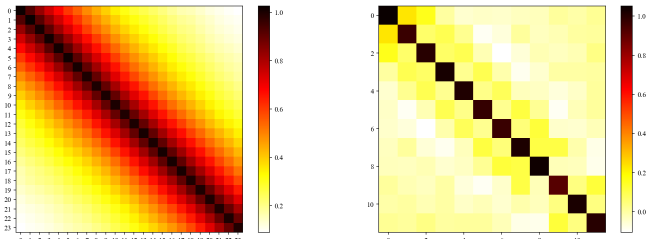
Fig. 5: Performance of JSCC schemes for transmission of Gauss-Markov source with BW compression; $\rho = 0.9$.

B. Case II: Gauss-Markov Source

In this section, we show the experimental results for the transmission of Gauss-Markov sources over the AWGN channel with BW compression. We set the source vector to be 24 dimensional, with correlation coefficient as 0.9, and the channel symbol vector is 12 dimensional. HDA schemes in [4], employing either linear or nonlinear analog coding, are selected as our baseline.

As plot in Fig. 5, we train the network under fixed CSNRs (15 dB, and 20 dB) and test it over the entire CSNR range. Its performance is compared with the HDA schemes (we select the curves with linear analog part that have been commented in the literature as providing 'best overall performance' in [4]) trained at CSNR 15dB and 20dB, as well as the OPTA. The value of Δ controls the ratio of power assigned to analog part to the total input power of the channel in HDA. The plot reveals that the proposed scheme outperforms the baseline over most of the test CSNRs, and has a smoother degradation with respect to mismatch in the channel condition. Particularly in the middle CSNR range, our scheme performs close to the OPTA. The baseline HDA schemes consider sources with dimension 32, which is longer than the proposed scheme. And also the turbo coding at the digital part of the encoder costs the HDA scheme higher complexity and longer delay. Admittedly, the performance of the proposed scheme is compromised at very

high CSNRs, and improving the performance in this regime will be investigated in future work.



(a) Source symbol (b) Transmitted symbol

Fig. 6: Correlation matrix of the source and the transmitted signal w.r.t. a Gauss-Markov source; $\rho = 0.9$, CSNR 10dB.

In Fig. 6, we plot the correlation matrix for the source symbols and transmitted symbols. It can be seen that the transmitted signal is uncorrelated in time with slightly different power levels across the different channel uses. This is reminiscent of applying a decorrelator followed by some form of reverse waterfilling. It is interesting to note that our encoder naturally learns a mapping which can be justified by information theory.

We also examine the capability to gracefully degrade with discrepancy in source correlation. As in Fig. 7, the model is trained with $\rho = 0.5$, CSNR 15 dB, and tested with various correlation values. The plot demonstrates that within a reasonable range of correlation values, the trained model is robust to mismatch in the source correlation.

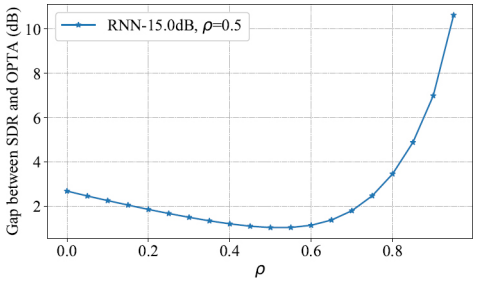


Fig. 7: Robustness of the proposed scheme to mismatch in the correlation of the Gauss-Markov source; CSNR 15dB.

V. Conclusion

We designed JSCC schemes for the transmission of correlated Gaussian sources over AWGN channels using an autoencoder model. We showed that SIRENs are a good choice for representing the encoder and decoder for the transmission of bivariate Gaussian sources. Our proposed model has a simple structure, is easy to train, and is extendable to sources of moderately higher dimension. Its performance matches the SOTA, and closely approaches the OPTA at low CSNRs. Further, our scheme is robust to source correlation discrepancy. We also proposed RNN-based schemes for the transmission of Gauss-Markov source. The learned system exhibits improved performance and smooth degradation in the presence of mismatch in the source correlation as well.

VI. Acknowledgement

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

References

[1] U. Mittal and N. Phamdo, "Hybrid digital-analog (HDA) joint source-channel codes for broadcasting and robust communications," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1082–1102, May 2002.

[2] F. Hekland, P. A. Floor, and T. A. Ramstad, "Shannon-Kotel'nikov mappings in joint source-channel coding," *IEEE Transactions on Communications*, vol. 57, no. 1, pp. 94–105, January 2009.

[3] Y. Hu, J. Garcia-Frias, and M. Lamarca, "Analog joint source-channel coding using non-linear curves and MMSE decoding," *IEEE Transactions on Communications*, vol. 59, no. 11, pp. 3016–3026, November 2011.

[4] M. Skoglund, N. Phamdo, and F. Alajaji, "Hybrid digital–analog source–channel coding for bandwidth compression/expansion," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3757–3763, 2006.

[5] Y. Wang, F. Alajaji, and T. Linder, "Hybrid digital-analog coding with bandwidth compression for Gaussian source-channel pairs," *IEEE Transactions on Communications*, vol. 57, no. 4, pp. 997–1012, 2009.

[6] 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.

[7] P. Suárez-Casal, Ó. Fresnedo, L. Castedo, and J. García-Frías, "Parametric analog mappings for correlated gaussian sources over awgn channels," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 3761–3765.

[8] S. Dörner, S. Cammerer, J. Hoydis, and S. Ten Brink, "Deep learning based communication over the air," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 132–143, 2017.

[9] H. Ye, G. Y. Li, B. F. Juang, and K. Sivanesan, "Channel agnostic end-to-end learning based communication systems with conditional GAN," in *IEEE Globecom Workshops*, 2018, pp. 1–5.

[10] Y. Jiang, H. Kim, H. Asnani, S. Kannan, S. Oh, and P. Viswanath, "LEARN codes: Inventing low-latency codes via recurrent neural networks," in *IEEE International Conference on Communications*, May 2019, pp. 1–7.

[11] Y. M. Saidutta, A. Abdi, and F. 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*, Sep. 2019, pp. 514–520.

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

[13] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2019, pp. 4774–4778.

[14] D. B. Kurka and D. Gündüz, "Deepjscc-f: Deep joint source-channel coding of images with feedback," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 1, pp. 178–193, 2020.

[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory 2nd Edition*. Wiley-Interscience, July 2006.

[16] T. Wiegand and H. Schwarz, *Source coding: Part I of fundamentals of source and video coding*. Now Publishers Inc, 2011.

[17] V. Sitzmann, J. N. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," in *arXiv*, 2020.

[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, pp. 1735–80, 12 1997.

[19] S. L. Smith, P.-J. Kindermans, C. Ying, and Q. V. Le, "Don't decay the learning rate, increase the batch size," *arXiv preprint arXiv:1711.00489*, 2017.

[20] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016.