Bandwidth Compression for a Bivariate Gaussian Source with Shannon-Kotel'nikov Mappings

Fei Liang*, Chong Luo†, Feng Wu* and Wenjun Zeng†
* University of Science and Technology of China, Hefei, China.
Email: lfbeyond@mail.ustc.edu.cn, fengwu@ustc.edu.cn
†Microsoft Research Asia, Beijing, China.
Email: cluo@microsoft.com, wezeng@microsoft.com

Abstract—The analog transmission of a bivariate Gaussian source over a white Gaussian channel with 2:1 bandwidth compression is addressed in this paper. We propose two nonlinear coding structures for bandwidth compression based on the Shannon-Kotel'nikov (S-K) mappings. For each coding structure, the closed-form expression of distortion at any given channel SNR is derived and the optimal energy scaling for each source is computed. We perform theoretical deduction and numerical evaluations to compare the performances of the proposed schemes with a baseline linear mapping scheme. Results show that nonlinear schemes are superior to the linear scheme when the channel condition is good and the difference between the two sources is not too large. Among the proposed nonlinear schemes, the design to combine the two sources for S-K mapping with the optimal energy scaling yields the best performance.

I. INTRODUCTION

The digital communication systems, although being widely deployed in the past a few decades, have their own disadvantages, including the well-known *threshold effect* and *leveling-off effect* [1]. In contrast, the analog systems do not suffer from these two effects, allowing the system performance to degrade gracefully when the channel signal-to-noise ratio (CSNR) decreases and to continuously improve when the CSNR increases. Recently, a scalable mobile video multicast system was proposed based on analog transmissions [2], which triggered tremendous research interests not only in mobile image/video systems but in analog communications as well. Our work to be presented in this paper is also motivated by this new trend.

In the design of analog communication systems, the mismatch of source-channel bandwidths has always been a great challenge. Unlike digital systems which could adjust the transmission rate through source coding, channel coding and/or modulation rates, analog systems could at most allocate one (complex) source sample to one (complex) wireless symbol. When the number of source samples exceeds the number of wireless symbols, it creates a *bandwidth compression* problem. The most well-known and best performed bandwidth compression approaches are Power Constrained Channel Optimized Vector Quantizers (PCCOVQ) [3] and Shannon-Kotel'nikov (S-K) mappings [4]. In general, they have comparable performance but S-K mappings have lower complexity [4].

The S-K mappings for a single variate Gaussian source have been extensively studied in the literature. Based on the general theory of S-K mappings [4], some practical schemes using spiral codes are proposed for 2:1 compression or 1:2

expansion systems [5], [6]. However, we noticed that multimedia communication systems usually need to cope with parallel Gaussian sources, as a result of source de-correlation [2], [7], [8]. A parallel Gaussian source consists of multiple independent Gaussian sources, which in general have different variances. The research on the bandwidth compression of parallel Gaussian sources is relatively few. To the best of our knowledge, there is no practical bandwidth compression scheme designed for parallel Gaussian sources with S-K mappings.

In this paper, we consider a 2:1 bandwidth compression system for a bivariate Gaussian source, where the variances of the two Gaussian components are strictly different. The solution to this basic system will be building blocks of more general $N : M$ bandwidth compression systems for parallel Gaussian sources. We investigate two coding structures. In Scheme-1, two sources are encoded separately using standard 2:1 compression method for single variate Gaussian source. In Scheme-2, instances from the two sources are combined into one codeword. The closed-form expressions of the distortions of both schemes are computed. Based on the distortion analysis, we propose optimization strategies for both schemes.

We conduct theoretical deductions and numerical evaluations to compare the performances of both schemes with a baseline scheme which simply discards all instances from the source with smaller variance. We discover that, although Scheme-2 does not perform as well as Scheme-1 in their plain forms, the optimized Scheme-2 is superior to the optimized Scheme-1. The performance gain increases as the ratio of the two variances increases. We also discover that, the S-K mapping-based schemes tend to outperform the baseline scheme when the CSNR is high and when the variances of the two sources do not differ too much. We use extensive simulations to verify our results.

The rest paper is organized as follow. We first provide background information in Section II, including the previous design and analysis of S-K mappings. In Section III, we describe two coding structures and perform optimization for both structures. Their performances are analyzed and compared. Then, in Section IV, we give numerical results to validate our analysis. We finally conclude this paper in Section V.
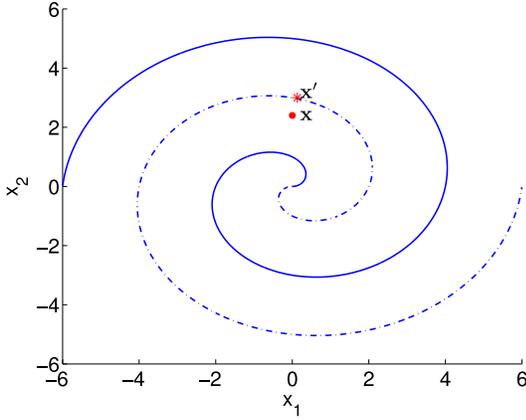
Fig. 1. One example of Archimedes' spiral. Δ=2.

II. BACKGROUNDS

A. A brief overview of related work

The S-K mappings are first proposed for bandwidth expansion [9] and then extended to bandwidth compression. The core idea of S-K mappings is to map the source samples to a lower (in the case of compression) or to a higher (in the case of expansion) dimensional space. Generally, the mappings are achieved through space-filling curves such as the Archimedes' spiral. Hekland et al. [4] build a general theory for bandwidth compression using S-K mappings. The authors also analyze in detail a concrete example for 2:1 bandwidth compression using Archimedes' spiral codes and show that the performance gap to the Shannon limit is only 1.1 dB. Hu et al. [5] further improves the performance of the spiral codes through the minimum mean squred error (MMSE) estimation and joint optimization of spiral size and mapping function. In [6], a low complexity decoding method is proposed by only performing MMSE estimation to the received coded symbol. Brante et al. [10] further propose to utilize the spatial diversity to improve the performance of S-K mappings in fading channels. In [11], the idea of S-K mappings are applied in Wyner-Ziv scenario, i.e. the side information is available in decoder. However, all these works only consider a single Gaussian source.

B. S-K mappings using spiral curve for 2:1 bandwidth compression

In order to make this paper self-contained, we now introduce how the S-K mappings with Archimedes' spiral can achieve efficient bandwidth compression. Similar analysis can also be found in [4]. The Archimedes' spiral can be presented as

$$\begin{cases} x_1 & = \frac{\Delta}{\pi}\theta\cos\theta \\ x_2 & = \frac{\Delta}{\pi}|\theta|\sin\theta. \end{cases} \tag{1}$$

where $\theta \in (-\infty, +\infty)$ and Δ is the distance between two spiral arms. One example is shown in Fig. 1 where $\theta > 0$ for the solid curve and $\theta < 0$ for the dashed curve.

The flow chart of a 2:1 compression system is shown in Fig. 2. Let x_1 and x_2 be two instances drawn from the memoryless Gaussian source X with $\mathbb{E}(X) = 0$ and $\mathbb{E}(X^2) = \sigma_X^2$, where $\mathbb{E}(\cdot)$ means expectation. In order to achieve 2:1 bandwidth compression, the source point $\mathbf{x} = (x_1, x_2)$ is quantized to the
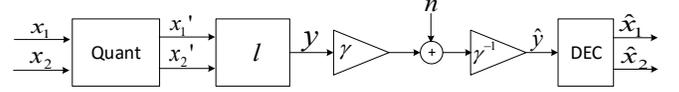

Fig. 2. The system model of spiral codes for 2:1 system.

closest point $\mathbf{x}'$ on the spiral as shown in Fig. 1. It is easy to show that the vector $\mathbf{x}' - \mathbf{x}$ is perpendicular to the tangent line at $\mathbf{x}'$ on the spiral. Then, the two dimensional point $\mathbf{x}'$ is mapped to a one-dimensional coded symbol y which could uniquely identify the point on the curve. In general, the arc length of spiral from the origin to $\mathbf{x}'$ is selected as the mapping function which can be calculated approximately as:

$$y = \pm l(\mathbf{x}') \approx \pm \eta \frac{\pi^2}{\Delta} \left((x_1')^2 + (x_2')^2 \right). \tag{2}$$

where $\eta \approx 0.16$ [4]. $+$ is valid when the point $\mathbf{x}'$ resides on the curve where $\theta \geq 0$ (the solid curve in Fig. 1) and $-$ is valid when the point $\mathbf{x}'$ resides where $\theta < 0$ (the dashed curve in Fig. 1). This kind of mapping function is selected because it ensures that the distortion caused by channel noise is independent from the source sample [4].

The coded symbol y is multiplied by a scaling factor γ to meet the power constraint. When the average symbol power is P, we have $\gamma = \sqrt{P/\sigma_y^2}$ where $\sigma_y^2 = \mathbb{E}(y^2)$. The scaled symbol is then transmitted over an additive white Gaussian noise (AWGN) channel. At the receiver, the noisy symbol $\gamma y + n$ is received where n is the channel noise with $\mathbb{E}(n^2) = \sigma_n^2$. After inverse scaling, we can get the noisy version of y, denoted by $\hat{y} = y + n/\gamma$. Thus the origin sample $\mathbf{x}$ could be reconstructed from $\hat{y}$ through maximum likelihood (ML) or MMSE decoding.

$$\hat{\mathbf{x}} = \begin{cases} l^{-1}(\hat{y}) & \text{for ML decoding,} \\ \mathbb{E}(\mathbf{x}|\hat{y}) & \text{for MMSE decoding.} \end{cases} \tag{3}$$

where $l^{-1}(\cdot)$ is the inverse function of mappings in (2).

With the mean squared error (MSE) as the optimization goal, MMSE decoding yields a better performance than ML decoding. However, the gap between MMSE and ML is negligible when CSNR is high. Besides, the theoretical analysis for MMSE decoding is difficult and its optimization is mainly based on simulations [5]. Therefore, most theoretical results are built on ML decoding.

The average distortion is defined as $D = \mathbb{E}\left(\|\hat{\mathbf{x}} - \mathbf{x}\|^2\right)/2$. The error vector $\hat{\mathbf{x}} - \mathbf{x}$ can be represented as

$$\hat{\mathbf{x}} - \mathbf{x} = (\mathbf{x}' - \mathbf{x}) + (\hat{\mathbf{x}} - \mathbf{x}'), \tag{4}$$

where the first term is the quantization error vector and the second term is the channel error vector. Since $\hat{\mathbf{x}}$ is on the spiral, when the channel noise is small, the channel error $\hat{\mathbf{x}} - \mathbf{x}'$ is in the tangent direction of $\mathbf{x}'$ on the spiral. Thus, the vectors $\hat{\mathbf{x}} - \mathbf{x}'$ and $\mathbf{x}' - \mathbf{x}$ are mutually orthogonal. Hence, the overall distortion can be calculated as follows.

$$\begin{aligned} D &= \frac{1}{2}\left(\mathbb{E}(\|\mathbf{x}' - \mathbf{x}\|^2) + \mathbb{E}(\|\hat{\mathbf{x}} - \mathbf{x}'\|)^2\right) \\ &= \frac{1}{2}\left(\frac{\Delta^2}{12} + \frac{\sigma_y^2}{P}\sigma_n^2\right), \end{aligned} \tag{5}$$

where the average quantization distortion is obtained by assuming $\Delta \ll \sigma_X$ and the quantization error follows the uniform distribution in the interval $[-\frac{\Delta}{2}, \frac{\Delta}{2}]$. In order to estimate the value of σ_y^2, we assume that the spiral is very dense and x_1', x_2' in (2) can be substituted by x_1, x_2 approximately. Therefore we have

$$\sigma_y^2 \approx \mathbb{E}\left(\frac{\eta\pi^2}{\Delta}\left((x_1)^2 + (x_2)^2\right)\right)^2 = \frac{8\eta^2\pi^4\sigma_X^4}{\Delta^2}. \quad (6)$$

Substitute (6) into (5) and apply the inequality ($\frac{a+b}{2} \geq \sqrt{ab}$). We thus have

$$D \geq \sqrt{\frac{\Delta^2\sigma_y^2\sigma_n^2}{12P}} = 2\eta\pi^2\sigma_X^2\sqrt{\frac{1}{6 \cdot \text{CSNR}}}. \quad (7)$$

where = holds when $\Delta = 2\pi\sigma_X \sqrt[4]{\frac{6\eta^2}{\text{CSNR}}}$. We define CSNR= P/σ_n^2. Equation (7) gives the minimum distortion the S-K mappings could achieve. Obviously, the optimal Δ balances the channel and quantization distortions, resulting in

$$\mathbb{E}\left(\|\mathbf{x}' - \mathbf{x}\|^2\right) = \mathbb{E}\left(\|\hat{\mathbf{x}} - \mathbf{x}'\|^2\right). \quad (8)$$

C. A baseline scheme

In this part, we introduce a baseline scheme for 2:1 compression for two independent sources X_1 and X_2 with $\mathbb{E}(X_1^2) = \sigma_1^2$ and $\mathbb{E}(X_2^2) = \sigma_2^2$. Without loss of generality, we assume $\sigma_1 > \sigma_2$. A straightforward method is to discard X_2, the source with lower variances and only transmit X_1. In this case, it is easy to calculate the final distortion D_{base} as:

$$D_{base} = \frac{1}{2}\left(\frac{\sigma_1^2\sigma_n^2}{P} + \sigma_2^2\right). \quad (9)$$

III. BANDWIDTH COMPRESSION OF A BIVARIATE GAUSSIAN SOURCE WITH S-K MAPPINGS

With above preliminary knowledge, we now begin to study the problem of 2:1 bandwidth compression with S-K mappings. We propose and analyze two coding structures, namely Scheme-1 and Scheme-2.

A. Two coding structures for 2:1 bandwidth compression

1) Scheme-1: In Scheme-1, we perform S-K mappings for the two sources separately. Assume $x_{i,1}, x_{i,2}$ are two source samples of X_i. In Scheme-1, the points $(x_{1,1}, x_{1,2})$ and $(x_{2,1}, x_{2,2})$ are mapped to y_1 and y_2 respectively using the compression method shown in Section II-B. Then the mapped symbols y_1 and y_2 are transmitted with power scaling. At the receiver, the two sources are also estimated separately. According to (5), if we define $\mu_i = 2\eta\pi^2\sigma_i^2\sqrt{\sigma_n^2/6}$, the distortion D_1 in this case can be simplified as:

$$D_1 = \frac{1}{2}\left(\frac{\mu_1}{\sqrt{P}} + \frac{\mu_2}{\sqrt{P}}\right). \quad (10)$$

where P is the average power of coded symbols.

2) Scheme-2: In Scheme-2, we combine the samples from the two sources for S-K mappings. Let $x_{1,1}$ and $x_{2,1}$ be source samples of source X_1 and X_2 respectively. In Scheme-2, we map $(x_{1,1}, x_{2,1})$ to a coded symbol y_1 for 2:1 bandwidth compression.

We still assume that Δ is very small compared to σ_1 and σ_2, so the quantization distortion is still approximated as $\Delta^2/12$. With regard to channel distortion, we need to calculate

σ_y^2 first according to (5). Since $x_{1,1}$ and $x_{2,1}$ both follow Gaussian distribution, it is easy to get

$$\sigma_y^2 \approx \mathbb{E}\left(\frac{\eta\pi^2}{\Delta}\left((x_{1,1})^2 + (x_{2,1})^2\right)\right)^2$$
$$= \frac{\eta^2\pi^4}{\Delta^2}\left(3\sigma_1^4 + 3\sigma_2^4 + 2\sigma_1^2\sigma_2^2\right). \quad (11)$$

Substitute (11) into (5) and apply the inequality ($\frac{a+b}{2} \geq \sqrt{ab}$), we get the minimum distortion D_2 and the optimal Δ as

$$D_2 = \eta\pi^2\sqrt{\frac{3\sigma_1^4 + 3\sigma_2^4 + 2\sigma_1^2\sigma_2^2}{12 \cdot \text{CSNR}}}, \quad (12)$$

$$\Delta_{opt} = \pi\sqrt[4]{\frac{12\eta^2(3\sigma_1^4 + 3\sigma_2^4 + 2\sigma_1^2\sigma_2^2)}{\text{CSNR}}}. \quad (13)$$

While the average distortion of the two sources is given in (12), the following theorem gives the average distortion of each source.

Theorem 1. *Let $D_{X,i}$ denote the average distortion of X_i, $i = 1, 2$. When Δ_{opt} is selected, we have $D_{X,1} = D_{X,2} = D_2$.*

The proof is in Appendix A. This theorem says that the two sources share the same distortion whatever their variances are.

Comparing the basic structures of Scheme-1 and Scheme-2, we have the following theorem.

Theorem 2. *For two sources X_1 and X_2 with $\sigma_1 > \sigma_2$, the basic structure of Scheme-1 outperforms Scheme-2, i.e. $D_1 < D_2$.*

The proof is in Appendix B. The unsatisfactory performance of this basic structure of Scheme-2 is mainly ascribed to that it does not balance the protection of the two sources. According to Theorem 1, two sources share the same distortions, which may be not the optimal results. However, as we will shown later, the optimized Scheme-2 will outperfom the optimized Scheme-1 although Scheme-2 cannot perform as well as Scheme-1 in their basic forms.

B. Optimization of Scheme-1

In Scheme-1, since y_1 and y_2 have different statistics, they need to be scaled differently before transmission. Let P_i be the average power allocated to source X_i. In order to achieve the minimum distortion, we formulate the optimization problem as:

$$\min_{P_1, P_2} \quad \frac{1}{2}\left(\frac{\mu_1}{\sqrt{P_1}} + \frac{\mu_2}{\sqrt{P_2}}\right), \quad (14)$$
$$\text{s.t.} \quad P_1 + P_2 = 2P.$$

This is a convex optimization problem and can be easily solved using Lagrangian multiplier. The optimal power allocation P_1^* and P_2^* are given as follows:

$$P_1^* = \frac{2P}{1 + \left(\frac{\mu_2}{\mu_1}\right)^{2/3}}, \quad P_2^* = \frac{2P}{1 + \left(\frac{\mu_1}{\mu_2}\right)^{2/3}}. \quad (15)$$

We use D_1^* to denote the minimum distortion of Scheme-1 after optimization and we thus have

$$D_1^* = \frac{\eta\pi^2}{\sqrt{12 \cdot \text{CSNR}}}\left(\sigma_1^2\sqrt{1 + \left(\frac{\sigma_2}{\sigma_1}\right)^{\frac{4}{3}}} + \sigma_2^2\sqrt{1 + \left(\frac{\sigma_1}{\sigma_2}\right)^{\frac{4}{3}}}\right). \quad (16)$$

C. Optimization of Scheme-2

Since that the two sources are combined for S-K mappings in Scheme-2, it is impossible to perform a general power allocation to improve its performance like Scheme-1. Hence, we try to optimize the performance of Scheme-2 in a different way.

In Scheme-2, $x_{1,1}$ and $x_{2,1}$ have different variances, so the source distributions need some shaping to adapt to the spiral. This is achieved by (source) energy scaling in our proposed optimization. We first transform source X_2 to αX_2 where $\alpha > 0$, so the energy of source X_2 becomes $\alpha^2 \sigma_2^2$. Then X_1 and αX_2 are encoded using the basic structure of Scheme-2. According to (12) and (13), we can directly write the average distortion for the sources $(X_1, \alpha X_2)$ as:

$$D_{(X_1, \alpha X_2)} = \eta \pi^2 \sqrt{\frac{(3\sigma_1^4 + 3\alpha^4 \sigma_2^4 + 2\alpha^2 \sigma_1^2 \sigma_2^2)}{12 \cdot \text{CSNR}}}, \quad (17)$$

and the optimal Δ as:

$$\Delta_{opt} = \pi \sqrt[4]{\frac{12\eta^2 (3\sigma_1^4 + 3\alpha^4 \sigma_2^4 + 2\alpha^2 \sigma_1^2 \sigma_2^2)}{\text{CSNR}}}. \quad (18)$$

According to Theorem 1, $D_{(X_1, \alpha X_2)}$ is also the average distortion of X_1 and αX_2. Therefore, the average distortion of the original sources X_1, X_2 under specific α can be computed:

$$
\begin{aligned}
D_{(X_1, X_2)} &= \frac{1}{2} D_{(X_1, \alpha X_2)} \left(1 + \frac{1}{\alpha^2}\right), \\
&= \frac{\eta \pi^2}{2} \left(1 + \frac{1}{\alpha^2}\right) \sqrt{\frac{3\sigma_1^4 + 3\alpha^4 \sigma_2^4 + 2\alpha^2 \sigma_1^2 \sigma_2^2}{12 \cdot \text{CSNR}}}.
\end{aligned}
\quad (19)
$$

In order to achieve the minimum distortion, we need to perform optimization for α. Note that $D_{(X_1, X_2)}$ is a continuously differentiable function of α, so it is easy to obtain the optimal α by letting $\frac{\mathrm{d}D_{(X_1, X_2)}}{\mathrm{d}\alpha} = 0$. Due to space limit, we directly give the final result. The optimal α is the solution of the following equation.

$$\alpha^8 + \left(1 + \frac{\sigma_1^2}{3\sigma_2^2}\right)\alpha^6 - \left(\frac{\sigma_1^4}{\sigma_2^4} + \frac{\sigma_1^2}{3\sigma_2^2}\right)\alpha^2 - \frac{\sigma_1^4}{\sigma_2^4} = 0. \quad (20)$$

The equation (20) is essentially a quartic equation of α^2 and has closed-form solution although it is very complicated.

In order to show the necessity of the energy scaling, we give the following theorem.

Theorem 3. *When $\sigma_1 > \sigma_2$, equation (20) has only one solution when $\alpha > 0$. Let α_{opt} denote this unique solution, which is also the optimal scaling factor. We have $1 < \alpha_{opt} < \frac{\sigma_1}{\sigma_2}$.*

The proof is in Appendix C. This theorem shows that $\alpha_{opt} \neq 1$ when two sources have different variances, meaning that the optimization will definitely yield a better performance. The minimum distortion D_2^* Scheme-2 can achieve is obtained by inserting α_{opt} to (19).

It is difficult to analytically compare the performance of Scheme-1 and Scheme-2 when they are both optimized. Thus we calculate the value of D_1^*/D_2^* under different σ_1/σ_2 and the results are shown in Fig .3. It is clearly shown that $D_1^* \geq D_2^*$ when $\sigma_1/\sigma_2 \geq 1$. Particularly, D_1^* is strictly equal to D_2^* if $\sigma_1 = \sigma_2$. Thus, we could give the following proposition.
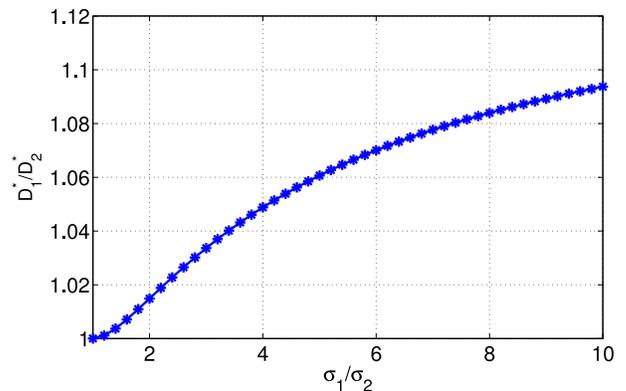

Fig. 3. The calculated value of $\frac{D_1^*}{D_2^*}$ under different $\frac{\sigma_1}{\sigma_2}$.

Proposition 1. *For the analog transmission of a bivariate Gaussian source with 2:1 bandwidth compression, the optimized Scheme-2 is superior to the optimized Scheme-1.*

IV. NUMERICAL RESULTS AND DISCUSSIONS

A. Verification of our analysis

In the theoretical analysis of the proposed nonlinear mapping schemes, we have made a few assumptions. In order to show that these assumptions are reasonable, we use simulations to verify our analysis. The objective evaluation criteria is the signal-to-distortion ratio (SDR) [4], [5] which is defined as

$$\text{SDR} = \left(\sigma_1^2 + \sigma_2^2\right) / \left(2 \cdot \text{MSE}\right), \quad (21)$$

where MSE is the mean square error of two sources.

In Fig. 4, we compare the simulated and calculated results of two schemes, with or without optimization. The source variances are set to $\sigma_1 = 1$ and $\sigma_2 = 0.5$. The spiral parameter Δ is optimized for each CSNR, and the simulated results are obtained by transmitting 10^5 source samples over the AWGN channel. The figure shows that the analytical results match very well with the simulation results in high CSNR regime. As the CSNR decreases, the analytical results may deviate from the actual performance. This is because that, in our analysis, the channel error vector is assumed to follow the tangent direction of the mapped point on the spiral, which may not be valid when the CSNR is low. However, our analytical results could still serve as a close approximation of the actual performance.

B. Comparison of all schemes

In this part, we compare the performance of all schemes. In Fig. 5, we show the simulated results under three source settings i.e. $\sigma = 1$ and $\sigma_2 = 0.3, 0.5, 0.8$ respectively. First, we can see that for Scheme-2, the optimization for energy scaling definitely brings performance gain, which is in accord with our theoretical analysis. This becomes obvious when σ_1/σ_2 is large. When $\sigma_1 = 1, \sigma_2 = 0.3$, the optimal energy scaling could improve the SDR by around 1.5 dB. For Scheme-1, the optimization of power allocation also improves the performance, although the gain is not as significant.

We can also see from Fig. 5 that Scheme-2 with optimal scaling factor outperforms Scheme-1 with optimal power allocation, which supports Proposition 1. For example, when
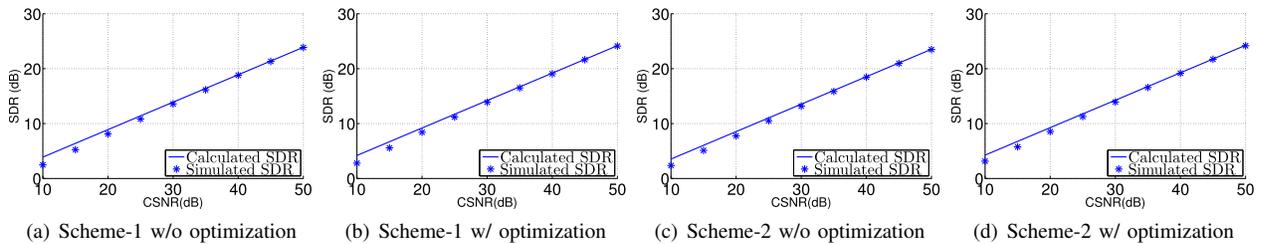
(a) Scheme-1 w/o optimization   (b) Scheme-1 w/ optimization   (c) Scheme-2 w/o optimization   (d) Scheme-2 w/ optimization

Fig. 4. The comparison of simulated results and calculated results. $\sigma_1 = 1, \sigma_2 = 0.5$.

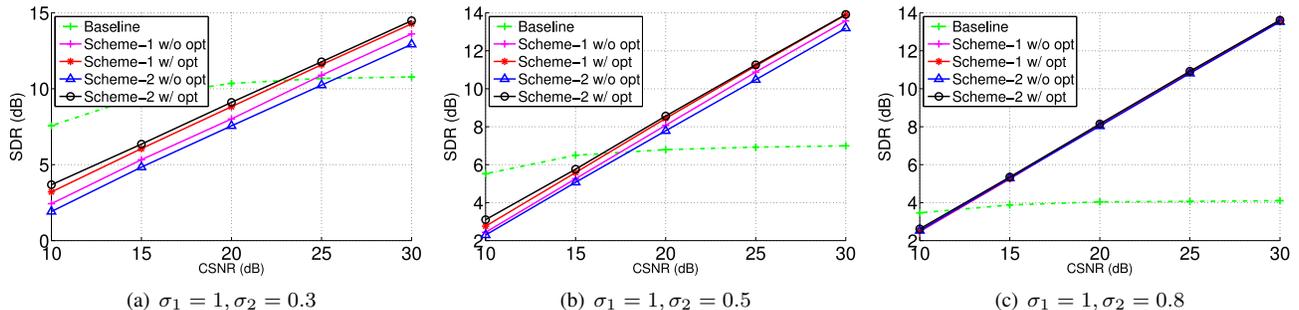(a) $\sigma_1 = 1, \sigma_2 = 0.3$ (b) $\sigma_1 = 1, \sigma_2 = 0.5$ (c) $\sigma_1 = 1, \sigma_2 = 0.8$

Fig. 5. Performance comparison of all schemes. The *opt* in the legend denotes the optimization for power allocation in Scheme-1 and source energy scaling in Scheme-2.

$\sigma_1 = 1, \sigma_2 = 0.3$, the performance gap is about 0.2-0.4dB. We do not expect a larger gap at other reasonable settings. As shown in Fig. 3, the value of D_1^*/D_2^* increases slowly with σ_1/σ_2. When $\sigma_1/\sigma_2 = 10$, D_1^*/D_2^* is only 1.09, which translates to 0.39dB of performance gap. The insight behind this phenomenon is that both schemes try to balance the protection of the two sources in the optimization, although it is achieved by power allocation in Scheme-1 and by energy scaling in Scheme-2.

Next, we compare the performance of the proposed nonlinear mapping schemes with the baseline linear scheme. From Fig. 5, we can see that the baseline scheme outperforms the nonlinear mapping schemes at low CSNRs, but becomes inferior to them when the CSNR is high. For example, when $\sigma_1 = 1, \sigma_2 = 0.5$, the turning point is at around 17 dB as shown in Fig. 5(b). This can be explained by comparing (9) and (19). When the CSNR increases, the distortion of the baseline scheme will be bounded by σ_2^2 while the distortion of Scheme-2 could approach zero. However, in low CSNR regime, Δ_{opt}, the optimal spiral size, becomes large as shown in (18). Hence, the quantization distortion will have a larger impact on the performance of Scheme-2.

For a given channel condition, we compare the performance of the baseline scheme and the best nonlinear mapping scheme (i.e. optimized Scheme-2) for different source statistics. In Fig. 6, CSNR is fixed to 25dB and the x-axis is σ_1/σ_2. When σ_1 is close to σ_2, Scheme-2 performs better; when σ_1/σ_2 is large, the baseline scheme yields better performance. It is also quite easy to understand. When σ_1/σ_2 is large, X_2 plays a negligible role and discarding it will not incur large distortion. However, if X_2 is of similar importance as X_1, discarding X_2 will bring severe performance loss. When CSNR=25dB, we can derive from our theoretical analysis that Scheme-2 will outperform the baseline scheme when $\sigma_1/\sigma_2 > 4.2$. This is very close to our simulated results as
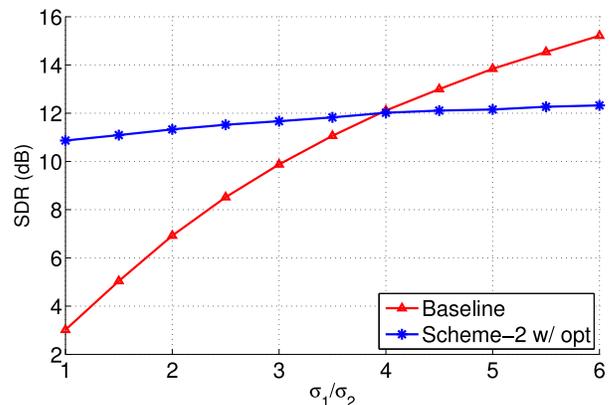

Fig. 6. Performance comparison of the baseline and optimized Scheme-2 under different σ_1/σ_2. CSNR=25dB.

shown in Fig. 6.

V. CONCLUSION

In this paper, we investigate the problem of 2:1 bandwidth compression for a bivariate Gaussian source with strictly different variances. Two nonlinear mapping schemes based on S-K mappings are analyzed, and their performances are compared with each other and with a baseline linear scheme. Both analytical and simulated results show that the optimized Scheme-2, which combines two sources for S-K mappings, yields the best performance when the CSNR is high and the two sources do not differ too much. We believe that our solution to the 2:1 bandwidth compression problem will be building blocks in solving more general N:M bandwidth compression problems.

APPENDIX A
PROOF OF THEOREM 1

Proof: According to (4), the total error can be split to two orthogonal parts, quantization error and channel error. For simplicity we use $\mathbf{q}$ and $\mathbf{c}$ to denote the quantization error vector and channel error vector respectively. The $\mathbf{q}$ and $\mathbf{c}$ can be presented as

$$\begin{aligned} \mathbf{q} &= (\|\mathbf{q}\| \cos \theta_q, \|\mathbf{q}\| \sin \theta_q), \\ \mathbf{c} &= (\|\mathbf{c}\| \cos \theta_c, \|\mathbf{c}\| \sin \theta_c). \end{aligned} \tag{22}$$

As analyzed in Section II-B, the channel error vector $\mathbf{c}$ is in the tangent direction at the mapped point $\mathbf{x}'$. Because the noise value does not depend on the position of $\mathbf{x}'$, we can conclude that $|\mathbf{c}|$ is independent with θ_c. With regard to $\mathbf{q}$, since it is assumed that the quantization error follows uniform distribution in $[-\Delta/2, \Delta/2]$ [4], thus $\|\mathbf{q}\|$ is also independent with θ_q.

The error for each source can be presented as

$$\begin{aligned} \hat{x}_1 - x_1 &= \|\mathbf{q}\| \cos \theta_q + \|\mathbf{c}\| \cos \theta_c, \\ \hat{x}_2 - x_2 &= \|\mathbf{q}\| \sin \theta_q + \|\mathbf{c}\| \sin \theta_c. \end{aligned} \tag{23}$$

We first calculate the average distortion of X_1.

$$\begin{aligned} D_{X,1} &= \mathbb{E}(\|\hat{x}_1 - x_1\|^2) = \mathbb{E}\left(\|\mathbf{q}\| \cos \theta_q + \|\mathbf{c}\| \cos \theta_c\right)^2, \\ &= \mathbb{E}\left(\|\mathbf{q}\|^2 \cos^2 \theta_q + \|\mathbf{c}\|^2 \cos^2 \theta_c + 2\|\mathbf{q}\|\|\mathbf{c}\| \cos \theta_q \cos \theta_c \right). \end{aligned} \tag{24}$$

For the third term in (24), we have

$$\mathbb{E}\left(\|\mathbf{q}\|\|\mathbf{c}\| \cos \theta_q \cos \theta_c\right) = \mathbb{E}\left(\|\mathbf{q}\|\|\mathbf{c}\|\right) \mathbb{E}\left(\cos \theta_q \cos \theta_c\right) = 0. \tag{25}$$

where the first equation follows the independence between $\|\mathbf{q}\|$ and θ_q, as well as $\|\mathbf{c}\|$ and θ_c. The second equation follows the symmetry of the probability distribution of $\cos \theta_q$ and $\cos \theta_c$. Therefore, we have

$$\begin{aligned} D_{X,1} &= \mathbb{E}\left(\|\mathbf{q}\|^2 \cos^2 \theta_q + \|\mathbf{c}\|^2 \cos^2 \theta_c\right), \\ &= \mathbb{E}(\|\mathbf{q}\|^2)\mathbb{E}(\cos^2 \theta_q) + \mathbb{E}(\|\mathbf{c}\|^2)\mathbb{E}(\cos^2 \theta_c), \\ &= \mathbb{E}(\|\mathbf{q}\|^2)\mathbb{E}\left(\cos^2 \theta_q + \cos^2 \theta_c\right) = \mathbb{E}(\|\mathbf{q}\|^2). \end{aligned} \tag{26}$$

where the third equation follows that $\mathbb{E}(\|\mathbf{q}\|^2) = \mathbb{E}(\|\mathbf{c}\|^2)$ when the minimum average distortion is achieved. The fourth equation follows that $\mathbf{q}$ and $\mathbf{c}$ are mutually orthogonal.

Similarly, we can obtain $D_{X,2} = \mathbb{E}(\|\mathbf{q}\|^2)$. Thus we have $D_{X,1} = D_{X,2} = D_2$. The theorem is proved. ∎

APPENDIX B
PROOF OF THEOREM 2

Proof: According to (10) and (12), we have

$$\begin{aligned} D_2^2 - D_1^2 &= \eta^2 \pi^4 \frac{\left(3\sigma_1^4 + 3\sigma_2^4 + 2\sigma_1^2 \sigma_2^2\right)}{12 \cdot \text{CSNR}} - \frac{\eta^2 \pi^4 \left(\sigma_1^2 + \sigma_2^2\right)^2}{6 \cdot \text{CSNR}} \\ &= \frac{\eta^2 \pi^4}{12 \cdot \text{CSNR}} \left(\sigma_1^2 - \sigma_2^2\right)^2. \end{aligned} \tag{27}$$

Obviously, $D_2^2 \geq D_1^2$ and $D_2^2 = D_1^2$ if and only if $\sigma_1 = \sigma_2$. Since we have assumed $\sigma_1 > \sigma_2$, we can conclude $D_2 > D_1$. The theorem is proved. ∎

APPENDIX C
PROOF OF THEOREM 3

Proof: For simplicity, let $\varphi = \alpha^2$ and $u = \sigma_1^2/\sigma_2^2$. Because we assume $\sigma_1 > \sigma_2$, we have $u > 1$. Then (20) can be simplified as

$$\varphi^4 + (1 + \frac{u}{3})\varphi^3 - (u^2 + \frac{u}{3})\varphi - u^2 = 0. \tag{28}$$

We first prove $\varphi > 1$. Let $f(\varphi)$ denote the left hand of (28). We have

$$f'(\varphi) = 4\varphi^3 + 3(1 + \frac{u}{3})\varphi^2 - (u^2 + \frac{u}{3}). \tag{29}$$

where $f'(\varphi)$ is the derivative of $f(\varphi)$. Because $u > 1$, it is easy to show $f'(\varphi) = 0$ has only one solution in $(0, +\infty)$. Let φ_0 denote this unique solution. Thus we have $f'(\varphi) < 0$ when $0 \leq \varphi < \varphi_0$ and $f'(\varphi) > 0$ when $\varphi > \varphi_0$. Therefore, we know that $f(\varphi)$ is a decreasing function when $0 \leq \varphi < \varphi_0$ and an increasing function when $\varphi > \varphi_0$. Meanwhile, it is easy to show $f(\varphi) < 0$, so we can conclude that the equation $f(\varphi) = 0$ has only one solution in $(0, +\infty)$. We can denote this unique solution as φ_{opt}. We also have $f(1) = 2 - 2u^2 < 0$ when $u > 1$. Thus the unique solution φ_{opt} must be in the interval $(1, +\infty)$. Hence, we have proved that $\alpha_{opt} > 1$.

In order to prove $\alpha_{opt} < \sigma_1/\sigma_2$, we denote $z = \varphi/u$. (28) is written as

$$u^2 z^4 + (1 + \frac{1}{3}u)u z^3 - (u + \frac{1}{3})z - 1 = 0. \tag{30}$$

Similarly to the previous proof, it is easy to prove $z_{opt} < 1$, where z_{opt} is the solution of (30). Thus we can obtain $\alpha_{opt} < \sigma_1/\sigma_2$. The theorem is proved. ∎

REFERENCES

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

[2] S. Jakubczak and D. Katabi, "A cross-layer design for scalable mobile video," in *proceedings of MobiCom*. ACM, 2011, pp. 289–300.

[3] A. Fuldseth, T. Ramstad *et al.*, "Bandwidth compression for continuous amplitude channels based on vector approximation to a continuous subset of the source signal space," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, 1997, pp. 3093–3096.

[4] F. Hekland, P. A. Floor, T. Ramstad *et al.*, "Shannon-kotel-nikov mappings in joint source-channel coding," *IEEE Trans. Commun.*, vol. 57, no. 1, pp. 94–105, 2009.

[5] Y. Hu, J. Garcia-Frias, and M. Lamarca, "Analog joint source-channel coding using non-linear curves and mmse decoding," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3016–3026, 2011.

[6] O. Fresnedo, F. J. Vazquez-Araujo, L. Castedo, and J. Garcia-Frias, "Low-complexity near-optimal decoding for analog joint source channel coding using space-filling curves," *IEEE Commun. Lett.*, vol. 17, no. 4, pp. 745–748, 2013.

[7] V. M. Prabhakaran, R. Puri, and K. Ramchandran, "Colored gaussian source–channel broadcast for heterogeneous (analog/digital) receivers," *Information Theory, IEEE Transactions on*, vol. 54, no. 4, pp. 1807–1814, 2008.

[8] H. Cui, C. Luo, C. W. Chen, and F. Wu, "Robust uncoded video transmission over wireless fast fading channel," in *Proceedings of IEEE INFOCOM*. IEEE, 2014, pp. 73–81.

[9] V. A. Kotel'nikov, R. A. Silverman, and G. L. Turin, "The theory of optimum noise immunity," *Physics Today*, vol. 13, p. 40, 1960.

[10] G. Brante, R. D. Souza, and J. Garcia-Frias, "Spatial diversity using analog joint source channel coding in wireless channels," *IEEE Trans. Commun.*, vol. 61, no. 1, pp. 301–311, 2013.

[11] 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 (CISS)*. IEEE, 2014, pp. 1–6.