To code or not to code: Revisited

Victoria Kostina, Sergio Verdú
Dept. of Electrical Engineering, Princeton University, NJ 08544, USA

Abstract—**We revisit the dilemma of whether one should or should not code when operating under delay constraints. In those curious cases when the source and the channel are probabilistically matched so that symbol-by-symbol coding is optimal in terms of the average distortion achieved, we show that it also achieves the dispersion of joint source-channel coding. Moreover, even in the absence of such probabilistic matching between the source and the channel, symbol-by-symbol transmission, though asymptotically suboptimal, might outperform not only separate source-channel coding but also the best known random-coding joint source-channel coding achievability bound in the finite blocklength regime.**

Index Terms—**Achievability, converse, finite blocklength regime, joint source-channel coding, lossy source coding, memoryless sources, rate-distortion theory, Shannon theory.**

I. Introduction

Shannon's fundamental limit on the maximal joint source-channel coding (JSCC) rate at a given fidelity is, in general, attainable only in the limit of long blocklength. Two conspicuous examples when symbol-by-symbol coding is, in fact, optimal in terms of average distortion are the transmission of a binary equiprobable source over a binary-symmetric channel provided the desired bit error rate is equal to the crossover probability of the channel [1, Sec.11.8], [2, Problem 7.16], and the transmission of a Gaussian source over an additive white Gaussian noise channel under the mean-square error distortion criterion, provided that the tolerable source signal-to-noise ratio attainable by an estimator is equal to the signal-to-noise ratio at the output of the channel [3]. More generally, Gastpar et al. [4] gave a set of necessary and sufficient conditions on the source, its distortion measure, the channel and its cost function in order for symbol-by-symbol transmission to be optimal in terms of average distortion.

In this paper, we revisit the dilemma of whether one should or should not code when operating under delay constraints. In the foregoing examples in which no coding attains the minimum average (over source realizations) distortion, the source and the channel are probabilistically matched. In this paper, we show that even in the absence of such a match between the source and the channel, symbol-by-symbol transmission, though asymptotically suboptimal, might outperform separate source-channel coding (SSCC) and even the best JSCC achievability bound from [5] in the finite blocklength regime. In addition, we prove (for channels with finite input alphabet without power constraints) that if the source and the channel are matched probabilistically in the sense of [4], then not only does symbol-by-symbol transmission achieve the minimum average distortion, but also the dispersion of JSCC. In other words, not only do such symbol-by-symbol codes attain the minimum average distortion but also the distortion variance of distortions at the decoder's output is the minimum achievable among all codes operating at that average distortion. Finally, we dissect the binary and the Gaussian examples mentioned above. One important conclusion in the binary case is that no coding is the best among all rate-1 blocklength-n codes, for all n, in terms of excess distortion, that is, it coincides with the converse exactly.

This work was supported in part by the National Science Foundation (NSF) under Grant CCF-1016625 and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under Grant CCF-0939370. The first author was supported in part by the Natural Sciences and Engineering Research Council of Canada.

II. Symbol-by-symbol vs. optimal source-channel codes

In the conventional block setting, the channel input and output alphabets are the $n-$fold Cartesian products of alphabets $\mathcal{A}$ and $\mathcal{B}$, and the source output and representation alphabets are the $k-$fold Cartesian products of alphabets $\mathcal{S}$ and $\hat{\mathcal{S}}$. A lossy source-channel k-to-n block code (of rate $\frac{k}{n}$) is a (possibly randomized) pair of mappings $\mathsf{f}\colon \mathcal{S}^k \mapsto \mathcal{A}^n$ and $\mathsf{g}\colon \mathcal{B}^n \mapsto \hat{\mathcal{S}}^k$. A distortion measure $\mathsf{d}^k\colon \mathcal{S}^k \times \hat{\mathcal{S}}^k \mapsto [0, +\infty)$ is used to quantify its performance. A cost function $\mathsf{c}^n\colon \mathcal{A}^n \mapsto [0, +\infty)$ is imposed on the channel inputs.

Definition 1. *An* $(k, n, d, \epsilon, \alpha)$ *code for* $\{\mathcal{S},\ \mathcal{A},\ \mathcal{B},\ \hat{\mathcal{S}},\ P_{\mathsf{S}},\ \mathsf{d},\ P_{\mathsf{Y}|\mathsf{X}},\ \mathsf{c}\}$ *is a source-channel code with* $\mathbb{P}\left[\mathsf{d}^k\left(S^k, \mathsf{g}(Y^k)\right) > d\right] \leq \epsilon$ *and either* $\mathbb{E}\left[\mathsf{c}^n(X^n)\right] \leq \alpha$ *(average cost constraint) or* $\sup_{x^n \in \mathcal{A}^n} \mathsf{c}^n(x^n) \leq \alpha$ *(maximal cost constraint), where* $\mathsf{f}(S^k) = X^n$.

A symbol-by-symbol code, formally defined next, has rate 1.

Definition 2. *An* (n, d, ϵ, α) *symbol-by-symbol code is an* $(n, n, d, \epsilon, \alpha)$ *code* (f, g) *(according to Definition 1) that satisfies*

$$\mathsf{f}(s^n) = (\mathsf{f}_1(s_1), \ldots, \mathsf{f}_1(s_n)) \tag{1}$$

$$\mathsf{g}(y^n) = (\mathsf{g}_1(y_1), \ldots, \mathsf{g}_1(y_n)) \tag{2}$$

where $\mathsf{f}_1\colon \mathcal{S} \mapsto \mathcal{A}$ *and* $\mathsf{g}_1\colon \mathcal{B} \mapsto \hat{\mathcal{S}}$.

Our goal in this section is to compare the excess distortion performance of the optimal code of rate 1 at channel blocklength n with that of the optimal symbol-by-symbol code, evaluated after n channel uses. The corresponding minimum achievable excess distortions are formally defined as follows.

Definition 3. *Fix ϵ, α, source blocklength k and channel blocklength n. The minimum achievable excess distortion is defined by*

$$D(k,n,\epsilon,\alpha) = \inf\{d\colon \exists (k,n,d,\epsilon,\alpha) \text{ code}\} \tag{3}$$

The minimum excess distortion achievable with symbol-by-symbol codes is defined by,

$$D_1(n,\epsilon,\alpha) = \inf\{d\colon \exists (n,d,\epsilon,\alpha) \text{ symbol-by-symbol code}\}\,. \tag{4}$$

If there is no cost constraint ($\mathsf{c}(a) = 0$ for all $a \in \mathcal{A}$), we will simplify the notation and write $D(k,n,\epsilon)$ and $D_1(n,\epsilon)$ for $D(k,n,\epsilon,\alpha)$ and $D_1(n,\epsilon,\alpha)$, respectively.

We assume the following basic conditions.

(i) The channel is stationary and memoryless, $P_{Y^n|X^n} = P_{\mathsf{Y}|\mathsf{X}} \times \ldots \times P_{\mathsf{Y}|\mathsf{X}}$, and its input cost function satisfies $\mathsf{c}^n(x^n) = \sum_{i=1}^n \mathsf{c}(x_i)$. The capacity-cost function is denoted by $C(\alpha)$.

(ii) The source is stationary and memoryless, $P_{S^n} = P_{\mathsf{S}} \times \ldots \times P_{\mathsf{S}}$, and the distortion measure is separable, $\mathsf{d}^n(s^n, z^n) = \frac{1}{n}\sum_{i=1}^n \mathsf{d}(s_i, z_i)$. The rate-distortion function is denoted by $R(d)$.

(iii) There exist $\bar{d}$ and α such that

$$R(\bar{d}) = C(\alpha) \tag{5}$$

and there exist $P_{\mathsf{X}^\star|\mathsf{S}}$ and $P_{\mathsf{Z}^\star|\mathsf{Y}}$ such that $P_{\mathsf{X}^\star}$ and $P_{\mathsf{Z}^\star|\mathsf{S}}$ generated by the joint distribution $P_{\mathsf{S}}P_{\mathsf{X}^\star|\mathsf{S}}P_{\mathsf{Y}|\mathsf{X}}P_{\mathsf{Z}^\star|\mathsf{Y}}$ achieve $C(\alpha)$ and $R(\bar{d})$, respectively.

(iv) The optimal distributions $P_{\mathsf{X}^\star}$ and $P_{\mathsf{Z}^\star|\mathsf{S}}$ are unique.

(v) $\mathbb{E}\left[\mathsf{d}^9(\mathsf{S},\mathsf{Z}^\star)\right] < \infty$ where the average is with respect to $P_{\mathsf{S}} \times P_{\mathsf{Z}^\star}$.

The essential condition (iii) ensures that there exists an optimal code operating at average distortion $\bar{d}$ and average cost α. Condition (iv) is imposed for clarity of presentation, while the technical condition (v) ensures applicability of the Gaussian approximation in Theorem 4 below.

Theorem 1 (Achievability, symbol-by-symbol code)**.** *Under restrictions* (i)-(iv), *if*

$$\mathbb{P}\left[\sum_{i=1}^n \mathsf{d}(S_i, Z_i^\star) > nd\right] \le \epsilon \tag{6}$$

where $P_{Z^{n\star}|S^n} = P_{\mathsf{Z}^\star|\mathsf{S}} \times \ldots \times P_{\mathsf{Z}^\star|\mathsf{S}}$, and $P_{\mathsf{Z}^\star|\mathsf{S}}$ achieves $R(\bar{d})$, then there exists an (n,d,ϵ,α) symbol-by-symbol code (average cost constraint).

Proof: As shown in [4], if (iii) holds there exist a symbol-by-symbol encoder and decoder such that the conditional distribution of the output of the decoder given the source outcome coincides with distribution $P_{\mathsf{Z}^\star|\mathsf{S}}$, so the excess-distortion probability of this symbol-by-symbol code is given by (6). ∎

Theorem 2 (Converse, symbol-by-symbol code)**.** *Under restriction* (i), *any (n,d,ϵ,α) symbol-by-symbol code (average cost constraint) must satisfy*

$$\epsilon \ge \inf_{\substack{P_{\mathsf{Z}|\mathsf{S}}\colon\\ I(\mathsf{S};\mathsf{Z}) \le C(\alpha)}} \mathbb{P}\left[\mathsf{d}^n(S^n, Z^n) > d\right] \tag{7}$$

where $P_{Z^n|S^n} = P_{\mathsf{Z}|\mathsf{S}} \times \ldots \times P_{\mathsf{Z}|\mathsf{S}}$.

Proof: The excess-distortion probability at blocklength n, distortion d and cost α achievable among all single-letter codes $P_{\mathsf{X}|\mathsf{S}}$, $P_{\mathsf{Z}|\mathsf{Y}}$ must satisfy

$$\epsilon \ge \inf_{\substack{P_{\mathsf{X}|\mathsf{S}}, P_{\mathsf{Z}|\mathsf{Y}}\colon\\ \mathsf{S}-\mathsf{X}-\mathsf{Y}-\mathsf{Z}\\ \mathbb{E}[\mathsf{c}(\mathsf{X})] \le \alpha}} \mathbb{P}\left[\mathsf{d}^n(S^n, Z^n) > d\right] \tag{8}$$

$$\ge \inf_{\substack{P_{\mathsf{X}|\mathsf{S}}, P_{\mathsf{Z}|\mathsf{Y}}\colon\\ \mathbb{E}[\mathsf{c}(\mathsf{X})] \le \alpha\\ I(\mathsf{S};\mathsf{Z}) \le I(\mathsf{X};\mathsf{Y})}} \mathbb{P}\left[\mathsf{d}^n(S^n, Z^n) > d\right] \tag{9}$$

where (9) holds since $\mathsf{S}-\mathsf{X}-\mathsf{Y}-\mathsf{Z}$ implies $I(\mathsf{S};\mathsf{Z}) \le I(\mathsf{X};\mathsf{Y})$ by the data processing inequality. The right side of (9) is lower bounded by the right side of (7) because $I(\mathsf{X};\mathsf{Y}) \le C(\alpha)$ holds for all P_{X} with $\mathbb{E}\left[\mathsf{c}(\mathsf{X})\right] \le \alpha$. ∎

Theorem 3 (Gaussian approximation, optimal symbol--by-symbol code)**.** *Assume that alphabets $\mathcal{S}$ and $\hat{\mathcal{S}}$ are finite. Under restrictions* (i)-(v) *and an average power constant,*

$$D_1(n,\epsilon,\alpha) = \bar{d} + \sqrt{\frac{\mathcal{W}_1(\bar{d},\alpha)}{n}}Q^{-1}(\epsilon) + O\left(\frac{1}{n}\right) \tag{10}$$

$$\mathcal{W}_1(\bar{d},\alpha) = \mathrm{Var}\left[\mathsf{d}(\mathsf{S},\mathsf{Z}^\star)\right] \tag{11}$$

Proof: Appendix. ∎

Theorem 4 (Gaussian approximation, optimal code)**.** *In addition to restrictions* (i)-(v)*, assume that the channel either has finite input and output alphabets and there is no cost constraint, or is AWGN with a maximal power constant. The excess distortion attained by the optimal code is given by*

$$D(n,n,\epsilon,\alpha) = \bar{d} + \sqrt{\frac{\mathcal{W}(\bar{d},\alpha)}{n}}Q^{-1}(\epsilon) + O\left(\frac{\log n}{n}\right) \tag{12}$$

$$\mathcal{W}(\bar{d},\alpha) = \frac{\mathrm{Var}\left[\jmath_{\mathsf{S}}(\mathsf{S},\bar{d})\right] + \mathrm{Var}\left[\imath^\star_{\mathsf{X};\mathsf{Y}}(\mathsf{X}^\star;\mathsf{Y}^\star)\right]}{\lambda^{\star 2}} \tag{13}$$

where $\lambda^\star = -R'(\bar{d})$,

$$\imath^\star_{\mathsf{X};\mathsf{Y}}(x;y) = \log\frac{dP_{\mathsf{Y}|\mathsf{X}=x}}{dP_{\mathsf{Y}^\star}}(y) \tag{14}$$

and $\jmath_{\mathsf{S}}(\mathsf{S},\bar{d})$ is the d-tilted information [6], [7] that can be computed using the identity

$$\jmath_{\mathsf{S}}(\mathsf{s},\bar{d}) = \imath_{\mathsf{S};\mathsf{Z}^\star}(s;z) + \lambda^\star\mathsf{d}(s,z) - \lambda^\star\bar{d} \tag{15}$$

which holds for $P_{\mathsf{Z}^\star}$−almost every z.

We call $\mathcal{W}(\bar{d},\alpha)$ ($\mathcal{W}(\bar{d})$ if there is no cost constraint) the distortion-dispersion function of JSCC, and $\mathcal{W}_1(\bar{d},\alpha)$ ($\mathcal{W}_1(\bar{d})$ if there is no cost constraint) the single-letter distortion-dispersion function of JSCC.

Remark 1. In the finite-alphabet channel case without cost constraints, because the set of all (n, n, d, ϵ) codes includes all (n, d, ϵ) symbol-by-symbol codes, we have $D_1(n, \epsilon) \geq D(n, n, \epsilon)$. Since $Q^{-1}(\epsilon)$ is positive or negative depending on whether $\epsilon < \frac{1}{2}$ or $\epsilon > \frac{1}{2}$, we must necessarily have

$$\mathcal{W}(\bar{d}) = \mathcal{W}_1(\bar{d}) \tag{16}$$

Note that (16) is a consequence of conditions (iii) and (iv).

III. Lossy transmission of BMS over BSC

The memoryless binary source with $P_{\mathsf{S}}(0) = p \leq P_{\mathsf{S}}(1)$ is transmitted over a binary symmetric channel with crossover probability δ, under the constraint that the bit error rate exceeds $0 \leq d \leq p$ with probability not greater than $0 < \epsilon < 1$.

If $p = \frac{1}{2}$, $C = 1 - h(\delta)$, $R(d) = 1 - h(d)$, and (5) is achieved at $\bar{d} = \delta$. If the encoder and the decoder are both identity mappings (uncoded transmission), the resulting joint distribution satisfies condition (iii). Using (11) and (13), it is easy to verify that

$$\mathcal{W}(\bar{d}) = \mathcal{W}_1(\bar{d}) = \delta(1 - \delta) \tag{17}$$

that is, uncoded transmission is optimal in terms of dispersion, as anticipated in Remark 1. Moreover, for an equiprobable source, regardless of the allowed epsilon, uncoded transmission attains the minimum distortion $D(n, n, \epsilon)$ achievable among all codes operating at blocklength n, as the following result demonstrates.

Theorem 5 (BMS-BSC, symbol-by-symbol code)**.** *At blocklength n and excess distortion probability ϵ, the uncoded scheme achieves*

$$D_1(n, \epsilon) = \min\left\{ d : \sum_{t=0}^{\lfloor nd \rfloor} \binom{n}{t} \delta^t (1-\delta)^{n-t} \geq 1 - \epsilon \right\} \tag{18}$$

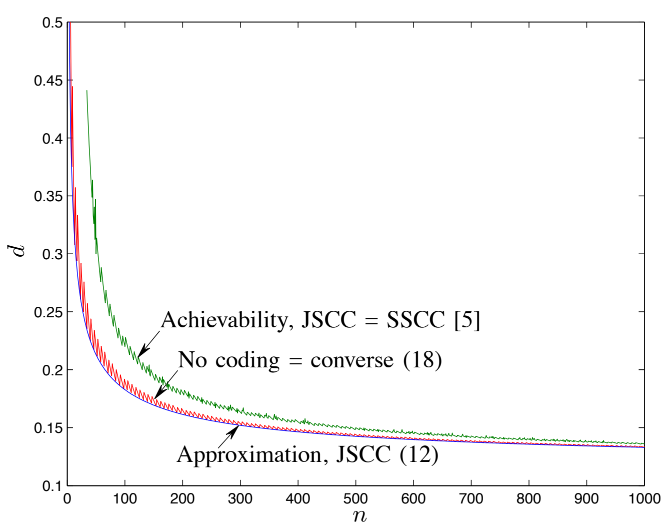

Fig. 1. Distortion-blocklength tradeoff for the transmission of a fair BMS over a BSC with crossover probability $\delta = 0.11$ and $R = 1$, $\epsilon = 10^{-2}$.

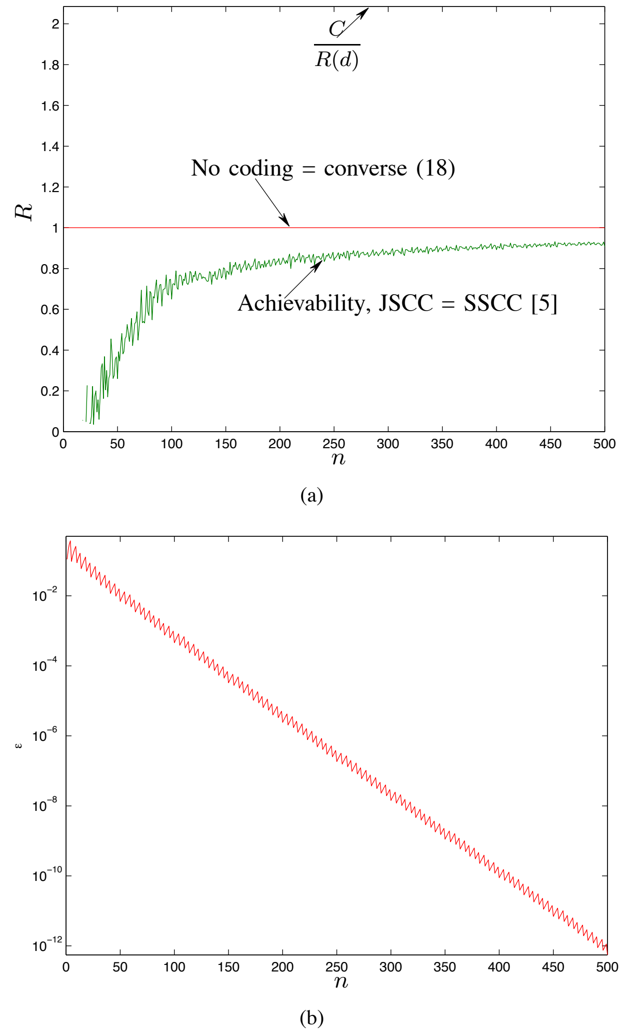

Fig. 2. Rate-blocklength tradeoff (a) for the transmission of a fair BMS over a BSC with crossover probability $\delta = 0.11$ and $d = 0.22$. The excess-distortion probability ϵ is set to be the one achieved by the uncoded scheme (b).

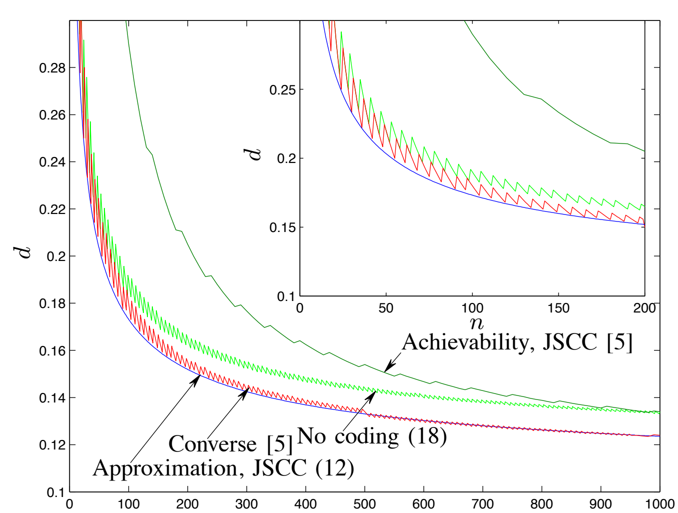

Fig. 3. Distortion-blocklength tradeoff for the transmission of a BMS with $p = \frac{2}{5}$ over a BSC with crossover probability $\delta = 0.11$ and $R = 1$, $\epsilon = 10^{-2}$.

Moreover, if the source is equiprobable,

$$D_1(n,\epsilon) = D(n,n,\epsilon) \tag{19}$$

To show (19), we need the converse result in Theorem 6 below, which is a particularization of [5, Theorem 5]. The idea behind its (omitted) proof is the observation that $1\{\mathsf{d}(S^n, Z^n) \le d\}$ is a (not necessarily optimal) binary hypothesis test between a carefully chosen auxiliary distribution and $P_S P_{X|S} P_{Y|X} P_{Z|Y}$ (where $P_{X|S}$ and $P_{Z|Y}$ are a given encoder/decoder pair). We adopt the following notation.

$$\left\langle {n \atop k} \right\rangle = \sum_{j=0}^{k} \binom{n}{j} \tag{20}$$

Theorem 6 (EBMS-BSC, Converse)**.** *If there exists a* (k,n,d,ϵ) *joint source-channel code, then*

$$\lambda \binom{n}{r^\star + 1} + \left\langle {n \atop r^\star} \right\rangle \le \left\langle {k \atop \lfloor kd \rfloor} \right\rangle 2^{n-k} \tag{21}$$

where

$$r^\star = \max \left\{ r : \sum_{t=0}^{r} \binom{n}{t} \delta^t (1-\delta)^{n-t} \le 1-\epsilon \right\} \tag{22}$$

and $\lambda \in [0,1)$ *is the solution to*

$$\sum_{j=0}^{r^\star} \binom{n}{t} \delta^t (1-\delta)^{n-t} + \lambda \delta^{r^\star+1} (1-\delta)^{n-r^\star-1} \binom{n}{r^\star+1} = 1-\epsilon \tag{23}$$

Proof of Theorem 5: Let us compare $d^\star = D_1(n,\epsilon)$ with the conditions imposed on d by Theorem 6. Comparing (18) to (22), we see that either

(a) equality in (18) is achieved, $r^\star = nd^\star$, $\lambda = 0$, and (plugging $k = n$ into (21))

$$\left\langle {n \atop nd^\star} \right\rangle \le \left\langle {n \atop \lfloor nd \rfloor} \right\rangle \tag{24}$$

thereby implying that $d \ge d^\star$, or

(b) $r^\star = nd^\star - 1$, $\lambda > 0$, and (21) becomes

$$\lambda \binom{n}{nd^\star} + \left\langle {n \atop nd^\star - 1} \right\rangle \le \left\langle {n \atop \lfloor nd \rfloor} \right\rangle \tag{25}$$

which also implies $d \ge d^\star$. To see this, note that $d < d^\star$ would imply $\lfloor nd \rfloor \le nd^\star - 1$ since $nd^\star$ is an integer, which in turn would require (according to (25)) that $\lambda \le 0$, which is impossible.

■

For the transmission of the fair binary source over a BSC, Figure 1 shows the distortion achieved by the uncoded scheme and the separated scheme versus n for a fixed excess-distortion probability $\epsilon = 0.01$. Figure 2(a) shows the rate achieved by separate coding when $d > \delta$ is fixed, and the excess-distortion probability ϵ is set to be the one achieved by uncoded transmission, namely, (18) (Figure 2(b)). Figure 2(a) highlights the fact that at short blocklengths ($n \le 100$) separate source/channel coding is vastly suboptimal. As the blocklength increases, the performance of the separated scheme approaches that of the no-coding scheme, but according to Theorem 5 it can never outperform it. Had we allowed the excess distortion probability to vanish sufficiently slowly, the JSCC curve would have approached the Shannon limit as $n \to \infty$. However, in Figure 2(a), the exponential decay in ϵ is such that there is indeed an asymptotic rate penalty as predicted in [8].

For the biased binary source with $p = \frac{2}{5}$ and BSC with crossover probability 0.11, Figure 3 plots the distortion achieved with probability 0.99 by the uncoded scheme, which in this case is asymptotically suboptimal. Nevertheless, uncoded transmission performs remarkably well in the displayed range of blocklengths, achieving the converse almost exactly at blocklengths less than 100, and outperforming the JSCC achievability result of [5] at blocklengths as long as 900.

IV. Lossy transmission of a GMS over an AWGN

In this section we analyze the setup where the Gaussian memoryless source $S_i \sim \mathcal{N}(0, \sigma_{\mathsf{S}}^2)$ is transmitted over an AWGN channel $\mathcal{N}(0, \sigma_{\mathsf{N}}^2)$ with average power P, under the constraint that the MSE distortion exceeds $0 \le d \le \sigma_{\mathsf{S}}^2$ with probability no greater than $0 < \epsilon < 1$.

Since $C(P) = \frac{1}{2} \log \left(1 + \frac{P}{\sigma_{\mathsf{N}}^2}\right)$ and $R(d) = \frac{1}{2} \log \left(\frac{\sigma_{\mathsf{S}}^2}{d}\right)$, we find that (5) is attained at

$$\bar{d} = \frac{\sigma_{\mathsf{N}}^2 \sigma_{\mathsf{S}}^2}{\sigma_{\mathsf{N}}^2 + P} \tag{26}$$

Theorem 7 (GMS-AWGN, single-letter code)**.** *Consider the following symbol-by-symbol transmission scheme in which the encoder and the decoder are amplifiers:*

$$\mathsf{f}_1(\mathsf{s}) = \alpha \mathsf{s}, \ \alpha^2 = \frac{P}{\sigma_{\mathsf{S}}^2} \tag{27}$$

$$\mathsf{g}_1(\mathsf{y}) = \beta \mathsf{y}, \ \beta = \frac{\alpha \sigma_{\mathsf{S}}^2}{\alpha^2 \sigma_{\mathsf{S}}^2 + \sigma_{\mathsf{N}}^2} \tag{28}$$

This is an (n,d,ϵ,P) *symbol-by-symbol code (with average cost constraint) such that*

$$\mathbb{P}\left[W\bar{d} > nd\right] = \epsilon \tag{29}$$

where $\bar{d}$ *is given by* (26)*, and* W *is chi-square distributed with* n *degrees of freedom.*

Note that (29) is a particularization of (7). Using (29), we find that

$$\mathcal{W}_1(\bar{d}, P) = 2\bar{d}^2 \log^2 e \tag{30}$$

On the other hand, using (13), we compute

$$\mathcal{W}(\bar{d}, P) = \bar{d}^2 (2 - \bar{d}^2) \log^2 e \tag{31}$$

which means that

$$\mathcal{W}_1(\bar{d}, P) > \mathcal{W}(\bar{d}, P) \tag{32}$$

for $\bar{d} < 1$. The difference between (32) and (16) is due to the fact that the optimal single-letter code in Theorem

7 obeys an average power constraint, rather than the more stringent maximal power constraint of Theorem 4, so it is not surprising that for $\epsilon > \frac{1}{2}$ the single-letter code outperforms the best code obeying the maximal power constraint. More interestingly, in the practically relevant case $\epsilon < \frac{1}{2}$, (32) implies that the symbol-by-symbol code of Theorem 7 is only suboptimal in terms of dispersion, even though it achieves the minimum average distortion. Nevertheless, in the range of blocklenghts displayed in Figure 4, the symbol-by-symbol code even outperforms the converse for codes operating under a maximal power constraint.

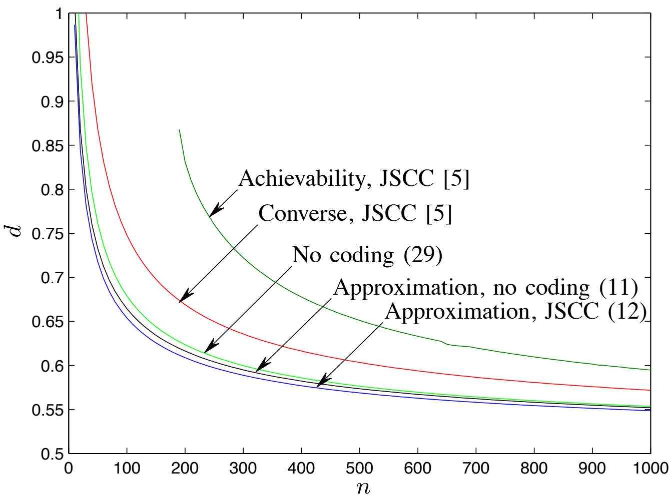

Fig. 4. Distortion-blocklength tradeoff for the transmission of a GMS over an AWGN channel with $\frac{P}{\sigma_{\mathsf{N}}^2} = 1$ and $R = 1$, $\epsilon = 10^{-2}$.

APPENDIX
PROOF OF THEOREM 3

Achievability. Similar to [6], note that restriction (v) implies that the third absolute moment of the random variables $\mathsf{d}(S_i, Z_i^\star)$ is finite, so the achievability part of (10) follows by a straightforward application of the Berry-Esseen bound to (6).

Before we show the converse, consider the following auxiliary result.

Lemma 1. *Let $\mathcal{D}$ be a compact metric space, and let $d\colon \mathcal{D}^2 \to \mathbb{R}^+$ be a metric. Fix $f\colon \mathcal{D} \mapsto \mathbb{R}$ and $g\colon \mathcal{D} \mapsto \mathbb{R}$. Let*

$$f^\star = \max_{x \in \mathcal{D}} f(x) \tag{33}$$

$$g^\star = \sup_{x \in \mathcal{D}^\star} g(x) \tag{34}$$

$$\mathcal{D}^\star = \{x \in \mathcal{D} \colon f(x) = f^\star\} \tag{35}$$

Suppose that for some constants $f_1 > 0, f_2 > 0$, we have

$$f^\star - f(x) \geq f_1 d(x, \mathcal{D}^\star) \tag{36}$$

$$|g(x) - g^\star| \leq f_2 d(x, \mathcal{D}^\star) \tag{37}$$

for all $x \in \mathcal{D}$, where

$$d(x, \mathcal{D}^\star) = \min_{y \in \mathcal{D}^\star} d(x, y) \tag{38}$$

For any positive scalars φ_1, φ_2 such that

$$f_2 \varphi_2 \leq f_1 \varphi_1 \tag{39}$$

we have

$$\max_{x \in \mathcal{D}} \left[\varphi_1 f(x) + \varphi_2 g(x)\right] = \varphi_1 f^\star + \varphi_2 g^\star \tag{40}$$

Proof: Let $x^\star$ achieve the maximum on the left side of (40). Using (36) and (37), we have

$$0 \leq \varphi_1 \left(f(x^\star) - f^\star\right) + \varphi_2 \left(g(x^\star) - g^\star\right) \tag{41}$$

$$\leq \left(-f_1 \varphi_1 + f_2 \varphi_2\right) d(x^\star, \mathcal{D}^\star) \tag{42}$$

$$\leq 0 \tag{43}$$

where where (43) follows from (39). ■

Converse. Observe that by restriction (iii) the constraint $I(\mathsf{S};\mathsf{Z}) \leq C(\alpha)$ in the right side of (7) can be replaced by $I(\mathsf{S};\mathsf{Z}) \leq R(\bar{d})$. Applying the Berry-Esseen bound to (7), we obtain

$$D_1(n, \epsilon, \alpha)$$

$$\geq \min_{\substack{P_{\mathsf{Z}|\mathsf{S}}\colon \\ I(\mathsf{S};\mathsf{Z}) \leq R(\bar{d})}} \left\{\mathbb{E}\left[\mathsf{d}(\mathsf{S},\mathsf{Z})\right] + \sqrt{\frac{\mathrm{Var}\left[\mathsf{d}(\mathsf{S},\mathsf{Z})\right]}{n}} Q^{-1}\left(\epsilon + \frac{B}{\sqrt{n}}\right)\right\}$$

$$= \bar{d} + \sqrt{\frac{\mathcal{W}_1(\bar{d}, \alpha)}{n}} Q^{-1}\left(\epsilon + \frac{B}{\sqrt{n}}\right) \tag{44}$$

where B is the Berry-Esseen constant, and (44) follows by the application of Lemma 1 with $\mathcal{D} = \left\{P_{\mathsf{SZ}} = P_{\mathsf{Z}|\mathsf{S}} P_{\mathsf{S}} \colon I(\mathsf{S};\mathsf{Z}) \leq R(\bar{d})\right\}$, $\varphi_1 = 1$, $\varphi_2 = \frac{1}{\sqrt{n}}$. Note that $\mathbb{E}\left[d(\mathsf{S},\mathsf{Z})\right]$ is a linear function of P_{SZ} and $\mathrm{Var}\left[d(\mathsf{S},\mathsf{Z})\right]$ is a quadratic function of P_{SZ}, so conditions (36) and (37) hold with the metric being the usual Euclidean distance between vectors in $\mathbb{R}^{|\mathcal{S}| \times |\hat{\mathcal{S}}|}$. So, (44) follows immediately upon observing that by the definition of the rate-distortion function, $\mathbb{E}\left[\mathsf{d}(\mathsf{S},\mathsf{Z})\right] \geq \mathbb{E}\left[\mathsf{d}(\mathsf{S},\mathsf{Z}^\star)\right] = \bar{d}$ for all $P_{\mathsf{Z}|\mathsf{S}}$ such that $I(\mathsf{S};\mathsf{Z}) \leq R(\bar{d})$.

REFERENCES

[1] F. Jelinek, *Probabilistic information theory: discrete and memoryless models*. McGraw-Hill, 1968.

[2] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge Univ Press, 2011.

[3] J. Goblick, T., "Theoretical limitations on the transmission of data from analog sources," *IEEE Transactions on Information Theory*, vol. 11, no. 4, pp. 558 – 567, Oct. 1965.

[4] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source-channel communication revisited," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1147 – 1158, May 2003.

[5] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," in *2012 IEEE International Symposium on Information Theory*, Cambridge, MA, July 2012.

[6] ——, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3309–3338, June 2012.

[7] ——, "A new converse in rate-distortion theory," in *46th Annual Conference on Information Sciences and Systems*, Princeton, NJ, Mar. 2012.

[8] I. Csiszár, "On the error exponent of source-channel transmission with a distortion threshold," *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 823–828, 1982.