

Nested Turbo Codes for the Costa Problem

Yong Sun, Momin Uppal, *Student Member, IEEE*, Angelos D. Liveris, *Member, IEEE*, Samuel Cheng, *Member, IEEE*, Vladimir Stanković, *Member, IEEE*, and Zixiang Xiong, *Fellow, IEEE*

Abstract—Driven by applications in data-hiding, MIMO broadcast channel coding, precoding for interference cancellation, and transmitter cooperation in wireless networks, Costa coding has lately become a very active research area. In this paper, we first offer code design guidelines in terms of source-channel coding for algebraic binning. We then address practical code design based on nested lattice codes and propose nested turbo codes using turbo-like trellis-coded quantization (TCQ) for source coding and turbo trellis-coded modulation (TTCM) for channel coding. Compared to TCQ, turbo-like TCQ offers structural similarity between the source and channel coding components, leading to more efficient nesting with TTCM and better source coding performance. Due to the difference in effective dimensionality between turbo-like TCQ and TTCM, there is a performance tradeoff between these two components when they are nested together, meaning that the performance of turbo-like TCQ worsens as the TTCM code becomes stronger and vice versa. Optimization of this performance tradeoff leads to our code design that outperforms existing TCQ/TTCM and TCQ/TTCM constructions and exhibits a gap of 0.94, 1.42 and 2.65 dB to the Costa capacity at 2.0, 1.0, and 0.5 bits/sample, respectively.

Index Terms—Costa coding, nested lattice codes, trellis-coded quantization (TCQ), turbo-like TCQ, and turbo trellis-coded modulation (TTCM).

I. INTRODUCTION

CHANNEL coding with side information (CCSI) refers to the problem of communicating over a noisy channel with some partial knowledge about the transmission channel in the form of side information available at the encoder but not at the decoder. Pioneering works were done by Gelfand and Pinsker [1] on general CCSI and by Costa on the special case of the so-called problem of “writing on dirty paper”

Paper approved by M. Skoglund, the Editor for Source/Channel Coding of the IEEE Communications Society. Manuscript received July 11, 2005; revised October 3, 2006 and March 6, 2007. This work was supported in part by a gift from Cisco Corporation.

Y. Sun was with the Dept of Electrical and Computer Engineering, Texas A&M University. He is now with Lemko Corporation, Schaumburg, IL 60194 USA (e-mail: yong.sun@lemko.com).

M. Uppal and Z. Xiong are with the Dept of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: {momin,zx}@ece.tamu.edu).

A. Liveris was with the Dept of Electrical and Computer Engineering, Texas A&M University. He is now with Microwave Networks Inc., Stafford, TX 77477 USA (e-mail: alive@ieee.org).

S. Cheng was with the Dept of Electrical and Computer Engineering, Texas A&M University. He is now with the School of Electrical and Computer Engineering, University of Oklahoma-Tulsa, Tulsa, OK 74135 USA (e-mail: samuel.cheng@ou.edu).

V. Stanković was with the Dept of Electrical and Computer Engineering, Texas A&M University. He is now with the Department of Electronic and Electrical Engineering University of Strathclyde, Glasgow G1 1XW, UK (e-mail: vladimir.stankovic@eee.strath.ac.uk).

Digital Object Identifier 10.1109/TCOMM.2008.050101.

[2] more than 20 years ago¹. These works have formed the information-theoretical foundation of modern steganography and data-hiding [5], [6].

In the multimedia data-hiding or watermarking problem, a message (or watermark) is to be embedded into a multimedia (e.g., audio, image or video) host signal that is present only at the encoder as the side information. The rules of data-embedding are usually that the host medium is minimally perturbed, i.e., the embedding processing is minimally intrusive, and that the embedded message can be reliably recovered by the intended decoder even in the presence of an attacker who might attempt to corrupt or erase the message while not rendering the embedded host signal unusable.

Although CCSI by association is naturally tied to covert communication problems such as data-hiding, the scope of its applicability is far more general, as it extends to non-covert communication systems as well. Specifically, it turns out that the most efficient way to digital broadcast [7] is to follow the principle of CCSI. Indeed, recent results [8], [9] on the capacity region of the MIMO Gaussian broadcast channel have their roots in the work of Marton [10], whose codebook construction is intimately connected to the methodology of CCSI. This connection between Costa coding and Marton’s achievable rate regions of general broadcast channels was first made in the introduction of Gelfand and Pinsker’s original paper [1]. Other applications of CCSI are precoding for inter-symbol interference channels [11], [12] and transmitter cooperation [13], [14] in wireless networks.

Driven by these applications, several research groups have recently proposed practical Costa code designs. The first was under the name of quantization index modulation [15]. Eggers *et al.* [16] studied Costa coding for information embedding based on the simplest scalar quantization, achieving a gap of 3.5 dB from the capacity at 1.0 bit per sample (b/s), Yu *et al.* [17] employed trellis-coded quantization (TCQ) [18] as the source code and trellis-coded modulation (TCM) [19] as the channel code. Due to the weakness of TCM, this TCQ/TCM scheme operates 3.75, 5.75, and 6.0 dB away from the capacity at 2.0, 1.0 and 0.5 b/s, respectively.

Chou *et al.* [20], [21] reported a turbo-coded trellis-based Costa coding scheme by nesting a TCQ source code inside a turbo TCM (TTCM) [22] channel code. However, owing to the structural dissimilarity between TCQ and TTCM, the actual performance of TCQ is severely degraded when it is coupled (or nested) with TTCM, especially at low rate. At 1.0

¹Shannon [3] considered the case of causal side information in 1958. Kuznetsov and Tsybakov [4] were the first to consider the case of non-causal side information, and Gelfand and Pinsker [1] obtained the capacity for this case. This paper considers code design for the Costa problem [2], where the side information is assumed to be non-causally available at the encoder.

b/s, the scheme by Chou [21] performs 2.07 dB away from the capacity.

Targeting the more challenging low rate regime, Erez and ten Brink [23] recently proposed an efficient code design within the framework of nested lattice codes [11] that performs 1.3 dB from the capacity at 0.25 b/s by using vector quantization (VQ) and irregular repeat-accumulate (IRA) codes [24]. Bennatan *et al.* [25] devised another practical scheme based on superposition coding [7], achieving the same performance as in [23] with TCQ and low-density parity-check (LDPC) codes [26], [27]. Using a combined source-channel coding approach, Sun *et al.* [28] reported the best result of 0.83 dB away from the capacity at 0.25 b/s by using TCQ and IRA codes. However, these systems cannot straightforwardly be applied to the high rate regime because it is much more involved to design good high rate LDPC/IRA codes for multi-level constellations, especially when shaping is used [29].

Aiming to design a system that performs well at both low and high transmission rates, we examine an alternative nested approach and provide practical nested turbo code designs for the Costa problem. Preliminary results from this work appeared in [30]. After reviewing the theoretical results on Gelfand-Pinsker coding and Costa coding in Section II, we offer code design guidelines in terms of source-channel coding for algebraic binning in Section III. Section IV discusses practical considerations in *nested* designs. Section V presents our proposed nested turbo codes using *turbo-like* TCQ for source coding and TTCM for channel coding. Since there exists no efficient turbo TCQ scheme yet², we employ turbo-like TCQ, which provides structural similarity to TTCM, hence enables better nesting with TTCM than TCQ. Because turbo-like TCQ is not turbo TCQ, the effective dimensionality of turbo-like TCQ is different from that of TTCM. This leads to a performance tradeoff between these two components when they are nested together in our Costa code design. Specifically, the performance of turbo-like TCQ steadily worsens when we make the TTCM code stronger; the opposite is true when TTCM is weakened (until it degenerates to TCM). Optimization of this performance tradeoff leads to our best code construction. Simulation results reported in Section VI show that our proposed Costa code design offers sizable performance gains at different transmission rates over the TCQ/TCM construction of [17] and TCQ/TTCM construction of [20], [21]. Section VII concludes the paper.

II. THE COSTA PROBLEM

CCSI at the encoder, or Gelfand-Pinsker coding [1], is schematically shown in Fig. 1. The transmitter wishes to send message $m \in \{1, \dots, M\}$ over a memoryless channel, which is defined by the transition probabilities $p(y|x, s)$, where X and Y are the channel input and output, respectively, and the random variable S , which is independent of X , is the state of the channel (the side information) known non-causally to the transmitter but not to the receiver. Based on the selected message m and the state of the channel S , the encoder

²On one hand, a recent attempt on turbo TCQ [31] resulted in worse results than TCQ; on the other hand, there are exciting works [32], [33] on graph-based codes for lossy coding of *binary* sources.

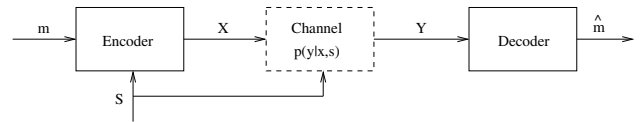


Fig. 1. CCSI at the encoder or Gelfand-Pinsker coding.

sends codeword X which must satisfy the power constraint $E[X^2] \leq P_X$. The capacity is given by [1]

$$C^* = \max_{p(u,x|s)} [I(U; Y) - I(U; S)], \quad (1)$$

where U is an auxiliary random variable such that $Y \rightarrow (X, S) \rightarrow U$ and $Y \rightarrow (U, S) \rightarrow X$ form Markov chains and $E[X^2] \leq P_X$. The proof of the Gelfand-Pinsker capacity is based on random coding and binning.

Gelfand-Pinsker coding in general suffers performance loss when compared to channel coding with side information available at both the transmitter and the receiver. For example, in the binary Gelfand-Pinsker problem [1], the channel output is $Y = X + S + Z$, where X , S , and Z are channel input, an interference binary-symmetric signal known to the transmitter but not to the receiver, and unknown i.i.d. Bernoulli- p channel noise, respectively. Under a Hamming power constraint $\frac{1}{n}E[w_H(X)] \leq \delta$, $0 < \delta < 1/2$, the capacity is given by $C^* = u.c.e.\{H(\delta) - H(p), (0, 0)\}$, where *u.c.e.* means upper concave envelope [34], [35]. C^* is strictly smaller than the capacity $C = H(p * \delta) - H(p)$ when the decoder also has access to the side information S .

In contrast to the binary case of Gelfand-Pinsker coding, in the Gaussian case, when we have the celebrated Costa problem [2], there is no performance loss with CCSI. Specifically, when S and Z are i.i.d. zero-mean Gaussian and the average channel input power constraint is $E[X^2] \leq P_X$, Costa showed that the capacity is given by [2]

$$C^* = \frac{1}{2} \log\left(1 + \frac{P_X}{P_Z}\right), \quad (2)$$

where P_Z is the noise power. Thus, although S is unknown to the decoder, the capacity remains the same as if S were available at the decoder. Costa's proof is again based on random coding and binning. The result in (2) was extended to arbitrarily distributed interference S in [12], [36].

III. CODE DESIGN GUIDELINES: SOURCE-CHANNEL CODING FOR ALGEBRAIC BINNING

Although Costa's proof shows the existence of capacity-achieving random binning schemes, it does not give any indication about practical code construction. Zamir *et al.* [11] suggested an algebraic binning scheme based on nested lattice codes. The scheme consists of a coarse lattice code nested within a fine lattice code. The fine lattice code needs to be a *good* channel code and the coarse lattice code needs to be a *good* source code to approach the capacity in (2).

Fig. 2 (a) illustrates 1-D nested lattice/scalar codes with an infinite uniform constellation, where Δ denotes the stepsize. The channel codewords are grouped into cosets/bins (labeled as 0, 1, 2, and 3) for source coding. At the encoder, the side information S is linearly scaled by α [2] and quantized to

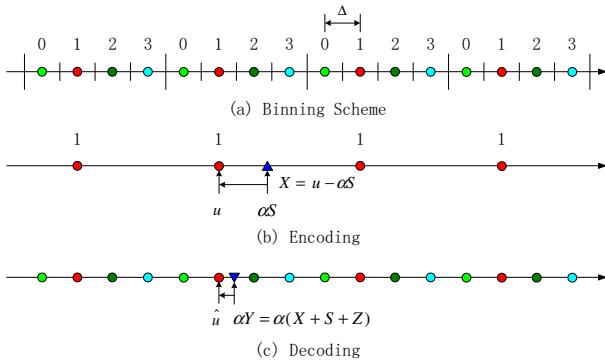


Fig. 2. 1-D nested lattice codes for Costa coding.

the closest codeword u by the source code selected by the message m to be transmitted (e.g., the coset/bin labeled 1 in Fig. 2 (b)), so that the obtained quantization error $X = u - \alpha S$ satisfies the power constraint $E[X^2] \leq P_X$. Then, X is transmitted over the additive white Gaussian noise channel with noise $Z \sim N(0, P_Z)$. According to [2], the optimal $\alpha = \frac{P_X}{P_X + P_Z} = \frac{\text{SNR}}{\text{SNR} + 1}$, with $\text{SNR} = \frac{P_X}{P_Z}$. The decoder (see Fig. 2 (c)) receives the signal $Y = X + S + Z$, scales it by α , and finds the codeword \hat{u} closest to αY . Finally the index of the bin containing \hat{u} is identified as the decoded message³.

It is shown in [11] that this nested scheme approaches the capacity in (2) as the dimensionality of the employed lattices approaches infinity. However, nested lattice coding calls for a joint source-channel code design, typically with the same dimensional coarse lattice source code and fine lattice channel code, which are difficult to implement in high dimensions.

We offer an algebraic message-based binning [30] interpretation of Costa coding in terms of source-channel coding and use it as the guiding principle for our proposed code designs. From an information-theoretical perspective, according to [37], there are granular gain and boundary gain in source coding, and packing gain⁴ and shaping gain in channel coding. Dirty-paper coding is primarily a channel coding problem (for transmitting messages), one should consider the packing gain and the shaping gain. In addition, the side information necessitates source coding to satisfy the power constraint, therefore, the constellation needs to be infinitely replicated so that one can quantize the side information to satisfy the power constraint [11]. Thus the source code in Costa coding is not conventional in the sense that there is only granular gain, but no boundary gain. One needs to establish the equivalence between the shaping gain in channel coding and the granular gain in source coding (e.g., via nested lattice codes [11]) for Costa coding. Then one can shoot for the shaping gain via source coding and the packing gain via channel coding. In

³In their information-theoretic works [11], [12], the authors subtract a common random dither from αS (to decorrelate X from the quantizer input) before quantization and add the dither back to the estimated channel output αY at the decoder. Together with scaling with the optimal α , dithering theoretically assures near-capacity operation of nested lattice coding. In our practical code designs, dithering plays a much less important role than scaling in terms of code performance. Hence we do not mention dithering any more in the sequel.

⁴In [37], packing gain in channel coding was also referred to as coding gain. In this paper, we use packing gain exclusively since we view coding gain as a generic term in the context of both source coding and channel coding.

practice, the former should be done with quantizers (e.g., TCQ) having almost spherical Voronoi regions in a high-dimensional Euclidean space, and the latter with near-capacity channel codes (e.g., turbo [38] and LDPC codes).

This points to a nested approach based on TCQ and TTCM for message-based algebraic binning, where the punch line is to group channel codewords corresponding to the same message into a bin, and within each bin, choose the codeword according to the side information. In other words, we adapt the codeword to the side information. This is much like adapting to the “dirt” when writing on dirty paper – hence the whimsical title of [2].

In practice, when the dimension of the coarse lattice Λ for source coding (or quantization) is finite but high, Erez and ten Brink [23] further show the capacity of the modulo lattice channel [11] induced by the lattice quantizer Λ is lower bounded by

$$C = \frac{1}{2} \log_2(1 + \text{SNR}) - \frac{1}{2} \log_2 2\pi e G(\Lambda), \quad (3)$$

where $G(\Lambda)$ is the normalized second moment of Λ . Since $G(\Lambda)$ starts from $\frac{1}{12}$ in the 1-D case and asymptotically approaches $\frac{1}{2\pi e}$ when the dimensionality of Λ goes to infinity [39], the granular gain $g(\Lambda) = -10 \log_{10} 12G(\Lambda)$ of Λ is maximally 1.53 dB [37]. Equation (3) indicates that with ideal channel coding, the loss in rate due to high-dimensional lattice quantization is maximally $\frac{1}{2} \log_2 2\pi e G(\Lambda)$ b/s. With practical channel coding, there is an additional packing loss $Loss_{CC}$ (in dB). In order to measure the losses from both source coding and channel coding (in dB), we equate the lower bound C in (3) with $C^* = \frac{1}{2} \log_2(1 + \text{SNR}^*)$ and define $Loss_{SC}$ (in dB) [28] due to source coding as

$$Loss_{SC} \triangleq 10 \log_{10} \frac{\text{SNR}}{\text{SNR}^*} = 10 \log_{10} \frac{2\pi e G(\Lambda) 2^{2C^*} - 1}{2^{2C^*} - 1}, \quad (4)$$

where $\text{SNR}^* = 2^{2C^*} - 1$ is the capacity-achieving SNR. Then we compute the total performance loss (in dB) in practical Costa coding as

$$Loss_{Total} = Loss_{SC} + Loss_{CC}. \quad (5)$$

When the capacity C^* is high, $Loss_{SC} \approx 10 \log_{10} 2\pi e G(\Lambda) = 1.53 - g(\Lambda)$ dB, i.e., $Loss_{SC}$ is approximately equal to the granular loss from source coding in this case. But as C^* decreases, Fig. 3 indicates that the granular loss is magnified more and more to become $Loss_{SC}$ (see also [12, Fig. 2]). To reduce $Loss_{SC}$ it is imperative to use high-dimensional lattice quantizers (or VQ in general) to reduce the granular loss as much as possible. This automatically precludes the scalar Costa scheme [16] from approaching the capacity. Fig. 3 also highlights the fact that the Costa code design problem is more challenging when the rate is low.

With (5), the aim of Costa code design is now clear: one needs to employ both strong source and channel codes so that the total loss is minimized. Once the source and channel codes are chosen, one can obtain the expected performance of the resulting Costa code. Conversely, once the performance of a Costa code is known, one can separately measure $Loss_{SC}$ due to source coding from (4), where $G(\Lambda)$ is replaced by the

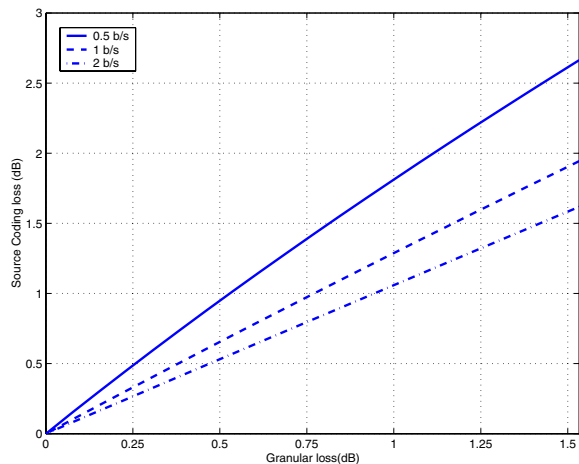


Fig. 3. Source coding loss $Loss_{SC}$ (in dB) in practical Costa coding at three different rates due to the granular loss $1.53 - g(\Lambda)$ (in dB) in source coding.

normalized version of the mean square error (MSE) $E[X^2]$ introduced by the quantizer, and the packing loss $Loss_{CC}$ due to channel coding. These are the guidelines we follow in constructing practical Costa codes.

IV. PRACTICAL CONSIDERATIONS IN NESTED DESIGNS

According to [11] and (5), a nested lattice code can asymptotically approach the capacity of Costa coding in (2) when the dimensionality of the employed lattices (for source coding and channel coding) goes to infinity. However, whereas recent progress [40] in iterative decoding of graph-based (e.g., LDPC) codes has made it possible to implement equivalent lattice channel codes of very high effective dimensions (e.g., in the thousands), such progress has not yet been mirrored in practical source coding (notwithstanding [32], [33]). For example, turbo TCQ [31] is worse than TCQ, which is currently the most efficient practical scheme for quantization. A 256-state TCQ (with 1.32 dB granular gain) can only outperform lattice source codes of up to 69 dimensions [18]. The lack of practically efficient graph-based codes for quantization of continuous (e.g., Gaussian) sources in general (and turbo TCQ in particular) makes it almost impossible to implement nested codes with the same but very high effective dimensionality.

To further see the performance difference [37] between lattice codes for source and channel coding, using results from [41, the sphere bound][42, Fig. 2] we plot in Fig. 4 as functions of lattice dimensionality the upper bound on the granular gain (in dB) of lattice quantization of Gaussian sources and the upper bound on the packing gain (in dB of normalized SNR) of lattice channel codes for AWGN channels (assuming $BER=10^{-5}$). With nested scalar lattices for Costa coding, the fine source code (uniform scalar quantization) leaves unexploited the maximum granular gain of only 1.53 dB [37], whereas the coarse channel code (scalar coset code) gives up the maximum 8.13 dB packing gain. With nested trellis-based codes, the effective dimensionality of TCQ or

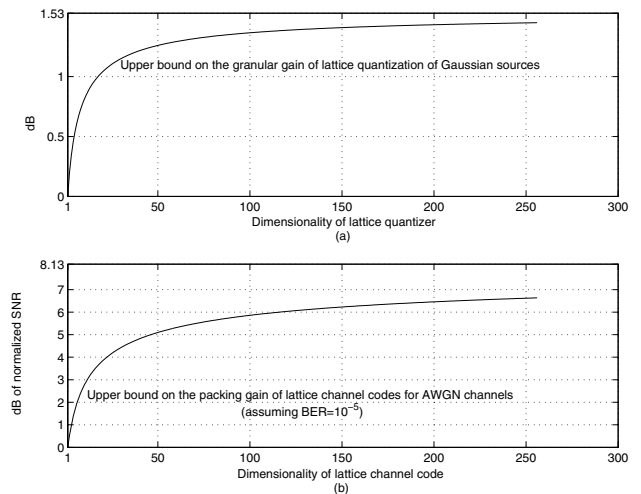


Fig. 4. (a) Upper bound on the granular gain (in dB) of lattice quantization of Gaussian sources and (b) upper bound on the packing gain (in dB of normalized SNR) of lattice channel codes for AWGN channels (with $BER=10^{-5}$), both as functions of lattice dimensionality.

TCM is usually less than 300 in practice⁵, the upper bounds in Fig. 4 and (5) can be used to predict and explain the performance of TCQ/TCM code constructions of [17], [20], [21]. In addition, Fig. 4 offers a one-to-one correspondence between the granular/packing gain of any source/channel code and the effective dimensionality of its equivalent lattice code. More importantly, it is seen that the granular loss of lattice quantization at dimension 256 is less than 0.1 dB, but lattice channel codes at this dimension still suffers more than 1 dB packing loss. This means that the effective dimensionality of capacity-approaching turbo or LDPC codes is indeed much higher than 256. Thus, when a strongest source code (e.g., TCQ) and a strong channel code (e.g., TTCM) are nested together for efficient Costa coding in practice (as prescribed in Section III), we are using two codes with very different effective dimensions⁶.

We argue that this *dimensionality mismatch*, i.e., the difference in the effective dimensions of strong source and channel codes, leads to a fundamental performance tradeoff between the source and channel codes in *any* efficient nested design⁷. Specifically, owing to the coupling between the two component codes, this tradeoff manifests itself in decreased source coding performance as the channel code is made

⁵256-state TCQ was used in [21] with 1.33 dB granular gain (or an effective dimension of 76); 512-state TCQ was employed in [17] with 1.36 dB granular gain (or effective dimension of 93). Under the constraint of current computing power, Yang *et al.* [43] were able to obtain a granular gain of 1.45 dB (or an effective dimension of 298) with 65,536-state TCQ.

⁶First, for two lattices to be nested, they do not have to be of the same dimensionality (e.g., a Z -lattice can be nested in any construction-A lattice as the coarse-fine lattice pair [39]). Second, since turbo TCQ [31] does not perform better than TCQ or has the same effective dimensionality as a good turbo channel code, for the best source coding performance some form of TCQ or other source code of similar effective dimensionality needs to be used. In the TCQ/TTCM construction of [20], [21] at 1.0 b/s, the 0.406 dB granular gain of TCQ barely exceeds that of a four-dimensional lattice quantizer, but the 7.51 dB packing gain of TTCM leads to an effective dimension of much higher than 256.

⁷A better alternative is to use superposition coding as done in [25] or combined source-channel coding as advocated in [28] *without nesting* (or *such a performance tradeoff*).

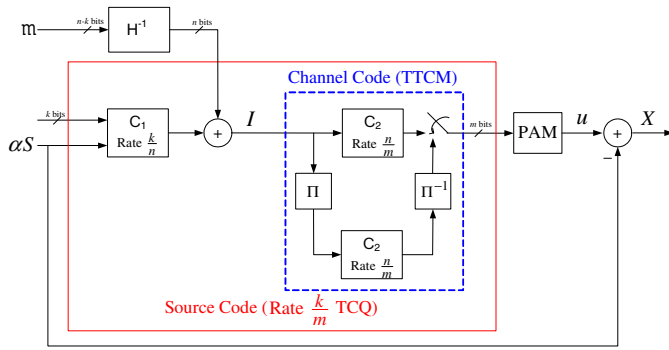


Fig. 5. Block diagram of the TCQ/TTCM encoder [20], [21].

stronger, and vice versa. For example, in the work of Erez and ten Brink [23] with VQ and IRA codes, the nested design leads to strong channel code with subpar source coding performance. As another example, we will see in Section V-A that, in migrating from TCQ/TCM to TCQ/TTCM for Costa coding, the performance of TCQ is severely degraded in [20], [21] when TCQ is nested inside the much stronger TTCM code than the similarly structured TCM code. Thus, our main task in efficient Costa code design is to use the strongest practical source and channel codes and additionally find the *best* nesting between them in terms of optimizing their performance tradeoff.

V. NESTED TURBO CODES BASED ON TURBO-LIKE TCQ AND TTCM

We first briefly review the related nested code construction of [20], [21], which uses TCQ for source coding and TTCM for channel coding.

A. The TCQ/TTCM Code Construction of [20], [21]

The trellis structure in the TCQ/TTCM scheme of [20], [21] is constructed via a rate- $k/n/m$ concatenated code (denoted by C_1+C_2 , with C_1 being the rate- k/n convolutional code and C_2 being the rate- n/m convolutional code) as shown in the encoder block diagram in Fig. 5. TCQ essentially relies on the trellis Γ_1 formed by C_1+C_2 ; the TTCM code consists of a parallel concatenated code with C_2 in both branches. C_2 in the bottom branch is preceded by an n -bit symbol interleaver and followed by an m -bit symbol deinterleaver. The two branches are multiplexed by taking even/odd-indexed symbols (of m bits each) from the top/bottom branch before PAM or QAM. It is seen from Fig. 5 that this code construction is nested since the TTCM code is part of the overall rate- k/m TCQ source code.

At the encoder, every $(n-k)$ -bit segment of the message \mathbf{m} is mapped to an n -bit symbol by the pseudo inverse of the parity-check matrix H of C_1 before being added to an output n -bit symbol of C_1 . This way, the codewords of C_1+C_2 are shifted by a fixed amount as determined by the message \mathbf{m} . Consequently one coset of TTCM codewords is selected by \mathbf{m} to be used for TCQ, which uses the Viterbi algorithm to search for its input sequence of k -bit symbols so that αS is quantized to u and the resulting quantization error $X = u - \alpha S$ satisfies the power constraint $E[X^2] \leq P_X$.

At the decoder, the received signal $Y = X + S + Z$ is first scaled by α , resulting in $\alpha Y = u + (1-\alpha)(-X) + \alpha Z$. Then the input symbols (of n bits each) to TTCM, i.e., the codewords of C_1 , are recovered from αY by an iterative BCJR decoder. Finally, the transmitted message \mathbf{m} is reconstructed by calculating the syndromes of the recovered codewords of C_1 .

Denote the input sequence of n -bit symbols to the TTCM encoder as $\mathbf{I} = [I(0), \dots, I(L-1)]$, where L is the sequence length (or trellis size) and $I(t)$ the t -th input symbol ($0 \leq t \leq L-1$). Whereas the presence of an interleaver greatly boosts the performance (and the effective dimensionality) of the TTCM code over TCM by reducing the number of nearest neighbors (or the probability of error) [44], the TCQ source code suffers because the interleaver significantly increases the number of paths that need to be searched, making the Viterbi algorithm no longer a viable solution to finding the closest codeword u to αS . In [20], [21], the bottom branch of TTCM is simply ignored during TCQ, i.e., \mathbf{I} is only computed from the L symbols passing through the top branch of TTCM during TCQ. But the actual average quantization error $E[X^2]$ includes contributions from both even-indexed symbols from the top branch and odd-indexed symbols from the bottom branch, i.e., $L/2$ symbols from each of the two branches. Specifically, when the rate- n/m code C_2 is systematic (as is chosen in simulations), the samples from the top branch are different than the ones from the bottom in only the $m-n$ parity bits (the n -bit systematic part of each symbol is the same for both branches). This leads to an *extra* quantization error in $E[X^2]$ that is responsible for the degradation of the source code performance in TCQ/TTCM. For example, at $C^* = 1.0$ b/s, Chou reported in [21] a gap of 5.23 dB and 2.07 dB to the capacity in (2) with TCQ/TCM and TCQ/TTCM, respectively. Close examination reveals that the granular gain $g(\Lambda)$ of 256-state TCQ in TCQ/TCM is the normal 1.33 dB (hence $Loss_{SC} = 0.28$ dB according to (4)), but it reduces to only 0.406 dB (with $Loss_{SC} = 1.45$ dB) in TCQ/TTCM. In the mean time, $Loss_{CC}$ equals to 4.95 dB and 0.62 dB in TCQ/TCM and TCQ/TTCM, respectively. These results are included in Table II for comparison purposes.

As the rate gets smaller, the power constraint P_X (hence the quantization error $E[X^2]$) is smaller, the impact of the extra quantization error on $E[X^2]$ becomes more severe. For example, at $C^* = 0.5$ b/s, the extra quantization error (even with the minimum $m-n=1$) causes the granular gain of the source code of Chou *et al.*'s TCQ/TTCM construction to be *negative*, leading to 4.00 dB loss from the corresponding capacity in (2). This result is given in Table III as a benchmark.

B. Turbo-like TCQ/TTCM code construction

The encoder block diagram of our proposed nested turbo code construction is depicted in Fig. 6, where the obvious difference from Fig. 5 is the employment of a new turbo-like TCQ in place of TCQ. The same iterative BCJR decoder for TCQ/TTCM [20], [21] is adopted in our design.

Turbo-like TCQ is chosen as the source code because it has a similar parallel concatenated structure as used in the TTCM channel code. This structure facilitates the nesting of the source code inside the channel code by enabling both

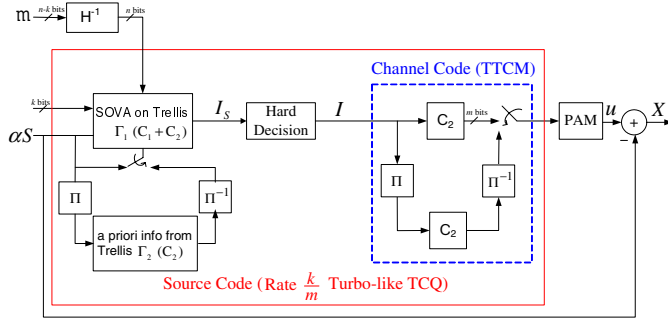


Fig. 6. Block diagram of the proposed turbo-like TCQ/TTCM encoder.

parallel branches of the source code to be taken into account in quantizing αS , hence leading to better source coding performance than in TCQ/TTCM. In this sense, turbo-like TCQ is better suited than TCQ in fulfilling the need for strong source code in our nested turbo code design.

However, we shall see soon that turbo-like TCQ is not turbo TCQ. Its effective dimensionality is still lower than that of TTCM. Dimensionality mismatch thus exists in our turbo-like TCQ/TTCM code design, and there is a performance tradeoff between these two components, meaning that the performance of turbo-like TCQ worsens as the TTCM code becomes stronger and vice versa. Another novelty of our work is optimization of this performance tradeoff by choosing the best percentage (between 50% and 100%) of samples processed by the top branch of the parallel concatenated structure in Fig. 6. The rest of this section gives details on the two new elements in our design, namely, turbo-like TCQ and performance tradeoff between turbo-like TCQ and TTCM, on top and beyond that of [20], [21].

1) *Turbo-Like TCQ*: Recall from Section V-A that the problem with the source code in [20], [21] is that the bottom branch of TTCM is ignored during TCQ. Turbo-like TCQ aims to alleviate this problem by also taking into account the bottom branch in source coding. Thus the main difference between turbo-like TCQ in Fig. 6 and TCQ in Fig. 5 lies in the computation of the input sequences of symbols $\mathbf{I} = [I(0), \dots, I(L-1)]$ to the TTCM encoder. Specifically, we compute the soft-output version of \mathbf{I} , denoted as \mathbf{I}_S , using a soft-output Viterbi algorithm (SOVA) [45] for the TCQ in the top branch; assuming even/odd multiplexing, in the even positions the TCQ metrics are computed from the top branch alone, while in the odd positions the *a priori* information from the bottom branch determines the TCQ metrics as described in more detail next.

Were it not for the bottom branch, turbo-like TCQ would degenerate to TCQ based on Γ_1 . In such case, SOVA-based computation of \mathbf{I}_S would proceed by first setting the n -bit input symbol $I(t)$ to a specific codeword c_2 of C_2 , i.e., $I(t) = c_2 \in \mathcal{C} = \{0, 1, \dots, 2^n - 1\}$, and then computing the soft-output $I_S(t, c_2)$ as the minimal total distortion corresponding to all possible input sequences $\mathbf{I} \in C_1^m$, which denotes the coset of C_1 indexed by the message m . That is

$$I_S(t, c_2) = \min_{\mathbf{I} \in C_1^m, I(t)=c_2} \sum_{l=0}^{L-1} \underbrace{\left\{ |u(l) - \alpha S(l)|^2 \right\}}_{\rho(l)},$$

$$0 \leq t \leq L-1, 0 \leq c_2 \leq 2^n - 1, \quad (6)$$

where $\mathbf{S} = [S(0), \dots, S(L-1)]$ is the length- L sequence of side information, $\mathbf{u} = [u(0), \dots, u(L-1)]$ is the sequence of trellis codewords corresponding to a certain input sequence \mathbf{I} with $I(t) = c_2$, and $\rho(l)$ denotes the distortion metric in TCQ. After computing $I_S(t, c_2)$ for all t ($0 \leq t \leq L-1$) and all $c_2 \in \mathcal{C}$, \mathbf{I}_S takes the matrix form of

$$\begin{bmatrix} I_S(0, 0) & \cdots & I_S(L-1, 0) \\ \vdots & \ddots & \vdots \\ I_S(0, 2^n - 1) & \cdots & I_S(L-1, 2^n - 1) \end{bmatrix}. \quad (7)$$

With turbo-like TCQ, calculation of \mathbf{I}_S in our nested turbo code design is based on both parallel branches. Trellis Γ_1 for the top TCQ source code is constructed by $C_1 + C_2$, while trellis Γ_2 for the bottom branch contains only C_2 . This parallel concatenated structure is necessary for more efficient message transmission (or embedding of the message m in trellis Γ_1), because the message is better protected by the powerful TTCM channel code. In this structure, code C_1 can only be merged on the top branch with C_2 , creating the equivalent Γ_1 trellis, but not in the bottom branch where the interleaver does not allow similar merging.

SOVA-based computation of \mathbf{I}_S requires a new composite distortion metric that takes both branches into account. Assuming even-odd multiplexing in the turbo-like TCQ/TTCM encoder, we puncture the systematic bits at odd indices in trellis Γ_1 and set the distortion metric $\rho_1(t)$ at index t in trellis Γ_1 to

$$\rho_1(t) = \begin{cases} |u(t) - \alpha S(t)|^2, & \text{if } t \text{ is even,} \\ 0, & \text{if } t \text{ is odd.} \end{cases} \quad (8)$$

The distortion from odd indices is provided by trellis Γ_2 in the form of *a priori* information. Borrowing ideas from the initialization step in TTCM decoding, for a systematic C_2 , we compute this *a priori* information at index t , denoted as $\rho_2(t, c_2)$, as the minimal distortion corresponding to the systematic input symbol $I(t) = c_2$ of C_2 and all possible parity symbols $B(t) \in \mathcal{B} = \{0, 1, \dots, 2^{m-n} - 1\}$. That is

$$\rho_2(t, c_2) = \begin{cases} 0, & \text{if } t \text{ is even,} \\ \min_{I(t)=c_2, B(t) \in \mathcal{B}} |u(\Pi(t)) - \alpha S(\Pi(t))|^2, & \text{if } t \text{ is odd,} \end{cases} \quad (9)$$

where $\Pi(t)$ is the same symbol interleaver as used in the TTCM encoder. The *a priori* information $\rho_2(t, c_2)$ is deinterleaved before being fed into trellis Γ_1 . To incorporate both $\rho_1(t)$ and $\rho_2(t, c_2)$ into the computation of $I_S(t, c_2)$, we set $\rho(t) = \rho_1(t) + \rho_2(t, I(t))$ in (6) and end up with

$$I_S(t, c_2) = \min_{\mathbf{I} \in C_1^m, I(t)=c_2} \sum_{l=0}^{L-1} \left\{ \rho_1(l) + \rho_2(l, I(l)) \right\}, \quad 0 \leq t \leq L-1, 0 \leq c_2 \leq 2^n - 1. \quad (10)$$

Note that the deinterleaving needed for the last equation has already been taken into account in equation (9). After running the SOVA with (10) on trellis Γ_1 , we output \mathbf{I}_S before hard thresholding \mathbf{I}_S to \mathbf{I} with

$$I_S(t) = \arg \min_{c_2 \in \mathcal{C}} I_S(t, c_2), \quad 0 \leq t \leq L-1. \quad (11)$$

Turbo-like TCQ is motivated by the need to take into account distortion from quantizers in both parallel branches of Fig. 6. It is not turbo TCQ mainly because quantization is not done iteratively (so as to avoid performance degradation as observed in [31] after the first iteration). Without iterative quantization (or source encoding), the distortion from the bottom-branch TCQ can only be included in the form of *a priori* information as done in (9). This in turn limits the improvement of turbo-like TCQ/TTCM over TCQ/TTCM in terms of source coding performance. The effective dimensionality of the turbo-like TCQ source code is still much lower than that of the TTCM channel code.

We call our source code turbo-like TCQ because it has the parallel concatenated structure with interleavers Π and Π^{-1} , and more importantly, the operation in (10) essentially implements the first iteration of turbo TCQ, which takes advantage of the fact [31] that turbo TCQ generally improves upon TCQ at the first iteration before losing ground at subsequent iterations.

2) *Performance Tradeoff Between Turbo-Like TCQ and TTCM*: Let T be the percentage of samples chosen by the multiplexer from the top branch of the parallel concatenated structure (for both turbo-like TCQ and TTCM). With the default setting of even-odd multiplexing in Fig. 6, $T = 50\%$. But T could be varied from 50% to 100%. Note that the distortion metric $\rho_1(t)$ in (8) and the *a priori* information $\rho_2(t, c_2)$ in (9) can be easily modified when $T \neq 50\%$.

On one hand, as T is increased from 50% to 100%, the turbo effect due to the presence of the interleaver is gradually reduced, causing the performance of the TTCM code to deteriorate. In fact, when $T = 100\%$, TTCM degrades to TCM, leading to the worst channel coding performance. On the other hand, increasing T guarantees improved source coding performance in our nested turbo code design. This is because the *a priori* information $\rho_2(t, c_2)$ accounted in turbo-like TCQ for samples from the bottom branch is still not as reliable as the actual distortion contributed by these samples to the final average quantization error $E[X^2]$. Higher T means less samples from the bottom branch, and hence less unreliable information in the distortion metric of turbo-like TCQ. When $T = 100\%$, turbo-like TCQ degenerates to TCQ, giving the best source coding performance; in this case, our turbo-like TCQ/TTCM code becomes a TCQ/TCM code.

Thus, even with the inclusion of $\rho_2(t, c_2)$, the extra quantization error also exists in turbo-like TCQ, although it is smaller than that in TCQ/TTCM. Increasing T reduces the number of samples contributing to this extra quantization error, making it even smaller. Overall, by increasing T from 50% to 100%, we are making the TTCM channel code weaker but the turbo-like TCQ source code stronger. The parameter T hence offers a means of trading off the performance of the source code and that of the channel code in our nested design. The best performance tradeoff can be reached by searching for the optimal percentage T^* between 50% and 100% that gives the minimal gap from the capacity-achieving SNR.

Remarks:

- Because the above performance tradeoff is rooted in dimensionality mismatch between the source and channel coding components in any nested design for Costa cod-

ing, it also applies to the TCQ/TTCM code construction. This means results in [20], [21] with TCQ/TTCM and $T = 50\%$ can be improved by simply searching for the best T^* between 50% and 100%. With $T = 100\%$, the TCQ/TTCM code also becomes the simple TCQ/TCM code.

With $T = 50\%$ (or default odd-even multiplexing), because turbo-like TCQ gives better source coding performance in our nested turbo code design than the TCQ/TTCM code construction, meaning dimensionality mismatch is less severe, we expect the optimal T^* in our design to be less than that for the TCQ/TTCM construction. In fact, if there were turbo TCQ (with the same effective dimensionality as TTCM), the optimal T^* would be 50% in an ideal nested turbo code design, i.e., no performance tradeoff would be needed!

VI. SIMULATION RESULTS

Picking the appropriate code rate parameters (n, k, m) , we simulate our code design for transmission rates of 2.0, 1.0, and 0.5 b/s. For these transmission rates, both convolutional codes C_1 and C_2 are chosen as the constraint-length four Ungerboeck code [19]. C_2 needs to be systematic to fit the turbo algorithm. If C_1 were also systematic, there would exist error propagation when recovering the original message \mathbf{m} via computing the syndromes, since the parity-check polynomials would have infinite weights. Therefore, non-systematic C_1 is chosen.

The code C_2 is mapped to a finite constellation, which we call the basic constellation. The side information S can have an arbitrary large magnitude, therefore we replicate the basic constellation infinitely so that S never lies in the overload region of the quantizer (so as to satisfy the power constraint). The quantizer thus selects a copy of the basic constellation codeword which lies nearest to S .

We evaluate our Costa code's performance by its BER at a certain SNR. First we look into the effect of varying the uniform quantization stepsize q [18] in TCQ. Our experiments indicate little performance difference by using different q 's, and this is true for different T 's and transmission rates. Thus, for results reported in the following, q is set to 1.0 for all transmission rates. In addition, all results are based on 256-state TCQ and a BER of 10^{-5} .

A. Simulation Results at 2.0 b/s

For a transmission rate of $C^* = 2.0$ b/s, the rate $\frac{k}{n}$ of code C_1 has to be chosen such that $n - k = 2$. Hence C_1 is chosen as a non-systematic rate- $\frac{1}{3}$ code. The polynomials are found using a computer search over all possible rate- $\frac{1}{3}$ codes [19]. The generator polynomials for C_1 are $g_0(D) = 2$, $g_1(D) = 4$ and $g_2(D) = 23$ in octal notation. Code C_2 is chosen as a systematic rate- $\frac{3}{4}$ code with feedback polynomial $h_0(D) = 23$, and feedforward polynomials $h_1(D) = 10$, $h_2(D) = 0$, and $h_3(D) = 0$. The codewords of C_2 are mapped to a 16-PAM constellation.

Fig. 7 shows the performance gap (in dB) of our turbo-like TCQ/TCM code to the capacity-achieving SNR vs. the percentage T when the trellis (or interleaver) length is $L = 50,000$. It also depicts the performance tradeoff between the

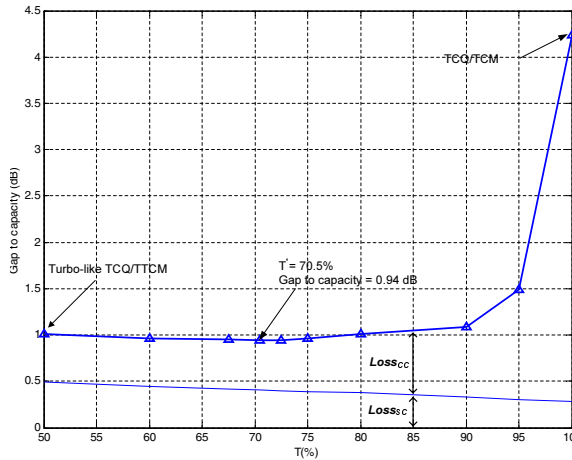


Fig. 7. Performance gap (in dB) of our turbo-like TCQ/TTCM code to the capacity-achieving SNR vs. the percentage T when $C^* = 2.0$ b/s and $L = 50,000$. The performance tradeoff between the source and channel coding components is depicted by partitioning the total performance gap into the sum of $Loss_{CC}$ and $Loss_{SC}$.

source and channel coding components by partitioning the total performance gap into the sum of $Loss_{SC}$ and $Loss_{CC}$. It is seen that $Loss_{SC}$ in (4) decreases as T increases, but it is offset by the increase in $Loss_{CC}$. Our best results indicate a performance gap to capacity of only 0.94 dB, which is achieved with $T^* = 70.5\%$.

Simulations are also conducted with $L = 10,000$, and results with $T = 50\%$ and the corresponding T^* s for both $L = 10,000$ and $L = 50,000$ are tabulated in Table I, where the performance of the simple TCQ/TTCM scheme of [21] is also included as a benchmark⁸. In addition, for comparison purposes, we give the performance of the TCQ/TTCM scheme of [20], [21] with $T = 50\%$ and with percentage optimization for both $L = 10,000$ and $L = 50,000$. We note that when $T = 100\%$, both our nested turbo code and the TCQ/TTCM code degenerate to TCQ/TCM.

Both Fig. 7 and Table I reveal the sizable performance gain obtained by using T^* (with optimal performance tradeoff between turbo-like TCQ and TTCM in our nested turbo code) over $T = 100\%$ (or TCQ/TCM). Using T^* offers a small performance gain over using $T = 50\%$ at the relatively high transmission rate of $C^* = 2.0$ b/s.

Table I also shows that our nested turbo code always outperforms the TCQ/TTCM code [20], [21], although the improvement margin is small. In addition, for the same L , the optimal T^* is smaller in our nested turbo code than that in the TCQ/TTCM code. This is consistent with the fact that, when nested inside TTCM, turbo-like TCQ (with 1.15 dB granular gain) performs better than TCQ (with 0.99 dB granular gain). The granular loss due to nesting (or the extra quantization error) in TCQ/TTCM is $1.27 - 0.99 = 0.28$ dB, but it is reduced to $1.27 - 1.15 = 0.12$ dB in turbo-like TCQ/TTCM (with optimal T^*).

⁸We note that for the TCQ/TCM scheme reported in [17], $g(\Lambda) = 1.36$ dB with $Loss_{SC} = 0.18$ dB, $Loss_{CC} = 3.57$ dB, leading to $Loss_{Total} = 3.75$ dB.

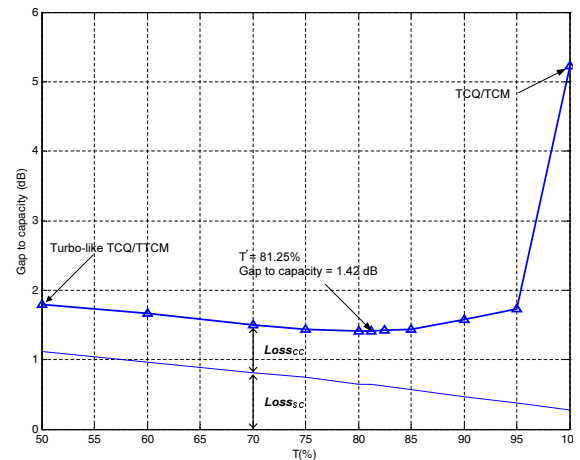


Fig. 8. Performance gap (in dB) of our turbo-like TCQ/TCM code to the capacity-achieving SNR vs. the percentage T when $C^* = 1.0$ b/s and $L = 50,000$. The performance tradeoff between the source and channel coding components is also depicted by partitioning the total performance gap into the sum of $Loss_{CC}$ and $Loss_{SC}$.

B. Simulation Results at 1.0 b/s

For the transmission rate of $C^* = 1.0$ b/s, C_1 is chosen as a non-systematic rate- $\frac{1}{2}$ code with generator polynomials $g_0(D) = 10$ and $g_1(D) = 23$ in octal notation. Code C_2 is a rate- $\frac{2}{3}$ systematic code with feedback polynomial $h_0(D) = 23$, and feedforward polynomials $h_1(D) = 10$, and $h_2(D) = 0$. The codewords of C_2 are mapped to an 8-PAM constellation.

Fig. 8 shows the performance gap (in dB) of our turbo-like TCQ/TCM code to the capacity-achieving SNR vs. the percentage T when the trellis (or interleaver) length is $L = 50,000$. The performance tradeoff between the source and channel coding components is also depicted by partitioning the total performance gap into the sum of $Loss_{CC}$ and $Loss_{SC}$. Our best results indicate a performance gap to capacity of only 1.42 dB, which is achieved with $T^* = 81.25\%$.

Table II with $C^* = 1.0$ b/s is the counterpart of Table I with $C^* = 2.0$ b/s. Compared to using $T = 50\%$, using the optimal percentage T^* improves the performance of our nested turbo code at $C^* = 1.0$ b/s by more than 0.3 dB for both $L = 10,000$ and $L = 50,000$. Moreover, it is seen from Table II that at $T = 50\%$, our turbo-like TCQ/TTCM code outperforms the TCQ/TTCM code [20], [21] by more than 0.2 dB. The improvements from using turbo-like TCQ with optimal T^* in our nested turbo code lead to 0.54-0.60 dB overall gain over the TCQ/TTCM code.

Results in Table II also indicate that optimizing T is more effective at lower rates. This is because the extra quantization error in the quantizer has higher impact on its granular gain at lower rates. A comparison of $g(\Lambda)$ in Tables I and II indeed reveals that the granular gain at $T = 50\%$ is much lower when $C^* = 1.0$ b/s than when $C^* = 2.0$ b/s. The granular loss due to nesting (or the extra quantization error) in TCQ/TTCM is $1.33 - 1.04 = 0.29$ dB in turbo-like TCQ/TTCM (with optimal T^*).

TABLE I
PERFORMANCE GAP (IN dB) TO THE CAPACITY-ACHIEVING SNR FOR DIFFERENT CODE DESIGNS WHEN THE RATE IS 2.0 B/S. A RATE-1/3/4
CONCATENATED CODE AND A 1-D PAM CONSTELLATION WITH STEPSIZE $q = 1.0$ IS USED.

	L = 10,000					L = 50,000			
	TCQ/TCM (T=100%)	TCQ/TTCM T=50% T*=72.5%		turbo-like TCQ/TTCM T=50% T*=70%		TCQ/TTCM T=50% T*=73%		turbo-like TCQ/TTCM T=50% T*=70.5%	
$g(\Lambda)$	1.27	0.99	1.12	1.07	1.15	0.99	1.12	1.07	1.15
$Loss_{SC}$	0.28	0.58	0.44	0.49	0.41	0.58	0.44	0.49	0.41
$Loss_{CC}$	3.96	0.64	0.65	0.60	0.61	0.56	0.58	0.52	0.53
$Loss_{Total}$	4.24	1.22	1.09	1.09	1.02	1.14	1.02	1.01	0.94

TABLE II
PERFORMANCE GAP (IN dB) TO THE CAPACITY-ACHIEVING SNR FOR DIFFERENT CODE DESIGNS WHEN THE RATE IS 1.0 B/S. A RATE-1/2/3
CONCATENATED CODE AND A 1-D PAM CONSTELLATION WITH STEPSIZE $q = 1.0$ IS USED.

	L = 10,000					L = 50,000			
	TCQ/TCM (T=100%)	TCQ/TTCM T=50% T*=82.5%		turbo-like TCQ/TTCM T=50% T*=80%		TCQ/TTCM T=50% T*=83.5%		turbo-like TCQ/TTCM T=50% T*=81.25%	
$g(\Lambda)$	1.33	0.406	0.95	0.67	1.04	0.406	0.97	0.67	1.04
$Loss_{SC}$	0.28	1.45	0.76	1.12	0.65	1.45	0.74	1.12	0.65
$Loss_{CC}$	4.95	0.62	0.87	0.71	0.88	0.57	0.79	0.67	0.77
$Loss_{Total}$	5.23 [21]	2.07 [21]	1.63	1.86	1.53	2.02	1.53	1.79	1.42

C. Simulation Results at 0.5 b/s

In order to get a fractional transmission rate, we use a 2-D QAM constellation. C_1 is chosen to be Ungerboeck's systematic rate- $\frac{2}{3}$ code for a QAM constellation, with the feedback polynomial being $h_0(D) = 23$ and the feedforward polynomials being $h_1(D) = 4$ and $h_2(D) = 16$. In order to avoid infinite weight, the parity-check matrix of C_1 used for calculating the syndromes is pre-multiplied by $h_0(D)$. Code C_2 is chosen as a rate- $\frac{3}{4}$ with $h_0(D) = 23$, $h_1(D) = 4$, $h_2(D) = 16$, and $h_3(D) = 2$. The mapping from the codewords of C_2 to the constellation is based on Ungerboeck's partitioning of a 16-QAM constellation [19].

Fig. 9 shows the performance gap (in dB) of our turbo-like TCQ/TCM code to the capacity-achieving SNR vs. the percentage T when the trellis (or interleaver) length is $L = 50,000$. The performance tradeoff between the source and channel coding components is again depicted by partitioning the total performance gap into the sum of $Loss_{CC}$ and $Loss_{SC}$. Our best results indicate a performance gap to capacity of 2.65 dB, which is achieved with $T^* = 85\%$.

From Table III, it is seen that compared to using $T = 50\%$, the performance gain from using the optimal T^* is more than 0.5 dB. Optimizing the percentage T is thus more effective as the rate gets lower. In fact, it significantly improves the TCQ/TTCM code, whose source coding performance at $T = 50\%$ is so bad that the corresponding granular gain is negative (worse than that of a scalar quantizer). The granular loss due to nesting (or the extra quantization error) in TCQ/TTCM is $1.32+0.35=1.67$ dB, but it is reduced to $1.32-0.92=0.40$ dB in turbo-like TCQ/TTCM (with optimal T^*).

Despite the effectiveness with using the optimal T^* , the gap to the capacity at $C^* = 0.5$ b/s remains at 2.65 dB. Hence, in the low rate regime, our nested lattice scheme cannot compete with recent code designs [23], [25], [28] geared towards lower

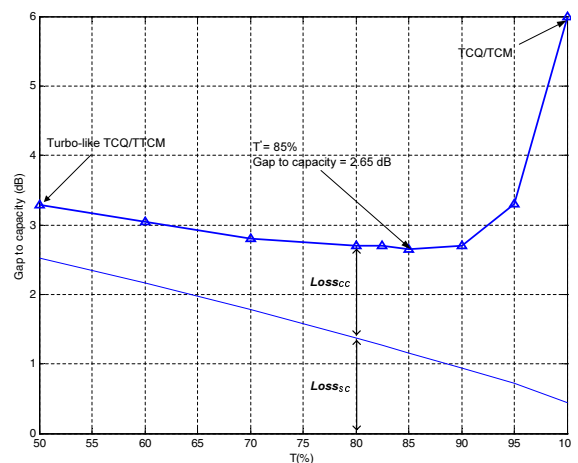


Fig. 9. Performance gap (in dB) of our turbo-like TCQ/TCM code to the capacity-achieving SNR vs. the percentage T when $C^* = 0.5$ b/s and $L = 50,000$. The performance tradeoff between the source and channel coding components is again depicted by partitioning the total performance gap into the sum of $Loss_{CC}$ and $Loss_{SC}$.

transmission rates. This is due to the coupling between the source and channel codes in our nested design of Fig. 6, where the TTCM channel code is part of the whole turbo-like TCQ source code. The inherent performance tradeoff between the source and channel codes is the main weakness that is responsible for the non-competitive performance of our nested design at low rates.

VII. CONCLUSIONS

We have introduced a practical nested turbo scheme for the Costa problem. Compared with the TCQ/TTCM scheme of [20], [21], where TCQ is weakened by the employment of TTCM, our turbo-like TCQ/TTCM scheme improves the

TABLE III
PERFORMANCE GAP (IN dB) TO THE CAPACITY-ACHIEVING SNR FOR DIFFERENT CODE DESIGNS WHEN THE RATE IS 0.5 B/S. A RATE-2/3/4
CONCATENATED CODE AND A 2-D QAM CONSTELLATION WITH STEPSIZE $q = 1.0$ IS USED.

	L = 10,000					L = 50,000			
	TCQ/TCM ($T=100\%$)	TCQ/TTCM $T=50\%$ $T^*=90\%$		turbo-like TCQ/TTCM $T=50\%$ $T^*=82.5\%$		TCQ/TTCM $T=50\%$ $T^*=92.5\%$		turbo-like TCQ/TTCM $T=50\%$ $T^*=85\%$	
$g(\lambda)$	1.32	-0.35	0.96	0.09	0.85	-0.35	1.05	0.09	0.92
$Loss_{SC}$	0.42	3.19	1.08	2.52	1.27	3.19	0.92	2.52	1.15
$Loss_{CC}$	5.58	0.81	1.86	0.86	1.53	0.73	1.87	0.77	1.50
$Loss_{Total}$	6.00 [21]	4.00	2.94	3.38	2.80	3.92	2.79	3.29	2.65

source coding performance by using a parallel concatenated structure and taking into account quantization errors from its top and bottom branches. In addition, aiming to balance the performance tradeoff between turbo-like TCQ and TTCM and hence obtain the best overall performance, we optimize the percentage of samples chosen by the multiplexer from the top branch of our nested turbo code design. Simulations show that our nested turbo code provides the best performance so far at medium-to-high rates (e.g., ≥ 1.0 b/s). For example, at 1.0 b/s, Eggers *et al.*'s scalar scheme [16], Chou *et al.*'s TCQ/TTCM code [20], [21], and our nested turbo code perform 3.5 dB, 2.07 dB, and 1.42 dB, respectively, away from the Shannon capacity.

Whenever dimensionality mismatch is present in practical nested Costa design, the tradeoff between the source and channel coding components limits its overall performance. However, in the high rate regime, our turbo-like TCQ/TTCM code works well and it seems to be the best option. In the low rate regime (e.g., < 1.0 b/s), Costa code design becomes more challenging (as seen from Section III and results in Table III); it is better to adopt Erez and ten Brink's design [23] in the case; the best alternative is to use superposition coding as done in [25] or combined source-channel coding as advocated in [28] without nesting. Overall, because our work and those of [23], [25], [28] target at different rate regimes, they are complementary to each other.

Finally, we note that application of Costa coding has been addressed in [15-16,44-47] for data-hiding and in [50] for digital broadcast, and that the design presented in this work has been applied to coding for MIMO broadcast channels in [51].

ACKNOWLEDGEMENT

The authors gratefully acknowledge Dr. Jim Chou for helping them understand the details of [20], [21] and Prof. Sandeep Pradhan for fruitful discussions. They also sincerely thank the reviewers for offering constructive comments, which have significantly improved the quality of this paper.

REFERENCES

- [1] S. Gelfand and M. Pinsker, "Coding for channel with random parameters," *Probl. Contr. Inform. Theory*, vol. 9, pp. 19-31, 1980.
- [2] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, pp. 439-441, May 1983.
- [3] C. Shannon, "Channels with side information at transmitter," *IBM J. Res. & Dev.*, pp. 289-293, 1958.
- [4] A. Kuznetsov and B. Tsybakov, "Coding in a memory with defective cells," *Probl. Pered. Inform.*, vol. 10, pp. 52-60, Apr.-June 1974.

- [5] F. Petitcolas, R. Anderson, and M. Kuhn, "Information hiding—A survey," *Proc. IEEE*, vol. 87, pp. 1062-1078, July 1999.
- [6] I. Cox, M. Miller, and A. McKellips, "Watermarking as communications with side information," *Proc. IEEE*, vol. 87, pp. 1127-1141, July 1999.
- [7] T. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. 18, pp. 2-14, Jan. 1972.
- [8] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1691-1706, July 2003.
- [9] H. Weingarten, U. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inform. Theory*, vol. 52, pp. 3936-3964, Sep. 2006.
- [10] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inform. Theory*, vol. 25, pp. 306-311, May 1979.
- [11] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1250-1276, June 2002.
- [12] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice-strategies for cancelling known interferences," *IEEE Trans. Inform. Theory*, vol. 51, pp. 3820-3833, Nov. 2005.
- [13] A. Host-Madsen, "Capacity bounds for cooperative diversity," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1522-1544, Apr. 2006.
- [14] C. Ng and A. Goldsmith, "Transmitter cooperation in ad-hoc wireless networks: Does dirty-paper coding beat relaying?" in *Proc. ITW*, Oct. 2004, pp. 277-282.
- [15] B. Chen and G. Wornell, "Achieving performance of digital watermarking systems," in *Proc. IEEE Int. Conf. Multimedia Comput. Syst.*, June 1999, vol. 1, pp. 13-18.
- [16] J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod, "Scalar costa scheme for information embedding," *IEEE Trans. Signal Processing*, vol. 51, pp. 1003-1019, Apr. 2003.
- [17] W. Yu, D. Varodayan, and J. Cioffi, "Trellis and convolutional precoding for transmitter-based interference pre-subtraction," *IEEE Trans. Commun.*, vol. 53, pp. 1220-1230, July 2005.
- [18] M. Marcellin and T. Fischer, "Trellis-coded quantization of memoryless and Gauss-Markov sources," *IEEE Trans. Commun.*, vol. 38, pp. 82-93, Jan. 1990.
- [19] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inform. Theory*, vol. 28, pp. 55-67, Jan. 1982.
- [20] J. Chou, S. Pradhan, and K. Ramchandran, "Turbo coded trellis-based constructions for data embedding: Channel coding with side information," in *Proc. 35th Asilomar Conf. Signals, Syst., Computers*, Nov. 2001, pp. 305-309.
- [21] J. Chou, "Channel coding with side information: Theory, practice and applications," Ph.D. dissertation, University of California at Berkeley, Berkeley, CA, 2002.
- [22] P. Robertson and T. Würz, "Bandwidth-efficient turbo trellis-coded modulation using punctured component codes," *IEEE J. Selected. Areas Commun.*, vol. 16, pp. 206-218, Feb. 1998.
- [23] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Inform. Theory*, vol. 51, pp. 3417-3432, Oct. 2005.
- [24] H. Jin, A. Khandekar and R. McEliece, "Irregular repeat-accumulate codes," in *Proc. 2nd Intl. Symp. Turbo Codes Related Topics*, Sep. 2000, pp. 1-8.
- [25] A. Bennatan, D. Burshtein, G. Caire, and S. Shamai, "Superposition coding for side-information channels," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1872-1889, May 2006.
- [26] R. Gallager, *Low Density Parity Check Codes*. Cambridge, MA: MIT Press, 1963.

- [27] D. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory*, vol. 45, pp. 399-431, Mar. 1999.
- [28] Y. Sun, A. Liveris, V. Stankovic, and Z. Xiong, "Near-capacity dirty-paper code designs based on TCQ and IRA codes," in *Proc. ISIT*, Sep. 2005, pp. 184-188.
- [29] A. Bennatan and D. Burshtein, "Design and analysis of nonbinary LDPC codes for arbitrary discrete-memoryless channels," *IEEE Trans. Inform. Theory*, vol. 52, pp. 549-583, Feb. 2006.
- [30] Z. Xiong, V. Stanković, S. Cheng, A. Liveris, and Y. Sun, "Source-channel coding for algebraic multiterminal binning," in *Proc. ITW*, Oct. 2004, pp. 318-323.
- [31] V. Chappelier, C. Guillemot, and S. Marinkovic, "Turbo trellis-coded quantization," in *Proc. 3th Intl. Symp. Turbo Codes*, Sep. 2003, pp. 51-54.
- [32] T. Murayama, "Thouless-Anderson-Palmer approach for lossy compression," *Physical Review E*, vol. 69, pp. 035105(1)-035105(4), 2004.
- [33] M. Wainwright and E. Maneva, "Lossy source coding via message-passing and decimation over generalized codewords of LDGM codes," in *Proc. ISIT*, Sep. 2005, pp. 1493-1497.
- [34] R. Barron, B. Chen, and G. Wornell, "The duality between information embedding and source coding with side information and some applications," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1159-1180, May 2003.
- [35] S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding with side information," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1181-1203, May 2003.
- [36] A. Cohen and A. Lapidot, "The Gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1639-1667, June 2002.
- [37] M. Eyuboglu and D. Forney Jr., "Lattice and trellis quantization with lattice- and trellis-bounded codebooks—High-rate theory for memoryless sources," *IEEE Trans. Inform. Theory*, vol. 39, pp. 46-59, Jan. 1993.
- [38] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Trans. Commun.*, vol. 44, pp. 1261-1271, Oct. 1996.
- [39] J. Conway and N. Sloane, *Sphere Packings, Lattices and Groups*. New York: Springer Verlag, 1998.
- [40] "Special issue on codes on graphs and iterative algorithms," *IEEE Trans. Inform. Theory*, vol. 47, Feb. 2001.
- [41] P. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. Inform. Theory*, vol. 28, pp. 139-149, Mar. 1982.
- [42] V. Tarokh, A. Vardy, and K. Zeger, "Universal bounds on the performance of lattice codes," *IEEE Trans. Inform. Theory*, vol. 45, pp. 670-681, Mar. 1999.
- [43] Y. Yang, S. Cheng, Z. Xiong, and W. Zhao, "Wyner-Ziv coding based on TCQ and LDPC codes," to be published.
- [44] D. Forney Jr. "Performance and complexity," Shannon Lecture, 1995.
- [45] J. Hagenauer and P. Hoeher, "A Viterbi algorithm with soft-decision outputs and its applications," in *Proc. Globecom*, Nov. 1989, pp. 1680-1686.
- [46] J. Chou and K. Ramchandran, "Robust turbo-based data hiding for image and video sources," in *Proc. ICIP*, Sep. 2002, vol. 2, pp. 133-136.
- [47] M. Miller, G. Doerr, and I. Cox, "Applying informed coding and embedding to design a robust high-capacity watermark," *IEEE Trans. Image Processing*, vol. 13, pp. 792-807, June 2004.
- [48] P. Comesana, F. Perez-Gonzalez, and F. Willems, "Applying Erez and ten Brink's dirty paper codes to data-hiding," in *Proc. SPIE EI Security, Steganography, Watermarking Multimedia Contents VII*, Jan. 2005, pp. 298-307.
- [49] Y. Yang, Y. Sun, V. Stanković, and Z. Xiong, "Image data-hiding based on capacity-approaching dirty-paper coding," in *Proc. SPIE EI: Security, Steganography, Watermarking Multimedia Contents VIII*, Jan. 2006, vol. 6072, pp. 607214-1-607214-11.
- [50] J. Kusuma and K. Ramchandran, "Communicating by cosets and applications to broadcast," in *Proc. CISS*, Mar. 2002 [CD-ROM].
- [51] M. Uppal, V. Stankovic, and Z. Xiong, "Code design for MIMO broadcast channels," in *Proc. ISIT*, July 2006, pp. 2627-2631.



Yong Sun received the B.S. and M.S. degrees in Department of Electrical Engineering, Tsinghua University, Beijing, China, in 1998 and 2001, respectively, and the Ph.D. degree in Electrical Engineering from Texas A&M University, College Station, TX, in 2005. He is now a Lead R&D Engineer in Lemko Corporation, Schaumburg, IL. His research interests include source-channel coding, wireless communications/networks, and signal processing.



Momin Uppal received the B.S. degree in Electronic Engineering with highest distinction from GIK Institute of Engineering Sciences and Technology, Pakistan, in 2002. He received the M.S. degree in Electrical Engineering from Texas A&M University in 2006, where he is currently pursuing the Ph.D. degree. His research interests include dirty-paper coding, broadcast channels, and cooperative communications.



tion, and wireless communications.

Angelos D. Liveris (S'00-M'05) received the Diploma in electrical and computer engineering from the National Technical University of Athens, Athens, Greece in 1998, and the Ph.D. degree in electrical engineering from Texas A&M University, College Station, TX, in 2004. Since 2005 he has been with Microwave Networks Inc, 4000 Greenbriar, Stafford, TX 77477. He is co-recipient of the 2006 IEEE Signal Processing Magazine Best Paper Award. His research interests include distributed source coding, channel coding with side information, and wireless communications.



Samuel Cheng (S'01-M'04) received the B.S. degree in Electrical and Electronic Engineering from the University of Hong Kong in 1995, and the M.Phil. degree in Physics and the M.S. degree in Electrical Engineering from Hong Kong University of Science and Technology and the University of Hawaii, Honolulu, in 1997 and 2000, respectively. He received the Ph.D. degree in Electrical Engineering from Texas A&M University in 2004.

Dr. Cheng worked in Microsoft Asia, China, and Panasonic Technologies Company, New Jersey, in the areas of texture compression and digital watermarking during the summers of 2000 and 2001. In 2004, he joined Advanced Digital Imaging Research, a research company based near Houston, Texas, as a Research Engineer to perform biomedical imaging research and was promoted to Senior Research Engineer the next year. He has several patent submissions and has been awarded three US patents in the area of digital watermarking. Since 2006, he has been with the Department of Electrical and Computer Engineering at the University of Oklahoma, where he is an Assistant Professor. He is co-recipient of the 2006 IEEE Signal Processing Magazine Best Paper Award. His research interests include information theory, image/signal processing, and pattern recognition.



Vladimir Stanković (M'03) received the Dr.-Ing. degree in 2003 from the University of Leipzig, Germany. From 2003 to 2006, he was with the Department of Electrical and Computer Engineering at Texas A&M University, College Station, first as a Postdoctoral Research Associate and then as a Research Assistant Professor. He is currently a Lecturer at the Department of Electronic and Electrical Engineering at the University of Strathclyde, Glasgow. His research interests include multimedia networking, network information theory, and wire-

less communications.



Zixiang Xiong (S'91-M'96-SM'02-F'07) received the Ph.D. degree in Electrical Engineering in 1996 from the University of Illinois at Urbana-Champaign.

From 1995 to 1997, he was with Princeton University, first as a visiting student, then as a research associate. From 1997 to 1999, he was with the University of Hawaii. Since 1999, he has been with the Department of Electrical and Computer Engineering at Texas A&M University, where he is a professor. He spent the summers of 1998 and 1999

at Microsoft Research, Redmond, WA. His research interests are network information theory, code designs and applications, networked multimedia, and genomic signal processing.

Dr. Xiong received a National Science Foundation Career Award in 1999, an Army Research Office Young Investigator Award in 2000, and an Office of Naval Research Young Investigator Award in 2001. He also received the 2006 IEEE Signal Processing Magazine best paper award. He served as associate editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (1999-2005), the IEEE TRANSACTIONS ON IMAGE PROCESSING (2002-2005), and the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2002-2006). He is currently an associate editor for the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS (Part B) and a member of the multimedia signal processing technical committee of the IEEE Signal Processing Society. He is the publications chair of GENSIPS'06 and ICASSP'07 and the technical program committee co-chair of ITW'07.