A Lego-Brick Approach to Coding for Asymmetric Channels and Channels with State

Nadim Ghaddar*, Shouvik Ganguly†, Lele Wang‡, and Young-Han Kim*
*University of California San Diego, La Jolla, CA 92093, USA, {nghaddar, yhk}@ucsd.edu
†XCOM Labs, Inc., San Diego, CA 92121, USA, sganguly@xcom-labs.com
‡University of British Columbia, Vancouver, BC V6T1Z4, Canada, lelewang@ece.ubc.ca

2021 IEEE International Symposium on Information Theory (ISIT) | 978-1-5386-8209-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISIT45174.2021.9518194

***Abstract*—Coding schemes for asymmetric channels and channels with state are developed starting from a pair of linear codes designed for symmetric channels. Guarantees on the block error rate performance of the coding schemes are derived in terms of the parameters of the constituent codes. Assuming the constituent codes satisfy some properties on the rate, the error probability, and the distribution of the Hamming distance to decoded sequences, the performance guarantees hold irrespective of other properties of the codes. This would allow one to leverage commercial off-the-shelf codes for point-to-point symmetric channels to design codes for asymmetric channels and channels with state known noncausally at the encoder.**

I. Introduction

The problem of coding for channels with a state that is known noncausally only at the encoder was initially formulated by Gelfand and Pinsker in [1], and its capacity was derived in the same paper. Their random coding scheme based on multicoding and joint typicality encoding and decoding has had implications for several information-theoretic problems, most notably the extension to Marton coding for two-user broadcast channels with both private and common messages [2], which yields tight inner bounds for all classes of discrete memoryless broadcast channels with known capacity regions. Gelfand and Pinsker's random coding scheme did not assume any structure in the code ensembles. In [3], Padakandla and Pradhan show that the same capacity can be achieved using random *nested linear* codes with joint typicality encoder and decoder. Another important implication of Gelfand and Pinsker's coding scheme, which we exploit in this paper, is its specialization to a capacity-achieving code for asymmetric channels by setting the state as the all-zero sequence.

The formulation of the Gelfand–Pinsker problem and its Gaussian counterpart, dirty paper coding [4], spurred the quest for practical codes to solve the Gelfand–Pinsker problem as well as the two-user broadcast problem. Mondelli et al. [5] presented a scheme to achieve Marton's inner bound for broadcast channels using polar codes based on a chaining or block Markov construction, extending earlier work on the lossy source coding problem [6] and on coding for compound channels [7]. Bennatan et al. [8] studied the binary-input binary-state as well as the Gaussian-input Gaussian-state Gelfand–Pinsker problems and applied nested lattice codes. Other approaches explored for practical Gelfand–Pinsker codes include variants of LDPC codes [9], [10], trellis codes [11], and lattice codes [12]. In the context of asymmetric channel coding, the "distribution shaping" problems have been studied among others by [13]–[15]. In particular, Mondelli et al. [16] performed a comparative study of various coding approaches for asymmetric channels using polar codes and spatially coupled codes.

In this paper, we design a coding scheme for the Gelfand–Pinsker problem by taking a *Lego-brick approach*: We treat encoders and decoders for symmetric channels as *black boxes* or *Lego bricks* and assemble them with other simple Lego bricks, such as dithers and interleavers, to form new coding schemes. We ask the following question: Without knowing the details of the encoding-decoding operations, can we express the performance of the proposed coding scheme in terms of simple properties of the constituent codes, such as the code rate, the blocklength, the probability of error, and certain easy-to-verify properties of the codewords? Such an approach was explored for the problems of coding for the binary symmetric channel and Slepian–Wolf coding [17], for a doubly symmetric binary source by Wyner [18], for general channel coding and general Slepian–Wolf coding in [19], [20], and for broadcast and multiple access channels in [21].

More specifically, the proposed coding scheme for the Gelfand–Pinsker problem starts from a pair of linear channel codes. For one code, we consider its block error rate performance in a desired symmetric channel. For the other code, we consider the distribution of the Hamming distance to the decoded sequences, a property referred to as the *decoding distance spectrum* of the code. These properties can be easily verified for off-the-shelf channel codes by simulations, and are sufficient to provide performance guarantees of the proposed scheme without knowing the specifics of the encoders and decoders. The proposed scheme can be specialized to a coding scheme for asymmetric channels by setting the state sequence to be the all-zero sequence. Let us now set up the notation for our discussion.

Channel Coding Problem. Consider a binary-input discrete memoryless channel (B-DMC) $p(y|x)$ with an input alphabet $\mathcal{X} = \{0, 1\}$ and a finite output alphabet $\mathcal{Y}$. A (k, n, ϵ) code (f, ϕ) for this channel consists of

- a codebook $\mathcal{C} \subseteq \{0, 1\}^n$ of size $|\mathcal{C}| = 2^k$,
- an encoder $f : [2^k] \to \mathcal{C}$ that maps each message $m \in$

$[2^k]$ to a codeword $x^n = f(m)$.
- a decoder $\phi : \mathcal{Y}^n \to \mathcal{C}$ that assigns a codeword estimate $\hat{x}^n = \phi(y^n)$ to each received sequence y^n.

The rate of the code is $R = k/n$. Assuming the codeword X^n is uniformly distributed over the codebook $\mathcal{C}$, the average probability of error of the code is $\mathsf{P}\{\hat{X}^n \neq X^n\} = \epsilon$.

A *linear* channel code can be defined by its parity check matrix $H_{(n-k)\times n}$ and its decoding function ϕ. We will refer to such a linear code as a (k, n, ϵ) code (H, ϕ). For notational convenience, we introduce the augmented parity check matrix defined by

$$\overline{H}_{n\times n} = \begin{bmatrix} \mathbf{0} \\ H \end{bmatrix},$$

where $\mathbf{0}$ corresponds to a $k \times n$ all-zero matrix.

We say a B-DMC $p(y|x)$ is *symmetric* if there exists a permutation $\pi : \mathcal{Y} \to \mathcal{Y}$ such that $p(y|x) = p(\pi(y)|x \oplus 1)$ for all $y \in \mathcal{Y}$ and $x \in \{0, 1\}$. Given a symmetric channel under permutation π, we define the $\odot$ operation between two sequences $y^n \in \mathcal{Y}^n$ and $s^n \in \{0, 1\}^n$ as follows [19]. For $y \in \mathcal{Y}$ and $s \in \{0, 1\}$, define

$$y \odot s = \begin{cases} y & \text{if } s = 0, \\ \pi(y) & \text{if } s = 1, \end{cases} \tag{1}$$

and let $y^n \odot s^n$ be the element-wise $\odot$ operation.

Slepian–Wolf Problem. A (binary) Slepian–Wolf problem $p(x, y)$ consists of two finite alphabets $\mathcal{X} = \{0, 1\}$, $\mathcal{Y}$ and a joint pmf $p(x, y)$ over $\mathcal{X} \times \mathcal{Y}$. The source X with side information Y generate a jointly i.i.d. random process $\{(X_i, Y_i)\}$ with $(X_i, Y_i) \sim p_{X,Y}(x_i, y_i)$. An (ℓ, n, ϵ) code (g, ψ) for the Slepian–Wolf problem $p(x, y)$ consists of

- an index set $\mathcal{I}$ of size $|\mathcal{I}| = 2^\ell$,
- an encoder $g : \{0, 1\}^n \to \mathcal{I}$ that maps a source sequence x^n to an index $m = g(x^n)$, and
- a decoder $\psi : \mathcal{I} \times \mathcal{Y}^n \to \{0, 1\}^n$ that assigns a source estimate $\hat{x}^n = \psi(m, y^n)$ to each index m and side information y^n.

The rate of the code is $R = \ell/n$. The average probability of error of the code is $\mathsf{P}\{\hat{X}^n \neq X^n\} = \epsilon$.

A Slepian–Wolf code is *linear* if the encoding function can be defined as a matrix multiplication, i.e., $g(x^n) = Hx^n$ for some $\ell \times n$ matrix H. In this case, we refer to the code as an (ℓ, n, ϵ) code (H, ψ).

We say a Slepian–Wolf problem $p(x, y)$ is *symmetric* under permutation π if $X \sim \text{Bern}(\frac{1}{2})$ and the channel $p(y|x)$ is symmetric under permutation π.

Gelfand–Pinsker Problem. A binary-input binary-state discrete memoryless channel $p(y|x, s)p(s)$ consists of finite alphabets $\mathcal{X} = \mathcal{S} = \{0, 1\}$ and $\mathcal{Y}$, a collection of conditional probability mass functions $p(y|x, s)$ on $\mathcal{Y}$ for each $x \in \mathcal{X}$ and $s \in \mathcal{S}$, and a probability mass function $p(s)$ on $\mathcal{S}$, where the state sequence $(S_1, S_2, \ldots)$ is i.i.d. with $S_i \sim p(s_i)$ and is available noncausally only at the encoder [1]. A (k, n, ϵ) code (h, ξ) for the channel $p(y|x, s)p(s)$ consists of

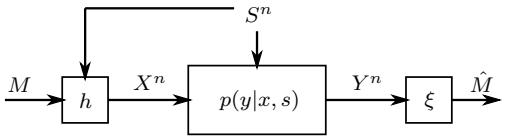

Fig. 1: Gelfand–Pinsker code.

- an encoder $h : [2^k] \times \mathcal{S}^n \to \mathcal{X}^n$ that assigns a codeword $x^n = h(m, s^n)$ to each message m and state sequence s^n, and
- a decoder $\xi : \mathcal{Y}^n \to [2^k]$ that assigns an estimate $\hat{m} = \xi(y^n)$ to each received sequence y^n.

The average probability of error of the code is $\mathsf{P}\{\hat{M} \neq M\} = \epsilon$. A Gelfand–Pinsker code is depicted in Fig. 1. With similarity to the point-to-point channel coding problem, where there is a target input pmf $p(x)$, for the Gelfand–Pinsker problem, there is a target conditional pmf $p(x|s)$.

II. Main Result

Consider a binary-input binary-state Gelfand–Pinsker problem $p(y|x, s)p(s)$ with $S \sim \text{Bern}(\theta)$ for some $\theta \in [0, 1]$. The main contribution of this paper is a coding scheme for the Gelfand–Pinsker problem when the target conditional distribution $p(x|s)$ is a $\text{BSC}(\alpha)$, a binary symmetric channel with crossover probability α, for some $\alpha \in (0, 1)$. Starting from a pair of linear codes that are designed for symmetric channels, the encoder guarantees that the channel input distribution is appropriately biased with respect to the state sequence, and the decoder is able to reconstruct the transmitted messages with some fidelity. Before we state our assumptions on the constituent codes, we define the *decoding distance spectrum* of a binary-input binary-output channel code.

Definition 1. Consider a code (f, ϕ) designed for a B-DMC $\widetilde{p}(y|x)$ with binary output alphabet $\mathcal{Y} = \{0, 1\}$. We define the *decoding distance spectrum* of the code (f, ϕ) as the random variable W such that

$$W = d_H\big(V^n, \phi(V^n)\big),$$

where $V^n \overset{\text{iid}}{\sim} \text{Bern}(1/2)$ and $d_H(.,.)$ is the Hamming distance.

Notice that, for any $w \in \{0, 1, \ldots, n\}$, we have

$$\mathsf{P}\{W = w\} = \tfrac{1}{2^n}\big|\{v^n \in \mathbb{F}_2^n : d_H\big(v^n, \phi(v^n)\big) = w\}\big|.$$

Moreover, since the alphabet size of W is linear in n, the distribution of W can be estimated for any off-the-shelf channel code (f, ϕ) via simulations, which aligns with the spirit of this work, aiming to design coding schemes starting from basic building blocks with easy-to-verify properties.

Now, we are ready to state the properties of the constituent codes, as highlighted in the following three "axioms".

Axiom 1. *Suppose that there exists a (k_1, n, ϵ_1) linear code (H_1, ϕ_1) with codebook $\mathcal{C}_1$ for the channel $p(y, u|x)$ defined over the input alphabet $\mathcal{X} = \{0, 1\}$ and output alphabet $\mathcal{Y} \times \{0, 1\}$ by*

$$p(y, u|x) = \sum_{s\in\mathcal{S}} p_S(s)p_{X|S}(x \oplus u|s)p_{Y|X,S}(y|x \oplus u, s),$$

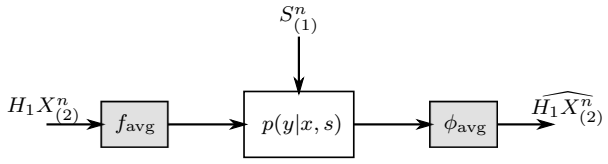

Fig. 2: Coding for the Gelfand–Pinsker channel in the first block using the code $(f_{\rm avg}, \phi_{\rm avg})$.

where $p_{X|S}$ is the target conditional distribution $\mathrm{BSC}(\alpha)$.

Axiom 2. *Suppose that there exists a (k_2, n, ϵ_2) linear code (H_2, ϕ_2) with codebook $\mathcal{C}_2$ for* $\mathrm{BSC}(\alpha)$ *and decoding distance spectrum W_2 satisfying*

$$\sum_{w=0}^{n} \left| \mathsf{P}\{W_2 = w\} - \binom{n}{w}\alpha^w(1-\alpha)^{n-w} \right| \le \delta, \qquad (2)$$

for some $\delta > 0$. We assume that $k_2 < k_1$.

Axiom 3. *Suppose that there exists a $(k_{\rm avg}, n, \epsilon_{\rm avg})$ code $(f_{\rm avg}, \phi_{\rm avg})$ for the "average" channel $p_{\rm avg}(y|x) = \sum_s p(s)p(y|x,s)$.*[1]

Remark 2. Without loss of generality, we assume that the parity check matrices H_1 and H_2 are in systematic form, i.e., $H_i = \begin{bmatrix} A_i & I_{n-k_i} \end{bmatrix}$, $i \in \{1, 2\}$, where A_i is an $(n-k_i) \times k_i$ matrix and I_{n-k_i} is the $(n-k_i) \times (n-k_i)$ identity matrix.[2]

Notice that the channel $p(y, u|x)$ in Axiom 1 is symmetric under the permutation $\pi(y, u) = (y, u \oplus 1)$. On the other hand, the condition (2) in Axiom 2 says that the distribution of the decoding distance spectrum is at most $\delta/2$-away in total variation distance from a $\mathrm{Binom}(n, \alpha)$ distribution[3].

The main result of this paper can be stated as follows. Starting from *any* pair of codes that satisfy Axioms 1 and 2 along with the code $(f_{\rm avg}, \phi_{\rm avg})$, the paper presents a construction of a coding scheme for the Gelfand–Pinsker problem $p(y|x,s)p(s)$. The coding scheme is defined upon the transmission of b blocks of information, and achieves an average probability of error over the b blocks that can be bounded as

$$P_e \le (b-1)(\delta + \epsilon_1) + \epsilon_{\rm avg}.$$

III. Gelfand–Pinsker Coding

A. Encoding

Fig. 2 and Fig. 3 show the block diagrams of the encoder to the Gelfand–Pinsker problem. Inspired by the chaining construction of universal polar codes introduced in [7], our encoding scheme is defined upon the transmission of b blocks of information. For each $t \in [b]$, let $V^n_{(t)} \overset{\rm iid}{\sim} \mathrm{Bern}(1/2)$ be a random dither shared with the decoder, and let $\Gamma_{(t)} : [n] \to [n]$ be a permutation chosen uniformly at random and shared with the decoder. Also, let $\widetilde{S}^n_{(t)} = \Gamma^{-1}_{(t)}(S^n_{(t)})$ (see Fig. 3)[4].

We now describe the encoding procedure starting from the b-th block. Given a message $M_{(b)} \in \mathbb{F}_2^{n-k_2}$, the encoder computes the sequence $Z^n_{(b)}$ as follows.

$$Z^n_{(b)} = \begin{bmatrix} \mathbf{0} \\ M_{(b)} \end{bmatrix}, \qquad (3)$$

where $\mathbf{0}$ consists of k_2 zeros. Then, for each $t = b, \dots, 2$, the encoder computes the sequences $X^n_{(t)}$ and $Z^n_{(t-1)}$ as follows.

$$\begin{aligned} X^n_{(t)} &= \Gamma_{(t)}\Big(\phi_2\big(Z^n_{(t)} \oplus \widetilde{S}^n_{(t)} \oplus V^n_{(t)}\big) \oplus Z^n_{(t)} \oplus V^n_{(t)}\Big), \\ Z^n_{(t-1)} &= \begin{bmatrix} \mathbf{0} \\ H_1X^n_{(t)} \\ M_{(t-1)} \end{bmatrix}, \end{aligned} \qquad (4)$$

where $M_{(t-1)} \in \mathbb{F}_2^{k_1-k_2}$ for each $t = 2, \dots, b$. Since H_2 is in systematic form, the sequence $Z^n_{(t)}$ satisfies that $H_2Z_{(b)} = M_{(b)}$ and

$$H_2Z^n_{(t-1)} = \begin{bmatrix} H_1X^n_{(t)} \\ M_{(t-1)} \end{bmatrix}$$

for each $t = 2, \dots, b$. Finally, in the first block, the transmitter uses the encoder $f_{\rm avg}$ to encode the syndrome vector $H_1X^n_{(2)}$, where $X^n_{(2)}$ is the transmitted sequence in the second block (see Fig. 2)[5][6]. Note that a loss in the overall achievable rate is incurred in the first block. However, this loss decays as $1/b$, and, thus, by choosing b large enough, the rate loss becomes negligible.

Notice that, for each $t = 2, \dots, b$, since $Z^n_{(t)} \oplus \widetilde{S}^n_{(t)} \oplus V^n_{(t)} \overset{\rm iid}{\sim} \mathrm{Bern}(1/2)$, we have that

$$\widetilde{X}^n_{(t)} \oplus \widetilde{S}^n_{(t)} = \phi_2\big(Z^n_{(t)} \oplus \widetilde{S}^n_{(t)} \oplus V^n_{(t)}\big) \oplus Z^n_{(t)} \oplus \widetilde{S}^n_{(t)} \oplus V^n_{(t)}$$

satisfies that $\mathrm{wt}(\widetilde{X}^n_{(t)} \oplus \widetilde{S}^n_{(t)}) \overset{d}{=} W_2$, where $\mathrm{wt}(.)$ denotes the Hamming weight function, $\overset{d}{=}$ denotes equality in distribution and W_2 denotes the decoding distance spectrum of the code (H_2, ϕ_2). The next lemma highlights the fact that when a random permutation $\Gamma_{(t)}$ is applied to $\widetilde{X}^n_{(t)} \oplus \widetilde{S}^n_{(t)}$, the total variation distance between the distribution of the resulting sequence $X^n_{(t)} \oplus S^n_{(t)} = \Gamma_{(t)}(\widetilde{X}^n_{(t)} \oplus \widetilde{S}^n_{(t)})$ and the i.i.d. Bernoulli(α) distribution is at most $\delta/2$.

Lemma 3. *For the arrangement shown in Fig. 3, we have*

$$\sum_{u^n} \left| \mathsf{P}\{X^n_{(t)} \oplus S^n_{(t)} = u^n\} - \alpha^{\mathrm{wt}(u^n)}(1-\alpha)^{n-\mathrm{wt}(u^n)} \right| \le \delta,$$

where $X^n_{(t)} \oplus S^n_{(t)} = \Gamma_{(t)}(\widetilde{X}^n_{(t)} \oplus \widetilde{S}^n_{(t)})$ and $\Gamma_{(t)}$ is a permutation chosen uniformly at random and independent from $\widetilde{X}^n_{(t)} \oplus \widetilde{S}^n_{(t)}$.

[1] Here, we assume that the average channel has a non-zero capacity. Note that this code will be used only for the first few transmission blocks.

[2] Note that a parity check matrix that is in systematic form can be obtained for any linear code by basic row operations and column permutations.

[3] As an example, polar codes designed for $\mathrm{BSC}(\alpha)$ with successive cancellation decoding using the "randomized rounding" rule satisfy condition (2), where δ decays exponentially fast with the block length [6].

[4] Note that $\widetilde{S}^n_{(t)}$ is independent of $\Gamma_{(t)}$.

[5] The assumption here is that $n - k_1 < k_{\rm avg}$. Otherwise, one should send the syndrome vector $H_1X^n_{(2)}$ over multiple blocks.

[6] Note that the sequence $Z^n_{(1)}$ is not used in this construction.

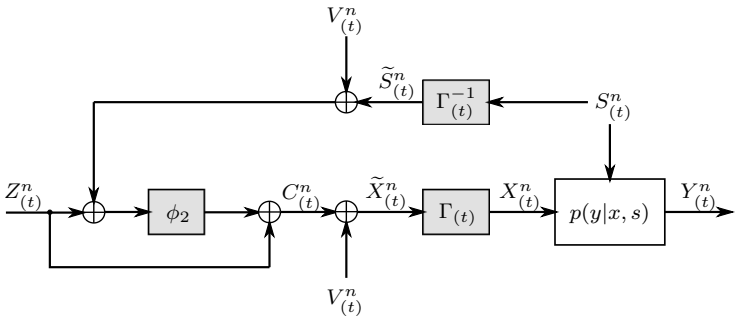

Fig. 3: Encoder in the t-th block ($2 \le t \le b$).

Proof. For any sequence u^n with $\mathrm{wt}(u^n) = w$, we have that

$$\begin{aligned}
&\mathsf{P}\{X^n_{(t)} \oplus S^n_{(t)} = u^n\} \\
&= \sum_{\widetilde{u}^n:\mathrm{wt}(\widetilde{u}^n)=w} \mathsf{P}\{\Gamma(\widetilde{u}^n) = u^n\}\, \mathsf{P}\{\widetilde{X}^n_{(t)} \oplus \widetilde{S}^n_{(t)} = \widetilde{u}^n\} \\
&= \sum_{\widetilde{u}^n:\mathrm{wt}(\widetilde{u}^n)=w} \frac{w!(n-w)!}{n!}\, \mathsf{P}\{\widetilde{X}^n_{(t)} \oplus \widetilde{S}^n_{(t)} = \widetilde{u}^n\} \\
&= \tfrac{1}{\binom{n}{w}}\, \mathsf{P}\{\mathrm{wt}\big(\widetilde{X}^n_{(t)} \oplus \widetilde{S}^n_{(t)}\big) = w\} = \tfrac{1}{\binom{n}{w}}\, \mathsf{P}\{W_2 = w\}.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\sum_{u^n} \Big| \mathsf{P}\{X^n_{(t)} \oplus S^n_{(t)} = u^n\} - \alpha^{\mathrm{wt}(u^n)}(1-\alpha)^{n-\mathrm{wt}(u^n)} \Big| \\
&= \sum_{u^n} \Big| \frac{\mathsf{P}\{W_2 = \mathrm{wt}(u^n)\}}{\binom{n}{\mathrm{wt}(u^n)}} - \alpha^{\mathrm{wt}(u^n)}\big(1-\alpha\big)^{n-\mathrm{wt}(u^n)} \Big| \\
&= \sum_{w=0}^{n} \Big| \mathsf{P}\{W_2 = w\} - \binom{n}{w}\alpha^w(1-\alpha)^{n-w} \Big| \le \delta,
\end{aligned}$$

where the last step holds since W_2 satisfies Axiom 2. □

Corollary 4. *For the arrangement shown in Fig. 3, we have*

$$\sum_{x^n} \Big| \mathsf{P}_{X^n_{(t)}|S^n_{(t)}}(x^n|s^n) - \alpha^{\mathrm{wt}(x^n \oplus s^n)}(1-\alpha)^{n-\mathrm{wt}(x^n \oplus s^n)} \Big| \le \delta$$

for each $s^n \in \{0,1\}^n$.

Proof. The result follows from Lemma 3 and the fact that $S^n_{(t)}$ is independent of $X^n_{(t)} \oplus S^n_{(t)}$. To see the latter, notice that

$$X^n_{(t)} \oplus S^n_{(t)} = \Gamma_{(t)}\Big(\phi_2\big(Z^n_{(t)} \oplus \widetilde{S}^n_{(t)} \oplus V^n_{(t)}\big) \oplus Z^n_{(t)} \oplus \widetilde{S}^n_{(t)} \oplus V^n_{(t)}\Big),$$

which is independent of $S^n_{(t)}$ since $Z^n_{(t)} \oplus \widetilde{S}^n_{(t)} \oplus V^n_{(t)}$ is independent of $S^n_{(t)}$. Note that this result holds for each block $t = 2, \ldots, b$. □

Lemma 3 and Corollary 4 say that the conditional distribution of $X^n_{(t)}$ given $S^n_{(t)}$ is $\delta/2$-away in total variation distance from that corresponding to a $\mathrm{BSC}(\alpha)$.

B. Decoding

At the decoder side, the key point is to view (X^n, Y^n) as realizations of the Slepian–Wolf problem $p(x,y)$, where $p(x,y)$ represents the desired joint distribution between the channel inputs and outputs, i.e.,

$$p(x,y) = \textstyle\sum_{s \in \mathcal{S}} p(s)p(x|s)p(y|x,s).$$

The goal, therefore, is to construct a Slepian–Wolf decoder to recover an estimate of the channel input $X^n_{(t)}$ in the t-th block. For this purpose, we utilize an implementation of a Slepian–Wolf decoder using point-to-point channel codes introduced in [19], which we review in the following two lemmas. The first lemma states that any Slepian–Wolf problem can be symmetrized by scrambling.

Lemma 5 (Lemma 3 [19])**.** *Consider a general Slepian–Wolf problem $\widetilde{p}(x,y)$. Let $U \sim \mathrm{Bern}(1/2)$ be independent of (X,Y). Let $\overline{X} = X \oplus U$ and $\overline{Y} = (Y,U)$. Then, the Slepian–Wolf problem $\widetilde{p}(\bar{x},\bar{y})$ is symmetric under permutation $\pi(y,u) = (y, u \oplus 1)$. Furthermore, we have*

$$\widetilde{p}_{\overline{Y}|\overline{X}}(y,u|x) = \widetilde{p}_{X,Y}(x \oplus u, y)$$

for each $x, u \in \{0,1\}$ and $y \in \mathcal{Y}$.

The next lemma states that a decoder for a symmetric Slepian–Wolf problem can be implemented using a symmetric point-to-point channel code.

Lemma 6 (Lemmas 1,2 [19])**.** *Consider a symmetric Slepian–Wolf problem $\widetilde{p}(x,y)$ under permutation π, as defined in Section I. Let (H,ϕ) be a channel code for $\widetilde{p}(y|x)$ with codebook $\mathcal{C}$ such that H is in systematic form. Let $(X^n, Y^n) \overset{\mathrm{iid}}{\sim} \widetilde{p}(x,y)$. If $C^n = X^n \oplus \overline{H}X^n$ and $R^n = Y^n \odot \overline{H}X^n$ (where $\odot$ is as defined in (1)), then*

$$\mathsf{P}(C^n = c^n, R^n = r^n) = \tfrac{1}{2^k} \textstyle\prod_{i=1}^n \widetilde{p}_{Y|X}(r_i|c_i)$$

for every $c^n \in \mathcal{C}$ and $r^n \in \mathcal{Y}^n$. That is, R^n has the same distribution as the output of the channel $\widetilde{p}(y|x)$ when the input to the channel is a uniformly distributed codeword C^n. It follows that, given the syndrome vector HX^n, one can get an estimate of the source vector X^n as

$$\hat{X}^n = \phi(Y^n \odot \overline{H}X^n) \oplus \overline{H}X^n.$$

When combined together, the previous two lemmas provide a construction of a decoder to *any* Slepian–Wolf problem using point-to-point *symmetric* channel codes. This hints to a possible construction of a decoder for the Gelfand–Pinsker problem. In particular, we look at the symmetrized version of the Slepian–Wolf problem $p(x,y)$, as described in Lemma 5. Then, a Slepian–Wolf decoder for the corresponding channel using a point-to-point channel code is utilized, as outlined in Lemma 6. Note that the code (H_1, ϕ_1) in Axiom 1 is designed precisely for this channel.

Therefore, the decoding proceeds as follows. In the first block, the receiver uses the decoder ϕ_{avg} to recover an estimate of $H_1 X^n_{(2)}$. In the t-th block ($2 \le t \le b$), the receiver uses the code (H_1, ϕ_1) to implement a decoder for the symmetrized

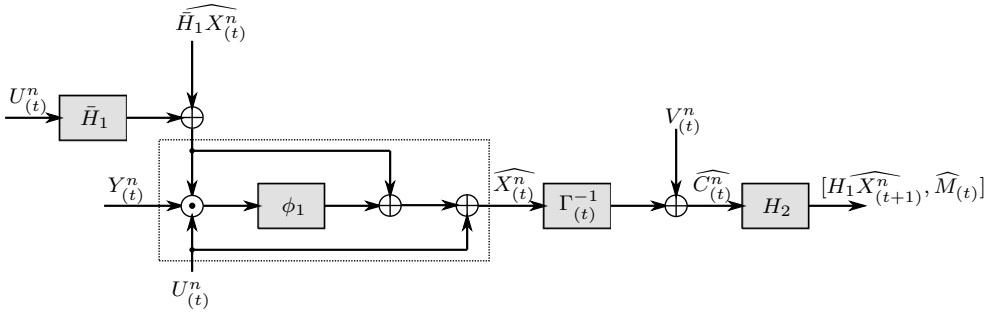

Fig. 4: Decoder in the t-th block $(2 \le t \le b)$. The dotted box represents an implementation of a Slepian–Wolf decoder using a channel code.

Slepian–Wolf problem. More precisely, the decoder utilizes the estimate of $H_1 X^n_{(t)}$ from previous blocks and computes

$$\widehat{X}^n_{(t)} = \phi_1\Big((Y^n_{(t)}, U^n_{(t)}) \odot (\overline{H}_1 U^n_{(t)} \oplus \widehat{\overline{H}_1 X^n_{(t)}})\Big) \oplus (\overline{H}_1 U^n_{(t)} \oplus \widehat{\overline{H}_1 X^n_{(t)}}) \oplus U^n_{(t)},$$

where $U^n_{(t)} \overset{\text{iid}}{\sim} \text{Bern}(1/2)$ is generated independently at the decoder. The decoder then computes $\widehat{C}^n_{(t)} = \Gamma^{-1}_{(t)}(\widehat{X}^n_{(t)}) \oplus V^n_{(t)}$, and declares the first $(n-k_1)$ bits of $H_2\widehat{C}^n_{(t)}$ as an estimate of the subsequent syndrome vector $H_1 X^n_{(t+1)}$, and the last $k_1 - k_2$ bits of $H_2\widehat{C}^n_{(t)}$ as an estimate of the message $M_{(t)}$ in the t-th block[7]. Fig. 4 shows the block diagram of the decoder.

C. Analysis of the probability of error

To gain an intuitive understanding of the probability of error of our coding scheme, consider a genie-aided decoder which recovers an estimate of $M_{(t)}$ in the t-th block based on the channel output $Y^n_{(t)}$ and the syndrome vector $H_1 X^n_{(t)}$ (which is supplied correctly by a genie regardless of any decoding errors in previous blocks). Notice that such a decoder would have the same probability of error over the b blocks as our decoder[8]. A similar argument has been made in the analysis of the successive cancellation decoding of polar codes [22]. Therefore, it suffices to analyze the error probability of the genie-aided decoder over the b blocks of transmission.

To this end, let q_{S^n,X^n,Y^n} be the true distribution of the tuple (S^n, X^n, Y^n). Let us focus first on one block. The average probability of error of the genie-aided decoder in the t-th block $(2 \le t \le b)$ can be bounded as[9]

$$\begin{aligned}
&\mathsf{P}\{\hat{M} \neq M\} \le \mathsf{P}\{\hat{X}^n \neq X^n\} \\
&= \sum_{x^n, s^n} \mathsf{P}\{\hat{X}^n \neq X^n | X^n = x^n, S^n = s^n\} q_{X^n,S^n}(x^n, s^n) \\
&= \sum_{x^n, s^n} \mathsf{P}\{\hat{X}^n \neq X^n | X^n = x^n, S^n = s^n\} \\
&\qquad \cdot \big(q_{X^n,S^n}(x^n, s^n) - p_{X^n,S^n}(x^n, s^n)\big) \\
&+ \sum_{x^n, s^n} \mathsf{P}\{\hat{X}^n \neq X^n | X^n = x^n, S^n = s^n\} p_{X^n,S^n}(x^n, s^n) \\
&\overset{(a)}{\le} \sum_{x^n, s^n} \big|q_{X^n,S^n}(x^n, s^n) - p_{X^n,S^n}(x^n, s^n)\big| + \epsilon_1 \\
&\overset{(b)}{\le} \delta + \epsilon_1,
\end{aligned}$$

where (a) follows since $\mathsf{P}\{\hat{X}^n \neq X^n | X^n = x^n, S^n = s^n\} \le 1$, $\mathsf{P}\{\hat{X}^n \neq X^n | X^n = x^n, S^n = s^n\}$ depends only on the channel $p(y|x,s)$, and the average probability of error of the Slepian–Wolf decoder using the code (H_1, ϕ_1) is exactly ϵ_1, and (b) follows by Corollary 4. By the union bound, it follows that the average probability of error of the genie-aided decoder (and, therefore, our decoder) over the b blocks can be bounded as

$$P_e \le (b-1)(\delta + \epsilon_1) + \epsilon_{\text{avg}},$$

where ϵ_{avg} is the average probability of error of the code used in the first block. The rate of the coding scheme is

$$R = \frac{(b-1)(k_1 - k_2) + n - k_1}{nb}.$$

Remark 7. Note that the case $\theta = 0$ (or, equivalently, setting the state sequence to the constant all-zero sequence) gives a coding scheme for the asymmetric channel.

Remark 8. If the pair of codes in Axioms 1 and 2 have rates close to the capacity of their respective symmetric channels, then, by choosing a large block length n and a large (but fixed) number of transmission blocks b, our coding scheme can approach a rate arbitrarily close to

$$H(X|S) - H(X|Y) = I(X;Y) - I(X;S),$$

evaluated for the conditional pmf $p(x|s)$ that is a $\text{BSC}(\alpha)$.

IV. Conclusion

In this paper, a construction of a coding scheme for the Gelfand–Pinsker problem and for the asymmetric channel using a pair of linear channel codes was developed. The parameters of the constituent channel codes (i.e., the probability of error of the first code and the decoding distance spectrum of the second code) can be easily estimated for any linear code through simulations. Therefore, our construction enables one to leverage existing codes for symmetric point-to-point channels to build coding schemes for both the Gelfand–Pinsker and the asymmetric channel problems.

Acknowledgement

This work was supported in part by the Institute for Information & Communication Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No. 2018-0-01659, 5G Open Intelligence-Defined RAN (ID-RAN) Technique based on 5G New Radio), in part by the NSERC Discovery Grant No. RGPIN-2019-05448, and in part by the NSERC Collaborative Research and Development Grant CRDPJ 543676-19.

[7]Note that $H_2 C^n_{(t)} = H_2 Z^n_{(t)}$ (see Fig. 3).

[8]To see this, observe that a decoding error can propagate from one block to another only through an error in the syndrome vector $H_1 X^n_{(t)}$. Consider the first block where such an error happens. Both decoders would make an error in that block, which is precisely an error event over the b blocks, irrespective of decisions made in subsequent blocks.

[9]For notational convenience, the subscript "(t)" corresponding to the t-th block is dropped.

REFERENCES

[1] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Control Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[2] ——, "Capacity of a broadcast channel with one deterministic component," *Problems Inform. Transmission*, vol. 16, no. 1, pp. 24–34, 1980.

[3] A. Padakandla and S. S. Pradhan, "Nested linear codes achieve Marton's inner bound for general broadcast channels," in *Proc. IEEE Int. Symp. Inf. Theory*, 2011, pp. 1554–1558.

[4] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[5] M. Mondelli, S. H. Hassani, I. Sason, and R. L. Urbanke, "Achieving Marton's region for broadcast channels using polar codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 783–800, Feb. 2015.

[6] S. B. Korada and R. L. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1751–1768, 2010.

[7] S. H. Hassani and R. Urbanke, "Universal polar codes," in *Proc. IEEE Int. Symp. Inf. Theory*, 2014, pp. 1451–1455.

[8] A. Bennatan, D. Burshtein, G. Caire, and S. Shamai, "Superposition coding for side-information channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1872–1889, May 2006.

[9] E. Martinian and M. J. Wainwright, "Low-density constructions can achieve the Wyner–Ziv and Gelfand–Pinsker bounds," in *Proc. IEEE Int. Symp. Inf. Theory*, 2006, pp. 484–488.

[10] K. M. Rege, K. Balachandran, J. H. Kang, and M. K. Karakayali, "A practical dirty paper coding scheme based on LDPC codes," in *Proc. IEEE Wireless Comm. and Networking Conf.*, 2014, pp. 188–193.

[11] U. Erez and S. ten Brink, "A close-to-capacity dirty paper coding scheme," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3417–3432, Oct. 2005.

[12] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3820–3833, Nov. 2005.

[13] Y. Liu and P. H. Siegel, "On the performance of direct shaping codes," *arXiv:2007.05638v1 [cs.IT]*, Jul. 2020.

[14] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, Jan. 2016.

[15] T. Philosof, U. Erez, and R. Zamir, "Combined shaping and precoding for interference cancellation at low SNR," in *Proc. IEEE Int. Symp. Inf. Theory*, 2003, pp. 68–.

[16] M. Mondelli, S. H. Hassani, and R. L. Urbanke, "How to achieve the capacity of asymmetric channels," *IEEE Trans. Inf. Theory*, vol. 64, no. 5, pp. 3371–3393, May 2018.

[17] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell Syst. Tech. J.*, vol. 52, no. 7, pp. 1037–1076, Sep. 1973.

[18] A. D. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 1, pp. 2–10, Jan. 1974.

[19] L. Wang and Y.-H. Kim, "Linear code duality between channel coding and Slepian–Wolf coding," in *Proc. 53rd Ann. Allerton Conf. Comm. Control Comput.*, Sep.–Oct. 2015, pp. 147–152.

[20] L. Wang, "Channel coding techniques for network communication," Ph.D. dissertation, University of California, San Diego, 2015.

[21] S. Ganguly, L. Wang, and Y.-H. Kim, "A functional construction of codes for multiple access and broadcast channels," in *Proc. IEEE Int. Symp. Inf. Theory*, 2020, pp. 1581–1586.

[22] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.