Sliding-Window Superposition Coding: Two-User Interference Channels

Lele Wang, *Member, IEEE*, Young-Han Kim, Chiao-Yi Chen, Hosung Park, *Member, IEEE*, and Eren Şaşoğlu

***Abstract*— A low-complexity coding scheme is developed to achieve the rate region of maximum likelihood decoding for interference channels. As in the classical rate-splitting multiple access scheme by Grant, Urbanke, and Whiting, the proposed coding scheme uses superposition of multiple codewords with successive cancellation decoding, which can be implemented using standard point-to-point encoders and decoders. Unlike rate-splitting multiple access, which is not rate-optimal for multiple receivers, the proposed coding scheme transmits codewords over multiple blocks in a staggered manner and recovers them successively over sliding decoding windows, achieving the single-stream optimal rate region as well as the more general Han–Kobayashi inner bound for the two-user interference channel. The feasibility of this scheme in practice is verified by implementing it using commercial channel codes over the two-user Gaussian interference channel.**

***Index Terms*— Coded modulation, coding technique, interference management, simultaneous decoding.**

I. Introduction

FOR high data rates and massive connectivity, next-generation cellular networks are expected to deploy many small base stations. While such dense deployment provides the benefit of bringing radio closer to end users, it also increases the amount of interference from neighboring cells. Consequently, efficient and effective management of interference is expected to become one of the main challenges for high-spectral-efficiency, low-power, broad-coverage wireless communications.

Over the past few decades, several techniques at different protocol layers [1]–[3] have been proposed to mitigate adverse effects of interference in wireless networks. One important conceptual technique at the physical layer is *simultaneous decoding* [4, Section 6.2], [5]. In this decoding method, each receiver attempts to recover both the intended and a subset of the interfering codewords at the same time. When the interference is strong [6], [7] and weak [8]–[11], simultaneous decoding of random code ensembles achieves the capacity of the two-user interference channel. In fact, for any given random code ensemble, simultaneous decoding achieves the same rates achievable by the optimal maximum likelihood decoding [10], [12], [13]. The celebrated Han–Kobayashi coding scheme [14] also relies on simultaneous decoding as a crucial component. As a main drawback, however, each receiver in simultaneous decoding (or maximum likelihood decoding) has to employ some form of multiuser sequence detection, which usually has high computational complexity. This issue has been tackled recently by a few approaches based on emerging spatially coupled and polar codes [15], [16], but these solutions involve very long block lengths.

For this reason, most practical communication systems use conventional single-user point-to-point decoding. The simplest method is *treating interference as noise*, in which only statistical properties (such as the distribution and power), rather than the actual codebook information, of the interfering signals, are used. In *successive cancellation decoding*, similar single-user decoding is performed in steps, first recovering interfering codewords and then incorporating them as part of the channel output for decoding of desired codewords. Successive cancellation decoding is particularly well suited when the messages are split into multiple parts by rate splitting, encoded into separate codewords, and transmitted via superposition coding. In particular, when there is only one receiver (i.e., for a multiple access channel), this rate-splitting coding scheme with successive cancellation decoding was proposed by Rimoldi and Urbanke [17] for the Gaussian case and Grant *et al.* [18] for the discrete case, and achieves the optimal rate region of the polymatroidal shape (the pentagon for two senders). When there are two or more receivers—as in the two-user interference channel or the compound multiple access channel—the rate-splitting multiple access scheme fails to achieve the optimal rate region as demonstrated earlier

Manuscript received January 9, 2017; revised December 16, 2018; accepted July 4, 2019. Date of publication February 20, 2020; date of current version May 20, 2020. This work was supported in part by the Natural Science and Engineering Research Council of Canada under Discovery Grant RGPIN-2019-05448, in part by the Natural Science and Engineering Research Council of Canada under Collaborative Research and Development Grant CRDPJ 543676-19, in part by Rogers Communications Inc. under the Rogers-UBC Collaborative Research Grant: Interference Mitigation for 5G Networks, in part by the Swiss National Science Foundation under grant PBELP2-137726, and in part by the National Science Foundation under Grant CCF-1320895. This article was presented in part at the 2014 IEEE International Symposium on Information Theory (ISIT) and in part at the 2014 IEEE Globecom Workshops (GC Wkshps). *(Corresponding author: Lele Wang.)*

Lele Wang is with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: lelewang@ece.ubc.ca).

Young-Han Kim is with the Department of Electrical and Computer Engineering, University of California at San Diego, San Diego, CA 92093 USA (e-mail: yhk@ucsd.edu).

Chiao-Yi Chen is with Broadcom Ltd., Sunnyvale, CA 94086 USA (e-mail: uscychen@gmail.com).

Hosung Park is with the School of Electronics and Computer Engineering, Chonnam National University, Gwangju 61186, South Korea (e-mail: hpark1@jnu.ac.kr).

Eren Şaşoğlu is with Apple Inc., Cupertino, CA 95014 USA (e-mail: eren.sasoglu@gmail.com).

Communicated by A. Tchamkerten, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2020.2975162

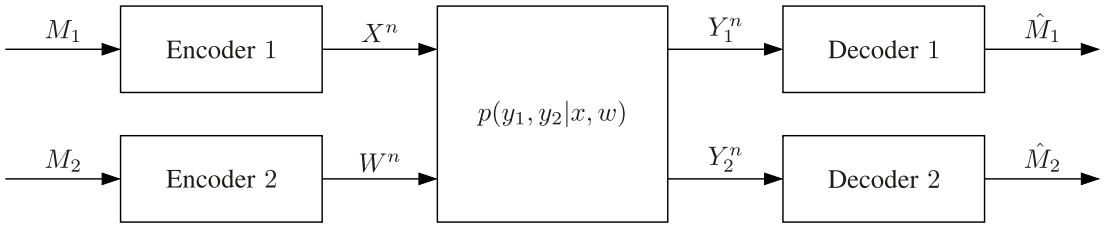


Fig. 1. The interference channel with two sender–receiver pairs.

in [19] for Gaussian codes and in Section III-B of this paper (and [20]) for general codes.

A natural question is whether single-user point-to-point coding techniques, which could achieve capacity for multiple access and single-antenna Gaussian broadcast channels, are fundamentally deficient for the interference channel, and high-complexity simultaneous decoding would be critical to achieve the capacity in general. In this paper, we develop a new coding scheme, called *sliding-window superposition coding*, that overcomes the limitations of single-user decoding through a new diagonal superposition structure. The main ingredients of the scheme are block Markov coding, sliding-window decoding (both commonly used for multihop relaying and feedback communication), superposition coding, and successive cancellation decoding (crucial for low-complexity implementation using standard point-to-point codes). Each message is encoded into a single long codeword that are transmitted diagonally over multiple blocks and multiple signal layers, which helps avoid the performance bottleneck for the aforementioned rate-splitting multiple access scheme. Receivers recover the desired and interfering codewords over a decoding window spanning multiple blocks. Successive cancellation decoding is performed within each decoding window as well as across a sequence of decoding windows for streams of messages. When the number and distribution of signal layers are properly chosen, the sliding-window superposition coding scheme can achieve every rate pair in the rate region of maximum likelihood decoding for the two-user interference channel with single streams, providing a constructive answer to our earlier question. We develop a more complete theory behind the number and distribution of signal layers and the choice of decoding orders, which leads to an extension of this coding scheme that achieves the entire Han–Kobayashi inner bound.

For practical communication systems, the conceptual sliding-window superposition coding scheme can be readily adapted to a coded modulation scheme using binary codes and common signal constellations. We compare this *sliding-window coded modulation scheme* with two well-known coded modulation schemes, multi-level coding [21], [22] and bit-interleaved coded modulation [23], [24]. We implement the sliding-window coded modulation scheme for the two-user Gaussian channel using the 4G LTE turbo code and demonstrate its performance improvement over treating interference as noise. Following earlier conference versions [20], [25] of this paper, several practical implementations of sliding-window superposition coding have been investigated [26]–[28] and proposed to the 5G standards [29]–[34].

The rest of the paper is organized as follows. We first define the problem and the relevant rate regions in Section II. Then, we explain the rate-splitting scheme and demonstrate its fundamental deficiency for the interference channel in Section III. We introduce the new sliding-window superposition coding in Section IV, first by developing a simple scheme that achieves the corner points of simultaneous decoding region, and then extending it to achieve every point in the region. In Section V, we present a more complete theory of the sliding-window superposition coding scheme with a discussion on the number of superposition layers and alternative decoding orders. With further extensions and augmentations, we develop a scheme that achieves the Han–Kobayashi inner bound [14] for the two-user interference channel with point-to-point encoders and decoders in Section VI. We devote Section VII to sliding-window coded modulation and its application in a practical communication setting. We offer a couple of concluding remarks in Section VIII.

Throughout the paper, we closely follow the notation in [4]. In particular, for $X \sim p(x)$ and $\epsilon \in (0, 1)$, we define the set of ϵ-typical n-sequences x^n (or the typical set in short) [35] as

$$\mathcal{T}_\epsilon^{(n)}(X) = \big\{x^n : |\#\{i\colon x_i = x\}/n - p(x)| \le \epsilon p(x) \quad \text{for all } x \in \mathcal{X}\big\}.$$

We use X_k^n to denote the vector $(X_{k1}, X_{k2}, \ldots, X_{kn})$. For $n = 1, 2, \ldots, [n] = \{1, 2, \ldots, n\}$ and for $a \ge 0, [2^a] = \{1, 2, \ldots, 2^{\lceil a \rceil}\}$, where $\lceil a \rceil$ is the smallest integer greater than or equal to a. The probability of an event $\mathcal{A}$ is denoted by $\mathsf{P}(\mathcal{A})$.

II. Two-User Interference Channels

Consider the communication system model depicted in Fig. 1, whereby senders 1 and 2 wish to communicate independent messages M_1 and M_2 to their respective receivers over a shared channel $p(y_1, y_2|x, w)$. Here X and W are channel inputs from senders 1 and 2, respectively, and Y_1 and Y_2 are channel outputs at receivers 1 and 2, respectively. In network information theory, this model is commonly referred to as the *two-user interference channel*.

The Gaussian interference channel in Fig. 2 is an important special case with channel outputs

$$\begin{aligned} Y_1 &= g_{11}X + g_{12}W + N_1, \\ Y_2 &= g_{21}X + g_{22}W + N_2, \end{aligned} \tag{1}$$

where g_{jk} denotes the channel gain coefficient from sender k to receiver j, and N_1 and N_2 are independent $\mathrm{N}(0, 1)$

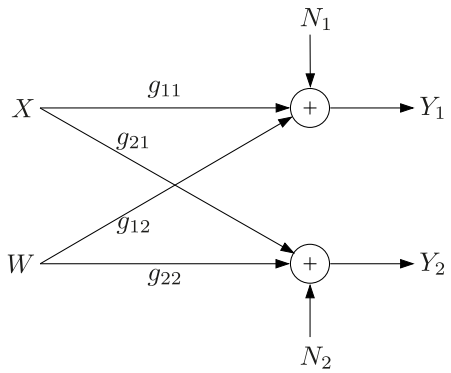


Fig. 2. The two-user Gaussian interference channel.

noise components. Under the average power constraint P on each input X and W, we denote the received signal-to-noise ratios (SNRs) as $S_1 = g_{11}^2\ P$ and $S_2 = g_{22}^2\ P$, and the received interference-to-noise ratios (INRs) as $I_1 = g_{12}^2\ P$ and $I_2 = g_{21}^2\ P$.

A $(2^{nR_1}, 2^{nR_2}, n)$ code $\mathcal{C}_n$ for the (two-user) interference channel consists of

- two message sets $[2^{nR_1}] := \{1, \ldots, 2^{\lceil nR_1 \rceil}\}$ and $[2^{nR_2}]$,
- two encoders, where encoder 1 assigns a codeword $x^n(m_1)$ to each message $m_1 \in [2^{nR_1}]$ and encoder 2 assigns a codeword $w^n(m_2)$ to each message $m_2 \in [2^{nR_2}]$, and
- two decoders, where decoder 1 assigns an estimate $\hat{m}_1$ or an error message e to each received sequence y_1^n and decoder 2 assigns an estimate $\hat{m}_2$ or an error message e to each received sequence y_2^n.

The performance of a given code $\mathcal{C}_n$ for the interference channel is measured by its average probability of error

$$P_e^{(n)}(\mathcal{C}_n) = \mathsf{P}\big\{(\hat{M}_1, \hat{M}_2) \neq (M_1, M_2)\big\},$$

where the message pair (M_1, M_2) is uniformly distributed over $[2^{nR_1}] \times [2^{nR_2}]$. A rate pair (R_1, R_2) is said to be *achievable* if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes $(\mathcal{C}_n)_{n=1}^{\infty}$ such that $\lim_{n\to\infty} P_e^{(n)}(\mathcal{C}_n) = 0$. A set of rate pairs, typically referred to as a *rate region*, is said to be achievable if every rate pair in the interior of the set is achievable. The *capacity region* is the closure of the set of achievable rate pairs (R_1, R_2), which is the largest achievable rate region and captures the optimal tradeoff between the two rates of reliable communication over the interference channel. The capacity region for the two-user interference channel is not known in general.

Let $p = p(x)p(w)$ be a given product pmf on $\mathcal{X} \times \mathcal{W}$. Suppose that the codewords $x^n(m_1), m_1 \in [2^{nR_1}]$, and $w^n(m_2), m_2 \in [2^{nR_2}]$, that constitute the codebook are generated randomly and independently according to $\prod_{i=1}^n p_X(x_i)$ and $\prod_{i=1}^n p_W(w_i)$, respectively. We refer to the codebooks generated in this manner collectively as the $(2^{nR_1}, 2^{nR_2}, n; p)$ *random code ensemble* (or the *p-distributed random code ensemble* in short).

Fixing the encoders as such, we now consider a few alternative decoding schemes. Here and henceforth, we assume $p = p(x)p(w)$ is fixed and write rate regions without p whenever it is clear from the context.

- *Treating interference as noise (IAN).* Receiver 1 recovers M_1 by treating the interfering codeword $W^n(M_2)$ as noise generated according to a given (memoryless) distribution $p(w)$. In other words, receiver 1 performs point-to-point decoding (either a specific decoding technique or a conceptual scheme) for the channel

$$\begin{aligned} p(y_1^n|x^n) &= \sum_{w^n} p(w^n) p(y_1^n|x^n, w^n) \\ &= \prod_{i=1}^n \sum_{w_i} p_W(w_i) p_{Y_1|X,W}(y_{1i}|x_i, w_i) \\ &= \prod_{i=1}^n p_{Y_1|X}(y_{1i}|x_i). \end{aligned}$$

  For example, if joint typicality decoding [36, Section 7.7] is used, the decoder finds $\hat{m}_1$ such that $(x^n(\hat{m}_1), y_1^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y_1)$. Similarly, receiver 2 can recover M_2 by treating X^n as noise. For the p-distributed random code ensemble, treating noise as interference achieves

$$\mathscr{R}_{\mathrm{IAN}} = \mathscr{R}_{1,\mathrm{IAN}} \cap \mathscr{R}_{2,\mathrm{IAN}}$$

  where $\mathscr{R}_{1,\mathrm{IAN}}$ and $\mathscr{R}_{2,\mathrm{IAN}}$ denote the sets of all rate pairs (R_1, R_2) such that $R_1 \le I(X; Y_1)$ and $R_2 \le I(W; Y_2)$, respectively; see Fig. 3a.
- *Successive cancellation decoding (SCD).* Receiver 1 recovers M_2 by treating X^n as noise and then recovers M_1 based on $W^n(M_2)$ (and Y_1^n). For example, in joint typicality decoding, the decoder finds a unique $\hat{m}_2$ such that $(w^n(\hat{m}_2), y_1^n) \in \mathcal{T}_\epsilon^{(n)}(W, Y_1)$ and then a unique $\hat{m}_1$ such that $(x^n(\hat{m}_1), w^n(\hat{m}_2), y_1^n) \in \mathcal{T}_\epsilon^{(n)}(X, W, Y_1)$. Receiver 2 operates in a similar manner. For the p-distributed random code ensemble, successive cancellation decoding achieves

$$\mathscr{R}_{\mathrm{SCD}} = \mathscr{R}_{1,\mathrm{SCD}} \cap \mathscr{R}_{2,\mathrm{SCD}},$$

  where $\mathscr{R}_{1,\mathrm{SCD}}$ consists of (R_1, R_2) such that

$$R_2 \le I(W; Y_1), \quad R_1 \le I(X; Y_1|W),$$

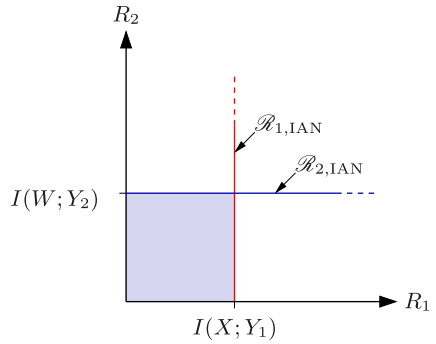
  and similarly $\mathscr{R}_{2,\mathrm{SCD}}$ consists of (R_1, R_2) such that

$$R_1 \le I(X; Y_2), \quad R_2 \le I(W; Y_2|X).$$

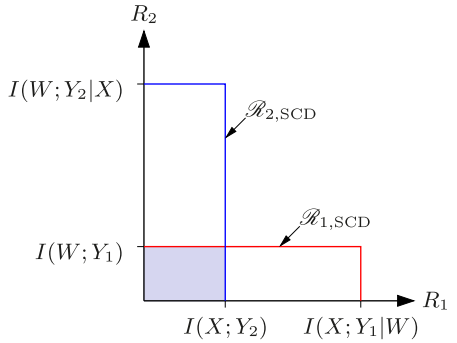
  See Fig. 3b for an illustration of $\mathscr{R}_{\mathrm{SCD}}$.
- *Mix and match.* Each receiver can choose between treating interference as noise and successive cancellation decoding. This mix-and-match achieves

$$(\mathscr{R}_{1,\mathrm{IAN}} \cup \mathscr{R}_{1,\mathrm{SCD}}) \cap (\mathscr{R}_{2,\mathrm{IAN}} \cup \mathscr{R}_{2,\mathrm{SCD}}). \tag{2}$$

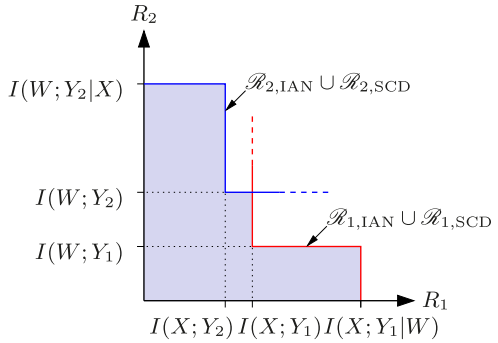
  The achievable rate region for mixing and matching is illustrated in Fig. 3c.
- *Simultaneous (nonunique) decoding (SND).* Receiver 1 recovers both the desired message M_1 and the interfering message M_2 simultaneously. It then keeps M_1 as the message estimate and ignores the error in estimating M_2. Receiver 2 operates in a similar manner. For example, in joint typicality decoding, receiver 1 finds a unique $\hat{m}_1$ such that $(x^n(\hat{m}_1), w^n(m_2), y_1^n) \in \mathcal{T}_\epsilon^{(n)}(X, W, Y_1)$ for some $m_2 \in [2^{nR_2}]$, and receiver 2 finds a unique $\hat{m}_2$



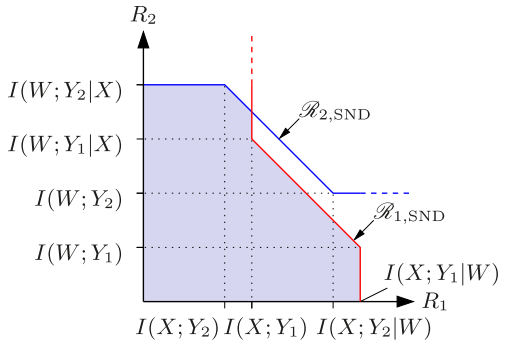
(a) $\mathscr{R}_{\mathrm{IAN}}$ is the intersection of the red-lined region $\mathscr{R}_{1,\mathrm{IAN}}$ and the blue-lined region $\mathscr{R}_{2,\mathrm{IAN}}$.



(b) $\mathscr{R}_{\mathrm{SCD}}$ is the intersection of the red-lined region $\mathscr{R}_{1,\mathrm{SCD}}$ and the blue-lined region $\mathscr{R}_{2,\mathrm{SCD}}$.



(c) The mix-and-match region is the intersection of the red-lined region $\mathscr{R}_{1,\mathrm{IAN}} \cup \mathscr{R}_{1,\mathrm{SCD}}$ and the blue-lined region $\mathscr{R}_{2,\mathrm{IAN}} \cup \mathscr{R}_{2,\mathrm{SCD}}$.



(d) $\mathscr{R}_{\mathrm{SND}}$ is the intersection of the red-lined region $\mathscr{R}_{1,\mathrm{SND}}$ and the blue-lined region $\mathscr{R}_{2,\mathrm{SND}}$. $\mathscr{R}_{\mathrm{SND}}$ is identical to the MLD region $\mathscr{R}^*$.

Fig. 3. Illustration of the MLD, IAN, SCD regions and their comparison.

such that $(x^n(m_1), w^n(\hat{m}_2), y_2^n) \in \mathcal{T}_\epsilon^{(n)}(X, W, Y_2)$ for some $m_1 \in [2^{nR_1}]$. For the p-distributed random code ensemble, simultaneous decoding achieves

$$\mathscr{R}_{\mathrm{SND}} = \mathscr{R}_{1,\mathrm{SND}} \cap \mathscr{R}_{2,\mathrm{SND}},$$

where $\mathscr{R}_{1,\mathrm{SND}}$ consists of (R_1, R_2) such that

$$R_1 \le I(X;Y_1) \tag{3}$$

or

$$\begin{aligned} R_1 &\le I(X;Y_1|W), \\ R_1 + R_2 &\le I(X,W;Y_1), \end{aligned} \tag{4}$$

and $\mathscr{R}_{2,\mathrm{SND}}$ is characterized by index substitution $1 \leftrightarrow 2$ and variable substitution $X \leftrightarrow W$ in (3) and (4), i.e.,

$$R_2 \le I(W;Y_2)$$

or

$$\begin{aligned} R_2 &\le I(W;Y_2|X), \\ R_1 + R_2 &\le I(X,W;Y_2). \end{aligned}$$

Note that $\mathscr{R}_{\mathrm{SND}}$ can be written as

$$\begin{aligned} \mathscr{R}_{\mathrm{SND}} &= (\mathscr{R}_{1,\mathrm{IAN}} \cup \mathscr{R}_{1,\mathrm{SD}}) \cap (\mathscr{R}_{2,\mathrm{IAN}} \cup \mathscr{R}_{2,\mathrm{SD}}) \\ &= (\mathscr{R}_{1,\mathrm{IAN}} \cap \mathscr{R}_{2,\mathrm{IAN}}) \cup (\mathscr{R}_{1,\mathrm{SD}} \cap \mathscr{R}_{2,\mathrm{IAN}}) \\ &\quad \cup (\mathscr{R}_{1,\mathrm{IAN}} \cap \mathscr{R}_{2,\mathrm{SD}}) \cup (\mathscr{R}_{1,\mathrm{SD}} \cap \mathscr{R}_{2,\mathrm{SD}}), \end{aligned} \tag{5}$$

where $\mathscr{R}_{1,\mathrm{SD}}$ is defined as the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\le I(X;Y_1|W), \\ R_2 &\le I(W;Y_1|X), \\ R_1 + R_2 &\le I(X,W;Y_1), \end{aligned} \tag{6}$$

and $\mathscr{R}_{2,\mathrm{SD}}$ is defined similarly by making the index substitution $1 \leftrightarrow 2$ and variable substitution $X \leftrightarrow W$ in $\mathscr{R}_{1,\mathrm{SD}}$.

As illustrated in Fig. 3d, $\mathscr{R}_{\mathrm{SND}}$ is in general strictly larger than the mix-and-match region in (2).

It turns out no decoding rule can improve upon $\mathscr{R}_{\mathrm{SND}}$. More precisely, given any codebook $\{(x^n(m_1), w^n(m_2))\}$, the probability of decoding error is minimized by the maximum likelihood decoding (MLD) rule

$$\begin{aligned} \hat{m}_1 &= \arg\max_{m_1} \sum_{m_2} \prod_{i=1}^{n} p_{Y_1|X,W}(y_{1i}|x_i(m_1), w_i(m_2)), \\ \hat{m}_2 &= \arg\max_{m_2} \sum_{m_1} \prod_{i=1}^{n} p_{Y_2|X,W}(y_{2i}|x_i(m_1), w_i(m_2)). \end{aligned} \tag{7}$$

The *optimal rate region* (or the MLD region) $\mathscr{R}^*(p)$ for the p-distributed random code ensembles is the closure of the set of rate pairs (R_1, R_2) such that the sequence of $(2^{nR_1}, 2^{nR_2}, n; p)$ random code ensembles satisfies

$$\lim_{n\to\infty} \mathsf{E}[P_e^{(n)}(\mathcal{C}_n)] = 0,$$

where the expectation is with respect to the randomness in codebook generation. It is established in [13] that SND is optimal for the p-distributed random code ensembles, i.e.,

$$\mathscr{R}^* = \mathscr{R}_{\mathrm{SND}}.$$

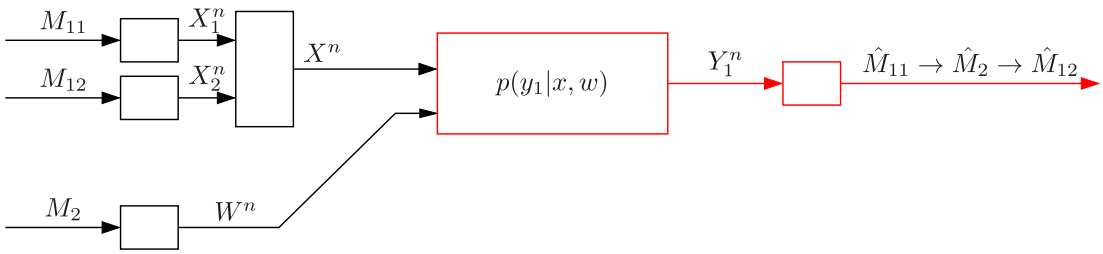


Fig. 4. Rate-splitting with successive cancellation for receiver 1.

As shown in Fig. 3d, $\mathscr{R}^* = \mathscr{R}_{\mathrm{SND}}$ is in general strictly larger than the mix-and-match region in (2), the gain of which may be attributed to high-complexity multiple sequence detection. The goal of this paper is to develop a coding scheme that achieves $\mathscr{R}^*$ using *low-complexity* point-to-point encoders and decoders.

III. Rate Splitting for the Interference Channel

In order to improve upon the mix-and-match scheme in the previous section at comparable complexity, one can incorporate the rate-splitting technique by Rimoldi and Urbanke [17] and Grant *et al.* [18].

A. Rate-Splitting Multiple Access

Consider the multiple access channel $p(y_1|x,w)$ with two inputs X and W and the common output Y_1. It is well-known that simultaneous decoding of the random code ensemble generated according to $p = p(x)p(w)$ achieves $\mathscr{R}_{1,\mathrm{SD}}(p)$ in (6). In the following, we show how to achieve this region via rate splitting with point-to-point decoders.

Suppose that the message $M_1 \in [2^{nR_1}]$ is split into two parts $(M_{11}, M_{12}) \in [2^{nR_{11}}] \times [2^{nR_{12}}]$ while the message $M_2 \in [2^{nR_2}]$ is not split. The messages m_{11} and m_{12} are encoded into codewords x_1^n and x_2^n, respectively, which are then symbol-by-symbol mapped to the transmitted sequence x^n, that is, $x_i(m_{11}, m_{12}) = x(x_{1i}(m_{11}), x_{2i}(m_{12}))$, $i \in [n]$, for some function $x(x_1, x_2)$. The message m_2 is mapped to w^n. For decoding, the receiver recovers $\hat{m}_{11}$, $\hat{m}_2$, and $\hat{m}_{12}$, successively, which is denoted as the decoding order

$$d_1 \colon \hat{m}_{11} \to \hat{m}_2 \to \hat{m}_{12}.$$

This rate-splitting scheme [17] with so-called homogeneous superposition coding [37] and successive cancellation decoding in Fig. 4 can be easily implemented by low-complexity point-to-point encoders and decoders.

Following the standard analysis for random code ensembles generated by $p'(x_1)p'(x_2)p'(w)$, decoding is successful if

$$\begin{aligned} R_{11} &< I(X_1; Y_1), \\ R_2 &< I(W; Y_1|X_1), \\ R_{12} &< I(X_2; Y_1|X_1, W) = I(X; Y_1|X_1, W). \end{aligned}$$

By setting $R_1 = R_{11} + R_{12}$, it follows that the scheme achieves the rate region $\mathscr{R}_{\mathrm{RS}}(p)$ consisting of (R_1, R_2) such that

$$\begin{aligned} R_1 &\le I(X_1; Y_1) + I(X; Y_1|X_1, W), \\ R_2 &\le I(W; Y_1|X_1). \end{aligned} \tag{8}$$

By varying $p'(x_1)p'(x_2)$ and $x(x_1, x_2)$, while maintaining $p'(x) = \sum_{x_1,x_2:x(x_1,x_2)=x} p'(x_1)p'(x_2) = p(x)$ and $p'(w) = p(w)$, which we compactly denote by $p' \simeq p$, the rectangular region (8) traces the boundary of rate region $\mathscr{R}_{1,\mathrm{SD}}(p)$. More precisely, we have the following identity; see Appendix A for the proof.

Lemma 1 (Layer-Splitting Lemma [18]):

$$\mathscr{R}_{1,\mathrm{SD}}(p) = \bigcup_{p' \simeq p} \mathscr{R}_{\mathrm{RS}}(p').$$

Remark 1: Simultaneous decoding of $\hat{M}_{11}, \hat{M}_{12}$, and $\hat{M}_2$ cannot achieve rates beyond $\mathscr{R}_{1,\mathrm{SD}}(p)$ and therefore it does not improve upon (the union of) successive cancellation decoding for the multiple access channel.

B. Rate Splitting for the Interference Channel

The main idea of rate splitting for the multiple access channel is to represent the messages by multiple parts and encode each into one of the superposition layers. Combined with successive cancellation decoding, this superposition coding scheme transforms the multiple access channel into a sequence of point-to-point channels, over which single-user encoders and decoders can be used. For the interference channel with multiple receivers, however, this rate-splitting scheme can no longer achieve the rate region of simultaneous decoding (cf. Remark 1). The root cause of this deficiency is not rate splitting per se, but suboptimal successive cancellation decoding. Indeed, proper rate splitting can achieve rates better than no splitting when simultaneous decoding is used (cf. Han–Kobayashi coding).

To understand the limitations of successive cancellation decoding, we consider the rate-splitting scheme with the same encoder structure as before and two decoding orders

$$\begin{aligned} d_1 &\colon \hat{m}_{11} \to \hat{m}_2 \to \hat{m}_{12}, \\ d_2 &\colon \hat{m}_{11} \to \hat{m}_{12} \to \hat{m}_2, \end{aligned}$$

as depicted in Fig. 5. Following the standard analysis, decoding is successful at receiver 1 if

$$R_{11} < I(X_1; Y_1), \tag{9a}$$
$$R_2 < I(W; Y_1|X_1), \tag{9b}$$
$$R_{12} < I(X; Y_1|X_1, W). \tag{9c}$$

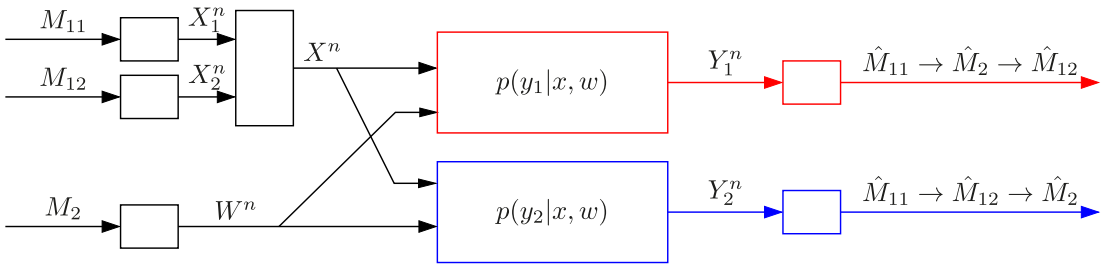


Fig. 5. Rate-splitting with successive cancellation in the two-user interference channel.

and at receiver 2 if

$$R_{11} < I(X_1; Y_2), \tag{9d}$$
$$R_{12} < I(X; Y_2|X_1), \tag{9e}$$
$$R_2 < I(W; Y_2|X). \tag{9f}$$

By Fourier–Motzkin elimination, this scheme achieves the rate region consisting of (R_1, R_2) such that

$$R_1 \le \min\{I(X_1; Y_1),\, I(X_1; Y_2)\} + \min\{I(X; Y_1|X_1, W),\, I(X; Y_2|X_1)\}, \tag{10a}$$
$$R_2 \le \min\{I(W; Y_1|X_1),\, I(W; Y_2|X)\}. \tag{10b}$$

Remark 2 (Min of the Sum vs. Sum of the Min): We note a common misconception in the literature, reported also in [38] (see the references therein), that the bounds on R_{11} and R_{12} in (9) would simplify to

$$R_1 \le \min\{I(X_1; Y_1) + I(X; Y_1|X_1, W),\, I(X; Y_2)\}, \tag{11}$$

which could be sufficient to achieve the MLD region $\mathscr{R}^*(p)$ in Section II. This conclusion is incorrect, since the bound in (10a) is strictly smaller than (11) in general. In fact, the rate region in (10), even after taking the union over all $p' \simeq p$ is strictly smaller than $\mathscr{R}^*(p)$. In order to ensure reliable communication over two different underlying multiple access channels $p(y_i|x, w), i = 1, 2$, the message parts in the rate-splitting scheme have to be loaded at the rate of the worse channel on each superposition layer, which in general incurs a total rate loss.

It turns out that this deficiency is fundamental and cannot be overcome by introducing more superposition layers and different decoding orders (which include treating interference as noise d_1: $\hat{m}_{11} \to \hat{m}_{12}$ and d_2: $\hat{m}_2$). To be more precise, we define the general (p', s, t, d_1, d_2) rate-splitting scheme. The message M_1 is split into s independent parts $M_{11}, M_{12}, \ldots, M_{1s}$ with rates $R_{11}, R_{12}, \ldots, R_{1s}$, respectively, and the message M_2 is split into t independent parts $M_{21}, M_{22}, \ldots, M_{2t}$ at rates $R_{21}, R_{22}, \ldots, R_{2t}$, respectively. These messages are encoded by the random code ensemble generated according to $p' = \left(\prod_{j=1}^{s} p'(x_j)\right)\left(\prod_{j=1}^{t} p'(w_j)\right)$ and the corresponding codewords are superimposed by symbol-by-symbol mappings $x(x_1, \ldots, x_s)$ and $w(w_1, \ldots, w_t)$. The receivers use successive cancellation decoding with decoding orders d_1 and d_2, where d_1 is an ordering of elements in $\{\hat{m}_{11}, \ldots, \hat{m}_{1s}, \hat{m}_{21}, \ldots, \hat{m}_{2k}\}$, $k \le t$, and d_2 is an ordering of elements in $\{\hat{m}_{11}, \ldots, \hat{m}_{1l}, \hat{m}_{21}, \ldots, \hat{m}_{2t}\}$, $l \le s$. The achievable rate region of this rate-splitting scheme is denoted by $\mathscr{R}_{\mathrm{RS}}(p', s, t, d_1, d_2)$. We establish the following statement in Appendix C.

Theorem 1: There exists an interference channel $p(y_1, y_2|x, w)$ and some input pmf $p = p(x)p(w)$ such that

$$\bigcup_{p' \simeq p} \mathscr{R}_{\mathrm{RS}}(p', s, t, d_1, d_2) \subsetneq \mathscr{R}^*(p) \tag{12}$$

for any finite s and t, and decoding orders d_1 and d_2. Moreover, strict inclusion holds even after taking union over all input pmfs p, i.e.,

$$\bigcup_{p} \bigcup_{p' \simeq p} \mathscr{R}_{\mathrm{RS}}(p', s, t, d_1, d_2) \subsetneq \left(\bigcup_{p} \mathscr{R}^*(p)\right). \tag{13}$$

Remark 3: It can be easily checked that the first three regions in the decomposition of $\mathscr{R}^*$ in (5) are achievable by properly chosen $(p', 2, 1, d_1, d_2)$ rate-splitting schemes. The fourth region $\mathscr{R}_{1,\mathrm{SD}} \cap \mathscr{R}_{2,\mathrm{SD}}$ is the bottleneck in achieving the entire $\mathscr{R}^*$ with rate splitting and successive cancellation.

IV. Sliding-Window Superposition Coding

In this section, we develop a new coding scheme, termed sliding-window superposition coding (SWSC), that overcomes the limitation of rate splitting by encoding the message to multiple superposition layers across consecutive blocks.

A. Corner Points

We first show how to achieve the rate region in (10b) and (11), which will be shown to be sufficient to achieve the corner points of $\mathscr{R}_{1,\mathrm{SD}} \cap \mathscr{R}_{2,\mathrm{SD}}$.

In SWSC, we consider a stream of messages, $(m_1(1), m_2(1)), (m_1(2), m_2(2)), \ldots$, to be communicated over multiple blocks. As before, $m_2(j)$ is encoded into a codeword w^n to be transmitted in block j. The message $m_1(j)$, which was split and transmitted in two layers X_1 and X_2 in the previous rate-splitting scheme, is now encoded into two sequences x_2^n and x_1^n to be transmitted in two consecutive blocks j and $j+1$, respectively; see Table I. The transmitted sequence x^n in block j is the symbol-by-symbol superposition of $x_1^n(m_1(j-1))$ and $x_2^n(m_1(j))$, which has the same superposition coding structure as in the rate-splitting scheme, but without actual splitting of message rates. Note that similar diagonal transmission of message streams has

block	$j-1$	j
X_1	$*$	$m_1(j-1)$
X_2	$m_1(j-1)$	$m_1(j)$
W	$*$	$m_2(j)$

(a) The initial state at the end of block j.

$j-1$	j
$*$	$*$
$*$	$m_1(j)$
$*$	$m_2(j)$

(b) Step 1: recover $\hat{m}_1(j-1)$.

$j-1$	j
$*$	$*$
$*$	$m_1(j)$
$*$	$*$

(c) Step 2: recover $\hat{m}_2(j)$.

Fig. 6. Illustration of the decoding process at receiver 1, where $*$ indicates a known or decoded codeword.

TABLE I
SLIDING-WINDOW SUPERPOSITION CODING SCHEME

block j	1	2	3	$\cdots$	$b-1$	b
X_1	1	$m_1(1)$	$m_1(2)$	$\ldots$	$\ldots$	$m_1(b-1)$
X_2	$m_1(1)$	$m_1(2)$	$\ldots$	$\ldots$	$m_1(b-1)$	1
W	$m_2(1)$	$m_2(2)$	$\ldots$	$\ldots$	$\ldots$	$m_2(b)$
Y_1	$\hat{m}_2(1)$	$\hat{m}_1(1)$ $\hat{m}_2(2)$	$\hat{m}_1(2)$ $\hat{m}_2(3)$	$\ldots$	$\ldots$	$\hat{m}_1(b-1)$ $\hat{m}_2(b)$
Y_2		$\hat{m}_1(1)$ $\hat{m}_2(1)$	$\hat{m}_1(2)$ $\hat{m}_2(2)$	$\ldots$	$\ldots$	$\hat{m}_1(b-1)$ $\hat{m}_2(b-1)\to\hat{m}_2(b)$

been already used in block Markov coding for relaying and feedback communication [39], [40]. For b blocks of communication, the scheme is initialized with $m_1(0)=1$ and terminated with $m_1(b)=1$, incurring a slight rate loss.

For decoding at receiver 1, $\hat{m}_1(j-1)$ and $\hat{m}_2(j)$ are recovered successively from the channel outputs $y_1^n(j-1)$ and $y_1^n(j)$, as shown in Fig. 6. In the language of typicality decoding, at the end of block j, it finds the unique message $\hat{m}_1(j-1)$ such that

$$(x_1^n(\hat{m}_1(j-2)), x_2^n(\hat{m}_1(j-1)), w^n(\hat{m}_2(j-1)), y_1^n(j-1)) \in \mathcal{T}_\epsilon^{(n)}(X_1, X_2, W, Y_1)$$

and

$$(x_1^n(\hat{m}_1(j-1)), y_1^n(j)) \in \mathcal{T}_\epsilon^{(n)}(X_1, Y_1)$$

simultaneously, where $\hat{m}_1(j-2)$ and $\hat{m}_2(j-1)$ are already known from the previous block. Then it finds the unique $\hat{m}_2(j)$ such that

$$(x_1^n(\hat{m}_1(j-1)), w^n(\hat{m}_2(j)), y_1^n(j)) \in \mathcal{T}_\epsilon^{(n)}(X_1, W, Y_1).$$

If any of the typicality checks fails, it declares an error. We represent this successive cancellation decoding operation compactly as

$$d_1\colon \hat{m}_1(j-1) \to \hat{m}_2(j), \tag{14}$$

which is performed at the end of block j. To recover the next pair of messages $\hat{m}_1(j)$ and $\hat{m}_2(j+1)$, receiver 1 slides the decoding window to $y_1^n(j)$ and $y_1^n(j+1)$ at the end of block $j+1$. This sliding-window decoding scheme is originally due to Carleial [41] and used in the network decode–forward relaying scheme [42], [43]. The overall schedule of message decoding is shown in Table I. As can be easily checked by inspection, decoding is successful if

$$\begin{aligned} R_1 &< I(X_1;Y_1) + I(X;Y_1|X_1,W), \\ R_2 &< I(W;Y_1|X_1). \end{aligned} \tag{15}$$

A formal proof of this error analysis along with a complete description of the corresponding random coding scheme is delegated to Appendix D. Receiver 2 similarly uses successive cancellation decoding at the end of each block j as

$$d_2\colon \hat{m}_1(j-1) \to \hat{m}_2(j-1).$$

In other words, at the end of block j, $2 \le j \le b$, it finds the unique $\hat{m}_1(j-1)$ such that

$$(x_1^n(\hat{m}_1(j-2)), x_2^n(\hat{m}_1(j-1)), y_2^n(j-1)) \in \mathcal{T}_\epsilon^{(n)}$$

and

$$(x_1^n(\hat{m}_1(j-1)), y_2^n(j)) \in \mathcal{T}_\epsilon^{(n)}$$

simultaneously. Then it finds the unique $\hat{m}_2(j-1)$ such that

$$(w^n(\hat{m}_2(j-1)), y_2^n(j-1), x_1^n(\hat{m}_1(j-2)), x_2^n(\hat{m}_1(j-1))) \in \mathcal{T}_\epsilon^{(n)}.$$

In the end, receiver 2 finds the unique $\hat{m}_2(b)$ such that

$$(w^n(\hat{m}_2(b)), y_2^n(b), x_1^n(\hat{m}_1(b-1)), x_2^n(m_1(b))) \in \mathcal{T}_\epsilon^{(n)}.$$

If any of the typicality checks fails, it declares an error. One can similarly check that the decoding is successful if

$$\begin{aligned} R_1 &< I(X_1;Y_2) + I(X_2;Y_2|X_1) = I(X;Y_2), \\ R_2 &< I(W;Y_2|X). \end{aligned}$$

When the nominal message rate pair of each block is (R_1, R_2), the scheme achieves $(\frac{b-1}{b}R_1, R_2)$ on average, which can be made arbitrarily close to (R_1, R_2) by letting $b \to \infty$. We summarize the performance of this SWSC scheme as follows.

Proposition 1: Let $p'(x_1)p'(x_2)p'(w)$ and $x(x_1,x_2)$ be fixed. Then the SWSC scheme in Table I achieves the rate region $\mathscr{R}_{\mathrm{SWSC}}(p', 2, 1, d_1, d_2)$ that consists of the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\le \min\{I(X_1;Y_1) + I(X;Y_1|X_1,W),\ I(X;Y_2)\}, \\ R_2 &\le \min\{I(W;Y_1|X_1),\ I(W;Y_2|X)\}. \end{aligned}$$

We now note that each corner point of $\mathscr{R}_{1,\mathrm{SD}} \cap \mathscr{R}_{2,\mathrm{SD}}$ is contained in one of the four regions

$$\begin{aligned} &\mathscr{R}_{1,\mathrm{SD}} \cap \mathscr{R}_{2,\mathrm{SCD1}\to 2}, \\ &\mathscr{R}_{1,\mathrm{SD}} \cap \mathscr{R}_{2,\mathrm{SCD2}\to 1}, \\ &\mathscr{R}_{1,\mathrm{SCD1}\to 2} \cap \mathscr{R}_{2,\mathrm{SD}}, \\ &\mathscr{R}_{1,\mathrm{SCD2}\to 1} \cap \mathscr{R}_{2,\mathrm{SD}}, \end{aligned} \tag{16}$$

where $\mathscr{R}_{j,\mathrm{SCD1}\to 2}, j = 1,2$, is the set of rate pairs (R_1, R_2) such that $R_1 \le I(X;Y_j), R_2 \le I(W;Y_j|X)$ and

TABLE II
SWSC SCHEME WITH DECODING ORDERS IN (18)

block j	1	2	3	4	$\cdots$	$b-1$	b
X_1	1	1	$m_1(1)$	$m_1(2)$	$\ldots$	$\ldots$	$m_1(b-2)$
X_2	1	$m_1(1)$	$m_1(2)$	$\ldots$	$\ldots$	$m_1(b-2)$	1
X_3	$m_1(1)$	$m_1(2)$	$\ldots$	$\ldots$	$m_1(b-2)$	1	1
W	$m_2(1)$	$m_2(2)$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$m_2(b)$
Y_1	$\hat{m}_2(1)$	$\hat{m}_2(2)$	$\hat{m}_1(1)$ $\hat{m}_2(3)$	$\hat{m}_1(2)$ $\hat{m}_2(4)$	$\ldots$	$\ldots$	$\hat{m}_1(b-2)$ $\hat{m}_2(b)$
Y_2	$\hat{m}_2(1)$		$\hat{m}_1(1)$ $\hat{m}_2(2)$	$\hat{m}_1(2)$ $\hat{m}_2(3)$	$\ldots$	$\ldots$	$\hat{m}_1(b-2)$ $\hat{m}_2(b-1) \to \hat{m}_2(b)$

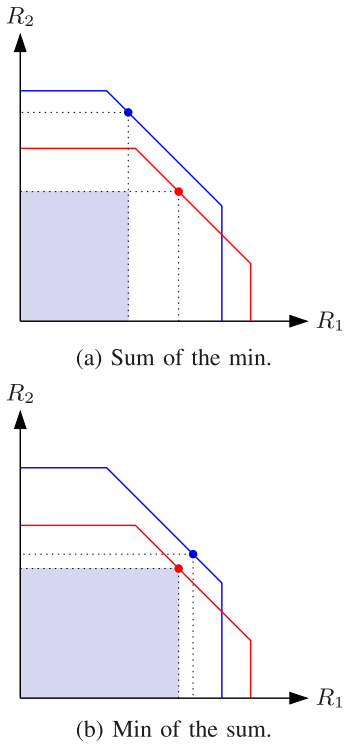
$\mathscr{R}_{j,\mathrm{SCD2}\to 1}$ is the set of rate pairs (R_1, R_2) such that $R_1 \le I(X;Y_j|W), R_2 \le I(W;Y_j)$. Since any boundary point in $\mathscr{R}_{1,\mathrm{SD}}$ can be expressed as (8) by Lemma 1, $\mathscr{R}_{1,\mathrm{SD}}(p) \cap \mathscr{R}_{2,\mathrm{SCD1}\to 2}(p)$ is contained in $\mathscr{R}_{\mathrm{SWSC}}(p', 2, 1)$ for some $p' \simeq p$ and is achieved by the SWSC scheme. The other three regions in (16) can be achieved similarly by using different decoding orders, and thus SWSC achieves every corner point of $\mathscr{R}_{1,\mathrm{SD}} \cap \mathscr{R}_{2,\mathrm{SD}}$.

Remark 4: In the SWSC scheme above, for *finite* b, there is a rate loss $(1/b)R_1$ for message M_1, since no message is scheduled via X_1^n in block 1 and via X_2^n in block b. The decoding delay of one block ($\hat{m}_1(j)$ recovered in block $j+1$) is independent of b, while the overall probability of error is, by the union-of-events bound, linear in b due to error propagation.

Remark 5: In order to reduce the rate loss, we can instead send message M_1 at the treating-interference-as-noise rate $\min\{I(X_1;Y_1), I(X_1;Y_2)\}$ for X_1^n in block 1 and at rate $\min\{I(X;Y_1|X_1,W), I(X;Y_2|X_1)\}$ for X_2^n in block b. This increases the overall R_1 by

$$\frac{1}{b}[\min\{I(X_1;Y_1), I(X_1;Y_2)\} + \min\{I(X;Y_1|X_1,W), I(X;Y_2|X_1)\}],$$

which is the same as $1/b$ times the achievable R_1 by rate-splitting in (10a).



(a) Sum of the min.

(b) Min of the sum.

Fig. 7. Rate loss of rate splitting in the interference channel.

B. General Rate Points

The SWSC scheme developed in the previous section cannot achieve the entire region of $\mathscr{R}_{1,\mathrm{SD}} \cap \mathscr{R}_{2,\mathrm{SD}}$ in general. As illustrated in Fig. 7a, the scheme can achieve any point on the dominant face of $\mathscr{R}_{1,\mathrm{SD}}$ or $\mathscr{R}_{2,\mathrm{SD}}$ at the respective receiver. (This is clearly an improvement over the rate-splitting multiple access scheme as noted in Remark 2.) In general, however, these two points are not aligned, which may result in a rate region strictly smaller than $\mathscr{R}_{1,\mathrm{SD}} \cap \mathscr{R}_{2,\mathrm{SD}}$. To overcome this deficiency, we introduce an additional layer to X while keeping W unsplit. The receivers now have the flexibility of merging three layers X_1, X_2, X_3 into two groups, for example, $(X_1), (X_2, X_3)$ at receiver 1 and $(X_1, X_2), (X_3)$ at receiver 2, which can align the two points on the dominant faces of $\mathscr{R}_{1,\mathrm{SD}}$ and $\mathscr{R}_{2,\mathrm{SD}}$ as illustrated in Fig. 7b.

To be more precise, we first present a coding scheme that achieves the rate region consisting of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\le \min\{I(X_1;Y_1) + I(X_2,X_3;Y_1|X_1,W), \\ &\qquad I(X_1,X_2;Y_2) + I(X_3;Y_2|X_1,X_2,W)\}, \\ R_2 &\le \min\{I(W;Y_1|X_1),\ I(W;Y_2|X_1,X_2)\}. \end{aligned} \tag{17}$$

In this SWSC scheme, the message $m_1(j)$ is encoded into three sequences x_3^n, x_2^n, and x_1^n to be transmitted in three consecutive blocks j, $j+1$, and $j+2$, respectively. The message $m_2(j)$ is encoded into a codeword w^n to be transmitted in block j. The encoding structure is illustrated in Table II. The transmitted sequence x^n in block j is the symbol-by-symbol superposition of $x_1^n(m_1(j-2))$, $x_2^n(m_1(j-1))$, and $x_3^n(m_1(j))$.

For decoding, the message $\hat{m}_1(j)$ is recovered via sliding-window decoding over three blocks. The decoding orders at two receivers are

$$d_1\colon \hat{m}_1(j-2) \to \hat{m}_2(j), \tag{18a}$$
$$d_2\colon \hat{m}_1(j-2) \to \hat{m}_2(j-1). \tag{18b}$$

The decoding process is illustrated in Table II. Following the standard analysis, the decoding is successful at receiver 1 if

$$R_1 < I(X_1;Y_1)+I(X_2;Y_1|X_1,W)+I(X_3;Y_1|X_1,X_2,W),$$
$$R_2 < I(W;Y_1|X_1),$$

and at receiver 2 if

$$R_1 < I(X_1;Y_2) + I(X_2;Y_2|X_1) + I(X_3;Y_2|X_1,X_2,W),$$
$$R_2 < I(W;Y_2|X_1,X_2),$$

which establishes the achievability of the rate region in (17). We denote this rate region by $\mathscr{R}_{\mathrm{SWSC}}(p',3,1,d_1,d_2)$.

By swapping the decoding orders between receivers 1 and 2, i.e.,

$$d_1'\colon \hat{m}_1(j-2) \to \hat{m}_2(j-1), \tag{19a}$$
$$d_2'\colon \hat{m}_1(j-2) \to \hat{m}_2(j), \tag{19b}$$

the SWSC scheme achieves the rate region $\mathscr{R}_{\mathrm{SWSC}}(p',3,1,d_1',d_2')$ characterized by

$$R_1 \le \min\{I(X_1,X_2;Y_1) + I(X_3;Y_1|X_1,X_2,W),$$
$$\qquad I(X_1;Y_2) + I(X_2,X_3;Y_2|X_1,W)\},$$
$$R_2 \le \min\{I(W;Y_2|X_1),\ I(W;Y_1|X_1,X_2)\}.$$

This SWSC scheme turns out to be sufficient to achieve any rate point in the simultaneous decoding region; see Appendix B for the proof.

Proposition 2:

$$\mathscr{R}_{1,\mathrm{SD}}(p)\cap\mathscr{R}_{2,\mathrm{SD}}(p) = \bigcup_{p'\simeq p}\ \bigcup_{(d_1,d_2)=(18)\ or\ (19)} \mathscr{R}_{\mathrm{SWSC}}(p',3,1,d_1,d_2).$$

C. SWSC Achieves the MLD Region $\mathscr{R}^*$

We now show that the other three component regions of $\mathscr{R}^*$ in (5), namely, $\mathscr{R}_{1,\mathrm{IAN}}\cap\mathscr{R}_{2,\mathrm{IAN}}$, $\mathscr{R}_{1,\mathrm{SD}}\cap\mathscr{R}_{2,\mathrm{IAN}}$, and $\mathscr{R}_{1,\mathrm{IAN}}\cap\mathscr{R}_{2,\mathrm{SD}}$, can be also achieved by the SWSC scheme in Table II (with the same encoding scheme, but with different decoding orders).

- $\mathscr{R}_{1,\mathrm{IAN}}\cap\mathscr{R}_{2,\mathrm{IAN}}$:

$$d_1\colon \hat{m}_1(j-2), \tag{20a}$$
$$d_2\colon \hat{m}_2(j). \tag{20b}$$

  The corresponding achievable rate region is the set of rate pairs (R_1,R_2) such that

$$R_1 \le I(X_1;Y_1) + I(X_2;Y_1|X_1) + I(X_3;Y_1|X_1,X_2) = I(X;Y_1),$$
$$R_2 \le I(W;Y_2).$$

- $\mathscr{R}_{1,\mathrm{SD}}\cap\mathscr{R}_{2,\mathrm{IAN}}$:

$$d_1\colon \hat{m}_1(j-2) \to \hat{m}_2(j), \tag{21a}$$
$$d_2\colon \hat{m}_2(j). \tag{21b}$$

  The corresponding achievable rate region is the set of rate pairs (R_1,R_2) such that

$$R_1 \le I(X_1;Y_1) + I(X;Y_1|X_1,W),$$
$$R_2 \le \min\{I(W;Y_1|X_1),\ I(W;Y_2)\},$$

  which, after taking the union over all $p'\simeq p$, is equivalent to $\mathscr{R}_{1,\mathrm{SD}}(p)\cap\mathscr{R}_{2,\mathrm{IAN}}(p)$ by Lemma 1.

- $\mathscr{R}_{1,\mathrm{IAN}}\cap\mathscr{R}_{2,\mathrm{SD}}$:

$$d_1\colon \hat{m}_1(j-2), \tag{22a}$$
$$d_2\colon \hat{m}_1(j-2) \to \hat{m}_2(j). \tag{22b}$$

  The corresponding achievable rate region is the set of rate pairs (R_1,R_2) such that

$$R_1 \le \min\{I(X;Y_1),\ I(X_1;Y_2) + I(X;Y_2|X_1,W)\},$$
$$R_2 \le I(W;Y_2|X_1),$$

  which, after taking the union over all $p'\simeq p$, is equivalent to $\mathscr{R}_{1,\mathrm{IAN}}(p)\cap\mathscr{R}_{2,\mathrm{SD}}(p)$ by Lemma 1.

In summary, the SWSC scheme in Table II, with $p'\simeq p$ and decoding orders (18)–(22), achieves the MLD region $\mathscr{R}^*$.

Theorem 2:

$$\mathscr{R}^*(p) = \bigcup_{p'\simeq p}\ \bigcup_{(d_1,d_2)=(18)\text{–}(22)} \mathscr{R}_{\mathrm{SWSC}}(p',3,1,d_1,d_2).$$

V. Superposition Layers and Decoding Orders

SWSC with a given encoder structure allows multiple decoding schemes, each with a different rate region. In this section, we provide a more systematic treatment of the relationship between superposition layers and decoding orders.

Suppose that we split X into K layers $(X_1,\ldots,X_K)$ and W into L layers $(W_1,\ldots,W_L)$. Consider a stream of messages, $(m_1(1),m_2(1)),(m_1(2),m_2(2)),\ldots$, to be communicated over multiple blocks. The message $m_1(j)$ is encoded into K sequences x_K^n, $x_{K-1}^n,\ldots,$ and x_1^n to be transmitted in K consecutive blocks j, $j+1,\ldots,$ and $j+K-1$, respectively. Similarly, the message $m_2(j)$ is encoded into L sequences w_L^n, $w_{L-1}^n,\ldots,$ and w_1^n to be transmitted in L consecutive blocks $j, j+1,\ldots,$ and $j+L-1$, respectively. The transmitted sequence x^n in block j is the symbol-by-symbol superposition of $x_1^n(m_1(j-K+1))$, $x_2^n(m_1(j-K+2)),\ldots,$ and $x_K^n(m_1(j))$. The transmitted sequence w^n in block j is the symbol-by-symbol superposition of $w_1^n(m_2(j-L+1))$, $w_2^n(m_2(j-L+2)),\ldots,$ and $w_L^n(m_2(j))$. We refer to such a layer split and message schedule as the *K-L split*. Table III illustrates the encoding of the 3-2 split.

As we saw in the previous section, different decoding orders may result in different achievable rate regions. A feasible decoding order for a K-L split is of the following form. At the end of block j, receiver $k=1,2$ either recovers

$$\hat{m}_1(j-K+1) \to \hat{m}_2(j-K+1-t_1),$$

TABLE III
SWSC ENCODING WITH A 3-2 SPLIT

block j	1	2	3	4	$\cdots$	$b-1$	b
X_1	1	1	$m_1(1)$	$m_1(2)$	$\cdots$	$\cdots$	$m_1(b-2)$
X_2	1	$m_1(1)$	$m_1(2)$	$\cdots$	$\cdots$	$m_1(b-2)$	1
X_3	$m_1(1)$	$m_1(2)$	$\cdots$	$\cdots$	$m_1(b-2)$	1	1
W_1	1	$m_2(1)$	$m_2(2)$	$\cdots$	$\cdots$	$\cdots$	$m_2(b-1)$
W_2	$m_2(1)$	$m_2(2)$	$\cdots$	$\cdots$	$\cdots$	$m_2(b-1)$	1

for some $t_1 = \min\{K, L\} - 1, \ldots, 1, 0$, or

$$\hat{m}_2(j - L + 1) \to \hat{m}_1(j - L + 1 - t_2),$$

for some $t_2 = 0, 1, \ldots, \max\{K, L\} - 1$. For the 3-2 split in Table III, there are five feasible decoding orders:

$$1\colon \hat{m}_1(j-2) \to \hat{m}_2(j-3) \quad (t_1 = 1) \tag{23}$$
$$2\colon \hat{m}_1(j-2) \to \hat{m}_2(j-2) \quad (t_1 = 0) \tag{24}$$
$$3\colon \hat{m}_1(j-2) \to \hat{m}_2(j-1) \quad (t_2 = 0)$$
$$4\colon \hat{m}_2(j-1) \to \hat{m}_1(j-2) \quad (t_2 = 1)$$
$$5\colon \hat{m}_2(j-1) \to \hat{m}_1(j-3) \quad (t_2 = 2)$$

In order to write the achievable rate region corresponding to each decoding order, we introduce the notion of *layer order*. Let $\lambda\colon Z_1 \to Z_2 \to \cdots \to Z_{K+L}$ be an ordering of the variables $\{X_1, \ldots, X_K, W_1, \ldots, W_L\}$ such that the relative orders $X_1 \to X_2 \to \cdots \to X_K$ and $W_1 \to W_2 \to \cdots \to W_L$ are preserved. We say that a layer order is *alternating* if it starts with either $X_1 \to \cdots \to X_{a_1}$, $a_1 = \max\{K, L\} - 1, \ldots, 1, 0$, or $W_1 \to \cdots \to W_{a_2}$, $a_2 = 1, 2, \ldots, \min\{K, L\}$, followed by one X and one W alternately until one of them is exhausted, and then by the remaining variables. As in the decoding orders, there are $K + L$ alternating layer orders. For the 3-2 split in Table III, the five alternating layer orders are listed as follows [1]

$$1\colon X_1 \to X_2 \to X_3 \to W_1 \to W_2, \tag{25}$$
$$2\colon X_1 \to X_2 \to W_1 \to X_3 \to W_2, \tag{26}$$
$$3\colon X_1 \to W_1 \to X_2 \to W_2 \to X_3,$$
$$4\colon W_1 \to X_1 \to W_2 \to X_2 \to X_3,$$
$$5\colon W_1 \to W_2 \to X_1 \to X_2 \to X_3.$$

A layer order indicates which variable (signal layer) is recovered first in successive cancellation decoding. For example, in decoding order $d = 1$ in (23), X_1, X_2, X_3 carrying $m_1(j-2)$ are recovered before W_1, W_2 carrying $m_2(j-3)$. In other words, all the X layers are recovered before the W layers in successive cancellation decoding, which corresponds to the layer order $\lambda = 1$ in (25). For another example, in decoding order $d = 2$ in (24), at the end of block j, $3 \le j \le b$, we alternately recover $\hat{m}_1(j-2)$ and $\hat{m}_2(j-2)$. The layers X_1 and X_2 are recovered before the layer W_1, while the layer X_3 is recovered after the layer W_1, which is followed by the layer W_2. This corresponds to the layer order $\lambda = 2$ in (26).

Given a layer order, the achievable rates R_1 and R_2 are given as sums of the corresponding mutual information terms. For example, for the layer order $\lambda = 1$ in (25), the achievable rate region at receiver $k = 1, 2$ is the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\le I(X_1; Y_k) + I(X_2; Y_k|X_1) + I(X_3; Y_k|X_1, X_2), \\ R_2 &\le I(W_1; Y_k|X_1, X_2, X_3) + I(W_2; Y_k|X_1, X_2, X_3, W_1). \end{aligned} \tag{27}$$

Similarly, for the layer order $\lambda = 2$ in (26), the achievable rate region at receiver k is characterized as

$$\begin{aligned} R_1 &\le I(X_1; Y_k) + I(X_2; Y_k|X_1) + I(X_3; Y_k|X_1, X_2, W_1), \\ R_2 &\le I(W_1; Y_k|X_1, X_2) + I(W_2; Y_k|X_1, X_2, X_3, W_1). \end{aligned} \tag{28}$$

Given a layer order $\lambda\colon Z_1 \to \cdots \to Z_{K+L}$, define

$$\begin{aligned} \mathcal{I}_1 &= \{i\colon Z_i \in \{X_1, \ldots, X_K\}\}, \\ \mathcal{I}_2 &= \{i\colon Z_i \in \{W_1, \ldots, W_L\}\}. \end{aligned} \tag{29}$$

Then the achievable rate region at receiver k with corresponding decoding order $d = \lambda$ is the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\le \sum_{i \in \mathcal{I}_1} I(Z_i; Y_k|Z^{i-1}), \\ R_2 &\le \sum_{i \in \mathcal{I}_2} I(Z_i; Y_k|Z^{i-1}). \end{aligned} \tag{30}$$

VI. HAN–KOBAYASHI INNER BOUND

The Han–Kobayashi coding scheme [14], illustrated in Fig. 8, is the most powerful among known single-letter coding techniques for the two-user interference channel. In this scheme, rate splitting is used for the messages $M_1 = (M_{10}, M_{11})$ and $M_2 = (M_{20}, M_{22})$. The messages $M_{10}, M_{11}, M_{20}, M_{22}$ are carried by codewords s^n, t^n, u^n, v^n, which are then superimposed into x^n and w^n by symbol-by-symbol mappings $x(s, t)$ and $w(u, v)$. Receiver 1 recovers $\hat{M}_{10}, \hat{M}_{20}, \hat{M}_{11}$ and receiver 2 recovers $\hat{M}_{10}, \hat{M}_{20}, \hat{M}_{22}$ using simultaneous decoding. If we consider S, T, U, V as the channel inputs, the original two-user interference channel can then be viewed as a four-sender two-receiver channel with conditional pmf

$$p(y_1, y_2|s, t, u, v) = p(y_1, y_2|x(s, t), w(u, v)).$$

For a fixed input pmf $p(s)p(t)p(u)p(v)$ and functions $x(s, t), w(u, v)$, the Han–Kobayashi coding scheme achieves the 4-dimensional auxiliary rate region

$$\mathscr{R}_{1,\mathrm{MAC}} \cap \mathscr{R}_{2,\mathrm{MAC}} \tag{31}$$

[1] There are layer orders that are not alternating, yet still preserve the relative orders $X_1 \to X_2 \to X_3$ and $W_1 \to W_2$. For example, $X_1 \to W_1 \to W_2 \to X_2 \to X_3$ and $W_1 \to X_1 \to X_2 \to X_3 \to W_2$. With the SWSC scheme introduced in this paper, such *nonalternating* layer orders do not admit corresponding decoding orders. However, it turns out nonalternating layer orders can be achieved if a new "dimension" is introduced in messages scheduling. The detail of this extension, which can be found in [28, Section 4.4], is beyond the scope of this paper.

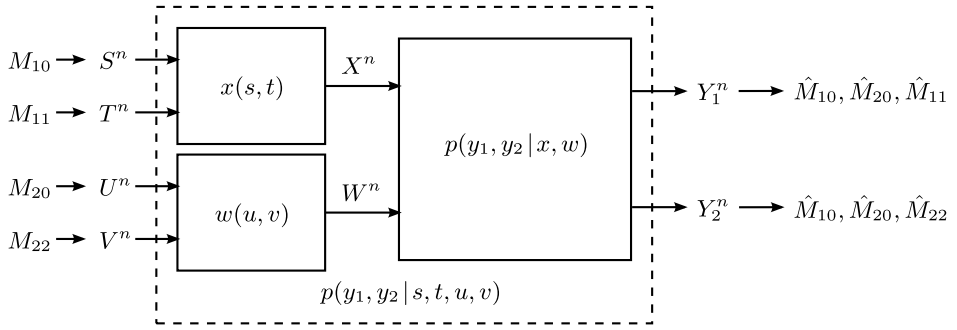


Fig. 8. Han–Kobayashi coding scheme.

TABLE IV
A SCHEME THAT ACHIEVES THE HAN–KOBAYASHI INNER BOUND WITH SINGLE-USER DECODING

block j	1	2	3	4	$\cdots$	$b-1$	b
S_1	1	1	$m_{10}(1)$	$m_{10}(2)$	$\ldots$	$\ldots$	$m_{10}(b-2)$
S_2	1	$m_{10}(1)$	$m_{10}(2)$	$\ldots$	$\ldots$	$m_{10}(b-2)$	1
S_3	$m_{10}(1)$	$m_{10}(2)$	$\ldots$	$\ldots$	$m_{10}(b-2)$	1	1
U	$m_{20}(1)$	$m_{20}(2)$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$m_{20}(b)$
T_1	$m'_{11}(1)$	$m'_{11}(2)$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$m'_{11}(b)$
T_2	$m''_{11}(1)$	$m''_{11}(2)$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$m''_{11}(b)$
V_1	$m'_{22}(1)$	$m'_{22}(2)$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$m'_{22}(b)$
V_2	$m''_{22}(1)$	$m''_{22}(2)$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$m''_{22}(b)$

where

$$\begin{aligned}\mathscr{R}_{1,\mathrm{MAC}} &= \{(R_{10}, R_{11}, R_{20}, R_{22}): \\ &\qquad (R_{10}, R_{11}, R_{20}) \in \mathscr{R}_{\mathrm{MAC}}(S, T, U; Y_1)\}, \\ \mathscr{R}_{2,\mathrm{MAC}} &= \{(R_{10}, R_{11}, R_{20}, R_{22}): \\ &\qquad (R_{10}, R_{20}, R_{22}) \in \mathscr{R}_{\mathrm{MAC}}(S, U, V; Y_2)\},\end{aligned}$$

and $\mathscr{R}_{\mathrm{MAC}}(A, B, C; Y)$ is the standard rate region for a three-user MAC $p(y|a, b, c)$ by random code ensemble $p(a)p(b)p(c)$. Recall that $\mathscr{R}_{\mathrm{MAC}}(A, B, C; Y)$ consists of rate triples (r_1, r_2, r_3) such that

$$\begin{aligned} r_1 &\le I(A; Y|B, C), \\ r_2 &\le I(B; Y|A, C), \\ r_3 &\le I(C; Y|A, B), \\ r_1 + r_2 &\le I(A, B; Y|C), \\ r_1 + r_3 &\le I(A, C; Y|B), \\ r_2 + r_3 &\le I(B, C; Y|A), \\ r_1 + r_2 + r_3 &\le I(A, B, C; Y). \end{aligned}$$

Finally, the Han–Kobayashi inner bound is the union over $p(s)p(t)p(u)p(v)$ and functions $x(s, t), w(u, v)$ of the rate region

$$\mathrm{Proj}_{4\to 2}\big(\mathscr{R}_{1,\mathrm{MAC}} \cap \mathscr{R}_{2,\mathrm{MAC}}\big), \tag{32}$$

where $\mathrm{Proj}_{4\to 2}$ denotes the projection of the 4-dimensional region of rate quadruples $(R_{10}, R_{11}, R_{20}, R_{22})$ to the 2-dimensional region of rate pairs $(R_1, R_2) = (R_{10} + R_{11}, R_{20} + R_{22})$.

Now we present a coding scheme that achieves the Han–Kobayashi inner bound with single-user decoding by showing the achievability of the 4-dimensional auxiliary region in (31). The two common messages M_{10} and M_{20} are transmitted using SWSC, with the 3-1 split in Section IV-B. The two private messages M_{11} and M_{22} are transmitted using the single-block rate-splitting scheme in Section III-A. The signal S is further split into three layers S_1, S_2, and S_3. For $j \in [b-2]$, the message $M_{10}(j)$ is carried by s_3^n, s_2^n, and s_1^n over blocks $j, j+1$, and $j+2$ respectively. Since the signal T is kept unsplit, the message $M_{20}(j)$ is carried by a single-block code u^n in block j. The private messages are further split into two parts $M_{11} = (M'_{11}, M''_{11})$ and $M_{22} = (M'_{22}, M''_{22})$. The four messages $M'_{11}, M''_{11}, M'_{22}, M''_{22}$ are carried by $t_1^n, t_2^n, v_1^n, v_2^n$, respectively, in a single block. The encoding is illustrated in Table IV.

At receiver 1, messages are recovered in the order d_1, which is one of the following six (trivial messages at the first and last blocks are skipped):

$$\begin{aligned}
&1\colon \hat{m}'_{11}(j-1) \to \hat{m}_{10}(j-2) \to \hat{m}_{20}(j-2) \to \hat{m}''_{11}(j-2), \\
&2\colon \hat{m}'_{11}(j) \to \hat{m}_{20}(j) \to \hat{m}_{10}(j-2) \to \hat{m}''_{11}(j), \\
&3\colon \hat{m}'_{11}(j) \to \hat{m}_{10}(j-2) \to \hat{m}''_{11}(j-2) \to \hat{m}_{20}(j), \\
&4\colon \hat{m}'_{11}(j-2) \to \hat{m}_{10}(j-2) \to \hat{m}_{20}(j-2) \to \hat{m}''_{11}(j-2), \\
&5\colon \hat{m}'_{11}(j) \to \hat{m}_{20}(j) \to \hat{m}_{10}(j-2) \to \hat{m}''_{11}(j-1), \\
&6\colon \hat{m}'_{11}(j) \to \hat{m}_{10}(j-2) \to \hat{m}''_{11}(j-2) \to \hat{m}_{20}(j-1).
\end{aligned}$$

Fig. 9 illustrates the decoding process for $d_1 = 1$, where $*$ indicates messages that were recovered previously. By the standard analysis, the achievable rate region for this decoding

block	$j-2$	$j-1$	j
S_1	$*$	$*$	$m_{10}(j-2)$
S_2	$*$	$m_{10}(j-2)$	$m_{10}(j-1)$
S_3	$m_{10}(j-2)$	$m_{10}(j-1)$	$m_{10}(j)$
U	$m_{20}(j-2)$	$m_{20}(j-1)$	$m_{20}(j)$
T_1	$*$	$m'_{11}(j-1)$	$m'_{11}(j)$
T_2	$m''_{11}(j-2)$	$m''_{11}(j-1)$	$m''_{11}(j)$

(a) The initial state at the end of block j.

$j-2$	$j-1$	j
$*$	$*$	$m_{10}(j-2)$
$*$	$m_{10}(j-2)$	$m_{10}(j-1)$
$m_{10}(j-2)$	$m_{10}(j-1)$	$m_{10}(j)$
$m_{20}(j-2)$	$m_{20}(j-1)$	$m_{20}(j)$
$*$	$*$	$m'_{11}(j)$
$m''_{11}(j-2)$	$m''_{11}(j-1)$	$m''_{11}(j)$

(b) Step 1: recover $\hat{m}'_{11}(j-1)$.

block	$j-2$	$j-1$	j
S_1	$*$	$*$	$*$
S_2	$*$	$*$	$m_{10}(j-1)$
S_3	$*$	$m_{10}(j-1)$	$m_{10}(j)$
U	$m_{20}(j-2)$	$m_{20}(j-1)$	$m_{20}(j)$
T_1	$*$	$*$	$m'_{11}(j)$
T_2	$m''_{11}(j-2)$	$m''_{11}(j-1)$	$m''_{11}(j)$

(c) Step 2: recover $\hat{m}_{10}(j-2)$ over blocks $j-2, j-1$, and j.

$j-2$	$j-1$	j
$*$	$*$	$*$
$*$	$*$	$m_{10}(j-1)$
$*$	$m_{10}(j-1)$	$m_{10}(j)$
$*$	$m_{20}(j-1)$	$m_{20}(j)$
$*$	$*$	$m'_{11}(j)$
$m''_{11}(j-2)$	$m''_{11}(j-1)$	$m''_{11}(j)$

(d) Step 3: recover $\hat{m}_{20}(j-2)$.

block	$j-2$	$j-1$	j
S_1	$*$	$*$	$*$
S_2	$*$	$*$	$m_{10}(j-1)$
S_3	$*$	$m_{10}(j-1)$	$m_{10}(j)$
U	$*$	$m_{20}(j-1)$	$m_{20}(j)$
T_1	$*$	$*$	$m'_{11}(j)$
T_2	$*$	$m''_{11}(j-1)$	$m''_{11}(j)$

(e) Step 4: recover $\hat{m}''_{11}(j-2)$.

Fig. 9. Illustration of the decoding process for $d_1 = 1$.

order is the set of rate quadruples $(R_{10}, R_{11}, R_{20}, R_{22})$ such that

$$\begin{aligned} R_{10} &\le I(S_1;Y_1) + I(S_2;Y_1|S_1,T_1) + I(S_3;Y_1|S_1,T_1,S_2),\\ R_{20} &\le I(U;Y_1|S_1,T_1,S_2,S_3),\\ R_{11} &\le I(T_1;Y_1|S_1) + I(T_2;Y_1|S_1,T_1,S_2,S_3,U), \end{aligned} \tag{33}$$

which is exactly the rate region corresponding to the layer order λ_1

$$1\colon S_1 \to T_1 \to S_2 \to S_3 \to U \to T_2.$$

One can similarly verify that the layer orders λ_1 corresponding to decoding orders $d_1 = 2, \ldots, 6$ are

$$\begin{aligned} 2\colon & T_1 \to U \to S_1 \to T_2 \to S_2 \to S_3,\\ 3\colon & T_1 \to S_1 \to U \to S_2 \to S_3 \to T_2,\\ 4\colon & S_1 \to S_2 \to T_1 \to S_3 \to U \to T_2,\\ 5\colon & T_1 \to U \to S_1 \to S_2 \to T_2 \to S_3,\\ 6\colon & T_1 \to S_1 \to S_2 \to U \to S_3 \to T_2. \end{aligned}$$

At receiver 2, the messages are recovered in the order d_2, which is one of the following six:

$$\begin{aligned} 7\colon & \hat{m}'_{22}(j-1) \to \hat{m}_{10}(j-2) \to \hat{m}_{20}(j-2) \to \hat{m}''_{22}(j-2),\\ 8\colon & \hat{m}'_{22}(j) \to \hat{m}_{20}(j) \to \hat{m}_{10}(j-2) \to \hat{m}''_{22}(j),\\ 9\colon & \hat{m}'_{22}(j) \to \hat{m}_{10}(j-2) \to \hat{m}''_{22}(j-2) \to \hat{m}_{20}(j),\\ 10\colon & \hat{m}'_{22}(j-2) \to \hat{m}_{10}(j-2) \to \hat{m}_{20}(j-2) \to \hat{m}''_{22}(j-2),\\ 11\colon & \hat{m}'_{22}(j) \to \hat{m}_{20}(j) \to \hat{m}_{10}(j-2) \to \hat{m}''_{22}(j-1),\\ 12\colon & \hat{m}'_{22}(j) \to \hat{m}_{10}(j-2) \to \hat{m}''_{22}(j-2) \to \hat{m}_{20}(j-1), \end{aligned}$$

with corresponding achievable layer orders λ_2

$$\begin{aligned} 7\colon & S_1 \to V_1 \to S_2 \to S_3 \to U \to V_2,\\ 8\colon & V_1 \to U \to S_1 \to V_2 \to S_2 \to S_3,\\ 9\colon & V_1 \to S_1 \to U \to S_2 \to S_3 \to V_2,\\ 10\colon & S_1 \to S_2 \to V_1 \to S_3 \to U \to V_2,\\ 11\colon & V_1 \to U \to S_1 \to S_2 \to V_2 \to S_3,\\ 12\colon & V_1 \to S_1 \to S_2 \to U \to S_3 \to V_2. \end{aligned}$$

(a) MLC: Short, nonuniversal. $R' < I(X_1;Y), R'' < I(X_2;Y|X_1)$.

(b) BICM: Treat other layers as noise. $R < I(X_1;Y) + I(X_2;Y)$.

(c) SWCM: Error propagation, rate loss. $R < I(X_1;Y) + I(X_2;Y|X_1)$.

Fig. 10. Comparison of three coded modulation schemes.

Let p' be the pmf $p'(s_1)p'(s_2)p'(s_3)p'(t_1)p'(t_2)p'(u)p'(v_1)$ $p'(v_2)$ along with $s(s_1,s_2,s_3), t(t_1,t_2)$, and $v(v_1,v_2)$. Let $\mathscr{R}_1(p',\lambda_1)$ be the rate region corresponding to the layer order $\lambda_1 = 1,\ldots,6$ at receiver 1. For example, $\mathscr{R}_1(p',1)$ is the set of rate quadruples $(R_{10}, R_{11}, R_{20}, R_{22})$ in (33). Similarly let $\mathscr{R}_2(p',\lambda_2)$ be the rate region corresponding to the layer order $\lambda_2 = 7,\ldots,12$ at receiver 2. This SWSC scheme achieves $\mathscr{R}_1(p',\lambda_1) \cap \mathscr{R}_2(p',\lambda_2)$ for any $\lambda_1 = 1,\ldots,6$ and $\lambda_2 = 7,\ldots,12$, which is sufficient to achieve the 4-dimensional auxiliary region in (31); see Appendix E for the proof.

Theorem 3: Let p denote the pmf $p(s)p(t)p(u)p(v)$ along with functions $x(s,t)$ and $w(u,v)$. Then

$$\mathscr{R}_{1,\mathrm{MAC}}(p) \cap \mathscr{R}_{2,\mathrm{MAC}}(p) = \bigcup_{p' \simeq p} \bigcup_{\lambda_1=1}^{6} \bigcup_{\lambda_2=7}^{12} [\mathscr{R}_1(p',\lambda_1) \cap \mathscr{R}_2(p',\lambda_2)].$$

Consequently, taking the union over all pmfs $p(s)p(t)$ $p(u)p(v)$ and functions $x(s,t), w(u,v)$, the coding scheme in Table IV achieves the Han–Kobayashi inner bound (32) for the two-user interference channel $p(y_1,y_2|x,w)$.

VII. Sliding-Window Coded Modulation

Coded modulation is the interface between channel coding and modulation, and specifies how (typically binary) codewords are mapped to sequences of constellation points. In this section, we show how the SWSC scheme can be specialized to a coded modulation scheme, termed *sliding-window coded modulation (SWCM)*, and demonstrate through practical implementation that conventional point-to-point encoders and decoders can be utilized to achieve the performance expected from high-complexity coding schemes. We also compare SWCM with existing coded modulation schemes, such as multilevel coding (MLC) [21], [22] and bit-interleaved coded modulation (BICM) [23], [24].

A. An Illustration of SWCM for 4PAM

Each coded modulation scheme is specified by two mappings: the symbol-level mapping and the block-level mapping. In SWCM, the symbol-level mapping is specified by the symbol-by-symbol mapping in superposition coding. For example, let $X_1, X_2 \in \{-1,+1\}$ be two BPSK symbols (throughout this section we assume the unit power constraint). Then a uniformly-spaced 4-PAM signal can be formed as

$$X = \frac{1}{\sqrt{5}}(X_1 + 2X_2) \in \{-\frac{3}{\sqrt{5}}, -\frac{1}{\sqrt{5}}, \frac{1}{\sqrt{5}}, \frac{3}{\sqrt{5}}\}. \quad (34)$$

The block-level mapping of SWCM is specified by the message scheduling of SWSC. For example, in the encoding scheme in Table I, each message is encoded to a length-$2n$ binary codeword (potentially with interleaving), the first n bits of which are carried by X_2 symbols in the current block, and the second n bits of which are carried by X_1 symbols in the next block. Accordingly, each transmission symbol X is then generated by (34), using a symbol X_2 from the current codeword and a symbol X_1 from the previous codeword. See Fig. 11 in Section VII-B for an illustration of the symbol-level and block-level mappings of the SWCM scheme that corresponds to Table I.

It is instructive to compare SWCM with two other popular coded modulation schemes, BICM and MLC. The key difference among the three lies in the block-level mapping; see Fig. 10. Assuming the same symbol-level mapping (34), in BICM, the two length-n parts x_1^n and x_2^n of a length-$2n$ codeword are transmitted in the same block. This contrasts the staggered transmission of x_1^n and x_2^n in SWCM. In MLC, instead of a single length-$2n$ codeword, two standalone length-n codewords x_1^n and x_2^n are generated by splitting the message (say M) into two parts (say M' and M''). When used for a point-to-point channel $p(y|x)$ under random coding, SWCM achieves

$$I(X_1;Y) + I(X_2;Y|X_1) = I(X;Y).$$

MLC achieves the same rate if individual rates of the two component codes are properly matched, while BICM achieves

$$I(X_1;Y) + I(X_2;Y) < I(X;Y),$$

the loss in which is due to self-interference between X_1 and X_2. Note that compared to the codeword length of n for MLC, the codeword length for SWCM and BICM is $2n$, which can potentially result in a better finite-blocklength performance with respect to the respective mutual information rate under random coding. More fundamentally, individual component codewords in MLC should be rate-controlled (which is difficult to be done optimally in practice) and reliably decoded (which results in rate loss under channel uncertainty or multiple receivers). The latter limitation is reflected in the deficiency of the rate-splitting scheme for the interference channel, as pointed out in Remark 2. In summary, SWCM has the advantage of high rate over BICM and the advantage of long block length and robustness over MLC, but at the same time suffers from error propagation over blocks and rate loss due to initialization/termination.

B. The Generalization to Other Constellations

The SWSC framework provides great flexibility in the symbol-level mapping and the number of layers, which results in a variety of practical coded modulation schemes.

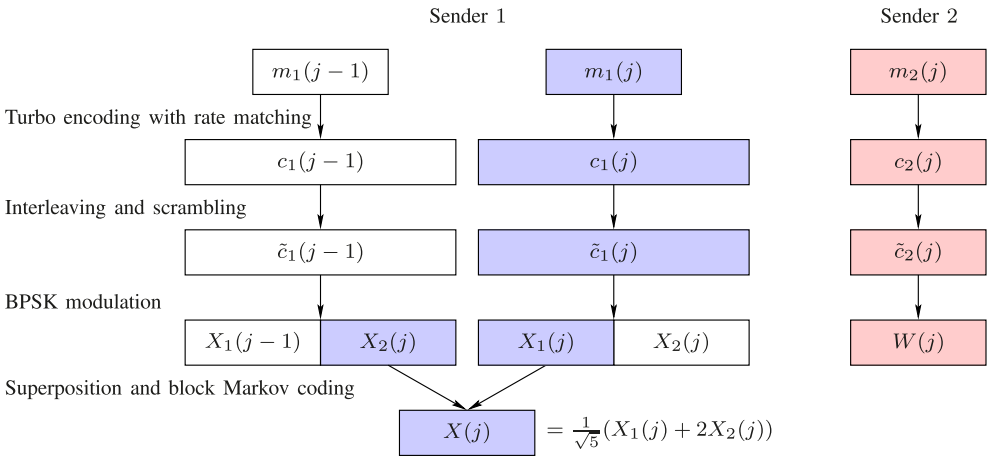


Fig. 11. Encoding diagram for the LTE-turbo implementation of SWCM.

For example, a Gray mapping from two BPSK symbols to the 4PAM constellation can be formed by a different symbol-level mapping

$$X = \frac{1}{\sqrt{5}}(X_1 + 2X_1 \cdot X_2). \tag{35}$$

There are four other symbol-level mappings for 4PAM.

Higher-order constellations have richer structures and allow for more diverse decompositions. For example, a uniformly-spaced 8PAM symbol can be decomposed as the superposition

$$X = \frac{1}{\sqrt{21}}(X_1 + 2X_2 + 4X_3) \tag{36}$$

of three BPSK layers X_1, X_2, X_3, or as the superposition

$$X = \frac{1}{\sqrt{21}}(X_1 + 2\sqrt{5}X_2) \tag{37}$$

of one BPSK layer X_1 and one 4PAM layer X_2. For the block-level mapping, each message is encoded into a length-$3n$ binary codeword. In case of (36), the three parts of the codeword, each of length n, are transmitted over three consecutive blocks. In case of (37), the first $2n$ bits of the codeword are carried by the 4PAM X_2 sequence (2 bits per symbol by the Gray or natural mapping) and the remaining n bits are carried by the BPSK X_1 sequence over two consecutive blocks.

As another example, consider the 16QAM coded modulation, which can be decomposed as the superposition

$$X = \frac{1}{\sqrt{5}}(X_1 + 2X_2) \tag{38}$$

of two QPSK symbols $X_1, X_2 \in \{e^{i\frac{\pi}{4}}, e^{i\frac{3\pi}{4}}, e^{-i\frac{3\pi}{4}}, e^{-i\frac{\pi}{4}}\}$, or as the superposition

$$X = \frac{1}{\sqrt{2}}(X_1 + iX_2) \tag{39}$$

of two 4PAM symbols $X_1, X_2 \in \{-\frac{3}{\sqrt{5}}, -\frac{1}{\sqrt{5}}, \frac{1}{\sqrt{5}}, \frac{3}{\sqrt{5}}\}$. For both cases, two halves of a length-$4n$ binary codeword are carried by x_1^n and x_2^n over two consecutive blocks. Alternatively, four BPSK layers can be used for staggered transmission over four consecutive blocks.

For multiple-input multiple-output (MIMO) transmission, there is a natural correspondence between the antenna ports and the symbol-level mapping. Suppose that there are t transmitting antennas. Then, each antenna port $X^{(k)}$ can transmit the codeword carried by the SWCM layer X_k, that is,

$$\begin{aligned} X &= (X^{(1)}, \ldots, X^{(t)}) \\ &= (X_1, \ldots, X_t). \end{aligned} \tag{40}$$

The SWCM scheme with the symbol-level mapping in (40) is in fact equivalent to the block-level diagonal Bell Labs layered space-time (D-BLAST) architecture [44]. Note that horizontal BLAST [45], [46] and vertical BLAST [47] correspond to MLC and BICM, respectively. In this sense, the encoder structure of sliding-window superposition coding may well be called *diagonal superposition coding* in contrast to the conventional *horizontal* superposition coding structure of MLC.

SWCM, however, can provide much greater flexibility than D-BLAST since the symbol-level mapping can be controlled at the constellation level, not just at the antenna level. For example, consider a MIMO system with two transmitting antennas, both of which use the 4PAM constellation as in (34)

$$\begin{aligned} X^{(1)} &= \frac{1}{\sqrt{5}}(A_{11} + 2A_{12}), \\ X^{(2)} &= \frac{1}{\sqrt{5}}(A_{21} + 2A_{22}), \end{aligned} \tag{41}$$

where $A_{11}, A_{12}, A_{21}, A_{22}$ are BPSK symbols. As in D-BLAST, we can use the symbol-level mapping in (40), or equivalently,

$$X_1 = (A_{11}, A_{12}), \quad X_2 = (A_{21}, A_{22}),$$

and communicate the two halves of a length-$4n$ binary codeword by x_2^n and x_1^n over two consecutive blocks. As an alternative, we can map the least significant bits in the two antennas to layer 1 and the remaining bits to layer 2, i.e.,

$$X_1 = (A_{11}, A_{21}), \quad X_2 = (A_{12}, A_{22}).$$

As another alternative, we can use 4 layers with symbols $A_{11}, A_{12}, A_{21}, A_{22}$, each carrying one fourth of the codewords

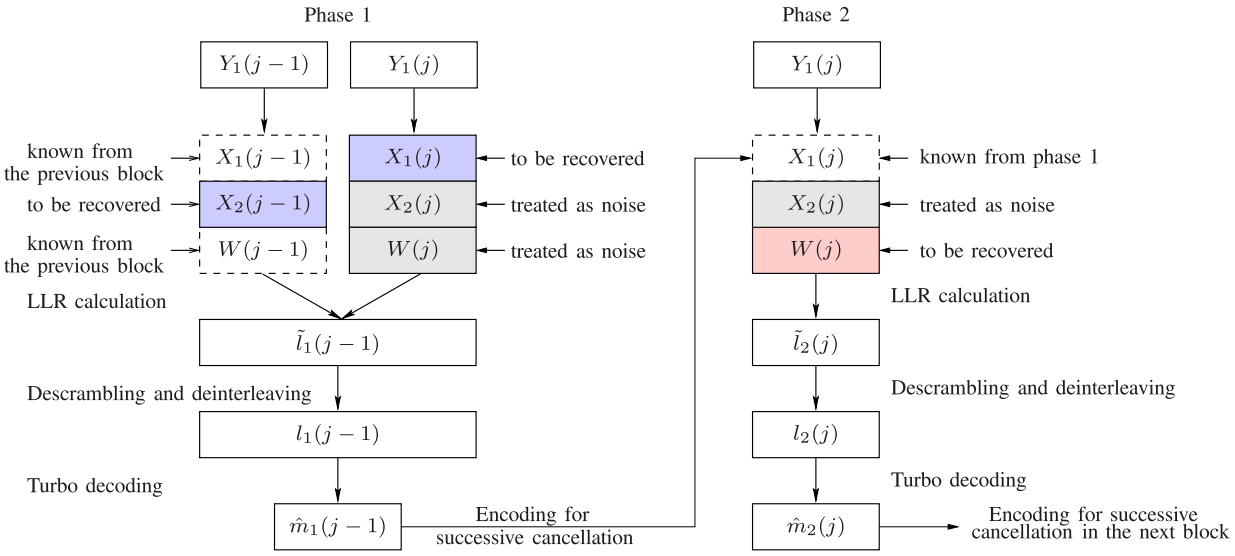


Fig. 12. Decoding diagram for the decoding order $\hat{m}_1(j-1) \to \hat{m}_2(j)$.

over four consecutive blocks. There can be other possibilities. This richness can be utilized for adaptive transmission for wireless fading channels, as demonstrated in [26].

C. Implementation With LTE Turbo Codes

We now demonstrate the feasibility of SWCM in practice by implementing the basic 4PAM coded modulation scheme in (34) for the Gaussian interference channel. More extensive studies for cellular networks are reported in [27].

Consider the 2-user Gaussian interference channel in (1), where sender 1 uses 4PAM as in (34) and sender 2 uses BPSK. Sender 1 uses a binary code of length $2n$ and rate $R_1/2$ to communicate $m_1(j)$ through x_2^n in block j and x_1^n in block $j+1$, while sender 2 uses a binary code of length n and rate R_2 to communicate $m_2(j)$ through w^n in block j; see Fig. 11.

We adopt the LTE standard turbo code [48], which has the flexibility in the code rate and the block length. In particular, we start with the rate $1/3$ mother code and adjust the rates and lengths according to the rate matching algorithm in the standard. Note that for $R_1 < 2/3$, some code bits are repeated and for $R_1 > 2/3$, some code bits are punctured. We set the block length $n = 2048$ and the number of blocks $b = 20$. We use the LOG-MAP algorithm with up to 8 iterations in each stage of turbo decoding. We assume that a rate pair (R_1, R_2) is achieved for a given channel if the resulting block-error rate (BLER) is below 0.1 over 200 independent sets of simulations. Sliding-window decoding is performed at both receivers. Fig. 12 illustrates the decoding operation at receiver 1, under decoding order $d_1 \colon \hat{m}_1(j-1) \to \hat{m}_2(j)$.

Fig. 13 plots the symmetric rate $(R_1 = R_2)$ against the INR for the symmetric Gaussian interference channel $(S_1 = S_2$ and $I_1 = I_2)$ when the SNR is held fixed at 8 dB. The solid lines represent theoretical achievable rates (mutual information) of MLD/SND, SWCM, and IAN. In IAN decoding, the interference is treated as *Gaussian* noise of the same power and the constellation information of interference is not

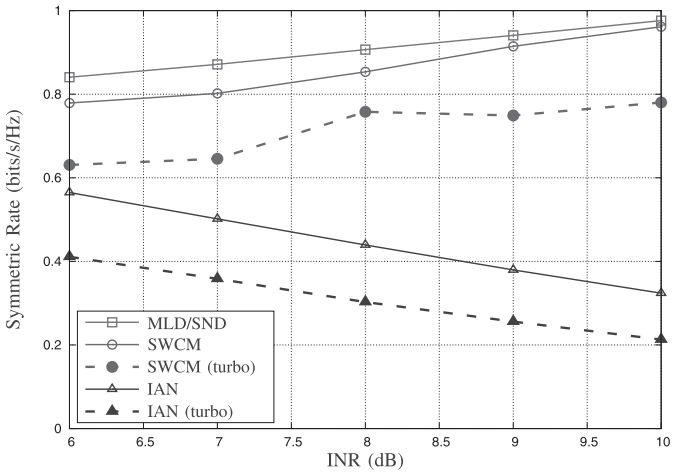


Fig. 13. Performance comparison in the symmetric Gaussian interference channel. The solid lines correspond to the theoretical performance. The dashed lines correspond to the simulation performance by the implementations using the LTE turbo codes.

used. In SWCM decoding, the optimal decoding orders are used at the given channel parameters. There is a gap between MLD/SND and SWCM (cf. Theorem 2), since the encoder is fixed using a symbol-level mapping (34) with only two layers $X_1, X_2 \sim \text{Unif}\{-1, +1\}$. The dashed lines represent the achievable rates of the actual implementation using the LTE turbo codes. The 4PAM encoding at sender 1 uses BICM for IAN. As the INR grows, the gain of SWCM over IAN increases from 53.44% (at the INR of 6 dB) to 150.32% (at 8 dB) and to 266.51% (at 10 dB).

VIII. Concluding Remarks

In this paper, we proposed the sliding-window superposition coding scheme (SWSC) as an implementable alternative to the rate-optimal simultaneous decoding. Combined with the conventional rate-splitting technique, the coding scheme can

be generalized to achieve the Han–Kobayashi inner bound on the capacity region of the two-user interference channel. Since the publication of the initial work [20] on SWSC, extensive simulations of the SWSC scheme have been performed in more practical communication scenarios, such as the Ped-B fading interference channel model [26], [27]. With several improvements in transceiver design, such as soft decoding, input bit-mapping and layer optimization, and power control, the performance figures presented here can be improved by another 10–20% [26]. System-level performance as well as requirements on the network operation for implementing SWSC in 5G cellular networks are discussed in [27]. These results indicate that SWSC is a promising candidate for interference management in future cellular networks.

Appendix A
Proof of Lemma 1

First, for any rate pair (R_1, R_2) in (8), we have

$$\begin{aligned} R_1 &\le I(X_1;Y_1) + I(X;Y_1|W,X_1) \\ &\overset{(a)}{\le} I(X_1;Y_1|W) + I(X;Y_1|W,X_1) \\ &\overset{(b)}{=} I(X;Y_1|W), \\ R_2 &\le I(W;Y_1|X_1) \\ &\overset{(c)}{\le} I(W;Y_1|X_1,X_2) \\ &\overset{(d)}{=} I(W;Y_1|X), \\ R_1+R_2 &\le I(X_1;Y_1) + I(X;Y_1|W,X_1) + I(W;Y_1|X_1) \\ &= I(X,W;Y_1), \end{aligned} \tag{42}$$

where (a) and (c) follow since W is independent of (X_1, X_2), and (b) and (d) follow since $X_1 \to X \to (W, Y_1)$ form a Markov chain. Thus, any rate point in $\mathscr{R}_{\mathrm{RS}}(p')$ with $p' \simeq p$ is also in $\mathscr{R}_{1,\mathrm{SD}}(p)$.

Now it suffices to show that for any rate point (I_1, I_2) on the *dominant face*, i.e., $I_1 + I_2 = I(X, W; Y_1)$, there exists a $p' \simeq p$ such that

$$\begin{aligned} I_1 &= I(X_1;Y_1) + I(X;Y_1|X_1,W), \\ I_2 &= I(W;Y_1|X_1). \end{aligned}$$

To this end, note that when $X_1 = X$ and $X_2 = \emptyset$, expression (8) attains one corner point $(I(X;Y_1), I(W;Y_1|X))$; when $X_1 = \emptyset$ and $X_2 = X$, expression (8) attains the other corner point $(I(X;Y_1|W), I(W;Y_1))$. Moreover, the rate pair in (8) and (I_1, I_2) share the same sum-rate as in (42). Hence, it suffices to show that for every $\alpha \in [0,1]$, there exists a choice of $p(x_1)p(x_2)$ and function $x(x_1, x_2)$ such that

$$I(W;Y_1|X_1) = I_2 = \alpha I(W;Y_1) + (1-\alpha) I(W;Y_1|X).$$

Let $p_{X_1}(x) = (1-\alpha)p_X(x)$ for $x \in \mathcal{X}$ and $p_{X_1}(\mathrm{e}) = \alpha$. Let X_2 be independent of X_1 and $p_{X_2}(x) = p_X(x)$ for $x \in \mathcal{X}$. Let

$$x(x_1,x_2) = \begin{cases} x_1, & \text{if } x_1 \ne \mathrm{e}, \\ x_2, & \text{otherwise.} \end{cases}$$

This choice of $p(x_1)p(x_2)$ and $x(x_1, x_2)$ induces a conditional pmf

$$p_{X_1|X}(x_1|x) = \begin{cases} 1-\alpha, & \text{if } x_1 = x, \\ \alpha, & \text{if } x_1 = \mathrm{e}, \\ 0, & \text{otherwise,} \end{cases} \tag{43}$$

which is an erasure channel with input X, output X_1, and erasure probability α. Define $E = \mathbb{1}_{\{X_1 = \mathrm{e}\}}$. It can be checked that $E \sim \mathrm{Bern}(\alpha)$ is independent of X and X_2. Thus, we have

$$\begin{aligned} I(W;Y_1|X_1) &= I(W;Y_1|X_1,E) \\ &= \alpha I(W;Y_1|X_1,E=1) \\ &\quad + (1-\alpha) I(W;Y_1|X_1,E=0) \\ &\overset{(a)}{=} \alpha I(W;Y_1|X_1=\mathrm{e},E=1) \\ &\quad + (1-\alpha) I(W;Y_1|X,X_1,E=0) \\ &\overset{(b)}{=} \alpha I(W;Y_1|X_1=\mathrm{e},E=1) \\ &\quad + (1-\alpha) I(W;Y_1|X) \\ &\overset{(c)}{=} \alpha I(W;Y_1|E=1) + (1-\alpha) I(W;Y_1|X) \\ &\overset{(d)}{=} \alpha I(W;Y_1) + (1-\alpha) I(W;Y_1|X), \end{aligned}$$

where (a) follows since when $E = 0$, $X_1 = X$, (b) follows since given X, (W, Y_1) are conditionally independent of (X_1, E), (c) follows since $E = 1$ is equivalent as $X_1 = \mathrm{e}$, and (d) follows since E is independent of (X, W, Y_1). Therefore, as α increases from 0 to 1, the rate pair in (8) moves continuously and linearly from one corner point to the other along the line $R_1 + R_2 = I(X, W; Y_1)$.

Appendix B
Proof of Proposition 2

By the proof of Lemma 1 (Appendix A), $\mathscr{R}_{1,\mathrm{SD}} \cap \mathscr{R}_{2,\mathrm{SD}}$ is equivalent to the set of (R_1, R_2) such that

$$\begin{aligned} R_1 &\le \min\{I(X_1';Y_1) + I(X;Y_1|X_1',W), \\ &\qquad I(X_1'';Y_2) + I(X;Y_2|X_1'',W)\} \\ R_2 &\le \min\{I(W;Y_1|X_1'), I(W;Y_2|X_1'')\} \end{aligned} \tag{44}$$

for erasure channels $p(x_1'|x)$ and $p(x_1''|x)$ with erasure probabilities α' and α'' respectively. Suppose that $\alpha' > \alpha''$. Then the channel $p(x_1'|x)$ is *degraded* with respect to the channel $p(x_1''|x)$. Since the rate expressions in (44) only depend on the marginal conditional pmfs of $p(x_1', x_1''|x)$, we assume without loss of generality that $X \to X_1'' \to X_1'$ form a Markov chain. By the functional representation lemma (twice), for $p(x_1'|x_1'')$, there exists an X_2 independent of X_1'' such that $X_1' = f(X_1'', X_2)$; for $p(x_1''|x)$, there exists an X_3 independent of (X_2, X_1') such that $X = g(X_1'', X_3) = g(f(X_1', X_2), X_3) \triangleq x(X_1', X_2, X_3)$. Renaming $X_1 \triangleq X_1'$ and plugging $X_1'' = f(X_1, X_2)$ into (44), we obtain the rate region $\mathscr{R}_{\mathrm{SWSC}}(p', 3, 1, d_1, d_2)$ with (d_1, d_2) given in (18). In the case when $\alpha' \le \alpha''$, we can assume that $X \to X_1' \to X_1''$ form a Markov chain. Then, using the functional representation similarly as above, the rate region in (44) can be reduced to the rate region $\mathscr{R}_{\mathrm{SWSC}}(p', 3, 1, d_1, d_2)$ with (d_1, d_2) given in (19).

APPENDIX C
PROOF OF THEOREM 1

We prove the stronger statement (13) in Theorem 1, which implies the weaker statement (12). Consider the symmetric Gaussian interference channel (cf. (1)) with $g_{11} = g_{22} = 1$, $g_{12} = g_{21} = g, S_1 = S_2 = S = P$ and $I_1 = I_2 = I = g^2P$. Assume that the interference channel has *strong, but not very strong,* interference, i.e., $S < I < S(S+1)$. The capacity region of this channel is characterized by the set of rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\le \mathsf{C}(S), \\ R_2 &\le \mathsf{C}(S), \\ R_1 + R_2 &\le \mathsf{C}(I+S), \end{aligned}$$

which is achieved by simultaneous decoding with a single input distribution $X_1, X_2 \sim N(0, P)$ [7], [14].

Given $\mathscr{R}(p, s, t, d_1, d_2)$, let $\mathscr{R}^*(s, t, d_1, d_2)$ be the closure of the union of $\mathscr{R}(p, s, t, d_1, d_2)$ over all p. Define

$$R_1^*(s, t, d_1, d_2) = \max\{R_1 \colon (R_1, \mathsf{C}(S)) \in \mathscr{R}^*(s, t, d_1, d_2)\}$$

as the maximal achievable rate R_1 such that R_2 is at individual capacity. In order to show the corner point of the capacity region is not achievable using any (p, s, t, d_1, d_2) rate-splitting scheme, it suffices to establish the following.

Proposition 3: For the symmetric Gaussian interference channel with $S < I < S(S+1)$,

$$R_1^*(s, t, d_1, d_2) < \mathsf{C}\Big(\frac{I}{1+S}\Big)$$

for any finite s, t and decoding orders d_1, d_2.

The remainder of this appendix is dedicated to the proof of Proposition 3. First, we find the optimal decoding order at receiver 2 of the (p, s, t, d_1, d_2) rate-splitting scheme that achieves $R_1^*(s, t, d_1, d_2)$. We note that in homogeneous superposition coding, message parts are encoded into independent codewords, and thus can be recovered in an arbitrary order in general (which is in sharp contrast to heterogeneous superposition coding, where $\hat{m}_{ij}$ has to be recovered before $\hat{m}_{ik}$ for $j < k, i = 1, 2$). Henceforth, by renaming the message parts, we assume without loss of generality that at receiver 2, the decoding order among message parts $\{\hat{m}_{11}, \ldots, \hat{m}_{1s}\}$ is $\hat{m}_{11} \to \hat{m}_{12} \to \cdots \to \hat{m}_{1s}$ and the decoding order among messages parts $\{\hat{m}_{21}, \ldots, \hat{m}_{2t}\}$ is $\hat{m}_{21} \to \hat{m}_{22} \to \cdots \to \hat{m}_{2t}$. Note that between message parts of m_1 and m_2, there are still flexibility for all possible permutations as long as the subsets $\{\hat{m}_{11}, \ldots, \hat{m}_{1s}\}$ and $\{\hat{m}_{21}, \ldots, \hat{m}_{2t}\}$ are in order. The next lemma states the optimal order among them.

Lemma 2: For any (p, s, t, d_1, d_2) rate-splitting scheme that achieves $R_1^*(s, t, d_1, d_2)$, the decoding order at receiver 2 is

$$d_2^* \colon \hat{m}_{11} \to \hat{m}_{12} \to \cdots \to \hat{m}_{1s} \to \hat{m}_{21} \to \hat{m}_{22} \to \cdots \to \hat{m}_{2t}.$$

Proof: Fix any (p, s, t, d_1, d_2) rate-splitting scheme that guarantees $R_2 = \mathsf{C}(S)$. Suppose that $\hat{m}_{2j}$ is recovered earlier than $\hat{m}_{1k}$ at receiver 2, that is,

$$d_2 \colon d_{21} \to \hat{m}_{2j} \to \hat{m}_{1k} \to d_{22}.$$

Now flip the decoding order of $\hat{m}_{2j}$ and $\hat{m}_{1k}$ in $\tilde{d}_2$ as

$$\tilde{d}_2 \colon d_{21} \to \hat{m}_{1k} \to \hat{m}_{2j} \to d_{22}$$

and construct $(p, s, t, d_1, \tilde{d}_2)$ rate-splitting scheme, where the message splitting, the underlying distribution, and decoding order d_1 remain the same. Let $\tilde{R}_{ij}$ be the rate of the message part m_{ij} in the $(p, s, t, d_1, \tilde{d}_2)$ rate-splitting scheme. Then we have that all the rates remain the same except

$$\begin{aligned} R_{2j} &= I(W_j; Y_2 | W^{j-1}, X^{k-1}), \\ \tilde{R}_{2j} &= I(W_j; Y_2 | W^{j-1}, X^{k}), \\ R_{1k} &= I(X_k; Y_2 | W^{j}, X^{k-1}), \\ \tilde{R}_{1k} &= I(X_k; Y_2 | W^{j-1}, X^{k-1}). \end{aligned}$$

Note that $R_{2j} \le \tilde{R}_{2j}$ since X_k is independent of (W^j, X^{k-1}). On the other hand, since R_{2j} already results in full rate at R_2, we must have $\tilde{R}_{2j} = R_{2j}$. It follows that

$$I(X_k; W_j | Y_2, W^{j-1}, X^{k-1}) = 0$$

and therefore $R_{1j} = \tilde{R}_{1j}$. ∎

Next, we discuss the structure of the decoding order at receiver 1.

Lemma 3: In order to show the insufficiency in achieving the corner point $(\mathsf{C}(I/(1+S)), \mathsf{C}(S))$ for any (p, s, t, d_1, d_2) rate-splitting scheme, it suffices to show the insufficiency of any (p, s, s, d_1^*, d_2^*) rate-splitting scheme with decoding orders

$$\begin{aligned} d_1^* \colon\ & \hat{m}_{1,\pi(1)} \to \hat{m}_{2,\sigma(1)} \to \hat{m}_{1,\pi(2)} \to \hat{m}_{2,\sigma(2)} \\ & \to \cdots \to \hat{m}_{1,\pi(s-1)} \to \hat{m}_{2,\sigma(s-1)} \to \hat{m}_{1,\pi(s)}, \\ d_2^* \colon\ & \hat{m}_{11} \to \hat{m}_{12} \to \cdots \to \hat{m}_{1s} \to \hat{m}_{21} \to \hat{m}_{22} \\ & \to \cdots \to \hat{m}_{2s}, \end{aligned} \tag{45}$$

where $\pi \colon [s] \to [s], \sigma \colon [s] \to [s]$ are two permutations on the index set $[s]$.

Proof: First, there is no loss of generality in assuming $s = t$, because any (p, s, t, d_1, d_2) scheme can be viewed as a special case of some $(p', \max\{s,t\}, \max\{s,t\}, d_1', d_2')$ scheme by nulling out the corresponding inactive variables and preserving the distribution and decoding orders of the active ones. For the alternating decoding order at receiver 1, we note that a (p, s, s, d_1, d_2^*) scheme with *arbitrary* decoding order d_1 can viewed as a special case of some $(\tilde{p}, 2s, 2s, \tilde{d}_1^*, \tilde{d}_2^*)$ scheme with *alternating* decoding order $\tilde{d}_1^*$. For example, for $s = 2$, any decoding order must be one of the following six forms

$$\begin{aligned} &\hat{m}_{1,\pi(1)} \to \hat{m}_{1,\pi(2)} \to \hat{m}_{2,\sigma(1)} \to \hat{m}_{2,\sigma(2)}, \\ &\hat{m}_{1,\pi(1)} \to \hat{m}_{2,\sigma(1)} \to \hat{m}_{1,\pi(2)} \to \hat{m}_{2,\sigma(2)}, \\ &\hat{m}_{2,\sigma(1)} \to \hat{m}_{1,\pi(1)} \to \hat{m}_{2,\sigma(2)} \to \hat{m}_{1,\pi(1)}, \\ &\hat{m}_{2,\sigma(1)} \to \hat{m}_{2,\sigma(2)} \to \hat{m}_{1,\pi(1)} \to \hat{m}_{1,\pi(2)}, \\ &\hat{m}_{1,\pi(1)} \to \hat{m}_{2,\sigma(1)} \to \hat{m}_{2,\sigma(2)} \to \hat{m}_{1,\pi(2)}, \\ &\hat{m}_{2,\sigma(1)} \to \hat{m}_{1,\pi(1)} \to \hat{m}_{1,\pi(2)} \to \hat{m}_{2,\sigma(2)}, \end{aligned}$$

which are all special cases of the alternating order

$$\begin{aligned} &\hat{m}_{1,\tilde{\pi}(1)} \to \hat{m}_{2,\tilde{\sigma}(1)} \to \hat{m}_{1,\tilde{\pi}(2)} \to \hat{m}_{2,\tilde{\sigma}(2)} \to \hat{m}_{1,\tilde{\pi}(3)} \\ &\to \hat{m}_{2,\tilde{\sigma}(3)} \to \hat{m}_{1,\tilde{\pi}(4)}. \end{aligned}$$

Moreover, because of the special structure of $\tilde{d}_2^*$, it remains the optimal decoding order (in the sense of Lemma 2) even after nulling out the inactive message parts. ∎

Now, we provide a necessary condition for a rate-splitting scheme to achieve the corner point of the capacity region.

Lemma 4: If a (p, s, t, d_1, d_2^*) rate-splitting scheme attains the corner point $(\mathsf{C}(I/(1+S)), \mathsf{C}(S))$, then p must satisfy

$$X \sim \mathrm{N}(0, P) \quad \textit{and} \quad W \sim \mathrm{N}(0, P).$$

Proof: No matter what d_1 is, because of the optimal order d_2^*, the rate constraints for R_2 must satisfy

$$\begin{aligned} R_2 &\le \sum_{j=1}^{t} I(W_j; Y_2 | X, W^{j-1}) \\ &= I(W; Y_2 | X) \\ &\le \mathsf{C}(S). \end{aligned} \tag{46}$$

Given X, the channel from W to Y_2 is a Gaussian channel with SNR S. Therefore the condition $W \sim \mathrm{N}(0, P)$ is necessary for (46) to hold with equality. Similarly, R_1 must satisfy

$$\begin{aligned} R_1 &\le \sum_{j=1}^{s} I(X_j; Y_2 | X^{j-1}) \\ &= I(X; Y_2) \\ &\le \mathsf{C}\left(\frac{I}{1+S}\right). \end{aligned} \tag{47}$$

Given $W \sim \mathrm{N}(0, P)$, the channel from X to Y_2 is a Gaussian channel with SNR $I/(1+S)$. Therefore, the condition $X \sim \mathrm{N}(0, P)$ is necessary for (47) to hold with equality. ∎

We also need the following technical lemma.

Lemma 5: Let $F(u, x)$ be any (continuous) distribution such that $X \sim \mathrm{N}(0, P)$ and $I(U; Y) = 0$, where $Y = X + N$ with $N \sim \mathrm{N}(0, 1)$ independent of X. Then, $I(U; X) = 0$.

Proof: For every $u \in \mathcal{U}$, we have

$$\begin{aligned} I(X; Y | U = u) &= h(Y | U = u) - h(Y | X, U = u) \\ &\overset{(a)}{=} h(Y) - h(Y | X) \\ &= \mathsf{C}(P), \end{aligned}$$

where (a) follows since Y is independent of U and $U \to X \to Y$ form a Markov chain. Suppose for some u, $\mathsf{E}(X^2 | U = u) < P$, i.e., the effective channel SNR is strictly less than P. Then $I(X; Y | U = u) < P$. As a result, we must have $\mathsf{E}(X^2 | U = u) \ge P$ for all $u \in \mathcal{U}$. On the other hand,

$$\begin{aligned} P &\le \int \mathsf{E}(X^2 | U = u) dF(u) \\ &= \mathsf{E}(X^2) \\ &= P, \end{aligned}$$

which forces $\mathsf{E}(X^2 | U = u) = P$ for almost all u. Since the Gaussian input $\mathrm{N}(0, P)$ is the unique distribution that attains the rate $\mathsf{C}(P)$ in the Gaussian channel with SNR P, the distribution $F(x|u)$ must be $\mathrm{N}(0, P)$ for almost all u. Therefore $I(U; X) = 0$. ∎

We are ready to establish the suboptimality of rate-splitting schemes.

By Lemma 3, it suffices to show the insufficiency of any (p, s, s, d_1^*, d_2^*) rate-splitting scheme with decoding orders given in (45). The achievable rate region of this scheme is characterized by

$$\begin{aligned} R_1 &\le \sum_{i=1}^{s} \min\{ I(X_{\pi(i)}; Y_2 | X^{\pi(i)-1}), \\ &\quad I(X_{\pi(i)}; Y_1 | X_{\pi(1)}, \ldots, X_{\pi(i-1)}, W_{\sigma(1)}, \ldots, W_{\sigma(i-1)}) \} \\ &\triangleq I_1 \\ R_2 &\le \sum_{i=1}^{s-1} \min\{ I(W_{\sigma(i)}; Y_2 | X, W^{\sigma(i)-1}), \\ &\quad I(W_{\sigma(i)}; Y_1 | X_{\pi(1)}, \ldots, X_{\pi(i)}, W_{\sigma(1)}, \ldots, W_{\sigma(i-1)}), \} \\ &\qquad + I(W_{\sigma(s)}; Y_2 | X, W_{\sigma(1)}, \ldots, W_{\sigma(s-1)}) \\ &\triangleq I_2 \end{aligned}$$

Assume that the corner point of the capacity region is achieved by this scheme, i.e.,

$$I_1 = \mathsf{C}(I/(1+S)), \tag{48}$$

$$I_2 = \mathsf{C}(S). \tag{49}$$

Then, by Lemma 4, we must have $X \sim \mathrm{N}(0, P)$ and $W \sim \mathrm{N}(0, P)$. Consider

$$\begin{aligned} I_1 &\le I(X_{\pi(1)}; Y_1) + \sum_{i \in [s] \setminus \pi(1)} I(X_i; Y_2 | X^{i-1}) \\ &\le I(X_{\pi(1)}; Y_1) + I(X^{\pi(1)-1}; Y_2) \\ &\qquad + I(X_{\pi(1)+1}^{s}; Y_2 | X^{\pi(1)}) \\ &\overset{(a)}{\le} I(X_{\pi(1)}; Y_1) + I(X^{\pi(1)-1}; Y_2 | X_{\pi(1)}) \\ &\qquad + I(X_{\pi(1)+1}^{s}; Y_2 | X^{\pi(1)}) \\ &= h(Y_1) - h(Y_1 | X_{\pi(1)}) + h(Y_2 | X_{\pi(1)}) - h(Y_2 | X^{\pi(1)}) \\ &\qquad + h(Y_2 | X^{\pi(1)}) - h(Y_2 | X) \\ &= h(Y_1) - h(Y_1 | X_{\pi(1)}) + h(Y_2' | X_{\pi(1)}) - h(Y_2' | X) \end{aligned} \tag{50}$$

where $Y_2' = Y_2/g = X + (W + N_2)/g$ and (a) follows since $X_{\pi(1)}$ is independent of $X^{\pi(1)-1}$. Since

$$\begin{aligned} \frac{1}{2}\log(2\pi e(S+1)/g^2) &= h(Y_2' | X) \\ &\le h(Y_2' | X_{\pi(1)}) \\ &\le h(Y_2') \\ &= \frac{1}{2}\log(2\pi e(I+S+1)/g^2), \end{aligned}$$

there exists an $\alpha \in [0, 1]$ such that $h(Y_2' | X_{\pi(1)}) = (1/2)\log(2\pi e\ (\alpha I + S + 1)/g^2)$. Moreover, since $W \sim \mathrm{N}(0, P)$ and $I < S(1+S)$, the channel $X \to Y_1$ is a degraded version of the channel $X \to Y_2'$, i.e., $Y_1 = Y_2' + N'$, where $N' \sim \mathrm{N}(0, I + 1 - (S+1)/g^2)$ is independent of X and W. By the entropy power inequality,

$$\begin{aligned} 2^{2h(Y_1 | X_{\pi(1)})} &\ge 2^{2h(Y_2' | X_{\pi(1)})} + 2^{2h(N' | X_{\pi(1)})} \\ &= 2\pi e(\alpha S + I + 1). \end{aligned}$$

Therefore, it follows from (50) that

$$\begin{aligned} I_1 &\le h(Y_1) - h(Y_1|X_{\pi(1)}) + h(Y_2'|X_{\pi(1)}) - h(Y_2'|X) \\ &\le \frac{1}{2}\log\left(\frac{(I+S+1)(\alpha I+S+1)}{(\alpha S+I+1)(1+S)}\right) \\ &\le \mathsf{C}(I/(1+S)), \end{aligned}$$

where the last step follows since $S < I$. To match the standing assumption in (48), we must have equality in (a), which forces $\alpha = 1$ and $h(Y_2'|X_{\pi(1)}) = (1/2)\log(2\pi e(I+S+1)/g^2) = h(Y_2')$, i.e., $I(X_{\pi(1)}; Y_2') = 0$. Note that $X, W \sim \mathrm{N}(0, P)$ and the channel from X to Y_2' is a Gaussian channel. Applying Lemma 5 yields

$$I(X_{\pi(1)}; X) = 0. \tag{51}$$

Now, I_2 can be simplified to

$$\begin{aligned} I_2 &\le I(W_{\sigma(1)}; Y_1|X_{\pi(1)}) + \sum_{i\in[s]\setminus\sigma(1)} I(W_i; Y_2|X, W^{i-1}) \\ &\overset{(b)}{=} I(W_{\sigma(1)}; Y_1) + I(W^{\sigma(1)-1}; Y_2|X) \\ &\qquad + I(W^s_{\sigma(1)+1}; Y_2|X, W^{\sigma(1)}) \\ &\overset{(c)}{\le} I(W_{\sigma(1)}; Y_1) + I(W^{\sigma(1)-1}; Y_2|X, W_{\sigma(1)}) \\ &\qquad + I(W^s_{\sigma(1)+1}; Y_2|X, W^{\sigma(1)}) \\ &= h(Y_1) - h(Y_1|W_{\sigma(1)}) + h(Y_2|X, W_{\sigma(1)}) \\ &\quad - h(Y_2|X, W^{\sigma(1)}) + h(Y_2|X, W^{\sigma(1)}) - h(Y_2|X, W) \\ &= h(\tilde{Y}_1) - h(\tilde{Y}_1|W_{\sigma(1)}) + h(\tilde{Y}_2|W_{\sigma(1)}) - h(\tilde{Y}_2|W) \end{aligned} \tag{52}$$

where $\tilde{Y}_1 = Y_1/g = W + (X+N_1)/g$ and $\tilde{Y}_2 = W + N_2$. Here (b) follows since $I(X_{\pi(1)}; Y_1|W_{\sigma(1)}) \le I(X_{\pi(1)}; Y_1|W) = I(X_{\pi(1)}; X + N_1) \le I(X_{\pi(1)}; X) = 0$ and $I(X_{\pi(1)}; Y_1) \le I(X_{\pi(1)}; X) = 0$, which implies

$$\begin{aligned} I(W_{\sigma(1)}; Y_1|X_{\pi(1)}) &= I(W_{\sigma(1)}; Y_1|X_{\pi(1)}) + I(X_{\pi(1)}; Y_1) \\ &= I(W_{\sigma(1)}; Y_1) + I(X_{\pi(1)}; Y_1|W_{\sigma(1)}) \\ &= I(W_{\sigma(1)}; Y_1), \end{aligned}$$

and (c) follows since $W_{\sigma(1)}$ and $(W^{\sigma(1)-1}, X)$ are independent. Since

$$\begin{aligned} \frac{1}{2}\log(2\pi e) &= h(\tilde{Y}_2|W) \\ &\le h(\tilde{Y}_2|W_{\sigma(1)}) \\ &\le h(\tilde{Y}_2) \\ &= \frac{1}{2}\log(2\pi e(1+S)), \end{aligned}$$

there exists a $\beta \in [0,1]$ such that $h(\tilde{Y}_2|W_{\sigma(1)}) = (1/2)\log(2\pi e\ (1+\beta S))$. Moreover, since $X \sim \mathrm{N}(0, P)$ and $I < S(1+S)$, $\tilde{Y}_1$ is a degraded version of $\tilde{Y}_2$, i.e., $\tilde{Y}_1 = \tilde{Y}_2 + \tilde{N}$, where $\tilde{N} \sim \mathrm{N}(0, (1+S)/g^2 - 1)$ is independent of X and W. Applying the entropy power inequality, we have

$$\begin{aligned} 2^{2h(\tilde{Y}_1|W_{\sigma(1)})} &\ge 2^{2h(\tilde{Y}_2|W_{\sigma(1)})} + 2^{2h(\tilde{N}|W_{\sigma(1)})} \\ &= 2\pi e(\beta S + (1+S)/g^2). \end{aligned}$$

Therefore, it follows from (52) that

$$\begin{aligned} I_2 &\le h(\tilde{Y}_1) - h(\tilde{Y}_1|W_{\sigma(1)}) + h(\tilde{Y}_2|W_{\sigma(1)}) - h(Y_2|X, W) \\ &\le \frac{1}{2}\log\left(\frac{(I+S+1)(1+\beta S)}{g^2(\beta S + (1+S)/g^2)}\right) \\ &\le \mathsf{C}(S), \end{aligned}$$

where the last step follows from the channel condition $I < (1+S)S$. To match the standing assumption in (49), we must have equality above, which forces $\beta = 1$ and $h(\tilde{Y}_2|W_{\sigma(1)}) = (1/2)\log(2\pi e(1+S)) = h(\tilde{Y}_2)$, i.e., $I(W_{\sigma(1)}; \tilde{Y}_2) = 0$. Note that $W \sim \mathrm{N}(0, P)$ and the channel from W to $\tilde{Y}_2$ is a Gaussian channel. Applying Lemma 5 yields

$$I(W_{\sigma(1)}; W) = 0. \tag{53}$$

To continue analyzing the dependency between $(X_{\pi(1)}, X_{\pi(2)})$ and X, we note that condition (53) implies that

$$\begin{aligned} &I(X_{\pi(2)}; Y_1|X_{\pi(1)}, W_{\sigma(1)}) \\ &= I(X_{\pi(2)}; Y_1|X_{\pi(1)}) + I(W_{\sigma(1)}; Y_1|X_{\pi(1)}, X_{\pi(2)}) \\ &\qquad - I(W_{\sigma(1)}; Y_1|X_{\pi(1)}) \\ &\overset{(d)}{=} I(X_{\pi(2)}; Y_1|X_{\pi(1)}), \end{aligned} \tag{54}$$

where (d) follows since

$$\begin{aligned} I(W_{\sigma(1)}; Y_1|X_{\pi(1)}) &\le I(W_{\sigma(1)}; Y_1|X_{\pi(1)}, X_{\pi(2)}) \\ &\le I(W_{\sigma(1)}; Y_1|X) \\ &\le I(W_{\sigma(1)}; W) \\ &= 0. \end{aligned}$$

Moreover, condition (51) implies that

$$\begin{aligned} I(X_{\pi(1)}; Y_2|X_{\pi(2)}) &\le I(X_{\pi(1)}; Y_2, X_{\pi(2)}) \\ &\le I(X_{\pi(1)}; Y_2, X) \\ &= I(X_{\pi(1)}; X, gX + W + N_2) \\ &= 0 \end{aligned}$$

and thus

$$h(Y_2|X_{\pi(2)}) = h(Y_2|X_{\pi(2)}, X_{\pi(1)}). \tag{55}$$

With (54) and (55), we can bound I_1 alternatively as

$$\begin{aligned} I_1 &\le I(X_{\pi(2)}; Y_1|X_{\pi(1)}, W_{\sigma(1)}) + \sum_{i\in[s]\setminus\pi(2)} I(X_i; Y_2|X^{i-1}) \\ &= I(X_{\pi(2)}; Y_1|X_{\pi(1)}) + I(X^{\pi(2)-1}; Y_2) \\ &\qquad + I(X^s_{\pi(2)+1}; Y_2|X^{\pi(2)}) \\ &\le I(X_{\pi(2)}; Y_1|X_{\pi(1)}) + I(X^{\pi(2)-1}; Y_2|X_{\pi(2)}) \\ &\qquad + I(X^s_{\pi(2)+1}; Y_2|X^{\pi(2)}) \\ &= h(Y_1|X_{\pi(1)}) - h(Y_1|X_{\pi(2)}, X_{\pi(1)}) + h(Y_2|X_{\pi(2)}) \\ &\qquad - h(Y_2|X) \\ &= h(Y_1) - h(Y_1|X_{\pi(2)}, X_{\pi(1)}) + h(Y_2|X_{\pi(2)}, X_{\pi(1)}) \\ &\qquad - h(Y_2|X) \\ &= h(Y_1) - h(Y_1|X_{\pi(2)}, X_{\pi(1)}) + h(Y_2'|X_{\pi(2)}, X_{\pi(1)}) \\ &\qquad - h(Y_2'|X). \end{aligned} \tag{56}$$

Note that the expression in (56) is in the same form of (50), except that $X_{\pi(1)}$ is replaced by the pair $(X_{\pi(1)}, X_{\pi(2)})$.

From this point on, following the identical argument as before with variable substitution $X_{\pi(1)} \leftrightarrow (X_{\pi(1)}, X_{\pi(2)})$, we conclude that

$$I(X_{\pi(1)}, X_{\pi(2)}; X) = 0.$$

Now, repeating this procedure, we can similarly show that

$$\begin{aligned} &I(X_{\pi(1)}, \ldots, X_{\pi(s-1)}; X) = 0, \\ &I(W_{\sigma(1)}, \ldots, W_{\sigma(s-1)}; W) = 0. \end{aligned} \tag{57}$$

However, condition (57) implies that for $i \in [s-1]$

$$\begin{aligned} &I(X_{\pi(i)}; Y_1 | X_{\pi(1)}, \ldots, X_{\pi(i-1)}, W_{\sigma(1)}, \ldots, W_{\sigma(i-1)}) \\ &\le I(X_{\pi(i)}; X, W, Y_1) \\ &= I(X_{\pi(i)}; X) \\ &= 0 \end{aligned}$$

and that

$$\begin{aligned} &I(X_{\pi(s)}; Y_1 | X_{\pi(1)}, \ldots, X_{\pi(s-1)}, W_{\sigma(1)}, \ldots, W_{\sigma(s-1)}) \\ &\quad - I(X; Y_1) \\ &= I(W_{\sigma(1)}, \ldots, W_{\sigma(s-1)}; Y_1 | X) \\ &\quad - I(X_{\pi(1)}, \ldots, X_{\pi(s-1)}; Y_1) \\ &\quad - I(W_{\sigma(1)}, \ldots, W_{\sigma(s-1)}; Y_1 | X_{\pi(1)}, \ldots, X_{\pi(s-1)}) \\ &\overset{(e)}{=} 0, \end{aligned}$$

where (e) follows since

$$\begin{aligned} &I(W_{\sigma(1)}, \ldots, W_{\sigma(s-1)}; Y_1 | X_{\pi(1)}, \ldots, X_{\pi(s-1)}) \\ &\le I(W_{\sigma(1)}, \ldots, W_{\sigma(s-1)}; Y_1 | X) \\ &\le I(W_{\sigma(1)}, \ldots, W_{\sigma(s-1)}; W) \\ &= 0 \end{aligned}$$

and

$$\begin{aligned} I(X_{\pi(1)}, \ldots, X_{\pi(s-1)}; Y_1) &\le I(X_{\pi(1)}, \ldots, X_{\pi(s-1)}; X) \\ &= 0. \end{aligned}$$

Therefore,

$$\begin{aligned} I_1 &= \sum_{i=1}^{s} \min\{I(X_{\pi(i)}; Y_2 | X^{\pi(i)-1}), \\ &\quad I(X_{\pi(i)}; Y_1 | X_{\pi(1)}, \ldots, X_{\pi(i-1)}, W_{\sigma(1)}, \ldots, W_{\sigma(i-1)})\} \\ &= \min\{I(X_{\pi(s)}; Y_2 | X^{\pi(s)-1}), \\ &\quad I(X_{\pi(s)}; Y_1 | X_{\pi(1)}, \ldots, X_{\pi(s-1)}, W_{\sigma(1)}, \ldots, W_{\sigma(s-1)})\} \\ &\le I(X_{\pi(s)}; Y_1 | X_{\pi(1)}, \ldots, X_{\pi(s-1)}, W_{\sigma(1)}, \ldots, W_{\sigma(s-1)}) \\ &= I(X; Y_1) \\ &= \mathsf{C}(S/(1+I)) \\ &< \mathsf{C}(I/(S+I)), \end{aligned}$$

which is a contradiction to the standing assumption in (48). This completes the proof of Proposition 3.

APPENDIX D
THE SWSC SCHEME IN TABLE I

Codebook generation. Fix a pmf $p'(x_1)p'(x_2)p'(w)$ and a function $x(x_1, x_2)$. Randomly and independently generate a codebook for each block. For notational convention, we assume $m_1(0) = m_1(b) = 1$. For $j \in [b]$, randomly and independently generate 2^{nR_1} sequences $x_1^n(m_1(j-1)), m_1(j-1) \in [2^{nR_1}]$, each according to $\prod_{i=1}^n p'_{X_1}(x_{1i})$. For $j \in [b]$, randomly and independently generate 2^{nR_1} sequences $x_2^n(m_1(j)), m_1(j) \in [2^{nR_1}]$, each according to $\prod_{i=1}^n p'_{X_2}(x_{2i})$. For $j \in [b]$, randomly and independently generate 2^{nR_2} sequences $w^n(m_2(j)), m_2(j) \in [2^{nR_2}]$, each according to $\prod_{i=1}^n p'_W(w_i)$. This defines the codebook

$$\begin{aligned} \mathcal{C}_j = \big\{ &x_1^n(m_1(j-1)), x_2^n(m_1(j)), w^n(m_2(j)) : \\ &m_1(j-1), m_1(j) \in [2^{nR_1}], m_2(j) \in [2^{nR_2}] \big\}, \quad j \in [b]. \end{aligned}$$

Encoding. In block $j \in [b]$, sender 1 transmits $x_i(x_{1i}(m_1(j-1)), x_{2i}(m_1(j)))$ at time $i \in [n]$ and sender 2 transmits $w^n(m_2(j))$. Table I reveals the scheduling of the messages.

Decoding. Let the received sequences in block j be $y_1^n(j)$ and $y_2^n(j)$, $j \in [b]$. For receiver 1, at the end of block 1, it finds the unique message $\hat{m}_2(1)$ such that

$$(w^n(\hat{m}_2(1)), y_1^n(1), x_1^n(m_1(0))) \in \mathcal{T}_\epsilon^{(n)}.$$

At the end of block j, $2 \le j \le b$, it finds the unique message $\hat{m}_1(j-1)$ such that

$$(x_1^n(\hat{m}_1(j-2)), x_2^n(\hat{m}_1(j-1)), w^n(\hat{m}_2(j-1)), y_1^n(j-1)) \in \mathcal{T}_\epsilon^{(n)}$$

and

$$(x_1^n(\hat{m}_1(j-1)), y_1^n(j)) \in \mathcal{T}_\epsilon^{(n)}$$

simultaneously. Then it finds the unique $\hat{m}_2(j)$ such that

$$(w^n(\hat{m}_2(j)), y_1^n(j), x_1^n(\hat{m}_1(j-1))) \in \mathcal{T}_\epsilon^{(n)}.$$

If any of the typicality checks fails, it declares an error.

We analyze the probability of decoding error averaged over codebooks. Assume without loss of generality that $M_1(j) = M_2(j) = 1$. We divide the error events as follows

$$\begin{aligned} \mathcal{E}_{11}(j-2) &= \{\hat{M}_1(j-2) \neq 1\}, \\ \mathcal{E}_{12}(j-1) &= \{\hat{M}_2(j-1) \neq 1\}, \\ \mathcal{E}_{13}(j-1) &= \{(X_1^n(\hat{M}_1(j-2)), X^n(1), W^n(\hat{M}_2(j-1)), \\ &\qquad Y_1^n(j-1)) \notin \mathcal{T}_\epsilon^{(n)} \text{ or } (X_1^n(1), Y_1^n(j)) \notin \mathcal{T}_\epsilon^{(n)}\}, \\ \mathcal{E}_{14}(j-1) &= \{(X_1^n(\hat{M}_1(j-2)), X^n(m_1(j-1)), \\ &\qquad W^n(\hat{M}_2(j-1)), Y_1^n(j-1)) \in \mathcal{T}_\epsilon^{(n)} \\ &\qquad \text{and } (X_1^n(m_1(j-1)), Y_1^n(j)) \in \mathcal{T}_\epsilon^{(n)} \\ &\qquad \text{for some } m_1(j-1) \neq 1\}, \\ \mathcal{E}_{15}(j) &= \{(W^n(1), Y_1^n(j), X_1^n(\hat{M}_1(j-1))) \notin \mathcal{T}_\epsilon^{(n)}\}, \\ \mathcal{E}_{16}(j) &= \{(W^n(m_2(j)), Y_1^n(j), X_1^n(\hat{M}_1(j-1))) \in \mathcal{T}_\epsilon^{(n)} \\ &\qquad \text{for some } m_2(j) \neq 1\}. \end{aligned}$$

We analyze by induction. By assumption $\mathcal{E}_{11}(0)=\emptyset$. Thus in block 1, the probability of error is upper bounded as

$$\begin{aligned}\mathsf{P}\{\hat{M}_2(1)\neq 1\} &= \mathsf{P}(\mathcal{E}_{12}(1))\\ &\le \mathsf{P}(\mathcal{E}_{15}(1))+\mathsf{P}(\mathcal{E}_{16}(1)).\end{aligned}$$

Now by the law of large numbers, $\mathsf{P}(\mathcal{E}_{15}(1))\to 0$ as $n\to\infty$. By the packing lemma, $\mathsf{P}(\mathcal{E}_{16}(1))\to 0$ as $n\to\infty$ if $R_2 < I(W;Y_1|X_1)-\delta(\epsilon)$. Now assume that the probability of error $\mathsf{P}(\mathcal{E}_{11}(j-2)\cup\mathcal{E}_{12}(j-1))$ in block $j-1$ tends to zero as $n\to\infty$. In block j, the probability of error is upper bounded as

$$\begin{aligned}&\mathsf{P}\{(\hat{M}_1(j-1),\hat{M}_2(j))\neq(1,1)\}\\ &\le \mathsf{P}(\mathcal{E}_{11}(j-2)\cup\mathcal{E}_{12}(j-1)\cup\mathcal{E}_{11}(j-1)\cup\mathcal{E}_{12}(j))\\ &\le \mathsf{P}(\mathcal{E}_{11}(j-2)\cup\mathcal{E}_{12}(j-1))\\ &\quad+\mathsf{P}(\mathcal{E}_{11}(j-1)\cap\mathcal{E}_{11}^c(j-2)\cap\mathcal{E}_{12}^c(j-1))\\ &\quad+\mathsf{P}(\mathcal{E}_{12}(j)\cap\mathcal{E}_{11}^c(j-1))\\ &\le \mathsf{P}(\mathcal{E}_{11}(j-2)\cup\mathcal{E}_{12}(j-1))\\ &\quad+\mathsf{P}(\mathcal{E}_{13}(j-1)\cap\mathcal{E}_{11}^c(j-2)\cap\mathcal{E}_{12}^c(j-1))\\ &\quad+\mathsf{P}(\mathcal{E}_{14}(j-1)\cap\mathcal{E}_{11}^c(j-2)\cap\mathcal{E}_{12}^c(j-1))\\ &\quad+\mathsf{P}(\mathcal{E}_{15}(j)\cap\mathcal{E}_{11}^c(j-1))+\mathsf{P}(\mathcal{E}_{16}(j)\cap\mathcal{E}_{11}^c(j-1)).\end{aligned}$$

By the induction assumption, the first term tends to zero as $n\to\infty$. By the independence of the codebooks, the law of large numbers, and the packing lemma, the second, fourth, and fifth terms tend to zero as $n\to\infty$ if $R_2 < I(W;Y_1|X_1)-\delta(\epsilon)$. The third term $\mathsf{P}(\mathcal{E}_{14}(j-1)\cap\mathcal{E}_{11}^c(j-2)\cap\mathcal{E}_{12}^c(j-1))$ requires a special care. We have

$$\begin{aligned}&\mathsf{P}(\mathcal{E}_{14}(j-1)\cap\mathcal{E}_{11}^c(j-2)\cap\mathcal{E}_{12}^c(j-1))\\ &= \mathsf{P}\{(X_1^n(1),X^n(m_1(j-1)),W^n(1),Y_1^n(j-1))\in\mathcal{T}_\epsilon^{(n)}\\ &\quad\text{and } (X_1^n(m_1(j-1)),Y_1^n(j))\in\mathcal{T}_\epsilon^{(n)}\\ &\quad\text{for some } m_1(j-1)\neq 1\}\\ &= \sum_{m_1(j-1)\neq 1}\mathsf{P}\{(X_1^n(m_1(j-1)),Y_1^n(j))\in\mathcal{T}_\epsilon^{(n)}\text{ and}\\ &\quad(X_1^n(1),X^n(m_1(j-1)),W^n(1),Y_1^n(j-1))\in\mathcal{T}_\epsilon^{(n)}\}\\ &\overset{(a)}{=} \sum_{m_1(j-1)\neq 1}\mathsf{P}\{(X_1^n(m_1(j-1)),Y_1^n(j))\in\mathcal{T}_\epsilon^{(n)}\}\\ &\quad\cdot\mathsf{P}\{(X_1^n(1),X^n(m_1(j-1)),W^n(1),Y_1^n(j-1)\in\mathcal{T}_\epsilon^{(n)}\}\\ &\overset{(b)}{\le} 2^{nR_1}2^{-n(I(X;Y_1|W,X_1)-\delta(\epsilon))}2^{-n(I(X_1;Y_1)-\delta(\epsilon))},\end{aligned}$$

which tends to zero if $R_1 < I(X_1;Y_1)+I(X;Y_1|W,X_1)-2\delta(\epsilon)$. Here (a) follows by the independence of the codebooks, which implies the events

$$\{(X_1^n(1),X^n(m_1(j-1)),W^n(1),Y_1^n(j-1))\in\mathcal{T}_\epsilon^{(n)}\}$$

and

$$\{(X_1^n(m_1(j-1)),Y_1^n(j))\in\mathcal{T}_\epsilon^{(n)}\}$$

are independent for each $m_1(j-1)\neq 1$, and (b) follows by the independence of the codebooks and the joint typicality lemma.

For receiver 2, at the end of block j, $2\le j\le b$, it finds the unique $\hat{m}_1(j-1)$ such that

$$(x_1^n(\hat{m}_1(j-2)),x_2^n(\hat{m}_1(j-1)),y_2^n(j-1))\in\mathcal{T}_\epsilon^{(n)}$$

and

$$(x_1^n(\hat{m}_1(j-1)),y_2^n(j))\in\mathcal{T}_\epsilon^{(n)}$$

simultaneously. Then it finds the unique $\hat{m}_2(j-1)$ such that

$$\begin{aligned}&(w^n(\hat{m}_2(j-1)),y_2^n(j-1),x_1^n(\hat{m}_1(j-2)),\\ &\qquad x_2^n(\hat{m}_1(j-1)))\in\mathcal{T}_\epsilon^{(n)}.\end{aligned}$$

In the end, receiver 2 finds the unique $\hat{m}_2(b)$ such that

$$(w^n(\hat{m}_2(b)),y_2^n(b),x_1^n(\hat{m}_1(b-1)),x_2^n(m_1(b)))\in\mathcal{T}_\epsilon^{(n)}.$$

If any of the typicality checks fails, it declares an error. With a similar analysis as above, the decoding is successful if

$$\begin{aligned}R_1 &< I(X_1;Y_2)+I(X_2;Y_2|X_1)-2\delta(\epsilon)\\ &= I(X;Y_2)-2\delta(\epsilon),\\ R_2 &< I(W;Y_2|X)-\delta(\epsilon).\end{aligned}$$

Appendix E
Proof of Theorem 3

We first extend the layer-splitting lemma (Lemma 1) to the three-user case and show that by splitting two inputs into two layers each and keeping one input unsplit, any rate triple on the dominant face of $\mathscr{R}_{\mathrm{MAC}}(A,B,C;Y)$ is achievable by successive cancellation decoding.

The dominant face of $\mathscr{R}_{\mathrm{MAC}}(A,B,C;Y)$ is illustrated in Figure 14. We label the six corner points by $\mathbf{I}_{ABC},\mathbf{I}_{BAC},\mathbf{I}_{BCA},\mathbf{I}_{CBA},\mathbf{I}_{CAB},\mathbf{I}_{ACB}$, corresponding to the following six rate vectors

$$\begin{aligned}\mathbf{I}_{ABC} &= (I(A;Y),I(B;Y|A),I(C;Y|A,B)),\\ \mathbf{I}_{BAC} &= (I(A;Y|B),I(B;Y),I(C;Y|B,A)),\\ \mathbf{I}_{BCA} &= (I(A;Y|B,C),I(B;Y),I(C;Y|B)),\\ \mathbf{I}_{CBA} &= (I(A;Y|C,B),I(B;Y|C),I(C;Y)),\\ \mathbf{I}_{CAB} &= (I(A;Y|C),I(B;Y|C,A),I(C;Y)),\\ \mathbf{I}_{ACB} &= (I(A;Y),I(B;Y|A,C),I(C;Y|A)).\end{aligned}$$

We partition this hexagon region into three subregions: two triangles $\triangle(\mathbf{I}_{ACB},\mathbf{I}_{ABC},\mathbf{I}_{BAC})$ and $\triangle(\mathbf{I}_{BCA},\mathbf{I}_{CBA},\mathbf{I}_{CAB})$, and a trapezoid $\square$ $(\mathbf{I}_{ACB},\mathbf{I}_{BAC},\mathbf{I}_{BCA},\mathbf{I}_{CAB})$. In order to achieve each region by successive cancellation decoding, we split A and B into (A_1,A_2) and (B_1,B_2) respectively. In other words, we consider p' of the form $p'(a_1)p'(a_2)p'(b_1)p'(b_2)p'(c)$ and functions $a(a_1,a_2)$ and $b(b_1,b_2)$ such that $p'\simeq p(a)p(b)p(c)$. Let $\mathscr{R}(p',\lambda)$, $\lambda=1,2,3$, be the set of achievable rate triples (R_1,R_2,R_3) associated with the following layer orders (defined in a similar manner as in Section V)

$$1\colon A_1\to B_1\to A_2\to C\to B_2, \tag{58a}$$

$$2\colon B_1\to C\to A_1\to B_2\to A_2, \tag{58b}$$

$$3\colon B_1\to A_1\to C\to A_2\to B_2. \tag{58c}$$

For example, $\mathscr{R}(p',1)$ is the set of rate triples (r_1,r_2,r_3) such that

$$\begin{aligned} r_1 &\le I(A_1;Y)+I(A_2;Y|A_1,B_1)\\ r_2 &\le I(B_1;Y|A_1)+I(B_2;Y|A_1,B_1,A_2,C)\\ r_3 &\le I(C;Y|A_1,B_1,A_2). \end{aligned} \tag{59}$$

We need to show for every point in $\mathscr{R}_{\mathrm{MAC}}(A,B,C;Y)$, there exists some choice of p' that achieves it. Similar to Lemma 1, we choose the conditional pmfs $p(a_1|a)$ and $p(b_1|b)$ as erasure channels with erasure probabilities α and β respectively. Then the rate expressions (59) can be further simplified as

$$\begin{aligned} r_1 &= (1-\alpha)(1-\beta)I(A;Y)+\alpha(1-\beta)I(A;Y|B)\\ &\quad +\beta I(A;Y),\\ r_2 &= (1-\alpha)(1-\beta)I(B;Y|A)+\alpha(1-\beta)I(B;Y)\\ &\quad +\beta I(B;Y|A,C),\\ r_3 &= (1-\alpha)(1-\beta)I(C;Y|A,B)+\alpha(1-\beta)I(C;Y|B,A)\\ &\quad +\beta I(C;Y|A). \end{aligned}$$

In other words, letting $\mathbf{r}:=(r_1,r_2,r_3)$, the achievable rate region $\mathscr{R}(\alpha,\beta,\lambda)$ for $\lambda=1$ is the set of rate vectors $\mathbf{r}$ such that

$$\mathbf{r}\le(1-\alpha)(1-\beta)\mathbf{I}_{ABC}+\alpha(1-\beta)\mathbf{I}_{BAC}+\beta\mathbf{I}_{ACB}.$$

This region covers every point in the triangle $\triangle(\mathbf{I}_{ACB},\mathbf{I}_{ABC},\mathbf{I}_{BAC})$ by varying $\alpha,\beta\in[0,1]$. Similarly, the achievable rate region $\mathscr{R}(\alpha,\beta,\lambda)$ for $\lambda=2$ is the set of rate vectors $\mathbf{r}$ such that

$$\mathbf{r}\le(1-\alpha)\beta\mathbf{I}_{CAB}+\alpha\beta\mathbf{I}_{CBA}+(1-\beta)\mathbf{I}_{BCA}.$$

This region covers every point in the triangle $\triangle(\mathbf{I}_{BCA},\mathbf{I}_{CBA},\mathbf{I}_{CAB})$ by varying $\alpha,\beta\in[0,1]$. For layer order $\lambda=3$, the achievable rate region $\mathscr{R}(\alpha,\beta,\lambda)$ is given by

$$\begin{aligned} \mathbf{r}\le{}&(1-\alpha)(1-\beta)\mathbf{I}_{BAC}+(1-\alpha)\beta\mathbf{I}_{ACB}+\alpha(1-\beta)\mathbf{I}_{BCA}\\ &+\alpha\beta\mathbf{I}_{CAB}. \end{aligned}$$

Note that for each fixed β, the trajectory of the achievable rate points when varying α from 0 to 1 is a line segment that is parallel to the two sides $(\mathbf{I}_{ACB},\mathbf{I}_{CAB})$ and $(\mathbf{I}_{BAC},\mathbf{I}_{BCA})$. By further varying β from 0 to 1, this layer order achieves every point in the trapezoid $\square\,(\mathbf{I}_{ACB},\mathbf{I}_{BAC},\mathbf{I}_{BCA},\mathbf{I}_{CAB})$.

Lemma 6 (Layer Splitting for a 3-User MAC [18]): For a 3-user MAC $p(y|a,b,c)$, the achievable rate region for input pmf $p=p(a)p(b)p(c)$ can be equivalently expressed as

$$\mathscr{R}_{\mathrm{MAC}}(A,B,C;Y)=\bigcup_{p'\simeq p}\bigcup_{\lambda=1}^{3}\mathscr{R}(p',\lambda).$$

where the layer orders $\lambda=1,2,3$ are given in (58)*. Moreover, let $p(a_1|a)$ and $p(b_1|b)$ be two erasure channels with erasure probabilities α and β respectively. Then,*

$$\mathscr{R}_{\mathrm{MAC}}(A,B,C;Y)=\bigcup_{\alpha\in[0,1],\beta\in[0,1]}\bigcup_{\lambda=1}^{3}\mathscr{R}(\alpha,\beta,\lambda). \tag{60}$$

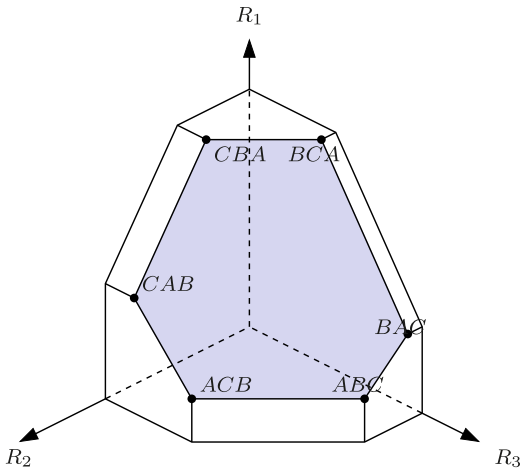


Fig. 14. Achievable rate region of the three-user MAC $p(y|a,b,c)$.

In order to express the 4-dimensional auxiliary region (31), we split S into three layers (S_1,S_2,S_3) and T,V into two layers each (T_1,T_2) and (V_1,V_2). At receiver 1, we consider layer orders λ_1 given by

$$\begin{aligned} &1\colon S_1\to T_1\to S_2\to S_3\to U\to T_2,\\ &2\colon T_1\to U\to S_1\to T_2\to S_2\to S_3,\\ &3\colon T_1\to S_1\to U\to S_2\to S_3\to T_2,\\ &4\colon S_1\to S_2\to T_1\to S_3\to U\to T_2,\\ &5\colon T_1\to U\to S_1\to S_2\to T_2\to S_3,\\ &6\colon T_1\to S_1\to S_2\to U\to S_3\to T_2. \end{aligned}$$

Let p' be the pmf $p(s_1)p(s_2)p(s_3)p(t_1)p(t_2)p(u)p(v_1)p(v_2)$ along with $s(s_1,s_2,s_3),t(t_1,t_2),v(v_1,v_2)$. Let $\mathscr{R}_1(p',\lambda_1)$ be the achievable rate region at receiver 1 for the layer order $\lambda_1\in[6]$. For example, $\mathscr{R}_1(p',1)$ is the set of rate quadruples $(R_{10},R_{11},R_{20},R_{22})$ such that

$$\begin{aligned} R_{10}&\le I(S_1;Y_1)+I(S_2;Y_1|S_1,T_1)+I(S_3;Y_1|S_1,T_1,S_2),\\ R_{20}&\le I(U;Y_1|S_1,T_1,S_2,S_3),\\ R_{11}&\le I(T_1;Y_1|S_1)+I(T_2;Y_1|S_1,T_1,S_2,S_3,U). \end{aligned}$$

At receiver 2, we consider layer orders $\lambda_2=7,8,\ldots,12$, which are obtained from $\lambda_1=1,2,\ldots,6$, respectively, by replacing T_1 by V_1 and T_2 by V_2. For example, the layer order $\lambda_2=7$ is obtained from the layer order $\lambda_1=1$ as

$$7\colon S_1\to V_1\to S_2\to S_3\to U\to V_2.$$

Let $\mathscr{R}_2(p',\lambda_2)$ be the achievable rate region at receiver 2 for the layer order $\lambda_2=7,8,\ldots,12$. For example, $\mathscr{R}_2(p',7)$ is the set of rate quadruples $(R_{10},R_{11},R_{20},R_{22})$ such that

$$\begin{aligned} R_{10}&\le I(S_1;Y_2)+I(S_2;Y_2|S_1,V_1)+I(S_3;Y_2|S_1,V_1,S_2),\\ R_{20}&\le I(U;Y_2|S_1,V_1,S_2,S_3),\\ R_{22}&\le I(V_1;Y_2|S_1)+I(V_2;Y_2|S_1,V_1,S_2,S_3,U). \end{aligned}$$

Lemma 7: Let p denote the pmf $p(s)p(t)p(u)p(v)$ along with functions $x(s,t)$ and $w(u,v)$. Then

$$\begin{aligned}&\mathscr{R}_{1,\mathrm{MAC}}(p)\cap\mathscr{R}_{2,\mathrm{MAC}}(p)\\&=\cup_{p'\simeq p}\big[\big(\cup_{\lambda_1=1}^{3}\cup_{\lambda_2=10}^{12}\left[\mathscr{R}_1(p',\lambda_1)\cap\mathscr{R}_2(p',\lambda_2)\right]\big)\\&\quad\cup\big(\cup_{\lambda_1=4}^{6}\cup_{\lambda_2=7}^{9}\left[\mathscr{R}_1(p',\lambda_1)\cap\mathscr{R}_2(p',\lambda_2)\right]\big)\big].\end{aligned}\tag{61}$$

Proof: By Lemma 6, the target rate region can be equivalently expressed as

$$\begin{aligned}&\mathscr{R}_{1,\mathrm{MAC}}(p)\cap\mathscr{R}_{2,\mathrm{MAC}}(p)\\&=\Big(\cup_{\alpha',\beta\in[0,1]}\cup_{\tilde{\lambda}_1=1}^{3}\mathscr{R}_1(\alpha',\beta,\tilde{\lambda}_1)\Big)\\&\qquad\cap\Big(\cup_{\alpha'',\gamma\in[0,1]}\cup_{\tilde{\lambda}_2=4}^{6}\mathscr{R}_2(\alpha'',\gamma,\tilde{\lambda}_2)\Big)\\&=\cup_{\alpha',\alpha'',\beta,\gamma\in[0,1]}\cup_{\tilde{\lambda}_1=1}^{3}\cup_{\tilde{\lambda}_2=4}^{6}[\mathscr{R}_1(\alpha',\beta,\tilde{\lambda}_1)\cap\mathscr{R}_2(\alpha'',\gamma,\tilde{\lambda}_2)]\end{aligned}$$

for some erasure channels $p(s_1'|s), p(s_1''|s), p(t_1|t), p(v_1|v)$ with erasure probabilities $\alpha', \alpha'', \beta, \gamma$, respectively, and the layer orders are

$$\begin{aligned}&1\colon S_1'\to T_1\to S_2'\to U\to T_2,\\&2\colon T_1\to U\to S_1'\to T_2\to S_2',\\&3\colon T_1\to S_1'\to U\to S_2'\to T_2,\\&4\colon S_1''\to V_1\to S_2''\to U\to V_2,\\&5\colon V_1\to U\to S_1''\to V_2\to S_2'',\\&6\colon V_1\to S_1''\to U\to S_2''\to V_2.\end{aligned}$$

Now following similar steps to the proof of Proposition 2, we can merge (S_1', S_2') and (S_1'', S_2'') into (S_1, S_2, S_3) as follows. When $\alpha' > \alpha''$, the channel $p(s_1'|s)$ is degraded with respect to $p(s_1''|s)$. We assume without loss of generality that $S \to S_1'' \to S_1'$ form a Markov chain. By the functional representation lemma (twice), for $p(s_1'|s_1'')$, there exists an S_2 independent of S_1' such that $S_1'' = f(S_1', S_2)$; for $p(s_1''|s)$, there exists an S_3 independent of (S_2, S_1') such that $S = g(S_1'', S_3) = g(f(S_1', S_2), S_3) \triangleq s(S_1', S_2, S_3)$. Renaming $S_1 \triangleq S_1'$ and plugging $S_1'' = f(S_1, S_2)$, the rate region $\mathscr{R}_1(\alpha', \beta, \tilde{\lambda}_1)$, $\tilde{\lambda}_1 = 1, 2, 3$, becomes $\mathscr{R}_1(p', \lambda_1)$, $\lambda_1 = 1, 2, 3$, respectively. The rate region $\mathscr{R}_2(\alpha'', \gamma, \tilde{\lambda}_2)$, $\tilde{\lambda}_2 = 4, 5, 6$, becomes $\mathscr{R}_2(p', \lambda_2)$, $\lambda_2 = 10, 11, 12$, respectively. Thus, we have

$$\begin{aligned}&\cup_{\tilde{\lambda}_1=1}^{3}\cup_{\tilde{\lambda}_2=4}^{6}[\mathscr{R}_1(\alpha',\beta,\tilde{\lambda}_1)\cap\mathscr{R}_2(\alpha'',\gamma,\tilde{\lambda}_2)]\\&=\cup_{\lambda_1=1}^{3}\cup_{\lambda_2=10}^{12}[\mathscr{R}_1(p',\lambda_1)\cap\mathscr{R}_2(p',\lambda_2)].\end{aligned}$$

When $\alpha' \le \alpha''$, we can assume that $S \to S_1' \to S_1''$ form a Markov chain. Following similar steps, the rate region $\mathscr{R}_1(\alpha', \beta, \tilde{\lambda}_1)$, $\tilde{\lambda}_1 = 1, 2, 3$, becomes $\mathscr{R}_1(p', \lambda_1)$, $\lambda_1 = 4, 5, 6$, respectively. The rate region $\mathscr{R}_2(\alpha'', \gamma, \tilde{\lambda}_2)$, $\tilde{\lambda}_2 = 4, 5, 6$, becomes $\mathscr{R}_2(p', \lambda_2)$, $\lambda_2 = 7, 8, 9$, respectively. Thus, we have

$$\begin{aligned}&\cup_{\tilde{\lambda}_1=1}^{3}\cup_{\tilde{\lambda}_2=4}^{6}[\mathscr{R}_1(\alpha',\beta,\tilde{\lambda}_1)\cap\mathscr{R}_2(\alpha'',\gamma,\tilde{\lambda}_2)]\\&=\cup_{\lambda_1=4}^{6}\cup_{\lambda_2=7}^{9}[\mathscr{R}_1(p',\lambda_1)\cap\mathscr{R}_2(p',\lambda_2)].\end{aligned}$$

□

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Associate Editor and anonymous reviewers for their helpful comments and suggestions, which improved the overall presentation of this article.

REFERENCES

[1] G. Boudreau, J. Panicker, N. Guo, R. Chang, N. Wang, and S. Vrzic, "Interference coordination and cancellation for 4G networks," *IEEE Commun. Mag.*, vol. 47, no. 4, pp. 74–81, Apr. 2009.

[2] C. Seol and K. Cheun, "A statistical inter-cell interference model for downlink cellular OFDMA networks under log-normal shadowing and multipath Rayleigh fading," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3069–3077, Oct. 2009.

[3] V. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[4] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[5] S. S. Bidokhti and V. M. Prabhakaran, "Is non-unique decoding necessary?" *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2594–2610, May 2014.

[6] M. Costa and A. E. Gamal, "The capacity region of the discrete memoryless interference channel with strong interference (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 5, pp. 710–711, Sep. 1987.

[7] H. Sato, "On the capacity region of a discrete two-user channel for strong interference (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 377–379, May 1978.

[8] X. Shang, G. Kramer, and B. Chen, "A new outer bound and the noisy-interference sum–rate capacity for Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 689–699, Feb. 2009.

[9] V. S. Annapureddy and V. V. Veeravalli, "Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3032–3050, Jul. 2009.

[10] A. S. Motahari and A. K. Khandani, "To decode the interference or to consider it as noise," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1274–1283, Mar. 2011.

[11] S. Liu, C. Nair, and L. Xia, "Interference channels with very weak interference," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 1031–1035.

[12] F. Baccelli, A. El Gamal, and D. N. C. Tse, "Interference networks with point-to-point codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2582–2596, May 2011.

[13] B. Bandemer, A. El Gamal, and Y.-H. Kim, "Optimal achievable rates for interference networks with random codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6536–6549, Dec. 2015.

[14] T. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.

[15] A. Yedla, P. S. Nguyen, H. D. Pfister, and K. R. Narayanan, "Universal codes for the Gaussian MAC via spatial coupling," in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep. 2011, pp. 1801–1808.

[16] L. Wang and E. Sasoglu, "Polar coding for interference networks," 2014, *arXiv:1401.7293*. [Online]. Available: http://arxiv.org/abs/1401.7293

[17] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 364–375, Mar. 1996.

[18] A. J. Grant, B. Rimoldi, R. L. Urbanke, and P. A. Whiting, "Rate-splitting multiple access for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 873–890, Mar. 2001.

[19] Y. Zhao, C. W. Tan, A. S. Avestimehr, S. N. Diggavi, and G. J. Pottie, "On the maximum achievable sum-rate with successive decoding in interference channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3798–3820, Jun. 2012.

[20] L. Wang, E. Sasoglu, and Y.-H. Kim, "Sliding-window superposition coding for interference networks," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 2749–2753.

[21] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 3, pp. 371–377, May 1977.

[22] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.
[23] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873–884, May 1992.
[24] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
[25] H. Park, Y.-H. Kim, and L. Wang, "Interference management via sliding-window superposition coding," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2014, pp. 972–976.
[26] K. T. Kim *et al.*, "Adaptive sliding-window coded modulation in cellular networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–7.
[27] K. T. Kim, S.-K. Ahn, Y.-S. Kim, J. Park, C.-Y. Chen, and Y.-H. Kim, "Interference management via sliding-window coded modulation for 5G cellular networks," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 82–89, Nov. 2016.
[28] L. Wang, "Channel coding techniques for network communication," Ph.D. dissertation, Univ. California, San Diego, CA, USA, 2015.
[29] *Vision and Schedule for 5G Radio Technologies*, document RWS-150039, 3GPP TSG RAN Workshop on 5G, Phoenix, AZ, USA, Sep. 2015. [Online]. Available: http://www.3gpp.org/ftp/workshop/2015-09-17_18_RAN_5G/Docs/RWS-150039.zip
[30] *Interference Coordination for 5G New Radio Interface*, document R1-162185, Apr. 2016. [Online]. Available: http://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_84b/Docs/R1-162185.zip
[31] *Discussion on Interference Management Based on Advanced Transceivers for NR*, document R1-164023, 3GPP TSG RAN WG1 #85, May 2016. [Online]. Available: http://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_85/Docs/R1-164023.zip
[32] *Discussion on Spatial Multiplexing for NR*, document R1-167887, 3GPP TSG RAN WG1 #86, Aug. 2016. [Online]. Available: http://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_86/Docs/R1-167887.zip
[33] *Discussion on Modulation for NR*, document R1-166776, 3GPP TSG RAN WG1 #86, Aug. 2016. [Online]. Available: http://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_86/Docs/R1-166776.zip
[34] *Discussion on Interference Management Based on Advanced Transceivers for NR*, document R1-166791, 3GPP TSG RAN WG1 #86, Aug. 2016. [Online]. Available: http://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_86/Docs/R1-166791.zip
[35] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.
[36] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.
[37] L. Wang, E. Sasoglu, B. Bandemer, and Y.-H. Kim, "A comparison of superposition coding schemes," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 2970–2974.
[38] O. Fawzi and I. Savov, "Rate-splitting in the presence of multiple receivers," 2012, *arXiv:1207.0543*. [Online]. Available: http://arxiv.org/abs/1207.0543
[39] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.
[40] T. Cover and C. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 27, no. 3, pp. 292–298, May 1981.
[41] A. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 841–850, Nov. 1982.
[42] L.-L. Xie and P. R. Kumar, "An achievable rate for the multiple-level relay channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1348–1358, Apr. 2005.
[43] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.
[44] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.
[45] X. Li, H. Huang, G. J. Foschini, and R. A. Valenzuela, "Effects of iterative detection and decoding on the performance of BLAST," in *Proc. Global Telecommun. Conf. (GLOBECOM)*, vol. 2, 2000, pp. 1061–1066.
[46] G. J. Foschini, D. Chizhik, M. J. Gans, C. Papadias, and R. A. Valenzuela, "Analysis and performance of some basic space-time architectures," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 303–320, Apr. 2003.
[47] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. URSI Int. Symp. Signals, Syst., Electron.*, Oct. 1998, pp. 295–300.
[48] *Multiplexing and Channel Coding*, document TS 36.212, 3GPP, Release 12, 2013.

Lele Wang (Member, IEEE) received the B.E. degree from Tsinghua University in 2009 and the Ph.D. degree from the University of California at San Diego (UCSD), San Diego, in 2015, all in electrical engineering. From 2015 to 2019, she was first a joint Post-Doctoral Fellow with Tel Aviv University and Stanford University and then an NSF Center for Science of Information (CSoI) Post-Doctoral Fellow. In 2019, she joined the University of British Columbia, Vancouver, where she is currently an Assistant Professor with the Department of Electrical and Computer Engineering. Her research interests include information theory, coding theory, communication theory, and mathematical data science. She was a recipient of the 2013 UCSD Shannon Memorial Fellowship, the 2013–2014 Qualcomm Innovation Fellowship, the 2017 IEEE Information Theory Society Thomas M. Cover Dissertation Award, and the 2017 NSF CSoI Postdoctoral Fellowship.

Young-Han Kim received the B.S. degree (Hons.) in electrical engineering from Seoul National University, Seoul, South Korea, in 1996, and the M.S. degree in electrical engineering, the M.S. degree in statistics, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 2001, 2006, and 2006, respectively. In 2006, he joined the University of California at San Diego, La Jolla, CA, USA, where he is currently a Professor with the Department of Electrical and Computer Engineering. He has coauthored the book *Network Information Theory* (Cambridge University Press, 2011) and the monograph *Fundamentals of Index Coding* (Now Publishers, 2019). His research interests include information theory, communication engineering, and data science. He was a recipient of the 2008 NSF Faculty Early Career Development Award, the 2009 U.S.–Israel Binational Science Foundation Bergmann Memorial Award, the 2012 IEEE Information Theory Paper Award, and the 2015 IEEE Information Theory Society James L. Massey Research and Teaching Award for Young Scholars. He served as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY and a Distinguished Lecturer for the IEEE Information Theory Society.

Chiao-Yi Chen received the B.S. degrees in electrical engineering and computer science and information engineering from National Taiwan University, Taipei, Taiwan, in 2006, and the M.S. degree in electrical and computer engineering from the University of California at San Diego, San Diego, in 2011. He is currently a Research and Development Engineer with Broadcom Inc., San Jose, CA, USA. His research interests include information theory, communication theory, and universal processing.

Hosung Park (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Seoul National University, Seoul, South Korea, in 2007, 2009, and 2013, respectively. He was a Post-Doctoral Researcher with the Institute of New Media and Communications, Seoul National University, in 2013, and the Qualcomm Institute, the California Institute for Telecommunications and Information Technology, University of California at San Diego, La Jolla, CA, USA, from 2013 to 2015. He has been an Associate Professor with the School of Electronics and Computer Engineering, Chonnam National University, Gwangju, South Korea, since 2015. His research interests include channel codes for communications systems, coding for memory/storage, coding for distributed storage, communication signal processing, compressed sensing, and network information theory.

Eren Şaşoğlu received the B.Sc. degree in electrical engineering from Boğaziçi University in 2005 and the M.Sc. and Ph.D. degrees in communication systems from EPFL in 2007 and 2011, respectively. He was a Post-Doctoral Scholar with UC San Diego and UC Berkeley, an Academic Visitor at Technion, and a Research Scientist with Intel. He is currently with Apple Inc. He was a recipient of the Best Doctoral Thesis Award at EPFL in 2013 and the STOC Best Paper Award in 2016.