On the Sum Capacity of Dual-Class Parallel Packet-Erasure Broadcast Channels

Sunghyun Kim, Soheil Mohajer, *Member, IEEE*, and Changho Suh, *Senior Member, IEEE*

***Abstract*— We investigate a K-user parallel packet-erasure broadcast channel. There is an ongoing effort to harness millimeter-wave bands, which are known to be unstable having high outage probabilities, by combining them with stable legacy bands. Motivated by this effort, we consider a *heterogeneous* scenario in which the parallel subchannels are categorized into two classes having different outage probabilities. For the two-user case, we characterize the sum capacity by developing an explicit achievable scheme and deriving a matching upper bound. In contrast to suboptimal schemes that apply coding on a per-subchannel basis only, our scheme applies coding *across* subchannels to exploit coding opportunities that arise from asymmetric outage probabilities more efficiently, thereby achieving optimality. By extending our scheme systematically to be applicable for the K-user case, we show that it can provide significant gains over existing schemes. Compared to the K-user scheme currently employed in practice, which allocates chunks of subchannels to users exclusively, we demonstrate the performance improvement attainable by our scheme to be substantial, as the multiplicative gain scales with K. Moreover, we find that our scheme outperforms a per-subchannel extension of state-of-the-art K-user schemes by large margins, further reducing the optimality gap. Our results suggest a potential coding scheme that can be employed in future wireless systems to meet ever-growing mobile data demands.**

***Index Terms*— Broadcast channels, communication systems, information theory, millimeter wave communication, state feedback.**

Manuscript received January 20, 2020; revised July 3, 2020 and October 23, 2020; accepted January 2, 2021. Date of publication January 13, 2021; date of current version April 16, 2021. The work of Soheil Mohajer was supported by the National Science Foundation under Grant CCF-1749981. The work of Changho Suh was supported by the Air Force Office of Scientific Research under Award number FA2386-19-1-4050, Korea Advanced Institute of Science and Technology (KAIST) Grand Challenge 30 (KC30) Project in 2019 (1711100603) funded by KAIST and Korea Government (MSIP), and National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIP) (NRF-2018R1A1A1A05022889). This article was presented in part at the 2017 IEEE International Symposium on Information Theory, Aachen, Germany, June 2017. The associate editor coordinating the review of this article and approving it for publication was G. Durisi. *(Corresponding author: Changho Suh.)*

Sunghyun Kim was with the Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea. He is now with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02142 USA (e-mail: sunghyun@csail.mit.edu).

Soheil Mohajer is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: soheil@umn.edu).

Changho Suh is with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea (e-mail: chsuh@kaist.ac.kr).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2021.3051392.

Digital Object Identifier 10.1109/TCOMM.2021.3051392

I. Introduction

MOBILE data demands are on the rise at an increasing pace. To meet the growing demands, the wireless communications industry is striving to tap into unexplored spectrum resources in high-frequency bands. One may think that employing established techniques directly into systems to be operating in the new bands will lead to its efficient use. However, a critical challenge arises with high-frequency carriers.

The challenge is that signals conveyed by high-frequency carriers (called millimeter waves, mmWave for short) suffer from high outage probabilities because of their vulnerability to blockage [2]. In dense urban areas such as downtown Manhattan, empirical research on mmWave signals has found that the outage probability is around 0.34 and 0.65 for transmitter-receiver pairs located within 200m and 425m respectively [3]. Worse yet, it could be even lower when they are placed farther apart and also when affordable devices with mediocre reception quality are used. This outage challenge will bring about poor performances unless addressed properly.

To see this expected drawback, let us consider a K-user parallel packet-erasure broadcast channel where the transmitter sends packets in M orthogonal subchannels and the users receive each packet with probability p.[1] In the systems currently deployed in practice, a range of spectrum is divided into multiple orthogonal bands, chunks of which are allocated to users exclusively [4]. Past channel state feedback is used for retransmission purposes to ensure packet delivery. In the channel we consider, when the M subchannels are divided into K distinct sets and each set is allocated to one user, an achievable rate is Mp. This approach leads to poor performances with low p, while it features a low implementation complexity.

This disappointing performance comes from the fact that the scheme employed in the current systems does not exploit multiuser diversity. In contrast, it is exploited by the schemes proposed in prior works [5], [6] for K-user single-input packet-erasure broadcast channels. The schemes harness channel state feedback. From feedback, the transmitter knows which users have received which packets, and generates coded packets that can benefit multiple users at once (details soon to be presented below). Applying this technique to each

[1] In practice, forward error correction protocols are implemented to enable the receiving end to recover some corrupted data. However, error correction cannot be achieved if the level of data corruption is too severe. For simplicity, we abstract this scenario with binary random states (see a detailed modeling in Section II).

subchannel independently, one can verify that an achievable rate is $M\frac{K}{\sum_{i=1}^{K}(1-(1-p)^i)^{-1}}$. There is a significant gain compared to the scheme employed in the current systems. Clearly, packets that are coded so as to benefit multiple users at once carry more information than packets intended for one user. Inspired by this observation that feedback helps to generate a packet useful for multiple users simultaneously, we propose a novel scheme to harness the mmWave bands.

The key element of our scheme is that instead of using the unstable mmWave bands alone, we use them in conjunction with the bands used in the current systems considered stable [7], building on a notion of carrier aggregation that combines discontiguous chunks of bands in serving users. As in the state-of-the-art schemes for the K-user single-input channel [5], [6], we devise our scheme to exploit multiuser diversity. More importantly, inspired by the state-of-the-art scheme for the two-user multi-input channel [8], we take a step further and specifically devise our scheme to exploit channel diversity as well. To this end, we apply coding techniques that take into consideration the fact that orthogonal subchannels have distinctive statistical characteristics. It turns out that by exploiting multiuser diversity together with channel diversity, our scheme provides gains over the state-of-the-art schemes and also outperforms the scheme employed in the current systems by significantly large margins.

A. *Main Contribution*

In this work, we investigate a K-user parallel packet-erasure broadcast channel where M and N orthogonal subchannels susceptible to packet-erasures receive packets with probabilities p and q respectively for $p < q \leq 1$. For the two-user case, our setting is a simplified version of the setting considered by prior work [8] where the capacity region has been characterized for arbitrary channel parameters via linear programming. By focusing on a simplified two-user setting, we develop an explicit achievable scheme, obtain an achievable sum rate in closed-form, and derive a matching upper bound to establish optimality. More importantly, building on explicit constructions of the two-user scheme, we extend our scheme systematically to be applicable for the K-user case. To our knowledge, our work is the first in the literature that develops an achievable scheme for a K-user heterogeneous parallel packet-erasure broadcast channel. As a result, we show that our scheme can provide significant gains over existing K-user schemes. For example, in scenarios where packet-erasures occur frequently in the unstable bands and modestly in the stable bands, e.g., $K = 25$, $M = 1000$, $N = 100$, $p = 0.1$ and $q = 0.8$, the multiplicative gain compared to the K-user scheme employed in the current systems reaches up to around a four-fold gain (see Fig. 9(a)). More importantly, the gain scales with K (see Fig. 9(b)). Moreover, our scheme outperforms a per-subchannel extension of state-of-the-art K-user schemes by large margins, further reducing the optimality gap (see Figs. 9(a) and 9(b)). All these results suggest that it can be significantly useful to employ advanced coding techniques in harnessing mmWave bands to meet ever-growing mobile data demands.

Our proposed scheme harmoniously integrates the ideas of exploiting side information, multiuser diversity, channel diversity, and harnessing channel state feedback. To see this, let us consider a simple two-user case example. Suppose each user has received one undesirable packet intended for the other user: User 1 wishes to decode a but has received b and User 2 wishes to decode b but has received a. Both users keep the "overheard" packets as side information for future use. From channel state feedback, the transmitter is aware of the situation, thus sending $a \oplus b$. If it is received by both users, then each user can decode its own desired packet by exploiting its side information. Note that two packets are decoded in a single transmission. This is called the butterfly effect in the literature. Creating such coding opportunities that exploit multiuser diversity from feedback has been essential in developing the state-of-the-art schemes for the K-user single-input packet-erasure broadcast channel [5], [6].

Exploiting channel diversity can take us one step further. To see this, let us consider one extreme case where the stable subchannels are ideal ($q = 1$). We can take two approaches. One is to send fresh packets (a's and b's) and coded packets ($a \oplus b$'s when such coding opportunities are present) using each subchannel individually. That is, we employ coding on a per-subchannel basis. The other is to send fresh packets using one subchannel and coded packets using another subchannel. That is, we employ coding *across* subchannels. In the first approach, coded packets achieve the butterfly effect only when both users receive the coded packets simultaneously. On the other hand, in the second approach, we can send fresh packets using the unstable subchannels to create coding opportunities and send coded packets using the perfectly stable subchannels to guarantee that both users receive them simultaneously. This is exploiting channel diversity and (in addition to multiuser diversity) it has been essential in developing the state-of-the-art scheme for the two-user multi-input packet-erasure broadcast channel [8].

To the best of our knowledge, our work is the first in the literature to explore a K-user parallel packet-erasure broadcast channel and demonstrates the benefits of applying the ideas of exploiting multiuser diversity and channel diversity by harnessing channel state feedback.

B. *Related Work*

A variety of erasure broadcast channels (BCs) have been investigated in the literature. Dana and Hassibi [9] have characterized the capacity region for a K-user parallel erasure BC without feedback. It has been shown therein that optimality is achieved by time-sharing between the users per each input. Channel state feedback can be useful in devising algorithms that are easy to implement in practice and also in improving their performance. Feedback indeed has proved beneficial in BCs. Georgiadis and Tassiulas [10] have investigated a two-user packet-erasure BC with feedback. They have characterized its capacity region and presented a couple of capacity-achieving schemes therein. Wang and Han [8] have considered a two-user parallel packet-erasure BC with feedback where the packet-erasure probabilities of subchannels can

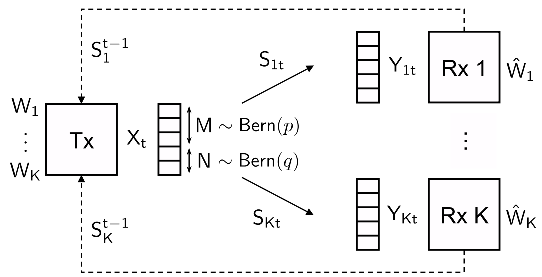

Fig. 1. K-user parallel packet-erasure broadcast channel.

be arbitrary, and established its capacity region in the form of a linear program. To this end, they have derived an outer bound, characterized the capacity region subject to linear network coding, and shown via algebraic arguments that they coincide. This work serves as a state-of-the-art scheme for the two-user and multi-input case. Gatzianas *et al.* [5] and Wang [6] have independently investigated a K-user packet-erasure BC with feedback. They have provided outer bounds for the capacity region and developed coding schemes therein, and established some optimality results under certain conditions on channel characteristics. These works serve as state-of-the-art schemes for the K-user and single-input case. On the other hand, Maddah-Ali and Tse [11] have considered a K-user Gaussian BC with feedback. They have developed a scheme that provides a degrees-of-freedom gain and showed that it is the optimal gain under certain scenarios.

Tracing back to the root of the key idea of our scheme, we find the notion of network coding in single-source multicast networks [12]–[15]. Network coding has since proved useful also in other networked settings, such as wireless networks [16], [17], distributed storage systems [18], and caching systems [19]–[21] to name just a few.

C. Paper Organization

We begin our paper by formally describing our problem setup in Section II. In Section III, we focus on the two-user case where we characterize the sum capacity. We propose two achievable schemes, which we call reactive and proactive schemes (the naming will be clear in Sections III-D and III-E where we describe them in detail), and show that the proactive scheme is optimal. In Section IV, we generalize our reactive scheme to the K-user case. We demonstrate that compared to the conventional scheme currently used in practice, coding across parallel broadcast channels provides a substantial gain that scales with K. We conclude our paper in Section V.

II. Problem Formulation

Fig. 1 describes the K-user parallel packet-erasure broadcast channel. The channel consists of $M + N$ orthogonal packet-erasure 1-to-K broadcast subchannels. We consider a time-slotted system where $M+N$ packets are broadcast from Transmitter to K Users through the subchannels per time-slot.

Let $S_{kt} \in \{0,1\}^{M+N}$ be the states of subchannels from Transmitter to User k at time t, representing packet-erasure events. $S_{kt}(m)$ follows $\text{Bern}(p)$ for $m = 1, \ldots, M$, and $\text{Bern}(q)$ for $m = M+1, \ldots, M+N$, and $p < q \leq 1$. $S_{kt}(m)$ is assumed to be independent over m. Also, S_{kt} is assumed to be i.i.d. over k and t.

Let $X_t \in \mathbb{F}^{M+N}$ be the transmitted packet of Transmitter at time t, where $\mathbb{F}$ is a finite field including all packets that can be represented in binary form in a single time slot. Transmitter generates X_t based on all messages and the feedback of all past channel states: $X_t = f_t(W_1, \ldots, W_K, S_1^{t-1}, \ldots, S_K^{t-1})$ where $S_k^{t-1} := \{S_{ki}\}_{i=1}^{t-1}$. Let $Y_{kt} \in \mathbb{F}^{M+N}$ be the received packet of User k at time t. $Y_{kt}(m) = S_{kt}(m)X_t(m)$ holds.

Over a communication session of T time slots, Transmitter wishes to deliver message $W_k := \{W_{ki} \in \mathbb{F}\}_{i=1}^{TR_k}$ (a collection of packets) reliably to User k, $\forall k = 1, \ldots, K$. User k decodes $\hat{W}_k$ based on all of its received packets and the global channel state information: $\hat{W}_k = g_{kt}(Y_k^t, S_1^t, \ldots, S_K^t)$.[2] An achievable rate region $\mathcal{R}_{\text{inner}}$ includes the set of $(R_1, \ldots, R_K)$ such that for any $\epsilon > 0$, there exists a transmission scheme that satisfies $\Pr(W_k \neq \hat{W}_k) < \epsilon, \forall k = 1, \ldots, K$. The capacity region $\mathcal{C}$ is the closure of all achievable rate regions. In this work, we focus on the sum rate $R_{\text{sum}} = \sum_{k=1}^{K} R_k$ where $(R_1, \ldots, R_K) \in \mathcal{R}_{\text{inner}}$ and the sum capacity $C_{\text{sum}} = \max_{(R_1,\ldots,R_K)\in\mathcal{C}} \sum_{k=1}^{K} R_k$.

Let us elaborate on the rationale behind our modeling of packet-erasures using binary random states. The current systems transfer data in packets. When data-carrying signals are transmitted in unstable wireless channels, some data in the received packet can be corrupted. In practice, various mechanisms, such as forward error correction and retransmission, are implemented to recover the corrupted data. However, in case of severe data corruption, the mechanisms do not work and the received packet results in an irrecoverable decoding-failure. We represent such a decoding-failure as a binary random state with some fixed packet-erasure probability.

We consider mmWave systems as a motivating example. The carrier frequencies in mmWave systems are between 30 and 300 GHz [7]. Assuming a user mobility of 5 to 50 km/h, the Doppler spread ranges from a few to hundreds of kHz. This leads to the coherence time ranging from a few to hundreds of μs, which requires mmWave systems to reduce the scheduling interval of 1 ms in the current systems by orders of magnitude [7]. Given such short coherence times of mmWave systems, it is plausible to model packet-erasures as i.i.d. binary random states.

III. Two-User Case

The work of [8] has characterized the capacity region of two-user parallel packet-erasure broadcast channels where the packet-erasure probabilities of the subchannels can be arbitrary. To obtain their results, they have first derived an outer bound on the capacity region, then established the *linear* network coding capacity region via linear programming, and finally demonstrated via algebraic arguments that it coincides with the derived outer bound.

[2] After we present details of our proposed scheme in Section IV-C, we discuss a method and its associated cost of conveying the global channel state information to all users. We also present a method to bypass the need of conveying it and use a decoder $\hat{W}_k = g_{kt}(Y_k^t)$ instead. See Remark 6 in Section IV-C.

Since our model is a special case of the model investigated in [8], the results of [8] subsume ours for the two-user case presented in this section. However, by focusing on a simplified setting, we seek to provide insights into the problem in a more accessible manner. To this end, (i) we develop a *systematic* scheme that explicitly constructs linear codes without relying on linear programs; and (ii) we obtain *closed-form* expressions. These insights further enable us to extend the proposed schemes for the two-user case to the general K-user case in Section IV.

We present our two-user case results in Section III-A and their proofs in Sections III-D, III-E and III-F. We discuss theoretical implications of the results in Section III-B, which can also be drawn in a different manner by examining the results of prior work [8] (see Appendix B for details).

A. Main Results

We propose two coding schemes. We call them reactive and proactive coding schemes (the naming will be clear in Sections III-D and III-E where we describe the schemes in detail).

Theorem 1 (Reactive Coding): The following sum rate is achievable by the proposed reactive coding scheme:

$$R_{\mathsf{sum}}^{\mathsf{react}} = \min\left\{ \begin{array}{l} \frac{1}{3-q}\big(Mp(4-p-q)+2Nq(2-q)\big), \\ \frac{2(2-p)}{3-p}\,(Mp+Nq) \end{array} \right\}. \tag{1}$$

Proof: See Section III-D. ■

Theorem 2 (Proactive coding): The following sum rate is achievable by the proposed proactive coding scheme:

$$R_{\mathsf{sum}}^{\mathsf{proact}} = \min\left\{ \begin{array}{l} \frac{2}{3-q}\big(Mp(2-p)+Nq(2-q)\big), \\ \frac{2(2-p)}{3-p}\,(Mp+Nq) \end{array} \right\}. \tag{2}$$

Proof: See Section III-E. ■

Also, we derive an upper bound on the sum rate. It turns out to match with the achievable rate by the proactive coding scheme, establishing optimality for the two-user case.

Theorem 3 (Optimality in the Two-User Heterogeneous Channel): The sum capacity for the two-user parallel packet-erasure broadcast channel that consists of M and N orthogonal subchannels whose packet-reception probabilities are p and q respectively where $p < q \leq 1$ is described as follows:

$$C_{\mathsf{sum}} = \min\left\{ \begin{array}{l} \frac{2}{3-q}\big(Mp(2-p)+Nq(2-q)\big), \\ \frac{2(2-p)}{3-p}\,(Mp+Nq) \end{array} \right\}. \tag{3}$$

Proof: Theorem 2 serves as the achievability proof and Section III-F proves the converse. ■

B. Discussion

From the results for the two-user case, we can make an important theoretical observation. Let us first briefly discuss broadcast channels with homogeneity. In the K-user *single-input* packet-erasure broadcast channel, optimality is achieved by state-of-the-art schemes [5], [6], [22] that employ coding over time, which seek to exploit multiuser diversity. In the K-user homogeneous parallel (multi-input) channel, it turns out that optimality can be achieved by employing the state-of-the-art-schemes developed for the single-input channel on a per-subchannel basis, which is applying them to each subchannel independently.

To see this, let us consider a K-user homogeneous parallel channel that consists of M orthogonal subchannels whose packet-reception probability is p. Let C_M be the sum capacity for the parallel channel and C_1 be that for the single-input channel ($M = 1$). Suppose $C_1 < \frac{C_M}{M}$. Consider each set of M packets that an optimal scheme for the parallel channel sends across M subchannels per time slot. In the homogeneous channels, all channel uses are subject to packet-erasures with the same probability p. Hence, in the single-input channel, one can send the same set of M packets using M time slots and achieve a rate of $\frac{C_M}{M}$. This result implies that there exists a scheme that achieves a rate above C_1, which is a contradiction. Therefore, $C_M = MC_1$. It can be shown from [5], [6], [22] that $C_1 = \frac{K}{\sum_{i=1}^{K}(1-(1-p)^i)^{-1}}$. Hence, the sum capacity for the K-user homogeneous parallel channel is:

$$M\frac{K}{\sum_{i=1}^{K}\frac{1}{1-(1-p)^i}}. \tag{4}$$

Note that this result indicates that the *separation principle* holds in the *homogeneous* parallel broadcast channel. In the *heterogeneous* parallel broadcast channel, however, it has been demonstrated by prior works [8], [23] that the separation principle *fails* to hold. We can make the same observation from our closed-form results obtained in a simplified version of the setting considered by [8]. To see this, let us consider the two-user sum capacity characterized by Theorem 3:

$$\min\left\{ \begin{array}{l} \frac{2}{3-q}\big(Mp(2-p)+Nq(2-q)\big), \\ \frac{2}{3-p}\big(Mp(2-p)+Nq(2-p)\big) \end{array} \right\}. \tag{5}$$

In contrast, according to (4), a per-subchannel extension of the state-of-the-art schemes developed for the single-input channel achieves:

$$M\frac{2}{\frac{1}{p}+\frac{1}{1-(1-p)^2}} + N\frac{2}{\frac{1}{q}+\frac{1}{1-(1-q)^2}} = 2\left(\frac{Mp(2-p)}{3-p}+\frac{Nq(2-q)}{3-q}\right). \tag{6}$$

We can easily check that (5) is strictly greater than (6) by showing that the following two always hold:

$$\frac{2}{3-q}\big(Mp(2-p)+Nq(2-q)\big) > 2\left(\frac{Mp(2-p)}{3-p}+\frac{Nq(2-q)}{3-q}\right), \tag{7}$$

$$\frac{2}{3-p}\big(Mp(2-p)+Nq(2-p)\big) > 2\left(\frac{Mp(2-p)}{3-p}+\frac{Nq(2-q)}{3-q}\right). \tag{8}$$

In (7), the first terms on both sides are different due to the factors of $\frac{1}{3-q}$ and $\frac{1}{3-p}$. In (8), the second terms on both sides are different due to the factors of $\frac{2-p}{3-p}$ and $\frac{2-q}{3-q}$. Since $p < q$, $\frac{1}{3-q} > \frac{1}{3-p}$ and $\frac{2-p}{3-p} > \frac{2-q}{3-q}$ hold. Hence, (7) and (8) always hold.

Remark 1: We have relied on our closed-form results obtained in a simplified setting to demonstrate that the separation principle fails to hold with heterogeneity. However, as mentioned, this conclusion has been drawn earlier from the results of prior work [8]. To give due credit, we describe the detailed steps in Appendix B. We also note that the same message has been delivered in a different, yet related setting where a Gaussian MIMO broadcast channel and a noiseless rate-limited multicast channel are present in parallel for two users [24].

C. Numerical Evaluation

To examine the gains attained by our schemes, we consider two existing schemes as baselines. First, we consider the scheme widely employed in practice: the transmitter assigns subchannels to users exclusively, and sends packets for each user only in the subchannels assigned to it. This simple scheme does not exploit multiuser diversity.

$$R_{\mathsf{sum}}^{\mathsf{noFB}} = Mp + Nq. \tag{9}$$

Second, we consider the scheme where the transmitter applies coding over time on a per-subchannel basis to exploit multiuser diversity. This is an extension of the schemes in [5], [6], [22] in a per-subchannel manner (no coding across subchannels).

$$\begin{aligned} R_{\mathsf{sum}}^{\mathsf{sep}} &= M\frac{2}{\frac{1}{p}+\frac{1}{1-(1-p)^2}} + N\frac{2}{\frac{1}{q}+\frac{1}{1-(1-q)^2}} \\ &= 2\left(\frac{Mp(2-p)}{3-p} + \frac{Nq(2-q)}{3-q}\right). \end{aligned} \tag{10}$$

We evaluate the gains attained by our proposed schemes from two perspectives. First, we compare them to the naive scheme to see the benefit of advanced coding schemes. To this end, we consider the following relative gains:

$$\mathsf{Gain}_{\mathsf{react}} = \frac{R_{\mathsf{sum}}^{\mathsf{react}} - R_{\mathsf{sum}}^{\mathsf{noFB}}}{R_{\mathsf{sum}}^{\mathsf{noFB}}} \times 100, \tag{11}$$

$$\mathsf{Gain}_{\mathsf{proact}} = \frac{R_{\mathsf{sum}}^{\mathsf{proact}} - R_{\mathsf{sum}}^{\mathsf{noFB}}}{R_{\mathsf{sum}}^{\mathsf{noFB}}} \times 100. \tag{12}$$

Second, we compare them to the separation scheme in order to see the additional benefit attainable by coding across subchannels. To this end, likewise, we consider the following relative gain:

$$\mathsf{Gain}_{\mathsf{sep}} = \frac{R_{\mathsf{sum}}^{\mathsf{sep}} - R_{\mathsf{sum}}^{\mathsf{noFB}}}{R_{\mathsf{sum}}^{\mathsf{noFB}}} \times 100. \tag{13}$$

Also, we consider two upper bounds on the achievable sum rate. The first is the cut-set bound [25]:

$$\overline{R}_{\mathsf{sum}}^{\mathsf{cut\text{-}set}} = M\left(1-(1-p)^2\right) + N\left(1-(1-p)^2\right). \tag{14}$$

The second is our upper bound. We know from Theorem 3 that it matches with the achievable sum rate by our proactive scheme. It is strictly tighter than the cut-set bound. As we obtain our upper bound by applying Fourier-Motzkin elimination to the outer bound derived in the work of Wang and Han [8], the two bounds coincide (see Section III-F for details). Naming after the authors of the existing bound, we denote our bound as follows:

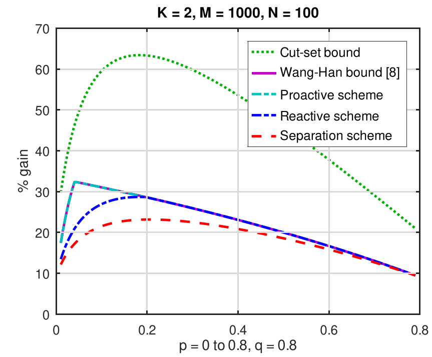

Fig. 2. Gain v.s. p: $K = 2$, $M = 1000$, $N = 100$, $q = 0.8$.

$$\overline{R}_{\mathsf{sum}}^{\mathsf{WH}} = \min\left\{ \begin{array}{l} \frac{2}{3-q}\big(Mp(2-p)+Nq(2-q)\big), \\ \frac{2}{3-p}\big(Mp(2-p)+Nq(2-p)\big) \end{array} \right\}. \tag{15}$$

To illustrate the above two bounds along with $\mathsf{Gain}_{\mathsf{react}}$, $\mathsf{Gain}_{\mathsf{proact}}$ and $\mathsf{Gain}_{\mathsf{sep}}$, we consider the following two bounds on the relative gains:

$$\overline{\mathsf{Gain}}_{\mathsf{cut\text{-}set}} = \frac{\overline{R}_{\mathsf{sum}}^{\mathsf{cut\text{-}set}} - R_{\mathsf{sum}}^{\mathsf{noFB}}}{R_{\mathsf{sum}}^{\mathsf{noFB}}} \times 100, \tag{16}$$

$$\overline{\mathsf{Gain}}_{\mathsf{WH}} = \frac{\overline{R}_{\mathsf{sum}}^{\mathsf{WH}} - R_{\mathsf{sum}}^{\mathsf{noFB}}}{R_{\mathsf{sum}}^{\mathsf{noFB}}} \times 100. \tag{17}$$

Fig. 2 shows our numerical evaluation result. We can see that (1) advanced coding schemes over the simple scheme used in the current systems lead to above 30% multiplicative gains; (2) applying coding across subchannels can achieve optimality, while applying coding in a per-subchannel manner is suboptimal.

It is natural to ask (1) how large the gains by applying advanced coding schemes can be for $K > 2$; and (2) whether the gains can scale with K. We will answer both questions in Section IV-F.

D. Proof of Theorem 1

One can verify that the reactive scheme by Theorem 1 underperforms the proactive scheme by Theorem 2 due to the fact that $4 - p - q \leq 2(2-p)$ for $p \leq q$. Despite its inferior performance compared to the proactive scheme in the two-user case, we here present its details, as our proposed scheme for the general K-user case to be presented in Section IV-C is an extension of it.

As illustrated previously, the transmitter can seek to deliver two packets in a single transmission when each user has received a packet intended for the other (User 1 wants a but has b and User 2 wants b but has a) by sending a coded packet ($a \oplus b$). Our reactive coding scheme builds on this idea.

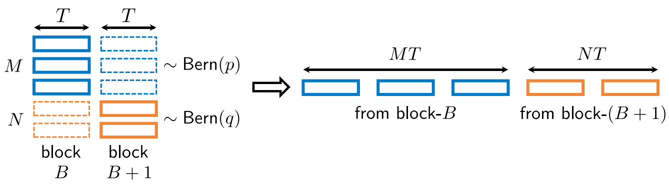

Fig. 3. Graphical illustration that depicts how we interpret channels. We consider the block Markov model.

The transmitter first sends uncoded packets, and then with the help of feedback, reactively sends coded packets to resolve any undecoded packets.

Let us present our reactive coding scheme in detail. There are two types of subchannels. They are different in terms of packet-erasure probability. Subject to packet-erasures, the subchannels receive packets with probability p in one type and q in the other type, where $p < q \leq 1$.

Fig. 3 illustrates how we interpret the channel. We adopt the block Markov model. Each block is of length T, sufficiently large to ensure the law of large numbers to hold. The transmitter exploits channel state feedback to generate coded packets, which opportunistically leads to efficient use of a single channel. More specifically, it exploits feedback from block-B, and uses the packets in block-B to generate coded packets for block-$(B+1)$.

We divide MT unstable channels equally into two sets. Each set of channels is used to send packets intended for each user: a's for User 1 and b's for User 2. Let us consider User 1's perspective. Due to symmetry, the same applies to User 2. Suppose the transmitter sends a packet a using one unstable channel. There are three possible cases:

- User 1 receives a. This case takes place with probability p. User 1 readily decodes it.
- User 2 receives a, but User 1 does not. This case takes place with probability $p(1-p)$. This packet can be coded with a packet that is intended for User 2, but received by User 1 only.
- No user receives a. This case takes place with probability $(1-p)^2$.

Then, the number of coded packets to generate to resolve all uncoded packets that have been sent but received by unintended users is $\frac{MT}{2} \times (1-p)p$.

We use NT stable channels to send the coded packets. Note that, from each user's perspective, receiving any packet during the retransmissions leads to resolving one of the previously sent (but received by the other user) uncoded packets on the user's side. Since the reception probability for each user is q, the number of stable channels required to ensure resolving all such packets is increased by a factor of $\frac{1}{q}$. Hence, the condition under which we are allowed to send uncoded packets in all stable channels is given as follows:

$$\frac{Mp(1-p)}{2q} < N. \tag{18}$$

When the condition (18) is met, we have some stable channels left after resolving all undecoded packets. Since they are homogeneous, using the remaining channels, we achieve $\frac{2q(2-q)}{(3-q)}$ per subchannel according to (4). Hence, the achievable sum rate is as follows:

$$\begin{aligned} &M\left(1-(1-p)^2\right) \times 1 \\ &\quad + \left(N - \frac{Mp(1-p)}{2q}\right) \times \frac{2q(2-q)}{3-q} \\ &= \frac{1}{3-q}\big(Mp(4-p-q) + 2Nq(2-q)\big). \end{aligned} \tag{19}$$

When the condition (18) is unmet, we use a fraction of unstable channels to send uncoded packets in order to guarantee their resolution. The fraction is given as follows:

$$f = \frac{N}{\frac{Mp(1-p)}{2q}} = \frac{2Nq}{Mp(1-p)}. \tag{20}$$

We have some unstable channels left. Since they are homogeneous, using the remaining channels, we achieve $\frac{2p(2-p)}{(3-p)}$ per subchannel according to (4). Hence, the achievable sum rate is as follows:

$$\begin{aligned} &fM\left(1-(1-p)^2\right) \times 1 + (1-f)M \times \frac{2p(2-p)}{3-p} \\ &\qquad = \frac{2(2-p)}{3-p}\left(Mp + Nq\right). \end{aligned} \tag{21}$$

This together with (19) proves Theorem 1.

E. Proof of Theorem 2

As illustrated previously, the transmitter can seek to deliver two packets in a single transmission by sending a coded packet $a \oplus b$ first. When at least one user (say User 1) receives it, the transmitter sends a resolving packet (say b) so that both users can decode their desired packet. Our proactive coding scheme builds on this idea. The transmitter first proactively sends coded packets, and then with the help of feedback, sends uncoded packets to resolve any undecoded packets. We use the same channel interpretation as in Section III-D and Fig. 3. However, in the proactive coding scheme, the order of sending uncoded and coded packets is reversed. We first send coded packets using unstable subchannels, and then send uncoded packets using stable subchannels. The idea of sending mixed information first and refining it later has also appeared in prior works [26]–[28] where it is called the poison-antidote approach, pre-mixing, and proactive coding respectively.

Let us present our proactive coding scheme in detail. We consider three cases: coded packets may be received by both users (*Case 1* below), by one user only (*Case 2* and *Case 3* below), or by none.

- *Case 1:* $a \oplus b$ is received by both users. We send a resolving packet a (by symmetry, b also works).
  - (i) a can be received by both users. Both users decode their desired packets. User 1 readily decodes a and User 2 decodes b by exploiting side information $a \oplus b$.
  - (ii) a can be received by User 1. User 1 can decode a and has b on the side (with previously received $a \oplus b$). User 2 has only $a \oplus b$. It additionally needs b (or a instead).
  - (iii) a can be received by User 2. User 2 can decode b and has a on the side (with previously received

$a \oplus b$). User 1 has only $a \oplus b$. It additionally needs a (or b instead).

Consider (ii) and (iii). For clarity, let the packets in question be a_2, b_2 in (ii) and a_3, b_3 in (iii). In (ii), User 2 needs b_2 (or a_2), and b_2 is available at User 1 as side information. In (iii), User 1 needs a_3 (or b_3), and a_3 is available at User 2 as side information. Hence, the retransmission protocol is to send $b_2 \oplus a_3$. If received by both users, resolving of previously sent $a_2 \oplus b_2$ or $a_3 \oplus b_3$ is done at both users. If not, we proceed in a similar manner encoding resolving packets by grouping two symmetrical occurrences: one such retransmission received by User 1 only as in (ii) above and another received by User 2 only as in (iii) above.

- *Case 2:* $a \oplus b$ is received by User 1 only. We send a resolving packet b. This intends to send b directly to User 2 and at the same time to help User 2 decode a.
  - (i) b is received by both users. Both users decode their desired packets.
  - (ii) b is received by User 1. User 1 can decode a and has b on the side. User 2 has nothing.
  - (iii) b is received by User 2. User 2 can decode b. User 1 still has $a \oplus b$ unresolved.

  Consider (ii) and (iii). For clarity, let the packets in question be a_2, b_2 in (ii) and a_3, b_3 in (iii). In (ii), User 2 needs b_2, and b_2 is available at User 1 as side information. In (iii), User 1 needs a_3, which can also be decoded by receiving b_3, and b_3 is available at User 2 as decoded information. Hence, the retransmission protocol is to send $b_2 \oplus b_3$. As in *Case 1*, we proceed until resolving of previously sent $a_2 \oplus b_2$ or $a_3 \oplus b_3$ is done at both users.
- *Case 3:* $a \oplus b$ is received by User 2 only. Due to symmetry, by reversing the roles of Users 1 and 2, the same retransmission protocol applies as in *Case 2*.

Note that, from each user's perspective, receiving any packet during the retransmissions leads to resolving one of the previously sent coded packets on the user's side. Since the reception probability for each user is q, the number of stable channels required to ensure resolving of all coded packets is increased by a factor of $\frac{1}{q}$. Hence, the condition under which we are allowed to send coded packets in all unstable channels is given as follows:

$$\frac{M\left(1-(1-p)^2\right)}{q} < N. \tag{22}$$

When the condition (22) is met, we have some stable channels left after resolving all coded packets. Since they are homogeneous, using the remaining channels, we achieve $\frac{2q(2-q)}{(3-q)}$ per subchannel according to (4). Hence, the achievable sum rate is as follows:

$$\begin{aligned} &M\left(1-(1-p)^2\right) \times 2 \\ &\quad + \left(N - \frac{M\left(1-(1-p)^2\right)}{q}\right) \times \frac{2q(2-q)}{3-q} \\ &= \frac{2}{3-q}\big(Mp(2-p) + Nq(2-q)\big). \end{aligned} \tag{23}$$

When the condition (22) is unmet, we use a fraction of unstable channels to send coded packets in order to guarantee their resolution. The fraction is given as follows:

$$f = \frac{N}{\frac{M(1-(1-p)^2)}{q}} = \frac{Nq}{M\left(1-(1-p)^2\right)}. \tag{24}$$

We have some unstable channels left. Since they are homogeneous, using the remaining channels, we achieve $\frac{2p(2-p)}{(3-p)}$ per subchannel according to (4). Hence, the achievable sum rate is as follows:

$$\begin{aligned} fM\left(1-(1-p)^2\right) \times 2 + (1-f)M \times \frac{2p(2-p)}{3-p} \\ = \frac{2(2-p)}{3-p}\left(Mp + Nq\right). \end{aligned} \tag{25}$$

This together with (23) proves Theorem 2.

Remark 2: In the literature, the idea to kill multiple birds with one stone is often called the butterfly effect. One packet is in effect worth multiple packets per transmission upon reception at multiple users. In all prior works [5], [6], [11], [22] that motivate our scheme, the butterfly effect plays a central role. The work of [11] has devised a scheme to achieve such gainful effects in a Gaussian broadcast channel. The works of [5], [6], [22] have sought to benefit from such effects by applying coding over time in a single-input packet-erasure broadcast channel. Due to its large influence, the butterfly effect can be found in a variety of contexts: caching [19]–[21], cache-aided networks [22], [29], [30], network coding [12], [13], [17], [18], delayed channel state information at the transmitter [5], [6], [11], feedback [31] and bursty channels with relays [32], [33] to name a few. In all contexts, exploiting side information is at the heart of notable gains.

Remark 3: In the reactive scheme, we send fresh uncoded packets first and then send coded packets intended for ensuring delivery of the uncoded packets that have not been received by the intended users. In contrast, in the optimal scheme, for some fraction of channels, we send coded packets first and then send uncoded packets intended for resolving the coded packets that cannot be immediately decoded even when received by the intended users (proactive coding). This shows that a balance between reactive coding and proactive coding is essential to achieving optimality. The idea of sending mixed information first and refining it later can also be found in prior works [26]–[28].

F. Proof of Theorem 3

We establish that our proactive coding scheme is optimal. To prove this, we adopt the outer bound to be derived in Section IV-E, which is applicable for $K \geq 2$. According to this bound, the capacity region for the two-user case is bounded by the intersection of the following outer bounds for (R_1, R_2):

$$R_1 \leq \sum_{m=1}^{M+N} R_1^{(\ell)}(m), \;\; R_2 \leq \sum_{m=1}^{M+N} R_2^{(\ell)}(m), \tag{26}$$

where $\ell \in \{1,2\}$, and $R_1^{(\ell)}(m)$ and $R_2^{(\ell)}(m)$ are subject to:

$$R_1^{(\ell)}(m) \geq 0, \text{ for } m \in [1:M+N], \quad R_2^{(\ell)}(m) \geq 0, \text{ for } m \in [1:M+N], \tag{27}$$

$$\frac{R_1^{(1)}(m)}{p} + \frac{R_2^{(1)}(m)}{1-(1-p)^2} \leq 1 \text{ for } m \in [1:M], \quad \frac{R_1^{(1)}(m)}{q} + \frac{R_2^{(1)}(m)}{1-(1-q)^2} \leq 1 \text{ for } m \in [M+1:M+N], \tag{28}$$

$$\frac{R_2^{(2)}(m)}{p} + \frac{R_1^{(2)}(m)}{1-(1-p)^2} \leq 1 \text{ for } m \in [1:M], \quad \frac{R_2^{(2)}(m)}{q} + \frac{R_1^{(2)}(m)}{1-(1-q)^2} \leq 1 \text{ for } m \in [M+1:M+N]. \tag{29}$$

We can see that the number of inequalities can be arbitrarily large depending on the channel parameters M and N. However, we can show that to find the maximum bound on R_1+R_2, it suffices to search for the case where $R_k^{(\ell)}(m)$'s are equal to, namely, $R_{k,p}^{(\ell)}$ for $m \in [1:M]$ and $R_{k,q}^{(\ell)}$ for $m \in [M+1:M+N]$. This reduces the number of inequalities to a constant regardless of the channel parameters. To show this, we exploit symmetry. Let us consider $\ell = 1$, $M = 2$ and $N = 1$. Suppose the maximum bound on $R_1 + R_2$ is obtained, omitting ℓ for simplicity, when $(R_1(1), R_2(1)) =: \boldsymbol{R}(1)$, $(R_1(2), R_2(2)) =: \boldsymbol{R}(2)$, $R_1(1) \neq R_1(2)$, and $R_2(1) \neq R_2(2)$. In words, for each user, the individual sub-rate is different across the two subchannels with the same packet-erasure probability. Now, let us consider $\boldsymbol{R} := \frac{1}{2}\boldsymbol{R}(1) + \frac{1}{2}\boldsymbol{R}(2)$ and let $(R_1'(1), R_2'(1)) = (R_1'(2), R_2'(2)) = \boldsymbol{R}$. These new equal sub-rate pairs of the two subchannels still satisfy (28) and lead to the same maximum bound on $R_1 + R_2$ according to (26). For $M \geq 2$ and $N = 1$, we can apply this argument to any number of different sub-rate pairs for $m \in [1:M]$ that leads to the maximum bound on R_1+R_2, by successively considering two such subchannels at a time. We can generalize this argument for any channel parameters M and N. This implies that we can always find the maximum bound on R_1+R_2 by searching only for the case where $R_k(m)$'s are equal to, namely, $R_{k,p}$ for $m \in [1:M]$, and $R_{k,q}$ for $m \in [M+1:M+N]$. This leads to one identical inequality for all $m \in [1:M]$, and another for all $m \in [M+1:M+N]$. Hence, we obtain (32) from (28). Similarly for $\ell = 2$, we obtain (33) from (29).

$$R_1 \leq MR_{1,p}^{(\ell)} + NR_{1,q}^{(\ell)}, \;\; R_2 \leq MR_{2,p}^{(\ell)} + NR_{2,q}^{(\ell)}, \tag{30}$$

where $R_{1,p}^{(\ell)}$, $R_{1,q}^{(\ell)}$, $R_{2,p}^{(\ell)}$, and $R_{2,q}^{(\ell)}$ are subject to:

$$R_{1,p}^{(\ell)} \geq 0, \;\; R_{1,q}^{(\ell)} \geq 0, \;\; R_{2,p}^{(\ell)} \geq 0, \;\; R_{2,q}^{(\ell)} \geq 0, \tag{31}$$

$$\frac{R_{1,p}^{(1)}}{p} + \frac{R_{2,p}^{(1)}}{1-(1-p)^2} \leq 1, \;\; \frac{R_{1,q}^{(1)}}{q} + \frac{R_{2,q}^{(1)}}{1-(1-q)^2} \leq 1, \tag{32}$$

$$\frac{R_{2,p}^{(2)}}{p} + \frac{R_{1,p}^{(2)}}{1-(1-p)^2} \leq 1, \;\; \frac{R_{2,q}^{(2)}}{q} + \frac{R_{1,q}^{(2)}}{1-(1-q)^2} \leq 1. \tag{33}$$

Finally, to obtain the closed-form expression in Theorem 3, we carry out Fourier-Motzkin elimination to the system of linear inequalities (30)–(33). We present the details in Appendix A.

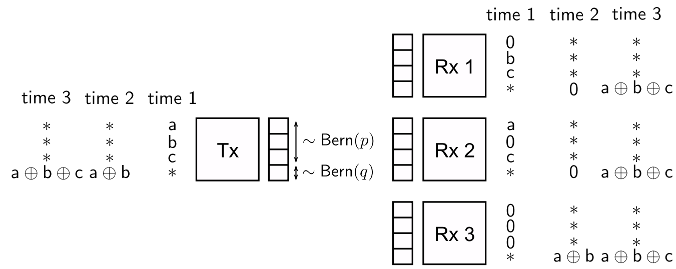

Fig. 4. A three-user example. Three subchannels are prone to packet-erasures hence receive packets with probability p and one subchannel is also prone to packet-erasures hence receives packets with probability q for $p < q < 1$. To highlight elements that matter, some packets are not specified. Time 1 corresponds to block-B, and Times 2 and 3 correspond to block-$(B+1)$ based on the block Markov model (see Fig. 3).

IV. K-User Case

A. Illustrative Example

Let us begin with illustrative examples for $K = 3$, before we extend the reactive coding scheme to the general K-user case. Fig. 4 highlights key elements.

- *Time 1*: the transmitter sends three fresh packets, one intended for each user, in the three unstable subchannels. A particular event takes place: none are received by the intended users. a and b are received by one unintended user, and c is received by two. For later use, they store the received packets as side information. Assuming channel state feedback is available, the transmitter knows which packets have been received by which users.
- *Time 2*: the transmitter sends a *coded* packet $a \oplus b$ in the stable subchannel. Notice that if the coded packet is received by User 1 and 2, it will lead to *two packets decoded per transmission*. User 1 decodes a by exploiting its side information b, and User 2 decodes b likewise. Unfortunately, it is received by User 3 only, and not received by the intended Users 1 and 2. For later use, User 3 stores its received packet as side information.
- *Time 3*: the transmitter sends a different coded packet $a \oplus b \oplus c$ in the stable subchannel. Notice that if the coded packet is received by Users 1, 2 and 3, it will lead to *three packets decoded per transmission*. Fortunately, the coded packet is received by all users. The users decode their desired packet by exploiting their side information.

Note two key coding strategies at play.

- The first is to seek maximal gains attainable by exploiting side information. At any moment, the transmitter generates coded packets in a way that they can benefit as many users as possible per transmission. At *Time 2*, it attempts to achieve two packets decoded at once by sending $a \oplus b$. At *Time 3*, it attempts to achieve three packets decoded at once by sending $a \oplus b \oplus c$.
- The second is to send coded packets, which are worth multiple packets in effect, in *stable* subchannels. The reason is clear. It is to increase the odds of delivery of the coded packets that are more valuable than fresh packets.

Remark 4: The first strategy can be found in prior work [5], [6], [11], [22] by which our reactive coding scheme is motivated. In particular, the work of [11] has devised a systematic way of coding packets by exploiting channel state feedback at the transmitter, introducing a notion of *order*. Packets are order-i when we can combine i of them, each intended for a distinct user, into a single packet that can benefit up to all i users at once. In the example, a and b are once order-2, and later become order-3 together with c. The order of a packet can change over time, and such transitions are analyzed in [11]. The second strategy can be linked to decentralized caching [20] operating in two phases. In the placement phase, without coordination, the users store random fragments of other users' contents in their cache for later use. In the delivery phase, encoded contents are transmitted in a shared link and multiple users can decode their desired contents by exploiting the side information (network coding [12], [13] being at play, which can also be found in other problems [17]–[19]). Our scheme can be viewed as the two phases taking place concurrently across multiple orthogonal subchannels. While placement (occurring randomly) is done mostly in unstable subchannels, delivery (aimed to exploit side information) is done mostly in stable subchannels.

B. Detailed Look at the Example

Building on the key elements presented in the example, we describe our scheme for the special case of $K = 3$ in further detail. This will make Section IV-C for the general K-user case more accessible. We use the same channel interpretation adopting the block Markov model illustrated in Fig. 3. Our strategy is to send as many high-order coded packets as possible in stable channels to increase the number of packets decoded per transmission. We have MT unstable channels and NT stable channels in total. The stable channels follow the unstable channels in the time domain. We further divide the $(M+N)T$ channels into three. We call the time-blocks of channels block-1, block-2 and block-3 respectively. T_i channels are in block-i. $T_1+T_2+T_3 = (M+N)T$ holds. The time-blocks are laid out in ascending order of indices: block-$(i+1)$ is available later than block-i. The transmitter exploits channel state feedback from block-i and uses the packets for block-i to generate coded packets for block-$(i+1)$.

Now, we explain how coded packets are generated. For clarity, we introduce some notation. Let $\mathcal{E}_{k,\mathcal{U}}$ be the set of packets that User k wants to receive and that the users in $\mathcal{U}$ can eventually cancel out in their received packets at the end of using all $(M+N)T$ channels. In block-1, the transmitter divides T_1 channels equally into three and uses each to send fresh packets intended for each user. These packets are initially in $\mathcal{E}_{k,\varnothing}$ for $k = 1, 2, 3$. For User 1 (by symmetry, the same applies to Users 2 and 3), on average (assuming the unstable channels are used),

- $\frac{T_1}{3} \times p$ fresh (order-1) packets are immediately decoded at User 1. They are discarded from $\mathcal{E}_{1,\varnothing}$.
- $\frac{T_1}{3} \times p(1-p)^2$ packets (expected size of $\mathcal{E}_{1,\{2\}}$ and $\mathcal{E}_{1,\{3\}}$) are received by one unintended user, and not by User 1. We can combine two such packets, each intended for a distinct user (say a from $\mathcal{E}_{1,\{2\}}$ and b from $\mathcal{E}_{2,\{1\}}$), and encode them into a single coded packet (say $a \oplus b$) that can benefit all two users. Hence, we call these packets order-2 packets intended for User 1. We add these packets to $\mathcal{E}_{1,\{2\}}$ and $\mathcal{E}_{1,\{3\}}$, and remove them from $\mathcal{E}_{1,\varnothing}$.
- $\frac{T_1}{3} \times p^2(1-p)$ packets (expected size of $\mathcal{E}_{1,\{2,3\}}$) are received by two unintended users, and not by User 1. We can combine three such packets (say $a \oplus b \oplus c$ where $a \in \mathcal{E}_{1,\{2,3\}}$, $b \in \mathcal{E}_{2,\{3,1\}}$ and $c \in \mathcal{E}_{3,\{1,2\}}$), and encode them into a single coded packet (say $a \oplus b \oplus c$) that can benefit all three users. Hence, we call these packets in $\mathcal{E}_{1,\{2,3\}}$ order-3 packets intended for User 1. We add these packets to $\mathcal{E}_{1,\{2,3\}}$, and remove them from $\mathcal{E}_{1,\varnothing}$.
- $\frac{T_1}{3} \times (1-p)^3$ packets are received by none of the users. They will be retransmitted. They remain as order-1 packets in $\mathcal{E}_{1,\varnothing}$.

At the end of block-1, from channel state feedback, the transmitter can know which packets have been received by which users, and update $\mathcal{E}_{1,\{2\}}$, $\mathcal{E}_{1,\{3\}}$ and $\mathcal{E}_{1,\{2,3\}}$ for User 1 (and likewise for Users 2 and 3). Assuming T_1 is large enough to accommodate all $|\mathcal{E}_{1,\varnothing}|$ packets, taking into account the packets that need to be retransmitted, all packets in $\mathcal{E}_{1,\varnothing}$ are either moved to one of $\mathcal{E}_{1,\{2\}}$, $\mathcal{E}_{1,\{3\}}$ and $\mathcal{E}_{1,\{2,3\}}$, or simply discarded due to successful decoding.[3]

At the beginning of block-2, the transmitter generates coded packets using the packets in $\mathcal{E}_{1,\{2\}}$, $\mathcal{E}_{1,\{3\}}$, and the like for Users 2 and 3 (namely, $\mathcal{E}_{2,\{1\}}$, $\mathcal{E}_{2,\{3\}}$, $\mathcal{E}_{3,\{1\}}$, and $\mathcal{E}_{3,\{2\}}$). Referring to $\mathcal{E}_{1,\{2\}}$ and $\mathcal{E}_{2,\{1\}}$, it can identify the packets intended for User 1 but received by User 2 (a's) and vice versa (b's). Then, in block-2, it attempts to send the XORs of the packets in $\mathcal{E}_{1,\{2\}}$ and $\mathcal{E}_{2,\{1\}}$ (collecting one from each set, say a and b, and generating one $a \oplus b$). If the XOR-coded packets are received by Users 1 and 2, both users can decode their desired packets by exploiting side information properly. This leads to two packets decoded per transmission. By symmetry, the same can happen for the other pairs of users.

At the beginning of block-3, the transmitter can apply similar techniques to the packets in $\mathcal{E}_{1,\{2,3\}}$, $\mathcal{E}_{2,\{3,1\}}$, and $\mathcal{E}_{3,\{1,2\}}$. Recall that at the end of block-1, from channel state feedback, it can identify the packets intended for User 1 but received by Users 2 and 3 (a's) and the like (b's at Users 3 and 1, and c's at Users 1 and 2), and send coded packets $a \oplus b \oplus c$'s. This can lead to three packets decoded per transmission.

So far, we have considered neat scenarios where coded packets are received by all intended users simultaneously. However, it is yet unclear how we should handle scenarios where coded packets generated are not received by all intended users simultaneously. For example, Users 1 and 2 may not receive coded packets $a \oplus b$'s in block-2, and only User 3, an unintended user, may receive them. Let us consider such a case where a coded packet intended for Users 1 and 2, say $a \oplus b$, is received only by User 3. Relying on symmetry, let us further consider the two other cases where $b \oplus c$ is received

[3] In a strict sense, except for the initialized $\mathcal{E}_{k,\varnothing}$ for all k, the cardinality of any set $\mathcal{E}$ is random. However, for simplicity, we resort to the law of large numbers and consider the cardinality of any set $\mathcal{E}$ to be equal to its expectation. We omit the expectation notation in the paper.

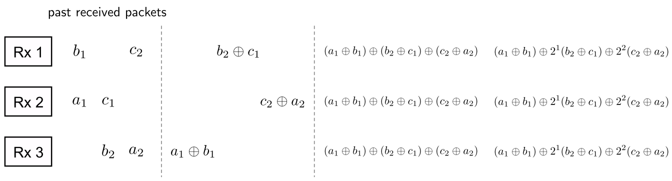

Fig. 5. Coded packets generated from order-2 packets become order-3 packets. Each coded packet is received by one additional unintended user. For example, $a_1 \oplus b_1$ for Users 1 and 2 is received by User 3. It is wanted by Users 1 and 2. $a_1 \oplus b_1$ belongs to $\mathcal{E}_{1,\{2,3\}}$ and $\mathcal{E}_{2,\{3,1\}}$. There exist two other symmetric packets $b_2 \oplus c_1$ belonging to $\mathcal{E}_{2,\{3,1\}}$ and $\mathcal{E}_{3,\{1,2\}}$, and $c_2 \oplus a_2$ belonging to $\mathcal{E}_{3,\{1,2\}}$ and $\mathcal{E}_{1,\{2,3\}}$. One order-3 packet from each of $\mathcal{E}_{1,\{2,3\}}$, $\mathcal{E}_{2,\{3,1\}}$ and $\mathcal{E}_{3,\{1,2\}}$ is chosen to generate a coded packet to benefit all three users.

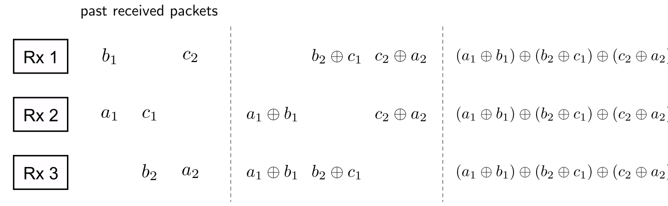

Fig. 6. Coded packets generated from order-2 packets become order-3 packets. Each coded packet is received by one additional unintended user. For example, $a_1 \oplus b_1$ for Users 1 and 2 is received by User 3. It is also received by User 2. Hence, it is wanted by User 1 only. $a_1 \oplus b_1$ belongs to $\mathcal{E}_{1,\{2,3\}}$. There exist two other symmetric packets $b_2 \oplus c_1$ belonging to $\mathcal{E}_{2,\{3,1\}}$, and $c_2 \oplus a_2$ belonging to $\mathcal{E}_{3,\{1,2\}}$. One order-3 packet from each of $\mathcal{E}_{1,\{2,3\}}$, $\mathcal{E}_{2,\{3,1\}}$ and $\mathcal{E}_{3,\{1,2\}}$ is chosen to generate a coded packet to benefit all three users.

only by User 1 and $c \oplus a$ is received only by User 2. For clarity, let us denote the coded packets by $a_1 \oplus b_1$, $b_2 \oplus c_1$ and $c_2 \oplus a_2$ respectively. Fig. 5 illustrates the scenario.

From User 1's perspective, it wants to receive $a_1 \oplus b_1$ and $c_2 \oplus a_2$ because it has b_1 and c_2. Note that User 1 also has $b_2 \oplus c_1$. Hence, the transmitter sends two XOR-sums, say, $(a_1 \oplus b_1) \oplus (b_2 \oplus c_1) \oplus (c_2 \oplus a_2)$ and $(a_1 \oplus b_1) \oplus 2^1(b_2 \oplus c_1) \oplus 2^2(c_2 \oplus a_2)$. Upon receiving them, User 1 can cancel out $b_2 \oplus c_1$ in them and decode a_1 and a_2 by exploiting side information. Similarly, User 2 and User 3 can decode b_1, b_2 and c_1, c_2 respectively.

The coded packet generation process is seemingly complicated. It is, however, conceptually simple when we view it as follows. The key is to keep establishing equivalency. Let us focus on User 1. It initially wants to decode a_1 and a_2. But in block-1, they are received by unintended users, and User 1 also receives undesired packets intended for other users (b_1 and c_2). Now, for User 1, receiving $a_1 \oplus b_1$ and $c_2 \oplus a_2$ is equivalent to decoding a_1 and a_2 directly, since it can cancel b_1 and c_2 out. Hence, in block-2, we may think as if User 1 wants $a_1 \oplus b_1$ and $c_2 \oplus a_2$, not a_1 and a_2. Let us denote $a_1 \oplus b_1$ and $c_2 \oplus a_2$ as A_1 and A_2. But unfortunately, in block-2, A_1 and A_2 are received by unintended users again, and User 1 also receives an undesired packet $b_2 \oplus c_1$, which we denote by B_2. User 1 needs to receive at least two extra packets to decode A_1 and A_2. Using the fact that User 1 can cancel B_2 out, in block-3, the transmitter generates two XOR-sums using A_1, A_2 and B_2, and sends them. In essence, receiving one XOR-sum is equivalent to decoding A_1 or A_2, hence we may think as if User 1 wants the two XOR-sums, not A_1 and A_2. Owing to symmetry, this coded packet generation process benefits all users equally, not User 1 only.

To understand it systematically, we need to describe it resorting to the definition of $\mathcal{E}$'s. $a_1 \oplus b_1$ is generated from a_1 in $\mathcal{E}_{1,\{2\}}$ and b_1 in $\mathcal{E}_{2,\{1\}}$, and is sent in block-2. Since User 1 can decode a_1 by receiving $a_1 \oplus b_1$, $a_1 \oplus b_1$ is intended for User 1. Since User 2 can decode b_1 by receiving $a_1 \oplus b_1$, $a_1 \oplus b_1$ is also intended for User 2. However, it is received by User 3. User 2 does not yet have $a_1 \oplus b_1$ per se. However, assuming User 2 will eventually receive $a_1 \oplus b_1$ to decode b_1, we can consider it to be available at User 2. Hence, we add $a_1 \oplus b_1$ to $\mathcal{E}_{1,\{2,3\}}$. For the same reason, we also add it to $\mathcal{E}_{2,\{3,1\}}$. Now, we discard a_1 in $\mathcal{E}_{1,\{2\}}$ and b_1 in $\mathcal{E}_{2,\{1\}}$ because they will be handled, in different forms, in $\mathcal{E}_{1,\{2,3\}}$ and $\mathcal{E}_{2,\{3,1\}}$. This conceptually corresponds to our aforementioned discussion where User 1 begins to want A_1 and A_2 instead of a_1 and a_2. The same holds for $b_2 \oplus c_1$ and $c_2 \oplus a_2$. To summarize, at the end of block-2, $\{a_1 \oplus b_1, c_2 \oplus a_2\} \subset \mathcal{E}_{1,\{2,3\}}$, $\{b_2 \oplus c_1, a_1 \oplus b_1\} \subset \mathcal{E}_{2,\{3,1\}}$ and $\{c_2 \oplus a_2, b_2 \oplus c_1\} \subset \mathcal{E}_{3,\{1,2\}}$. Also, $a_1 \notin \mathcal{E}_{1,\{2\}}$, $a_2 \notin \mathcal{E}_{1,\{3\}}$, $b_1 \notin \mathcal{E}_{2,\{1\}}$, $b_2 \notin \mathcal{E}_{2,\{3\}}$, $c_1 \notin \mathcal{E}_{3,\{2\}}$ and $c_2 \notin \mathcal{E}_{3,\{1\}}$. Now, it is clear why we send two XOR-sums of the same three packets in block-3: $(a_1 \oplus b_1) \oplus (b_2 \oplus c_1) \oplus (c_2 \oplus a_2)$ and $(a_1 \oplus b_1) \oplus 2^1(b_2 \oplus c_1) \oplus 2^2(c_2 \oplus a_2)$. We collect one packet from each of $\mathcal{E}_{1,\{2,3\}}$, $\mathcal{E}_{2,\{3,1\}}$ and $\mathcal{E}_{3,\{1,2\}}$, and encode them into a single XOR-sum. Since each of the three packets can be in multiple bins, as in this scenario, it is possible that the same set of three packets is collected for generating different XOR-sums. In this case, we construct distinct XOR-sums to enable the users to decode their desired packets (see Remark 5 below). As in block-1, assuming T_2 is large enough to accommodate all packets in $\mathcal{E}_{1,\{2\}}$, $\mathcal{E}_{1,\{3\}}$ and the like, taking into account some necessary retransmissions, all such packets are either moved to $\mathcal{E}_{1,\{2,3\}}$ and the like, or simply discarded due to successful delivery.

Yet in another scenario, coded packets intended for Users 1 and 2 may be received by one additional unintended user and one intended user. That is, $a \oplus b$ is received by Users 2 and 3, but not by User 1. Similar techniques exploiting symmetry are applicable. To see this, let us consider $a_1 \oplus b_1$, $b_2 \oplus c_1$ and $c_2 \oplus a_2$. $a_1 \oplus b_1$ is received by Users 2 and 3, $b_2 \oplus c_1$ by Users 3 and 1 and $c_2 \oplus a_2$ by Users 1 and 2. Fig. 6 illustrates the scenario. At the end of block-2, $a_1 \oplus b_1 \in \mathcal{E}_{1,\{2,3\}}$, $b_2 \oplus c_1 \in \mathcal{E}_{2,\{3,1\}}$ and $c_2 \oplus a_2 \in \mathcal{E}_{3,\{1,2\}}$. We collect one from each set and encode one coded packet (XOR-sum) that can benefit all three users. Upon reception of the XOR-sum, each user can decode its desired packet.

We finally note that we keep sending coded packets generated by combining three order-3 packets, each of which in $\mathcal{E}_{1,\{2,3\}}$, $\mathcal{E}_{2,\{3,1\}}$ and $\mathcal{E}_{3,\{1,2\}}$ respectively, until they are received by all three users. This is because they cannot be promoted to higher-order packets and handled later. This persistent retransmission ensures decoding of all fresh packets in $\mathcal{E}_{1,\varnothing}$, $\mathcal{E}_{2,\varnothing}$ and $\mathcal{E}_{3,\varnothing}$. T_3 should be large enough to ensure it.

Remark 5: Let us describe in greater detail the process of generating XOR-sums that can be generalized to arbitrary numbers of users. We apply *bit-shifting* in addition to XOR-ing (which corresponds to linear combinations in $\mathbb{F}$). Packets are bit strings, thus multiplication of 2^n to a packet corresponds to shifting of the bits in the packet to the left by n bits. The transmitter generates XOR-sums in the form of $\oplus_{k\in\mathcal{U}} s_k x_k$, where s_k's are bit-shifts and $x_k \in \mathcal{E}_{k,\mathcal{U}\setminus k}$. Choice of s_k's is subject to two main constraints. To present them, let us revisit the scenario in Fig. 5. Suppose we choose $a_1 \oplus b_1 \in \mathcal{E}_{1,\{2,3\}}$, $a_1 \oplus b_1 \in \mathcal{E}_{2,\{3,1\}}$ and $c_2 \oplus a_2 \in \mathcal{E}_{3,\{1,2\}}$, and naively generate an XOR-sum in block-3: $(a_1 \oplus b_1) \oplus (a_1 \oplus b_1) \oplus (c_2 \oplus a_2) = c_2 \oplus a_2$. Then, the information of $a_1 \oplus b_1$ is completely lost. Consequently, the scheme does not work as intended. Thus, we need to choose s_k's so as not to lose the information of any x_k in $\oplus_{k\in\mathcal{U}} s_k x_k$. Also, suppose we generate $(a_1 \oplus b_1) \oplus 2^1(b_2 \oplus c_1) \oplus (c_2 \oplus a_2)$ instead of $(a_1 \oplus b_1) \oplus 2^1(b_2 \oplus c_1) \oplus 2^2(c_2 \oplus a_2)$ in block-3 (2^2 is missing). Then, User 1 cannot decode its desired packets a_1 and a_2. Thus, we need to choose s_k's in such a way that all users can decode their desired packet(s) using their side information. One simple method to meet the two aforementioned requirements is to choose *distinct* s_k's for each $k \in \mathcal{U}$, say, from $\{1, 2^1, \ldots, 2^{|\mathcal{U}|-1}\}$. This method ensures that the information of each x_k is not lost in $\oplus_{k\in\mathcal{U}} s_k x_k$ by accident as an unintended consequence of XOR-ing. Also, provided that User $\ell \in \mathcal{U}$ knows s_k's, User ℓ can reconstruct the undesired part ($\oplus_{k\in\mathcal{U}\setminus\ell} s_k x_k$) by using the bit-shifts (s_k's for $k \in \mathcal{U}\setminus\ell$) and side information ($x_k$'s for $k \in \mathcal{U}\setminus\ell$), cancel it out, and decode the desired packet (x_ℓ). From a practical perspective, it is necessary to communicate this control information (s_k's) between the transmitter and receivers to make the proposed scheme work in a robust manner. In Remark 6 in Section IV-C, we present two methods to accomplish it and also discuss their associated costs which can be made negligible. Prior works have employed techniques such as maintaining exponentially many sub-queues [34], and having a sufficiently large finite field and employing linear coding [6], [8] to develop concrete coded packet generation processes. Although the employed techniques are slightly different, the principles of generating coded packets are essentially the same. Thus, we omit a rigorous proof of correctness for our process, which follows similar lines of arguments presented in [6], [8], [34]. In the following sections, we omit s_k's for the sake of brevity, yet any XOR-sum implies that proper bit-shifting is applied.

C. Achievable Scheme

To generalize the three-user examples to the K-user case with concreteness, we introduce some additional definitions. We let $[K] := \{1, 2, \ldots, K\}$. Extending the definition of $\mathcal{E}_{k,\mathcal{U}}$, we let $\mathcal{E}_{\mathcal{U},\mathcal{V}}$, where $\mathcal{U}\cap\mathcal{V} = \varnothing$, be the set of packets that the users in $\mathcal{U} \subseteq [K]$ want to receive and also that the users in $\mathcal{V} \subset [K]$ will eventually be able to cancel out in their received packets. Also, we extend our channel interpretation, previously illustrated in Fig. 3. It is now presented in further detail in Figs. 7 and 8.

Our reactive coding scheme for general K is a natural extension of that for $K = 3$ explained in Section IV-B. As shown in Figs. 7 and 8, we use MT unstable channels and NT stable channels in total. They collectively constitute one superblock (superblock-b) and coding techniques are applied within the superblock. From now, let us omit index b, as we focus on only one superblock. One superblock consists of K blocks and

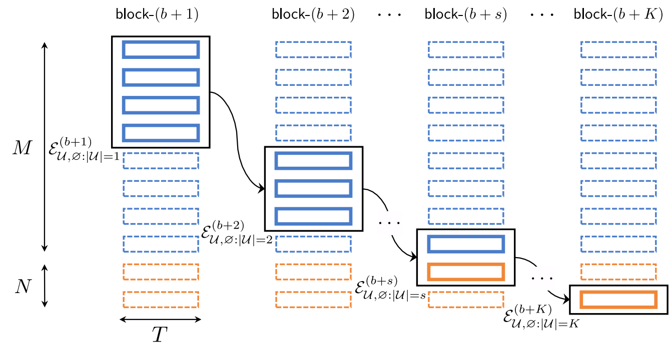

Fig. 7. Graphical illustration of our channel interpretation. Superblock-b consists of K blocks: the ones encapsulated in solid boxes connected by arrows from block-$(b+1)$ to block-$(b+K)$. In block-$(b+i)$, we send packets that can benefit up to i users simultaneously. In generating coded packets for block-$(b+i)$, we make use of some of the packets for all previous blocks.

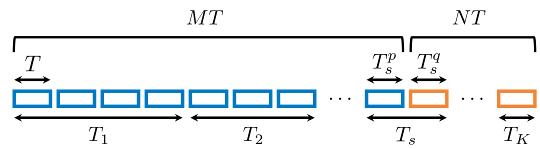

Fig. 8. Graphical illustration of our scheme. We have higher chances of delivering packets using q-channels than p-channels. We send high-order packets using q-channels to achieve a large number of packets decoded per transmission.

block-$(i+1)$ follows block-i for $i = 1, \ldots, K-1$. In block-i, coded packets that can benefit up to i users simultaneously are transmitted (uncoded, fresh packets are transmitted for $i = 1$). Since there are $\binom{K}{i}$ possible i-user sets, block-i is further divided into $\binom{K}{i}$ subblocks. In each subblock, coded packets intended for a specific set of i users are transmitted. Recall that we have two different channels. Thus, in some block, say block-s, a fraction of packets intended for s-user sets are transmitted in p-channels, and the rest in q-channels (see Fig. 8). All packets in block-i are transmitted in p-channels only when $i < s$ and in q-channels only when $i > s$. Denoting the number of channels for block-i by T_i and that for each subblock within block-i by t_i, $T_1 + T_2 + \cdots + T_K = (M+N)T$ and $T_i = \binom{K}{i} t_i$ hold.

In block-1, there are K subblocks. In each subblock, the transmitter sends fresh packets for a specific user. Some of the packets are received by the intended user and decoded immediately. Some are not received by the intended user, but received by some unintended user(s). These packets will be encoded into an XOR-sum and transmitted in later blocks. Some are received by none of the users, hence retransmitted until they are received by at least one user. Packet transmissions take place using T_1 channels where T_1 is determined in such a way that the transmitter can handle all fresh packets prepared at initialization, including some necessary retransmissions. At the end of block-1, the transmitter generates coded packets for block-2. That is, it searches for packets received by one unintended user but not by the intended user: for example, packet a intended for User 1 but received by User 2, and packet b intended for User 2 but received by User 1 qualify. The transmitter generates a coded packet $a \oplus b$ for block-2.

In block-2, there are $\binom{K}{2}$ subblocks. In each subblock, the transmitter sends coded packets intended for a specific set of two users. Consider Users 1 and 2, and corresponding desired packets a's and b's that have been received by the counterpart unintended user. The coded packets are in $a \oplus b$ form. User 1 can decode a when it receives $a \oplus b$, because it has side information b. Likewise, User 2 can decode b. Note that each user can decode its desired packet regardless of the reception outcome at the other user. Put differently, from each user's perspective, it is as if the transmitter treats them individually. From User 1's standpoint, if it receives $a \oplus b$, it can decode a, hence no longer needs to receive $a \oplus b$ again. If $a \oplus b$ is received by some user(s) other than Users 1 and 2, $a \oplus b$ will be encoded into another XOR-sum and transmitted in later blocks. Otherwise, $a \oplus b$ needs to be retransmitted. From User 2's standpoint, swapping the roles of a and b, the same holds true.

The retransmission protocol needs further inspection. Suppose $a_1 \oplus b_1$ is not received by User 1, but received by User 2 only. Symmetrically, $a_2 \oplus b_2$ is not received by User 2, but received by User 1 only. Both packets are not received by any other user(s), hence need to be retransmitted. However, User 1 already has $a_2 \oplus b_2$, hence does not need it again. Also, User 2 already has $a_1 \oplus b_1$, hence does not need it again, either. Do we retransmit them separately? The transmitter applies the same coding technique to avoid two transmissions; it generates $(a_1 \oplus b_1) \oplus (a_2 \oplus b_2)$. Yet again, from each user's perspective, it is as if the transmitter treats them individually. Packet transmissions take place using T_2 channels where T_2 is determined so as to handle all coded packets intended for all possible sets of two users, including some necessary retransmissions. That is, they are either decoded by intended users, or promoted to higher-order packets (hence, deferred to later blocks).

At the beginning of block-i (or at the end of block-$(i-1)$), the transmitter generates coded packets intended for all possible sets of i users by using the packets transmitted in the $i-1$ previous blocks. The key is that it generates a coded packet in such a way that each user in a specific set of i users can in effect decode one desired packet when the user receives the coded packet. Packet transmissions take place using T_i channels where T_i is determined so as to handle all coded packets intended for all possible sets of i users. In block-K, the coded packets need to be retransmitted until they are delivered to all intended users.

For concreteness, we need a systematic way of specifying the coded packet generation process. Algorithms 1 and 2 describe our scheme for general K in greater detail mathematically, how we generate coded packets in particular, based on the channel interpretation shown in Figs. 7 and 8. In Algorithm 1, the outer for-loop (**for** $i = 1 : K$ **do**) deals with K blocks within one superblock. The first inner for-loop (**for** $\forall \mathcal{U} \subseteq [K]$ *s.t.* $|\mathcal{U}| = i$ **do**) deals with $\binom{K}{i}$ subblocks within block-i. The second inner for-loop (**for** $\forall \mathcal{U}' \subseteq [K]$ *s.t.* $|\mathcal{U}'| = i+1$ **do**) generates coded packets for block-$(i+1)$ according to the function in Algorithm 2.

Algorithm 1 Reactive Coding Scheme for the General K-User Case

1 **initialize** $\mathcal{E}_{k,\varnothing}^{(b+1)}$ for $\forall k \in [K]$; `/* add an equal number of fresh packets */`
2 **for** $i = 1 : K$ **do** `/* for block-`$(b+i)$` of superblock-`b` */`
3 **for** $\forall \mathcal{U} \subseteq [K]$ *s.t.* $|\mathcal{U}| = i$ **do** `/* for each subblock of block-`$(b+i)$` */`
4 **while** $\mathcal{E}_{\mathcal{U},\varnothing}^{(b+i)} \neq \varnothing$ **do**
5 send $\texttt{pkt} \in \mathcal{E}_{\mathcal{U},\varnothing}^{(b+i)}$ using a p-channel or a q-channel depending on i with respect to s;
6 **for** $\forall k \in \mathcal{U}$ **do**
7 **if** pkt *is received by User* k **then**
8 do nothing; `/* decoded at User-`k` */`
9 **else if** pkt *is received by additional* $j-i$ *users in* $\mathcal{V}$ *where* $\mathcal{U} \cap \mathcal{V} = \varnothing$ **then**
10 add pkt to $\mathcal{E}_{k,\mathcal{U}\cup\mathcal{V}\setminus k}^{(b+j)}$; `/* promoted to order-`$j$` packet */`
11 **else**
12 add pkt to $\mathcal{E}_{k,\mathcal{U}\setminus k}^{(b+i)}$; `/* to be retransmitted */`
13 **end**
14 **end**
15 remove pkt from $\mathcal{E}_{\mathcal{U},\varnothing}^{(b+i)}$;
16 GENERATE_CODED_PACKETS($\mathcal{U}$, $b+i$); `/* prepare for retransmissions */`
17 **end**
18 **end**
19 **for** $\forall \mathcal{U}' \subseteq [K]$ *s.t.* $|\mathcal{U}'| = i+1$ **do** `/* generate coded packets for block-`$(b+i+1)$` */`
20 GENERATE_CODED_PACKETS($\mathcal{U}'$, $b+i+1$);
21 **end**
22 **end**

Algorithm 2 Coded Packet Generation Function

1 **function** GENERATE_CODED_PACKETS($\mathcal{U}, b+i$)
2 **while** $\mathcal{E}_{k,\mathcal{U}\setminus k}^{(b+i)} \neq \varnothing$ *for* $\forall k \in \mathcal{U}$ **do**
3 collect $\texttt{pkt}_k \in \mathcal{E}_{k,\mathcal{U}\setminus k}^{(b+i)}$ for each $k \in \mathcal{U}$;
4 generate one coded − pkt as an XOR-sum of $\texttt{pkt}_k$'s over $k \in \mathcal{U}$;
5 add coded − pkt to $\mathcal{E}_{\mathcal{U},\varnothing}^{(b+i)}$;
6 remove $\texttt{pkt}_k$ from $\mathcal{E}_{k,\mathcal{U}\setminus k}^{(b+i)}$ for each $k \in \mathcal{U}$;
7 **end**
8 **end**

D. Achievable Rate Analysis

To compute the achievable sum rate, we need to count the number of packets to be decoded and the required number of channels to guarantee their successful decoding. First, let us consider the number of packets. As described at line 5 in Algorithm 1, we send one coded packet in each iteration of the while loop of $\mathcal{E}_{\mathcal{U},\varnothing} \neq \varnothing$. Each coded packet in $\mathcal{E}_{\mathcal{U},\varnothing}$ is generated (except when initialized for $|\mathcal{U}| = 1$) by collecting one packet from $\mathcal{E}_{k,\mathcal{U}\setminus k}$ for each $k \in \mathcal{U}$. Hence, $|\mathcal{E}_{\mathcal{U},\varnothing}| = |\mathcal{E}_{k,\mathcal{U}\setminus k}|$ holds. Resorting to symmetry, we assume that $|\mathcal{E}_{k,\mathcal{U}\setminus k}|$ is equal for all $k \in \mathcal{U}$.

Next, let us consider the required number of channels to handle $|\mathcal{E}_{\mathcal{U},\varnothing}|$ coded packets. As described at line 12 in Algorithm 1, a coded packet may need to be retransmitted if it is neither received by the intended user (line 8) nor promoted to a higher-order packet (line 10). Hence, the required number of channels to handle $|\mathcal{E}_{\mathcal{U},\varnothing}|$ coded packets is strictly greater than $|\mathcal{E}_{\mathcal{U},\varnothing}|$. To compute the number of channels, we introduce $\beta_j(p)$. Assuming that we use a p-channel, $\beta_j(p)$ is the probability that each transmission of a coded packet in $\mathcal{E}_{\mathcal{U},\varnothing}$ with $|\mathcal{U}| = j$ (uncoded fresh packet for $j = 1$) causes the packet to be either received by the intended user, or promoted to a higher-order packet. This event happens except for the case where neither the intended user nor the $K - j$ remaining users receive the coded packet (w.p. $(1-p) \times (1-p)^{K-j}$) [22]:

$$\beta_j(p) := 1 - (1-p)^{K-j+1}. \tag{34}$$

With probability $\beta_j(p)$, each transmission of a coded packet in $\mathcal{E}_{\mathcal{U},\varnothing}$ with $|\mathcal{U}| = j$ removes the packet from the set (from line 8 or 10 to line 15). With probability $1 - \beta_j(p)$, each transmission of a coded packet fails, hence retransmissions of the packet are required. The packet is combined with other similar packets and retransmitted (line 12 to 16 to 5) in a different form in the same block. Therefore, the required number of p-channels to handle all coded packets in $\mathcal{E}_{\mathcal{U},\varnothing}$, where $\mathcal{U}$ is a set of j users, is (we generalize it to arbitrary channels in (41)):

$$t_j = \frac{1}{\beta_j(p)} |\mathcal{E}_{\mathcal{U},\varnothing}|, \text{ where } |\mathcal{U}| = j. \tag{35}$$

Now, we compute $|\mathcal{E}_{\mathcal{U},\varnothing}|$, where $|\mathcal{U}| = j$. Recall that some coded packets in $\mathcal{E}_{\mathcal{U},\varnothing}$ with $|\mathcal{U}| = i$ become order-j packets and are added to $\mathcal{E}_{k,\mathcal{U}'\setminus k}$ with $|\mathcal{U}'| = j$, where $i < j$. Also, one packet from $\mathcal{E}_{k,\mathcal{U}'\setminus k}$ for each $k \in \mathcal{U}'$ is collected to generate one coded packet in $\mathcal{E}_{\mathcal{U}',\varnothing}$, where $|\mathcal{U}'| = j$. To compute $|\mathcal{E}_{\mathcal{U}',\varnothing}|$, where $|\mathcal{U}'| = j$, we introduce $\alpha_{i\to j}(p)$. Assuming that we use a p-channel, $\alpha_{i\to j}(p)$ is the probability that each transmission of a coded packet in $\mathcal{E}_{\mathcal{U},\varnothing}$ with $|\mathcal{U}| = i$ (uncoded fresh packet for $i = 1$) causes the packet to be promoted to an order-j packet and added to $\mathcal{E}_{k,\mathcal{U}'\setminus k}$ with $|\mathcal{U}'| = j$. This event happens for the case where the coded packet is received by $j - i$ additional unintended users (w.p. p^{j-i}), but not received by the intended user and the remaining $K - j$ users (w.p. $(1-p) \times (1-p)^{K-j}$) [22]:

$$\alpha_{i\to j}(p) := p^{j-i}(1-p)^{K-j+1}. \tag{36}$$

With probability $\alpha_{i\to j}(p)$, each transmission of a coded packet in $\mathcal{E}_{\mathcal{U},\varnothing}$ with $|\mathcal{U}| = i$ adds the packet to $\mathcal{E}_{k,\mathcal{U}'\setminus k}$ with $|\mathcal{U}'| = j$, which in turn is collected with other similar packets to generate a coded packet in $\mathcal{E}_{\mathcal{U}',\varnothing}$ with $|\mathcal{U}'| = j$, where $i < j$ (from line 10 to line 20). Therefore, the number of coded packets in $\mathcal{E}_{\mathcal{U}',\varnothing}$ generated from the packets in $\mathcal{E}_{\mathcal{U},\varnothing}$, where $\mathcal{U}$ is a set of i users and $\mathcal{U}'$ is a set of j users, is (we generalize it to arbitrary channels in (40)):

$$\ell_{i\to j} = t_i \alpha_{i\to j}(p). \tag{37}$$

For a given set of j users, say $\mathcal{U}'$, there are $\binom{j-1}{i-1}$ proper subsets, say $\mathcal{U}$'s, from which coded packets in $\mathcal{E}_{\mathcal{U},\varnothing}$ can become order-j packets in $\mathcal{E}_{k,\mathcal{U}'\setminus k}$. Hence, to compute $|\mathcal{E}_{\mathcal{U}',\varnothing}|$, where $|\mathcal{U}'| = j$ (recall that this is equal to $|\mathcal{E}_{k,\mathcal{U}'\setminus k}|$), we need a binomial coefficient and a summation. Pick User $k \in \mathcal{U}'$. Exclude User k for later inclusion and choose $i - 1$ users among the remaining $j - 1$ users. This leads to $\binom{j-1}{i-1}$. Since any coded packet in $\mathcal{E}_{\mathcal{U},\varnothing}$ for $\mathcal{U} \subsetneq \mathcal{U}'$ can become order-j packet and be added to $\mathcal{E}_{k,\mathcal{U}'\setminus k}$, we sum over all $i < j$:

$$|\mathcal{E}_{\mathcal{U},\varnothing}| = \sum_{i=1}^{j-1} \binom{j-1}{i-1} \ell_{i\to j}, \text{ where } |\mathcal{U}| = j. \tag{38}$$

Putting this into (35), we get:

$$t_j = \frac{1}{\beta_j(p)} \sum_{i=1}^{j-1} \binom{j-1}{i-1} \ell_{i\to j}. \tag{39}$$

Finally, we conclude the packet and channel counting processes. Suppose we initialize $\mathcal{E}_{k,\varnothing}$ for each $k \in [K]$ with ℓ_1 fresh packets and all of them are decoded at the intended users in T time slots, using MT p-channels and NT q-channels. Then, the achievable sum rate is $R_{\text{sum}}^{\text{react}} = \frac{K\ell_1}{T}$. We send as many high-order coded packets as possible in q-channels. Coded packets of some order may be sent in both p-channels and q-channels. Let s be this order (see Fig. 8). Let λ be the fraction of order-s coded packets that are sent in T_s^p p-channels. The remaining order-s coded packets are sent in T_s^q q-channels. Let $f = \frac{\ell_1}{T}$. We have three parameters subject to design: f, s and λ.

Depending on the channels we use, we must apply different channel parameters. Generalizing $\ell_{i\to j}$ and t_j in (37) and (39) respectively, we have:

$$\ell_{i\to j} = \begin{cases} t_i \alpha_{i\to j}(p), & i < s; \\ t_i \alpha_{i\to j}(q), & i > s; \\ t_s^p \alpha_{s\to j}(p) + t_s^q \alpha_{s\to j}(q), & i = s, \end{cases} \tag{40}$$

$$t_j = \begin{cases} \dfrac{1}{\beta_j(p)} \sum_{i=1}^{j-1} \binom{j-1}{i-1} \ell_{i\to j}, & j < s; \\ \dfrac{1}{\beta_j(q)} \sum_{i=1}^{j-1} \binom{j-1}{i-1} \ell_{i\to j}, & j > s; \\ t_s^p + t_s^q, & j = s, \end{cases} \tag{41}$$

where t_s^p and t_s^q are given as:

$$t_s^p = \frac{\lambda}{\beta_s(p)} \sum_{i=1}^{s-1} \binom{s-1}{i-1} \ell_{i\to s}, \tag{42}$$

$$t_s^q = \frac{1-\lambda}{\beta_s(q)} \sum_{i=1}^{s-1} \binom{s-1}{i-1} \ell_{i\to s}. \tag{43}$$

We need initial conditions to compute (40) and (41). We set $\ell_{i\to j}$ as ℓ_1 for $j = 1$. Also, let T_j be the number of channels required to handle all order-j coded packets (fresh packets for $j = 1$). Since there are $\binom{K}{j}$ ways to construct sets of j users, the following holds:

$$T_j = \binom{K}{j} t_j. \tag{44}$$

We set f, s and λ, the three parameters subject to design in such a way that the following two conditions are met. This is to use all channels and guarantee decoding of all fresh packets:

$$\sum_{j=1}^{s-1} T_j + \binom{K}{s} t_s^p = MT, \sum_{j=s+1}^{K} T_j + \binom{K}{s} t_s^q = NT. \quad (45)$$

Let us denote the set parameters by f^*, s^* and λ^*. We send f^*T fresh packets for each user (which corresponds to line 1 in Algorithm 1), and all of them are guaranteed to be decoded in T time slots. This leads us to the following theorem.

Theorem 4 (Reactive coding): The following sum rate is achievable by the proposed reactive coding scheme:

$$R_{\mathsf{sum}}^{\mathsf{react}} = Kf^*, \quad (46)$$

where f^* is the solution of the equations from (36) to (45).

Remark 6: As mentioned in Footnote 2, our proposed scheme in Section IV-C requires each user to know the global channel state information. As mentioned in Remark 5, it also requires each user to know the bit-shifting information used to generate XOR-sums. After each communication session using T time slots ($(M + N)T$ packets are transmitted in total), we append an extra session of exchanging these types of control information using $T_{\rm ext}$ time slots. A rough cost can be calculated as follows. The total number bits to represent the global channel state information is $(M + N)T \times K$. The total number of bits to represent the bit-shifting information is $(M + N)T \times K \log_2 K$ because at most $K \log_2 K$ bits are required to describe the bit-shifts for one XOR-sum: up to K packets can be used to generate one XOR-sum and $\log_2 K$ bits can specify the bit-shift for each packet, which is chosen out of size-K pool $\{1, 2^1, \cdots, 2^{K-1}\}$. The naive scheme (9) achieves a broadcast rate of $Mp+Nq$ to all users. Suppose a single packet contains L bits. Then, the fraction of extra time slots required to convey all control information is $\frac{T_{\rm ext}}{T} = \frac{K(M+N)(1+\log_2 K)}{L(Mp+Nq)}$. For sufficiently large L, the cost can be made negligible. Similar arguments have appeared in prior work [6]. Alternatively, there is a way to avoid the need of conveying the global channel state information and use a decoder $\hat{W}_k = g_{kt}(Y_k^t)$ (instead of $\hat{W}_k = g_{kt}(Y_k^t, S_1^t, \ldots, S_K^t)$ described in Section II). Packets include a header which contains critical control information. Suppose we assign a K-bit chunk in the header to indicate for which set of users the packet is intended (hence, each user can tell if the received packet is intended for itself or an overheard packet intended for others), and assign another K-bit chunk to instruct each user to discard the latest overheard packet. Each user keeps overheard packets in its FIFO (First-In-First-Out) queues and discards the latest (i.e., "Last-In") overheard packet if instructed to do so. The transmitter gets past channel state feedback instantly, thus can ensure all users to keep necessary overheard packets only and keep them in order. Upon receiving a packet, each user can examine the first K-bit chunk in the header, and if it is intended for itself, determine which oldest (i.e., "First-Ins") overheard packets to cancel out from the packet to decode its desired packet. Also, the header can reserve $K \log_2 K$ bits to describe the bit-shifts used to generate the coded packet. For sufficiently large L, these costs of extra control information in packet headers can be made negligible. Similar arguments have appeared in prior works [5], [30].

Remark 7: Let us compare our proposed K-user scheme with the state-of-the-art [5], [6]. The works of [5], [6] consider a K-user *single-input* packet-erasure broadcast channel where packet-erasure probabilities are arbitrary and *time-invariant*. Thus, it is straightforward to extend the schemes developed therein either to the homogeneous multi-input channel or to the heterogeneous multi-input channel on a per-subchannel basis (separation scheme). However, it is not straightforward to generalize them in the heterogeneous case without significant modifications. Also, the two works do not assume prior knowledge of channel statistics such as packet-erasure probabilities. In contrast, our simplified setting, where two types of subchannels with two different packet-erasure probabilities are assumed, implicitly allows such heterogeneity to be leveraged. In fact, our scheme is specifically devised to take advantage of it. Our scheme builds on common ideas as in [5], [6] of generating coded packets by the transmitter according to past channel state feedback. In addition, it intentionally sends coded packets using stable subchannels to increase the chances of such high-utility packets being received simultaneously by as many users as possible. This idea of exploiting channel diversity is missing in the state-of-the-art schemes, for the justifiable reasons of the single-input and time-invariant model considered therein.

E. Outer Bound

We derive an outer bound $\mathcal{R}_{\mathsf{outer}}$ for $(R_1, \ldots, R_K)$ for the general K-user case. The proof closely follows the bounding techniques used in the works of [5], [6], [8]. For completeness, we present the details of the proof.

Theorem 5: The capacity region for the K-user packet-erasure broadcast channel is bounded by the intersection of the following outer bounds for $(R_1, \ldots, R_K)$ where $\ell \in [1 : K!]$ is an index of the permutations π_ℓ's of $\{1, \ldots, K\}$ whose entries indicate the users:

$$R_k \leq \sum_{m=1}^{M+N} R_k^{(\ell)}(m), \quad (47)$$

where $R_k^{(\ell)}(m)$ is subject to:

$$R_k^{(\ell)}(m) \geq 0, \text{ for } k \in [1 : K] \text{ and } m \in [1 : M+N], \quad (48)$$

$$\sum_{i=1}^{K} \frac{R_{\pi_\ell(i)}^{(\ell)}(m)}{1-(1-p)^i} \leq 1 \text{ for } m \in [1 : M], \quad (49)$$

$$\sum_{i=1}^{K} \frac{R_{\pi_\ell(i)}^{(\ell)}(m)}{1-(1-q)^i} \leq 1 \text{ for } m \in [M+1 : M+N]. \quad (50)$$

Proof: We can construct an augmented channel from the original channel of interest as in [35]. For example, we provide the output of User 1 to User 2 and in turn provide the output of User 2 to User 3, and so on. Then, we have a channel where the output of User k is the collection of the outputs of Users up to $k-1$ for $1 \leq k < K$.

There are $K!$ distinct ways in total to construct such augmented channels. Consider one permutation of $\{1, \ldots, K\}$,

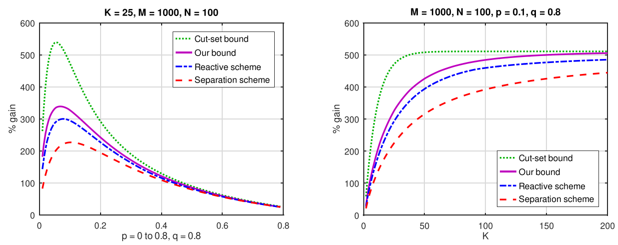

Fig. 9. Gain v.s. p: $K = 25$, $M = 1000$, $N = 100$, $q = 0.8$ (left); Gain v.s. K: $M = 1000$, $N = 100$, $p = 0.1$, $q = 0.8$ (right).

which we denote by π_ℓ where $\ell \in [1 : K!]$, and let the entries indicate the users. Then, we construct an augmented channel in such a way that we provide the output of the user indexed by the i-th entry in π_ℓ, which we denote by $\pi_\ell(i)$, to the user indexed by $\pi_\ell(i+1)$ for $1 \le i < K$. The effective reception for the user indexed by $\pi_\ell(i)$ is $1-(1-p)^i$ in the unstable subchannels and $1-(1-q)^i$ in the stable subchannels.

Note that the capacity region for the original channel is bounded by that of each augmented channel. Note also that the augmented channels are physically degraded since the channel input X is conditionally independent of the channel output of User $\pi_\ell(i+1)$ given the channel output of User $\pi_\ell(i)$ for $1 \le i < K$. It has been shown in [36] that feedback does not enlarge the capacity region for physically degraded broadcast channels. Also, in [9], they showed that the capacity region for parallel erasure broadcast channels without feedback is described by time-sharing between users at each subchannel. Hence, the capacity region for each augmented channel (with feedback) coincides with the capacity region for each augmented channel without feedback, which is described by time-sharing between users at each subchannel as in (49) and (50). We can obtain an outer bound for the capacity region for the original channel by taking the intersection of the $K!$ distinct outer bounds as in (47), which are obtained from $K!$ distinct augmented channels. This completes the proof. ■

F. Numerical Evaluation

Generalizing (14) and (15) to the K-user case, we consider the following two bounds[4]:

$$\overline{R}_{\text{sum}}^{\text{cut-set}} = M\left(1-(1-p)^K\right) + N\left(1-(1-p)^K\right), \quad (51)$$

$$\overline{R}_{\text{sum}}^{\text{ours}} = \max_{(R_1,R_2,\ldots,R_K)\in\mathcal{R}_{\text{outer}}} \sum_{k=1}^{K} R_k. \quad (52)$$

Generalizing (10) to the K-user case, the separation scheme in the K-user case achieves:

$$R_{\text{sum}}^{\text{sep}} = M\frac{K}{\sum_{i=1}^{K}\frac{1}{1-(1-p)^i}} + N\frac{K}{\sum_{i=1}^{K}\frac{1}{1-(1-p)^i}}. \quad (53)$$

Finally, we consider $R_{\text{sum}}^{\text{react}}$ in Theorem 4.

We examine $\overline{\text{Gain}}_{\text{cut-set}}$, $\overline{\text{Gain}}_{\text{ours}}$, Gain_{sep}, and $\text{Gain}_{\text{react}}$, which can be computed similarly as in the two-user case in Section III-A. Fig. 9(a) illustrates the relative gains with respect to p for $K = 25$, $M = 1000$, $N = 100$ and $q = 0.8$. As the range of spectrum is far wider in the mmWave bands than in the current system bands, we set M to be ten times N. The (green) dotted curve plots $\overline{\text{Gain}}_{\text{cut-set}}$, the (magenta) solid curve plots $\overline{\text{Gain}}_{\text{ours}}$, the (blue) dash-dotted curve plots $\text{Gain}_{\text{react}}$, and the (red) dashed curve plots Gain_{sep}. This setting could represent scenarios where cells with small coverage serve a reasonable number of users (e.g., pico/femto cells). We can see that for fairly large p (say 0.6), our scheme can achieve a reasonable gain of 57%. In adverse scenarios, that is, for small p (say 0.1), our scheme can achieve a huge gain of 293%, a four-fold increase, with a noticeable gap compared to a gain of 225% achieved by the separation scheme.

Fig. 9(b) illustrates the relative gains with respect to K for adverse scenarios: $M = 1000$, $N = 100$, $p = 0.1$ and $q = 0.8$. We can see that for a handful of users ($K > 5$), our scheme can achieve at least a two-fold gain of 104% and reach a five-fold gain of 394% at $K = 50$. We can see that the gains by our scheme and the separation scheme both scale as K increases. This scalability stems from the use of causal channel state feedback.

We evaluate $\overline{\text{Gain}}_{\text{ours}}$ (solid magenta curve) and $\overline{\text{Gain}}_{\text{cut-set}}$ (dotted green curve). Also, it is yet unclear whether our upper bound is in fact the sum capacity in the general K-user case. As shown in the two-user case, we anticipate that incorporating proactive coding techniques will lead to an improved achievable sum rate in the K-user case, further narrowing the current gap. We leave it as future work to characterize the sum capacity for the K-user case by developing a better scheme that incorporates reactive and proactive coding techniques, and if necessary by deriving a tighter upper bound.

Remark 8: We have developed our scheme without taking into account some implementation issues. Let us briefly present two such issues. First, each user keeps copies of overheard packets in its local memory. This clearly incurs a storage cost. In practice, each user has a limited memory budget, which is a critical real-world constraint. Second, packet-erasure probabilities are not necessarily identical across users and/or subchannels in real-world systems. This leads to an imbalance in number across the bins of packets that are used to generate coded packets. For example, $|\mathcal{E}_{1,\{2\}}| \neq |\mathcal{E}_{2,\{1\}}|$ in the two-user case, thus the number of coded packets is limited to $\min\left(|\mathcal{E}_{1,\{2\}}|, |\mathcal{E}_{2,\{1\}}|\right)$. We need a way to process the remaining packets in $\mathcal{E}_{1,\{2\}}$ or $\mathcal{E}_{2,\{1\}}$ that cannot be coded into XOR-sums. One heuristic scheme to deal with these issues can be as follows. Given K users, we divide them into multiple groups of k users where $1 < k \le K$. Then, we divide $M + N$ subchannels into K/k chunks and assign one chunk exclusively to each k-user group. We now consider the channel as K/k independent k-user channels and treat them separately. As k is a design parameter, we can choose k depending on the memory budget at the users and the targeted implementation complexity. Also, at the end of each communication session using T time slots, we can spare some extra time slots to retransmit the remaining packets in $\mathcal{E}$'s until they are received by the intended users. A simulation shows that the heuristic scheme achieves a sensible throughput gain. We consider $K = 60$, $M = 1000$ and $N = 100$. We also assume that the packet-erasure probabilities follow

[4]We abuse our notation used in the two-user case and re-use it in the K-user case, in order not to over-complicate it. The meaning should be clear from the context.

TABLE I

SIMULATED PERFORMANCES BY THE HEURISTIC SCHEME PROPOSED IN REMARK 8 WHERE WE DIVIDE A GIVEN K-USER CHANNEL INTO MULTIPLE k-USER CHANNELS ($1 < k \leq K$) AND TREAT THEM INDEPENDENTLY

k	2	3	4	5	6	10	12	15	20	30	60
$\mathsf{Gain}_{\mathsf{heur}}(k)$	20%	33%	44%	54%	62%	88%	95%	102%	109%	115%	122%
$\frac{\mathsf{Gain}_{\mathsf{heur}}(k)}{\mathsf{Gain}_{\mathsf{react}}(K=60)}$	16%	27%	36%	44%	50%	71%	77%	83%	88%	93%	99%

normal distributions. We consider a mean of 0.4 for unstable subchannels, 0.8 for stable channels, and a standard deviation of 0.05 for both. Table I shows our result. $\mathsf{Gain}_{\mathsf{heur}}(k)$ is defined as in (11) replacing $R^{\mathsf{react}}_{\mathsf{sum}}$ with the achievable sum rate by the heuristic scheme for k. $\frac{\mathsf{Gain}_{\mathsf{heur}}(k)}{\mathsf{Gain}_{\mathsf{react}}}$ measures the portion of the gain that the heuristic scheme for k can reap with respect to our proposed scheme for the K-user case. Our result demonstrates that for $k = 6$ (only one-tenth of $K = 60$), we can reap half the potential gain with an improvement of 62% over the scheme currently employed in practice.

V. CONCLUSION

Ongoing efforts in harnessing a large amount of mmWave spectrum motivated us to investigate the K-user parallel packet-erasure broadcast channel where some subchannels are susceptible to packet-erasures (mmWave bands) while some are less so (legacy system bands). We considered real-world settings where the unstable bands are prone to frequent packet-erasures while the stable bands are prone to modest packet-erasures. We showed that applying coding *across* subchannels spanning both types of bands can provide significant gains that scale with K over the scheme currently employed in practice, which simply allocates chunks of subchannels to users exclusively. We also demonstrated that it outperforms a per-subchannel extension of the state-of-the-art K-user schemes by large margins, reducing the optimality gap. Our results suggest a potential coding scheme that can meet ever-growing mobile data demands in future wireless systems.

APPENDIX A
FOURIER-MOTZKIN ELIMINATION

We aim to derive the bound on R_1+R_2 only for $\ell = 1$, since we obtain an identical bound for $\ell = 2$ with the swapped roles of the users. For simplicity, we omit superscript ℓ. We have 8 inequalities and 6 variables.

$$R_1 \leq MR_{1,p} + NR_{1,q}, \tag{54}$$
$$R_2 \leq MR_{2,p} + NR_{2,q}, \tag{55}$$
$$R_{1,p} \geq 0, \tag{56}$$
$$R_{1,q} \geq 0, \tag{57}$$
$$R_{2,p} \geq 0, \tag{58}$$
$$R_{2,q} \geq 0, \tag{59}$$
$$\frac{R_{1,p}}{p} + \frac{R_{2,p}}{1-(1-p)^2} \leq 1, \tag{60}$$
$$\frac{R_{1,q}}{q} + \frac{R_{2,q}}{1-(1-q)^2} \leq 1. \tag{61}$$

By Fourier-Motzkin elimination, we remove the following variables in order: $R_{1,p}$, $R_{1,q}$, $R_{2,p}$, and $R_{2,q}$. In the end, we have inequalities only in terms of R_1 and R_2 from which we obtain the desired bound on $R_1 + R_2$.

From (54), (56), and (60), we have

$$\max\left\{0, \frac{R_1 - NR_{1,q}}{M}\right\} \leq R_{1,p} \leq p\left(1 - \frac{R_{2,p}}{1-(1-p)^2}\right). \tag{62}$$

Matching each of the lower bounds on $R_{1,p}$ with its the upper bound eliminates $R_{1,p}$. For any set of the remaining variables that satisfies the consequential system of inequalities, there exists a value of $R_{1,p}$ that satisfies the original system. Hence, from

$$0 \leq p\left(1 - \frac{R_{2,p}}{1-(1-p)^2}\right), \tag{63}$$
$$\frac{R_1 - NR_{1,q}}{M} \leq p\left(1 - \frac{R_{2,p}}{1-(1-p)^2}\right), \tag{64}$$

we respectively obtain

$$R_{2,p} \leq 1-(1-p)^2, \tag{65}$$
$$R_{1,q} \geq \frac{R_1 - Mp\left(1 - \frac{R_{2,p}}{1-(1-p)^2}\right)}{N}. \tag{66}$$

From (57), (61), and (66), we have

$$\max\left\{0, \frac{R_1 - Mp\left(1 - \frac{R_{2,p}}{1-(1-p)^2}\right)}{N}\right\} \leq R_{1,q} \leq q\left(1 - \frac{R_{2,q}}{1-(1-q)^2}\right). \tag{67}$$

Matching each of the lower bounds on $R_{1,q}$ with its the upper bound eliminates $R_{1,q}$. Hence, from

$$0 \leq q\left(1 - \frac{R_{2,q}}{1-(1-q)^2}\right), \tag{68}$$
$$\frac{R_1 - Mp\left(1 - \frac{R_{2,p}}{1-(1-p)^2}\right)}{N} \leq q\left(1 - \frac{R_{2,q}}{1-(1-q)^2}\right), \tag{69}$$

we respectively obtain

$$R_{2,q} \leq 1-(1-q)^2, \tag{70}$$
$$R_{2,p} \leq \left(1-(1-p)^2\right)\left(1 - \frac{R_1 - Nq\left(1 - \frac{R_{2,q}}{1-(1-q)^2}\right)}{Mp}\right). \tag{71}$$

From (55), (58), (65), and (71) we have

$$\max\left\{\begin{matrix} 0, \\ \frac{R_2 - NR_{2,q}}{M} \end{matrix}\right\} \leq R_{2,p}$$

$$\leq \min\left\{\begin{array}{l}1-(1-p)^2,\\ \left(1-(1-p)^2\right)\left(1-\dfrac{R_1-Nq\left(1-\dfrac{R_{2,q}}{1-(1-q)^2}\right)}{Mp}\right)\end{array}\right\}. \quad (72)$$

Matching each of the lower bounds on $R_{2,p}$ with each of its upper bound eliminates $R_{2,p}$. Hence, from

$$\frac{R_2-NR_{2,q}}{M} \leq 1-(1-p)^2, \quad (73)$$

$$\frac{R_2-NR_{2,q}}{M} \leq \left(1-(1-p)^2\right)\left(1-\frac{R_1-Nq\left(1-\frac{R_{2,q}}{1-(1-q)^2}\right)}{Mp}\right), \quad (74)$$

$$0 \leq \left(1-(1-p)^2\right)\left(1-\frac{R_1-Nq\left(1-\frac{R_{2,q}}{1-(1-q)^2}\right)}{Mp}\right), \quad (75)$$

we respectively obtain

$$R_{2,q} \geq \frac{R_2-M\left(1-(1-p)^2\right)}{N}, \quad (76)$$

$$R_{2,q} \leq \frac{(2-p)(2-q)\left(Mp+Nq-R_1\right)-(2-q)R_2}{N(q-p)}, \quad (77)$$

$$R_{2,q} \leq \frac{(2-q)(Mp+Nq-R_1)}{N}. \quad (78)$$

From (59), (70), (76), (77), and (78), we have

$$\max\left\{\begin{array}{l}0,\\ \dfrac{R_2-M\left(1-(1-p)^2\right)}{N}\end{array}\right\} \leq R_{2,q}$$

$$\leq \min\left\{\begin{array}{l}1-(1-q)^2,\\ \dfrac{(2-q)(Mp+Nq-R_1)}{N},\\ \dfrac{(2-p)(2-q)\left(Mp+Nq-R_1\right)-(2-q)R_2}{N(q-p)}\end{array}\right\}. \quad (79)$$

Matching each of the lower bounds on $R_{2,q}$ with each of its upper bound eliminates $R_{2,q}$. Hence, from

$$\frac{R_2-M\left(1-(1-p)^2\right)}{N} \leq 1-(1-q)^2, \quad (80)$$

$$0 \leq \frac{(2-q)(Mp+Nq-R_1)}{N}, \quad (81)$$

$$\frac{R_2-M\left(1-(1-p)^2\right)}{N} \leq \frac{(2-q)(Mp+Nq-R_1)}{N}, \quad (82)$$

$$0 \leq \frac{(2-p)(2-q)\left(Mp+Nq-R_1\right)-(2-q)R_2}{N(q-p)}, \quad (83)$$

$$\frac{R_2-M\left(1-(1-p)^2\right)}{N} \leq \frac{(2-p)(2-q)\left(Mp+Nq-R_1\right)-(2-q)R_2}{N(q-p)}, \quad (84)$$

we respectively obtain

$$R_2 \leq Mp(2-p)+Nq(2-q), \quad (85)$$

$$R_1 \leq Mp+Nq, \quad (86)$$

$$(2-q)R_1+R_2 \leq Mp(2-p)+(Mp+Nq)(2-q), \quad (87)$$

$$(2-p)R_1+R_2 \leq (Mp+Nq)(2-p), \quad (88)$$

$$(2-q)R_1+R_2 \leq Mp(2-p)+Nq(2-q). \quad (89)$$

For $\ell = 2$, we obtain the following two inequalities with the swapped roles of the users, which are symmetric with (88) and (89).

$$R_1+(2-p)R_2 \leq (Mp+Nq)(2-p), \quad (90)$$

$$R_1+(2-q)R_2 \leq Mp(2-p)+Nq(2-q). \quad (91)$$

From (88) and (90), and also from (89) and (91), we obtain the desired bound on R_1+R_2.

$$R_1+R_2 \leq \frac{2(2-p)}{3-p}(Mp+Nq), \quad (92)$$

$$R_1+R_2 \leq \frac{2}{3-q}\left(Mp(2-p)+Nq(2-q)\right). \quad (93)$$

APPENDIX B
DERIVATION OF RESULTS IN SECTION III FROM [8]

The work of [8] has characterized the capacity region of two-user parallel packet-erasure broadcast channels where the packet-erasure probabilities of the subchannels can be arbitrary. Lemma 1 therein describes a rate region achievable by the separation scheme that employs coding in each subchannel separately. Lemma 2 therein describes the capacity region. Hence, by evaluating the rate regions of these lemmas and obtaining the respective sum rates, one can show that the separation principle fails in heterogeneous parallel packet-erasure broadcast channels.

Let us consider our setting where we have M unstable subchannels with packet-reception probability p and N stable subchannels with packet-reception probability q in parallel. To prove that the separation principle fails with heterogeneity, one needs to show that (i) in the homogeneous case $(p=q)$, the sum rates from Lemma 1 and Lemma 2 coincide for all channel parameters (M,N,p); (ii) in the heterogeneous case $(p<q)$, there exists a channel parameter (M,N,p,q) for which Lemma 2 yields a greater sum rate than Lemma 1.

First, let us show (i). The rate region achievable by Lemma 1 is described as follows [8]:

$$R_1 \leq \sum_{m=1}^{M+N} R_1(m),\ R_2 \leq \sum_{m=1}^{M+N} R_2(m), \quad (94)$$

where $R_1(m)$ and $R_2(m)$ are subject to:

$$\begin{array}{l}R_1(m) \geq 0,\ \text{for } m \in [1:M+N],\\ R_2(m) \geq 0,\ \text{for } m \in [1:M+N],\\ \dfrac{R_1(m)}{p}+\dfrac{R_2(m)}{1-(1-p)^2} < 1 \text{ for } m \in [1:M],\end{array} \quad (95)$$

$$\frac{R_1(m)}{q}+\frac{R_2(m)}{1-(1-q)^2} < 1 \text{ for } m \in [M+1:M+N], \quad (96)$$

$$\frac{R_2(m)}{p} + \frac{R_1(m)}{1-(1-p)^2} < 1 \text{ for } m \in [1:M],$$
$$\frac{R_2(m)}{q} + \frac{R_1(m)}{1-(1-q)^2} < 1 \text{ for } m \in [M+1:M+N]. \quad (97)$$

One can use the same arguments used in Section III-F, which exploits symmetry of the channel, to reduce the number of variables. Using the arguments, one can set $R_k(m)$'s to be equal to $R_{k,p}$ for all $m \in [1:M]$ and $R_{k,q}$ for all $m \in [M+1:M+N]$ for the sake of obtaining the maximum bound on R_1+R_2. Then the above system of inequalities can be re-expressed as follows:

$$R_1 \leq MR_{1,p} + NR_{1,q}, \; R_2 \leq MR_{2,p} + NR_{2,q}, \quad (98)$$

where $R_{1,p}$, $R_{1,q}$, $R_{2,p}$, and $R_{2,q}$ are subject to:

$$R_{1,p} \geq 0, \; R_{1,q} \geq 0, \; R_{2,p} \geq 0, \; R_{2,q} \geq 0, \quad (99)$$
$$\frac{R_{1,p}}{p} + \frac{R_{2,p}}{1-(1-p)^2} < 1, \; \frac{R_{1,q}}{q} + \frac{R_{2,q}}{1-(1-q)^2} < 1, \quad (100)$$
$$\frac{R_{2,p}}{p} + \frac{R_{1,p}}{1-(1-p)^2} < 1, \; \frac{R_{2,q}}{q} + \frac{R_{1,q}}{1-(1-q)^2} < 1. \quad (101)$$

One can obtain the achievable sum rate of

$$R_1 + R_2 \leq 2\left(\frac{Mp(2-p)}{3-p} + \frac{Nq(2-q)}{3-q}\right). \quad (102)$$

Lemma 2 uses the same techniques used in Section III-F to derive an outer bound on the capacity region. Hence, Lemma 2 yields an upper bound on achievable sum rates equal to (3). Setting $p = q$, one can verify that (3) and (102) coincide. This means that the separation scheme is optimal in the homogeneous case.

Next, let us show (ii). As mentioned, Lemma 2 yields the capacity region. The work of [8] has shown that it is achievable by exhaustively searching for the best linear network coding scheme(s). Thus, one can obtain the sum capacity in closed-form by applying Fourier-Motzkin elimination to Lemma 2 and compare it with (102) analytically. Appendix A presents the detailed steps. However, for the sake of checking if a scheme that employs coding across subchannels can outperform the separation scheme in the heterogeneous case, it suffices to find numerically that there exists an instance where Lemma 2 yields a greater sum rate than Lemma 1. One can verify that for $(K, M, N, p, q) = (2, 100, 10, 0.2, 0.8)$, Lemma 2 yields a sum rate of 16 and Lemma 1 yields a sum rate of around 12.88. This means that the separation principle fails to hold in the heterogeneous case.

References

[1] S. Kim, S. Mohajer, and C. Suh, "Coding across heterogeneous parallel erasure broadcast channels is useful," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2017, pp. 1883–1887.

[2] R. W. Heath, Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[3] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.

[4] X. Lin *et al.*, "5G new radio: Unveiling the essentials of the next generation wireless access technology," *IEEE Commun. Standards Mag.*, vol. 3, no. 3, pp. 30–37, Sep. 2019.

[5] M. Gatzianas, L. Georgiadis, and L. Tassiulas, "Multiuser broadcast erasure channel with feedback—Capacity and algorithms," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5779–5804, Sep. 2013.

[6] C.-C. Wang, "On the capacity of 1-to-K broadcast packet erasure channels with channel output feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 931–956, Feb. 2012.

[7] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[8] C.-C. Wang and J. Han, "The capacity region of two-receiver multiple-input broadcast packet erasure channels with channel output feedback," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5597–5626, Sep. 2014.

[9] A. F. Dana and B. Hassibi, "The capacity region of multiple input erasure broadcast channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Sep. 2005, pp. 2315–2319.

[10] L. Georgiadis and L. Tassiulas, "Broadcast erasure channel with feedback–Capacity and algorithms," in *Proc. Workshop Netw. Coding, Theory, Appl.*, Jun. 2009, pp. 54–61.

[11] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4418–4431, Jul. 2012.

[12] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[13] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.

[14] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[15] T. Ho *et al.*, "A random linear network coding approach to multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.

[16] A. F. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros, "Capacity of wireless erasure networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 789–804, Mar. 2006.

[17] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "XORs in the air: Practical wireless network coding," *IEEE/ACM Trans. Netw.*, vol. 16, no. 3, pp. 497–510, Jun. 2008.

[18] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4539–4551, Sep. 2010.

[19] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

[20] M. A. Maddah-Ali and U. Niesen, "Decentralized coded caching attains order-optimal memory-rate tradeoff," *IEEE/ACM Trans. Netw.*, vol. 23, no. 4, pp. 1029–1040, Aug. 2015.

[21] R. Pedarsani, M. Ali Maddah-Ali, and U. Niesen, "Online coded caching," *IEEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 836–845, Apr. 2016.

[22] A. Ghorbel, M. Kobayashi, and S. Yang, "Content delivery in erasure broadcast channels with cache and feedback," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6407–6422, Nov. 2016.

[23] Q. Geng, M. Yuan, H. T. Do, Y. Li, R. Wu, and W. Ding, "On the capacity region of broadcast packet erasure relay networks with feedback," in *Proc. IEEE Int. Conf. Commun.*, May 2019, pp. 1–6.

[24] J. Chen, S. Yang, A. Özgür, and A. Goldsmith, "Achieving full DoF in heterogeneous parallel broadcast channels with outdated CSIT," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 4154–4171, Jul. 2016.

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

[26] D. Traskov, N. Ratnakar, D. Lun, R. Koetter, and M. Medard, "Network coding for multiple unicasts: An approach based on linear optimization," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2006, pp. 1758–1762.

[27] W.-C. Kuo and C.-C. Wang, "Robust and optimal opportunistic scheduling for downlink two-flow network coding with varying channel quality and rate adaptation," *IEEE/ACM Trans. Netw.*, vol. 25, no. 1, pp. 465–479, Feb. 2017.

[28] M. Heindlmaier and S. Saeedi Bidokhti, "Capacity regions of two-receiver broadcast erasure channels with feedback and memory," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5042–5069, Jul. 2018.

[29] K. Shanmugam, M. Ji, A. M. Tulino, J. Llorca, and A. G. Dimakis, "Finite-length analysis of caching-aided coded multicasting," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5524–5537, Oct. 2016.

[30] S. Saeedi Bidokhti, M. Wigger, and R. Timo, "Noisy broadcast networks with receiver caching," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 6996–7016, Nov. 2018.

[31] C. Suh and D. N. C. Tse, "Feedback capacity of the Gaussian interference channel to within 2 bits," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2667–2685, May 2011.

[32] S. Kim, I.-H. Wang, and C. Suh, "A relay can increase degrees of freedom in bursty interference networks," *IEEE Trans. Inf. Theory*, vol. 64, no. 6, pp. 4581–4593, Jun. 2018.

[33] S. Kim, S. Mohajer, and C. Suh, "Role of a relay in bursty networks with correlated transmissions," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2016, pp. 2639–2643.

[34] S. Athanasiadou, M. Gatzianas, L. Georgiadis, and L. Tassiulas, "Stable XOR-based policies for the broadcast erasure channel with feedback," *IEEE/ACM Trans. Netw.*, vol. 24, no. 1, pp. 476–491, Feb. 2016.

[35] L. Ozarow and S. Leung-Yan-Cheong, "An achievable region and outer bound for the Gaussian broadcast channel with feedback (corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-30, no. 4, pp. 667–671, Jul. 1984.

[36] A. El Gamal, "The feedback capacity of degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 379–381, May 1978.

Sunghyun Kim received the B.S. and M.S. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology in 2013 and 2015, respectively. From 2019 to 2020, he was a Software Engineer with Samsung Electronics working on the development of neural processing units for mobile system-on-chips. From 2015 to 2019, he was a Research Associate with the Electronics and Telecommunications Research Institute working on the development of 4G and 5G small cell base station software. He is currently with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Soheil Mohajer (Member, IEEE) received the B.Sc. degree in electrical engineering from the Sharif University of Technology, Iran, in 2004, and the M.Sc. and Ph.D. degrees in communication systems from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 2005 and 2010, respectively.

He was affiliated as a Post-Doctoral Researcher with Princeton University from 2010 to 2011, and the University of California at Berkeley, Berkeley, CA, USA, from 2011 to 2013. He is currently a McKnight Land-Grant Associate Professor with the Department of Electrical and Computer Engineering, University of Minnesota, Twin Cities, Minneapolis, MN, USA. His research interests include information theory and its applications in distributed storage systems, wireless networks, bioinformatics, and statistical machine learning. He received the NSF CAREER Award in 2018. He currently serves as an editor for IEEE TRANSACTIONS ON COMMUNICATIONS.

Changho Suh (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from KAIST, in 2000 and 2002, respectively, and the Ph.D. degree in electrical engineering and computer sciences from UC Berkeley in 2011.

From 2002 to 2006, he was with Samsung Electronics. From 2011 to 2012, he was a Post-Doctoral Associate with the Research Laboratory of Electronics, MIT. He is currently an Associate Professor with the School of Electrical Engineering, KAIST. He was a recipient of numerous awards in research and teaching, such as the 2019/2020 AFOSR Grant, the 2019 Google Education Grant, the 2018 IEIE/IEEE Joint Award, the 2015 IEIE Haedong Young Engineer Award, the 2015 Bell Labs Prize finalist, the 2013 IEEE Communications Society Stephen O. Rice Prize, the 2011 David J. Sakrison Memorial Prize, the 2009 IEEE ISIT Best Student Paper Award, and the 2012/2019/2020 Department Teaching Awards. He is a Distinguished Lecturer of the IEEE Information Theory Society from 2020 to 2022, the General Co-Chair of the Inaugural East Asian School of Information Theory 2021, and a member of the Young Korean Academy of Science and Technology. He is also an Associate Editor of Statistical Learning for IEEE TRANSACTIONS ON INFORMATION THEORY, an Editor for IEEE INFORMATION THEORY NEWSLETTER, and a Senior Programme Committee of IJCAI from 2019 to 2021.