

[特別講演]

An Information-Theoretic Perspective on Interference Management

Young-Han KIM

Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA 92093-0407, USA
yhk@ucsd.edu

Abstract Two competing paradigms of interference management are introduced via a few recent developments in network information theory. In the first “distributed network” paradigm, the network consists of autonomous cells with minimal cooperation. For the corresponding mathematical model of the “interference channel,” advanced channel coding techniques are presented, focusing mainly on the sliding-window superposition coding scheme that achieves the performance of simultaneous decoding through point-to-point channel codes and low-complexity decoding. In the second “centralized network” paradigm, the network is a group of neighboring cells connected via noiseless links. For uplink and downlink communications over this “two-hop relay network,” two coding schemes in a dual relationship—noisy network coding and distributed decode-forward—are presented that achieve capacity universally within a finite number of bits per degree of freedom.

1. Introduction

Demand of wireless data is increasing exponentially. According to a recent report [1], the amount of mobile traffic is projected to grow 47% per year and the number of mobile devices is expected to grow 8% per year over the next 5 years. The existing cellular network architecture, which is based on the idea of spatial reuse of frequency among geometrically sparse base stations, does not seem to be sufficient to support a large number of devices, each requiring more and more data. Indeed, a simple information-theoretic argument shows that the achievable rate per user in a network with N users per base station can scale at most as $O((\log N)/N)$. It is hence inevitable that more base stations are deployed to satisfy the projected data demand.

Densification of wireless networks is in fact a historical norm. Cooper’s law [2], [3], which is regarded as a Moore’s law for wireless communications, tracks the number of conversations that can be conducted over a unit area in all of the available wireless spectrum since the days of Marconi. This “law” dictates that over the past 45 years, the areal throughput increased by a factor of one million, in which 1,600 is attributed to adding more base stations.

Accordingly the next-generation cellular networks are expected to deploy many small base stations. While dense

deployment of base stations provides the benefit of bringing broader radio spectrum closer to end users, it also increases the amount of interference from neighboring cells. Consequently, smart management of interference would become one of the key challenges in future wireless communication.

This paper aims to provide an accessible account of two competing paradigms of interference management for cellular networks. In the “distributed network” paradigm, intercell interference is to be mitigated via advanced channel coding techniques with minimal amount of coordination among cells (Section 2.). In the “centralized network” paradigm, multiple cells cooperate in a group via a dedicated network (Section 3.). For both network cases, simple mathematical models are introduced to capture the gist of the problem and information-theoretic analysis are presented for fundamental limits for such network models and coding schemes that achieve those limits. As a disclaimer, this paper provides neither a comprehensive survey nor an in-depth treatment of new developments on this broad topic of interference mitigation. The readers are referred to [4], [5] and the reference therein for a selection of recent developments.

2. Distributed Networks

2.1 Interference Channels

As a simple model of limited coordination among multiple

cells, we study the interference channel with two sender–receiver (user) pairs depicted in Fig 1. In this model, each sender $j = 1, 2$ wishes to communicate the message $M_j \in [1 : 2^{nR_j}] = \{1, 2, \dots, 2^{\lceil nR_j \rceil}\}$ to the desired receiver by encoding it into a codeword $X_j^n = (X_{j1}, \dots, X_{jn})$ and sending it via n transmissions over a channel $p(y_1, y_2 | x_1, x_2)$. Upon receiving the channel output Y_j^n , each receiver $j = 1, 2$ estimates the message M_j . The most well-known example of the two-user interference channel is the Gaussian interference channel, in which the channel outputs are

$$Y_1 = g_{11}X_1 + g_{12}X_2 + Z_1 \quad \text{and} \quad Y_2 = g_{21}X_1 + g_{22}X_2 + Z_2,$$

where g_{jk} is the channel gain from sender k to receiver j , X_1 and X_2 are power constrained inputs, and Z_1 and Z_2 are $N(0, 1)$ noise components. Although we focus on two users, similar analysis can be easily adapted to three or more users, provided that each receiver has one dominant interferer.

The capacity region captures the optimal tradeoff between the data rates R_1 and R_2 that can be reliably communicated over the interference channel when the block length n is arbitrarily large; see [6, Chapter 6] for a precise definition. A computable characterization of the capacity region for the general two-user interference channel is still open. The best known coding scheme and the corresponding single-letter inner bound on the capacity region for the general two-user interference channel are due to Han and Kobayashi [7]. A recent study [8], however, shows that the Han–Kobayashi coding scheme can be outperformed by its multiletter version, which strongly indicates that the quest of finding a computable single-letter characterization of the capacity region may well be a mission impossible.

2.2 Optimal Rate Region Under Random Codes

As a practical alternative to the capacity region, one can consider the highest rates achievable by *point-to-point random code ensembles* and the optimal *maximum likelihood decoding* (MLD) rule. More precisely, let $p = p_1(x_1)p_2(x_2)$ be a given product pmf for the channel input pair (X_1, X_2) . Suppose that the codewords $x^n(m_1)$, $m_1 \in [2^{nR_1}]$, and $x^n(m_2)$, $m_2 \in [2^{nR_2}]$, that constitute the codebook are generated randomly and independently according to $\prod_{i=1}^n p_{X_1}(x_{1i})$ and $\prod_{i=1}^n p_{X_2}(x_{2i})$, respectively. The *optimal rate region* (or the MLD region) $\mathcal{R}^*(p)$ for the p -distributed random code ensembles is the closure of the set of rate pairs (R_1, R_2) such that the sequence of such random code ensembles satisfies

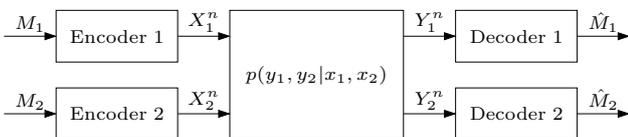


Fig 1 Interference channel with two sender–receiver pairs.

$\lim_{n \rightarrow \infty} \mathbf{E}[P_e^{(n)}(\mathcal{C}_n)] = 0$, where the expectation is with respect to the randomness in codebook generation.

Since the encoder is limited to a specific form, this MLD region is in general smaller than the capacity region. Nonetheless, this notion provides useful insights into optimal communication over interference channels. First, point-to-point random coding is optimal when interference is *weak* or *strong*; see, for example, [9], [10]. Second, many commercial off-the-shelf codes (such as turbo and LDPC codes) are designed with the aim of tracking the performance of point-to-point random codes, and consequently, the MLD region can be viewed as a theoretical performance bound for such commercial codes. Third, the Han–Kobayashi coding scheme [7] can be recast as an instance of point-to-point random coding with four senders and two receivers [11].

The MLD region was originally formulated and characterized for the Gaussian interference channel with Gaussian inputs in [12], [13]. For the general two-user interference channel, it can be characterized [11] as the intersection of two regions \mathcal{R}_1 and \mathcal{R}_2 , each of which characterizes the condition for successful decoding at each receiver; see Fig 2. Here, \mathcal{R}_1 consists of the rate pairs (R_1, R_2) such that

$$R_1 < I(X_1; Y_1 | X_2) \quad \text{and} \quad R_1 + R_2 < I(X_1, X_2; Y_1)$$

or

$$R_1 < I(X_1; Y_1),$$

and \mathcal{R}_2 can be written similarly with $1 \leftrightarrow 2$ substitution.

2.3 Sliding-Window Superposition Coding

Once equipped with the simple characterization of the MLD region for point-to-point random code ensembles, the natural next task is to find a simple scheme that achieves this performance limit. Paralleling Shannon-theoretic random codes and coding-theoretic practical implementations for point-to-point communication, conceptual schemes such as MLD or simultaneous (nonunique) decoding [11] involve exhaustive search over exponentially many codeword pairs and call for practical solutions.

As an alternative to high-complexity multisequence detection, one can restrict the attention to typical point-to-point decoding schemes that involve low-complexity single-sequence detection. The simplest among them is so-called

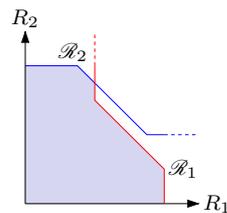


Fig 2 A typical shape of the MLD region for a given random code ensemble.

treating interference as (Gaussian) noise, in which only time-invariant statistics of interfering codewords such as the signal-to-interference noise ratio or the modulation information are incorporated into decoding of the output. Information-theoretically, reliable decoding at receiver 1 is guaranteed if

$$R_1 < I(X_1; Y_1).$$

Depending on the channel condition, successive cancellation decoding can be used, whereby the interfering codeword is first recovered and then cancelled to facilitate the decoding for the desired codeword. Note that in each step of decoding, low-complexity point-to-point decoding is used. Information-theoretically, reliable decoding at receiver 1 is guaranteed if

$$R_2 < I(X_2; Y_1) \quad \text{and} \quad R_1 < I(X_1; Y_1 | X_2).$$

Neither treating interference as noise nor successive cancellation decoding outperforms the other in general. Moreover, both are insufficient to achieve the MLD region in general.

Single-sequence detection can be further improved by changing the encoder design. One can decompose each message into multiple parts as with rate-splitting of Han–Kobayashi coding and recover them along with interfering parts successively; see Fig 3 for an illustration of this idea when M_1 is split into two parts M'_1 and M''_1 . By changing the superposition layers U and V , the decoding order of M'_1 , M''_1 , and M_2 , and the rates for these messages accordingly, this scheme recovers and outperforms both treating interference as noise and successive cancellation decoding.

When there is a single receiver (as in the multiple access channel), this scheme is referred to as *rate-splitting multiple access* [14] and achieves the standard pentagonal region (which is equivalent to the MLD region for a given random coding ensemble) that constitutes the multiple access channel capacity region. When there are multiple receivers, however, the scheme fails to achieve the MLD region. The root cause of this deficiency is suboptimal successive cancellation decoding. Each message part should be recovered correctly at multiple receivers, causing some rate loss that is accumulated over multiple message parts successively recovered. This deficiency can be somewhat remedied by increasing the number of message layers and optimizing over the decoding

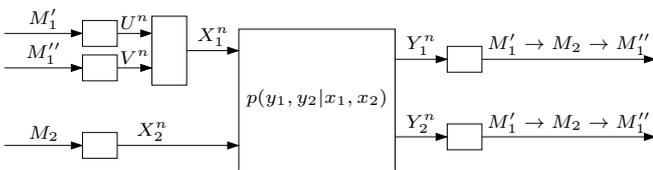


Fig 3 Rate-splitting with successive cancellation decoding.

orders [15], but it can be shown [4] that the scheme still cannot achieve the MLD region.

The *sliding-window superposition coding* scheme [4] overcomes this difficulty by adding an additional dimension to the coding scheme. As illustrated in Fig 4, the scheme has the same superposition layer structure as in the aforementioned rate-splitting scheme. Instead of splitting the messages, however, it uses block Markov coding commonly used in relaying [16] and feedback communication [17]. A stream of message pairs $(M_1(j), M_2(j))$ is communicated over b blocks. In block j , $M_2(j)$ is transmitted via X_2^n . The other message $M_1(j)$ is transmitted over two consecutive blocks j and $j+1$ via $U^n(j)$ and $V^n(j+1)$, as illustrated in Fig 5. In contrast to the conventional *horizontal* superposition coding scheme, this method is called *diagonal* superposition coding.

The receivers use successive cancellation decoding across both blocks and messages. Receiver 1 first recovers $M_2(j)$ from $Y_1^n(j)$ and then cancels it. It then recovers $M_1(j)$ from $Y_1^n(j)$ and $Y_1^n(j+1)$ and then cancels it for decoding of $M_2(j+1)$. The decoding process continues by sliding the decoding window to the next block. Note that this sliding-window decoding scheme traces back to [18] and is commonly used in network decode–forward relaying [19], [20]. It can be shown by the standard argument that decoding is successful at both receivers if

$$\begin{aligned} R_2 &< \min_{j=1,2} I(X_2; Y_j | U), \\ R_1 &< \min_{j=1,2} (I(U; Y_j) + I(V; Y_j | U, X_2)), \end{aligned} \quad (1)$$

which contrasts the stricter condition for successful decoding of the aforementioned rate-splitting scheme:

$$\begin{aligned} R_2 &< \min_{j=1,2} I(X_2; Y_j | U), \\ R_1 &< \min_{j=1,2} (I(U; Y_j) + \min_{j=1,2} I(V; Y_j | U, X_2)). \end{aligned}$$

By optimizing over the two superposition layers U and V , and the decoding orders for two messages, it can be shown [21] that sliding-window superposition coding can achieve every *corner* point of the MLD region. By adding additional superposition layers, a similar scheme can achieve every point of the MLD region, which can be further extended to achieve the entire Han–Kobayashi inner bound in point-to-point decoding [4].

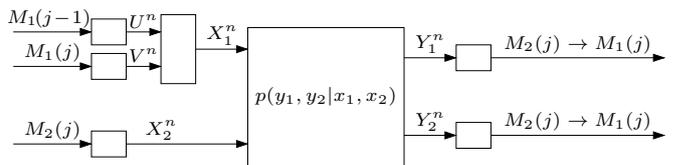


Fig 4 Sliding-window superposition coding.

| Block | 1 | 2 | 3 | 4 | 5 | 6 |
|-------|----------|----------|----------|----------|----------|----------|
| U | | $M_1(1)$ | $M_1(2)$ | $M_1(3)$ | $M_1(4)$ | $M_1(5)$ |
| V | $M_1(1)$ | $M_1(2)$ | $M_1(3)$ | $M_1(4)$ | $M_1(5)$ | |
| X_2 | $M_2(1)$ | $M_2(2)$ | $M_2(3)$ | $M_2(4)$ | $M_2(5)$ | $M_2(6)$ |

Fig 5 Transmission of messages over multiple blocks.

It is instructive to narrow our attention on how a single message is transmitted and compare it with other conventional coded modulation schemes. To be concrete, we consider a pulse amplitude modulation with four levels (4-PAM), which can be viewed as a superposition of two binary phase shift keying (BPSK) layers; see Fig 6. In multilevel coding (MLC) [22], [23], a message M is split into two parts that are separately encoded and transmitted over U and V layers, respectively. This scheme can achieve the point-to-point capacity under the 4-PAM input, but both codewords are short, whose rates should be carefully adapted. When there are multiple receivers, the scheme results in some rate loss since individual codewords should be reliably recovered. In bit-interleaved coded modulation (BICM) [24], [25], the message M is encoded into a single long (interleaved) codeword, which is then transmitted over both U and V layers. The long codeword is recovered as a whole so that there is no rate loss for multiple receivers, but the achievable rate is in general smaller than the point-to-point capacity due to self-interference between U and V layers. The coded modulation scheme from sliding-window superposition coding, which may well be called *sliding-window coded modulation* employs a single long codeword as in BICM and achieves the point-to-point capacity as in MLC, apparently taking the advantages of the two schemes. As a downside, however, communication over multiple blocks incurs error propagation issues and some rate loss due to initialization and finalization. The comparison among these three coded modulation schemes (see Fig 7) may well be captured by *horizontal*, *vertical*, and *diagonal* superposition coding, which parallels horizontal, vertical, and diagonal Bell Labs layered space-time (BLAST) schemes in multiple-input multiple-output (MIMO) communication.

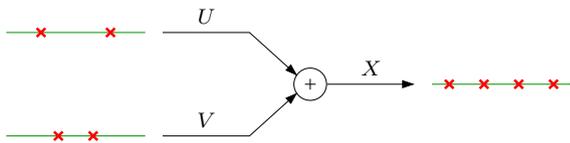


Fig 6 4-PAM as a superposition of BPSK layers.

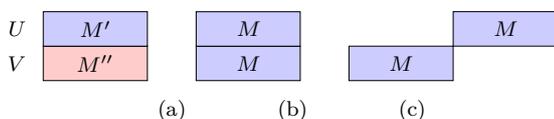


Fig 7 Comparison of MLC, BICM, and sliding-window coded modulation.

This sliding-window coded modulation scheme allows commercial off-the-shelf codes (both encoders and decoders) to be used to track the performance guarantee of the MLD region. Several experiment studies have been performed to test its practical feasibility [26], [27]. The outcomes of these experiments are quite encouraging. According to one of the system-level simulations [27], wireless networks using sliding-window coded modulation can achieve about 55% higher cell-average throughput and 72% higher cell-edge throughput compared to existing networks without interference-aware decoding at the same complexity and networking overhead.

3. Centralized Networks

3.1 Two-Hop Relay Networks

As a model for centralized networks, we consider the cloud radio access network (C-RAN) architecture [28] shown in Fig 8. In this model, several base stations are connected to a cloud-based central processor through wired or wireless fronthaul links. Conceptually, when the fronthaul link capacities are unbounded, this architecture can be interpreted as a “distributed” MIMO, whereby the base stations function as spatially distributed radio heads for the central processing node. For the more realistic situation of limited capacities, the optimal beamforming solution is typically computed, assuming infinite fronthaul capacities, and then compressed individually, which is then applied at the base stations.

As an alternative, we model the C-RAN as a relay network, in which base stations act as relays that summarize the received signals to the central processor (uplink) and transmit the prescribed signals from the central processor (downlink). To be concrete, we model the uplink C-RAN as the multiple access communication problem over a two-hop relay network depicted in Fig 9, where the first hop, namely, the (wireless) channel from K user devices to L radio heads, is a discrete memoryless network $p(y^L|x^K)$, and the second hop, namely, the channel from the radio heads to the central processor, consists of orthogonal (noiseless) links of capacities C_1, \dots, C_L bits per transmission, decoupled from the

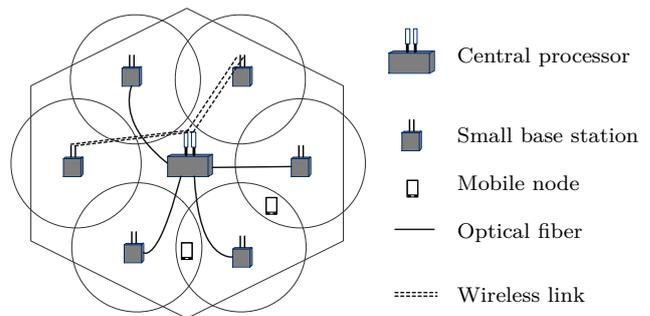


Fig 8 Cloud radio access networks.

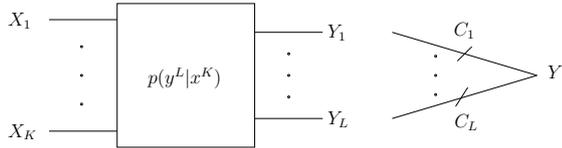


Fig 9 The uplink cloud radio access network model.

first hop. The channel output at the central processor (receiver) is (W_1, \dots, W_L) , where $W_l \in [1 : 2^{nC_l}]$ is a reliable estimate of what relay l communicates to the receiver over n transmissions. The wireless channel (first hop) will be often assumed to be Gaussian, namely,

$$Y^L = GX^K + Z^L,$$

where G is the channel gain matrix and each sender is subjected to average power constraint P . Similarly, the downlink C-RAN is modeled as the broadcast communication problem over a two-hop relay network with the noiseless first hop and the wireless second hop (as reversed from the uplink model).

In each of the uplink and downlink models, the ultimate goal is to characterize the capacity region as a function of the link capacities and find the optimal scheme that achieves the capacity region. Unfortunately, except for the trivial case of $K = L = 1$, the capacity region is not known. Hence, we instead focus on *approximate* capacity region and coding schemes that achieve tight gap that is independent of G and P .

3.2 Cutset Bound

The classical cutset bound on the capacity region [29], [30] can be specialized to the uplink C-RAN model and characterized as the set of rate tuples (R_1, \dots, R_K) such that

$$\sum_{k \in S_1} R_k \leq I(X(S_1); Y(S_2^c) | X(S_1^c), Q) + \sum_{l \in S_2} C_l$$

for all $S_1 \subseteq [1 : K]$ and $S_2 \subseteq [1 : L]$ for some pmf $p(q) \prod_{k=1}^K p(x_k | q)$. For the Gaussian case, this bound can be expressed as

$$\sum_{k \in S_1} R_k \leq \frac{1}{2} \log |PG_{S_2^c, S_1} G_{S_2^c, S_1}^T + I| + \sum_{l \in S_2} C_l \quad (2)$$

for all S_1 and S_2 . Here $G_{S_2^c, S_1}$ is the submatrix of G with row indices in S_2^c and column indices in S_1 .

Similarly, the cutset bound for the downlink C-RAN consists of the rate tuples (R_1, \dots, R_K) such that

$$\sum_{k \in S_2^c} R_k \leq I(X(S_1); Y(S_2^c) | X(S_1^c)) + \sum_{l \in S_1^c} C_l$$

for all S_1 and S_2 for some pmf $p(x^K)$. For the Gaussian case, this bound can be expressed as

$$\sum_{k \in S_2^c} R_k \leq \frac{1}{2} \log |G_{S_2^c, S_1} \Sigma_{S_1 | S_1^c} G_{S_2^c, S_1}^T + I| + \sum_{l \in S_1^c} C_l$$

for all S_1 and S_2 for some covariance matrix Σ with $\Sigma_{ii} \leq P$, $l \in [1 : L]$. Here $\Sigma_{S_1 | S_1^c}$ denotes the conditional covariance matrix of the indices in S_1 given the indices in S_1^c .

3.3 Uplink Multihop Relaying

The network compress-forward coding scheme [20] or the noisy network coding scheme [31] can be specialized to the uplink C-RAN model. In particular, each sender $k \in [1 : K]$ transmits a codeword $X_k^n(M_k)$ and each relay $l \in [1 : L]$ compresses the received sequence Y_l^n into a “compression” sequence $\hat{Y}_l^n(W_l, W_l')$ and forwards W_l . The receiver recovers (M_1, \dots, M_K) based on (W_1, \dots, W_L) . It can be shown [32] that this scheme achieves any rate tuple (R_1, \dots, R_K) such that

$$\sum_{k \in S_1} R_k < I(X(S_1); \hat{Y}(S_2^c) | X(S_1^c)) + \sum_{l \in S_2} C_l - \sum_{l \in S_2} I(Y_l; \hat{Y}_l | X^K)$$

for all S_1 and S_2 for some pmf $\prod_{k=1}^K p(x_k) \prod_{l=1}^L p(\hat{y}_l | y_l)$. For the Gaussian case, this inner bound on the capacity region can be expressed as the set of rate tuples (R_1, \dots, R_K) such that

$$\sum_{k \in S_1} R_k < \frac{1}{2} \log \left| \frac{P}{\sigma^2 + 1} G_{S_2^c, S_1} G_{S_2^c, S_1}^T + I \right| + \sum_{l \in S_2} C_l - \frac{|S_2|}{2} \log \left(1 + \frac{1}{\sigma^2} \right) \quad (3)$$

for all S_1, S_2 for some $\sigma^2 > 0$. A careful analysis [5] reveals that the gap between the cutset outer bound in (2) and the network compress-forward inner bound in (3) is at most $\Delta = (1/2) \log(eL)$ bits per user, regardless of G and P . Note that when $K \ll L$ (as in massive MIMO systems), $cK \log L \ll L$ and thus network compress-forward captures the correct capacity scaling of $O(L)$. Practical implementation of this scheme is yet to be seen.

3.4 Downlink Multihop Relaying

For downlink communication, we can adapt the distributed decode-forward scheme [33], which can be viewed as a “dual” of aforementioned noisy network coding [31] and extends network coding for graphical networks, Marton coding for broadcast channels, and partial decode-forward for relay channels to noisy relay networks. When specialized to the downlink C-RAN model, the sender in distributed decode-forward “precodes” $(X_1^n(W_1), \dots, X_L^n(W_L))$ together with $(U_1^n(M_1, M_1'), \dots, U_K^n(M_K, M_K'))$ as in Marton coding [34] before the actual transmission starts. It then propagates W_1, \dots, W_L to the relays, which then transmit $X_1^n(W_1), \dots, X_L^n(W_L)$, respectively. Receiver $k \in [1 : K]$ recovers M_k based on the received sequence Y_k^n .

It can be shown [5] that this scheme achieves the inner bound that consists of the rate tuples (R_1, \dots, R_K) such that

$$\sum_{k \in S_2^c} R_k < I(X(S_1); U(S_2^c) | X(S_1^c)) + \sum_{l \in S_1^c} C_l \\ - \sum_{k \in S_2^c} I(U_k; X^L | Y_k)$$

for all S_1 and S_2 for some pmf $\prod_{l=1}^L p(x_l) \prod_{k=1}^K p(u_k | x^L)$. For the Gaussian case, this inner bound becomes the set of rate tuples (R_1, \dots, R_K) such that

$$\sum_{k \in S_2^c} R_k < \frac{1}{2} \log \left| \frac{P}{\sigma^2} G_{S_2^c, S_1} G_{S_2^c, S_1}^T + I \right| + \sum_{l \in S_1^c} C_l \\ - \frac{|S_2^c|}{2} \log \left(1 + \frac{1}{\sigma^2} \right),$$

for S_1 and S_2 , which is within $\Delta = (1/2) \log(eKL)$ from the cutset bound. As in the uplink case, this scheme is yet to be implemented in practical settings.

Acknowledgment

The author would like to express his sincere gratitude to Professor Hirosuke Yamamoto for his encouragement and assistance in preparation of this manuscript.

Reference

- [1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2016–2021," Feb. 2017.
- [2] ArrayComm, "Cooper's law," 2011. Webpage available at <http://www.arraycomm.com/technology/coopers-law/>.
- [3] J. Zander and P. Mähönen, "Riding the data tsunami in the cloud: Myths and challenges in future wireless access," *IEEE Commun. Mag.*, vol.51, no.3, pp.145–151, March 2013.
- [4] L. Wang, Y.H. Kim, C.Y. Chen, H. Park, and E. Şaşıoğlu, "Sliding-window superposition coding: Two-user interference channels," 2017. Preprint available at <http://arXiv.org/abs/1701.02345>.
- [5] S. Ganguly and Y.H. Kim, "On the capacity of cloud radio access networks," *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, pp.2063–2067, June 2017.
- [6] A. El Gamal and Y.H. Kim, *Network Information Theory*, Cambridge University Press, Cambridge, 2011.
- [7] T.S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol.27, no.1, pp.49–60, 1981.
- [8] C. Nair, L. Xia, and M. Yazdanpanah, "Sub-optimality of the Han–Kobayashi achievable region for the interference channels," *Proc. IEEE Int. Symp. Inf. Theory*, Hong Kong, China, June 2015.
- [9] H. Satô, "An outer bound to the capacity region of broadcast channels," *IEEE Trans. Inf. Theory*, vol.24, no.3, pp.374–377, 1978.
- [10] S. Liu, C. Nair, and L. Xia, "Interference channels with very weak interference," *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, pp.1031–1035, June/July 2014.
- [11] B. Bandemer, A. El Gamal, and Y.H. Kim, "Optimal achievable rates for interference networks with random codes," *IEEE Trans. Inf. Theory*, vol.61, no.12, pp.6536–6549, Dec. 2015.
- [12] A.S. Motahari and A.K. Khandani, "To decode the interference or to consider it as noise," *IEEE Trans. Inf. Theory*, vol.57, no.3, pp.1274–1283, March 2011.
- [13] F. Baccelli, A. El Gamal, and D.N.C. Tse, "Interference networks with point-to-point codes," *IEEE Trans. Inf. Theory*, vol.57, no.5, pp.2582–2596, May 2011.
- [14] A.J. Grant, B. Rimoldi, R.L. Urbanke, and P.A. Whiting, "Rate-splitting multiple access for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol.47, no.3, pp.873–890, 2001.
- [15] Y. Zhao, C.W. Tan, S.A. Avestimehr, S.N. Diggavi, and G.J. Pottie, "On the maximum achievable sum-rate with successive decoding in interference channels," *IEEE Trans. Inf. Theory*, vol.58, no.6, pp.3798–3820, June 2012.
- [16] T.M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol.25, no.5, pp.572–584, Sept. 1979.
- [17] T.M. Cover and C.S.K. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol.27, no.3, pp.292–298, 1981.
- [18] A.B. Carleial, "Multiple-access channels with different generalized feedback signals," *IEEE Trans. Inf. Theory*, vol.28, no.6, pp.841–850, Nov. 1982.
- [19] L.L. Xie and P.R. Kumar, "An achievable rate for the multiple-level relay channel," *IEEE Trans. Inf. Theory*, vol.51, no.4, pp.1348–1358, 2005.
- [20] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol.51, no.9, pp.3037–3063, Sept. 2005.
- [21] L. Wang, E. Şaşıoğlu, and Y.H. Kim, "Sliding-window superposition coding for interference networks," *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, pp.2749–2753, June/July 2014.
- [22] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol.23, no.3, pp.371–377, May 1977.
- [23] U. Wachsmann, R.F.H. Fischer, and J.B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol.45, no.5, pp.1361–1391, 1999.
- [24] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Comm.*, vol.40, no.5, pp.873–884, May 1992.
- [25] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol.44, no.3, pp.927–946, May 1998.
- [26] H. Park, Y.H. Kim, and L. Wang, "Interference management via sliding-window superposition coding," *Proc. of the International Workshop on Emerging Technologies for 5G Wireless Cellular Networks*, IEEE GLOBECOM, Austin, TX, pp.1057–1061, Dec. 2014.
- [27] K.T. Kim, S.K. Ahn, Y.S. Kim, J. Park, C.Y. Chen, and Y.H. Kim, "Interference management via sliding-window coded modulation for 5G cellular networks," *IEEE Commun. Mag.*, vol.52, no.5, pp.52–60, May 2014.
- [28] T.Q.S. Quek, M. Peng, O. Simeone, and W. Yu, eds., *Cloud Radio Access Networks: Principles, Technologies, and Applications*, Cambridge University Press, 2017.
- [29] A. El Gamal, "The capacity of the physically degraded Gaussian broadcast channel with feedback," *IEEE Trans. Inf. Theory*, vol.27, no.4, pp.508–511, 1981.
- [30] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, second ed., Wiley, New York, 2006.
- [31] S.H. Lim, Y.H. Kim, A. El Gamal, and S.Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol.57, no.5, pp.3132–3152, May 2011.
- [32] A. Sanderovich, O. Somekh, H.V. Poor, and S. Shamai, "Uplink macro diversity of limited backhaul cellular network," *IEEE Trans. Inf. Theory*, vol.55, no.8, pp.3457–3478, Aug. 2009.
- [33] S.H. Lim, K.T. Kim, and Y.H. Kim, "Distributed decode-forward for relay networks," *IEEE Trans. Inf. Theory*, vol.63, no.7, pp.4103–4118, July 2017.
- [34] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol.25, no.3, pp.306–311, 1979.