

Achievable Rate Regions for the Gaussian Broadcast Channel with Fixed Blocklength and Per User Reliability

Paul Sheldon, Besma Smida, Natasha Devroye, and Daniela Tuninetti
Electrical and Computer Engineering Department,
University of Illinois Chicago, Chicago 60607, IL USA
psheld2, smida, devroye, danielat@uic.edu

Abstract—This paper considers second-order achievable rate regions for two-user, single antenna, static Gaussian Broadcast channels with individual user error constraints. Even though the channel is stochastically degraded, we surprisingly show that the “usual” infinite blocklength superposition order in which the user with the higher Signal to Noise Ratio (SNR) is superimposed on top of the lower SNR user, does not necessarily lead to the largest second-order achievable rates. This suggests that a careful consideration of user requirements, in addition to channel conditions, is required in designing broadcasting schemes with a fixed blocklength and per-user reliability requirements. This form of superposition coding combined with rate splitting gives the largest known second order region, as it also includes “concatenate and code”, the transmission of a single message containing bits intended for both users as a special case.

I. INTRODUCTION

The use cases for wireless connectivity continually develop and evolve, leading to network design optimizations over multiple parameters. A common requirement among many of these new applications is guaranteed expected levels of reliability coupled with low latency. The inclusion of Ultra-Reliable Low-Latency Communication (URLLC) by 3GPP as part of the 5G standards speaks to this need.

Traditional multi-user information theoretic analyses that characterize the largest supportable rate region as the packet size (a proxy for delay) tends to infinity and the probability of error vanishes, are of low utility for networks with hard deadlines and exact reliability requirements. This has motivated recent studies on the finite blocklength, or second-order analysis, of the (reliability, rate) trade-off, where the packet size is large but finite and error rate is small but not vanishing [1], [2]. Finite blocklength results often aim to derive approximations to the rate region in the spirit of the so-called “normal approximation,” [3], a refined analysis of how the mutual information density concentrates to its mean. This quantifies how many bits can be sent through the channel within a finite number of channel uses while maintaining a given reliability.

When developing “normal approximations” for multi-user networks, care must be taken to define how reliability is measured. It may be a “global” requirement of the network such that each message, regardless of intended recipient, must meet an expected reliability. Alternatively, it may be

a requirement on a per-user basis. The expected reliability of all messages delivered to a user must meet the reliability requirements of that user.

The global requirement is a good model for networks in which users are undifferentiated. However, in some networks, users might have known, varying reliability requirements. A transmitter that simultaneously sends entertainment information to one user while transmitting critical public safety information to another is an example. This network should not be constrained by a global error probability as a significantly greater error rate in entertainment data might be acceptable.

In this paper, we focus on extending our previous work [4] covering second order achievability results of several multiple access techniques for the downlink, or Additive White Gaussian Noise (AWGN) Broadcast Channel (BC), where one transmitter simultaneously sends different independent messages to various independently operating users under a hard deadline and a global error probability constraint by examining achievability regions for independent error constraints.

There are several well known strategies for creating a code for a BC. A subject of much recent interest is Non-Orthogonal Multiple-Access (NOMA), see e.g. [5]. Superposition coding is a form of NOMA, where codewords are superimposed, starting with the codeword for the user with the lowest SNR [6]. Performance is determined by the Signal to Interference and Noise Ratio (SINR), as each user treats users with better channel qualities as noise. In the capacity achieving scheme for the two-user AWGN BC, the codeword intended for the user with the lowest SNR is the “cloud center” while the codeword for the user with the largest SNR is the “satellite”. The intuitive reason for this ordering is that, when one considers vanishing probability of error, the user with the largest SNR can decode all messages without loss of generality. Surprisingly, our previous work showed that for the second order region this need not be the case. Indeed, we observed that a scheme known as *Concatenate and Code Protocol (CCP)* can outperform, in certain conditions, this form of superposition coding (SUP). In CCP, the transmitter concatenates the users’ message bits in a single data packet before encoding with a point-to-point code; each user decodes the entire CCP packet to extract its own bits. The performance is dictated by the channel with the lowest SNR, akin to the common message capacity of a BC [6, Problem 5.9.(c)]. In SUP, we mimicked the capacity-achieving

strategy by adopting the optimal second-order strategy for the multiple access channel with degraded message sets [7].

The main contribution of this paper is showing that for the two user AWGN BC with per-user error constraints, the capacity achieving ordering of SUP code construction (with the highest SNR user's codewords being chosen as satellites of the other user's codewords) is at times sub-optimal. In fact, the code construction ordering should consider the suitability of each channel for the intended use case. As an example, the user with lower SNR might require less reliability and thus its second-order point-to-point rate can be greater than that of the user with larger SNR.

We also show that by including a rate-splitting step such that cloud center messages can be allocated to either user, the largest known rate region can be achieved, which (as evaluated through second-order approximations) is simply the union of the SUP and CCP regions.

II. CHANNEL MODEL AND PROBLEM FORMULATION

We consider the memoryless two-user static Additive White Gaussian Noise (AWGN) Broadcast Channel (BC). The channel between the base-station sending signal X and the multiple users is modeled as $Y_i = h_i X + N_i$ for user $i \in \{1, 2\}$. We assume without loss of generality that $|h_1| \geq |h_2|$.

The noise $N_i \sim \mathcal{N}(0, 1)$, is the Gaussian random variable at receiver i (assumed to be independent of all other noises, and messages), and h_i is the static channel state at receiver i . The input X is subject to a maximum power constraint $|X|^2 \leq nP$. The SNR at receiver i is $\gamma_i := |h_i|^2$.

We are focused on the case where the base-station must convey information to the users within n channel uses, where n represents a hard deadline for message reception. We are thus interested in the so-called second-order regime [7], [8], where the blocklength n is assumed to be large, but not infinite, and probability of error ε is small but not vanishing in n . We consider a definition of reliability under a per-user error constraint, communication to each user must be accomplished with a probability of error less than a per-user threshold. Toward that end we make use of the following definitions.

Definition 1. Code with per-user reliability

An $(n, M_1, M_2, \varepsilon_1, \varepsilon_2)$ code for the AWGN BC consists of two independent and uniformly distributed messages $m_1 \in [M_1]$ and $m_2 \in [M_2]$ encoded via a single encoder $f_n : [M_1] \times [M_2] \rightarrow \mathbb{R}^n$ and two decoders $\phi_{1,n} : \mathbb{R}^n \rightarrow [M_1]$ and $\phi_{2,n} : \mathbb{R}^n \rightarrow [M_2]$ chosen such that

$$\|f_n(m_1, m_2)\|^2 = nP, \quad \forall (m_1, m_2) \in [M_1] \times [M_2], \quad (1)$$

$$\Pr[(\hat{m}_1) \neq (m_1), (m_1) \text{ sent}] \leq \varepsilon_1, \quad (2)$$

$$\Pr[(\hat{m}_2) \neq (m_2), (m_2) \text{ sent}] \leq \varepsilon_2, \quad (3)$$

where $\hat{m}_j = \phi_{j,n}(Y_j^n)$, $j \in \{1, 2\}$.

Definition 2. Achievability

A pair of non-negative numbers (R_1, R_2) is $(n, \varepsilon_1, \varepsilon_2)$ -achievable if there exists an $(n, M_1, M_2, \varepsilon_1, \varepsilon_2)$ code with

$$\frac{1}{n} \log M_i \geq R_i \quad i \in \{1, 2\}. \quad (4)$$

For the point-to-point (P2P) AWGN channel, $M^*(n, \varepsilon)$ (the largest number of messages that can be sent within n channel uses and with error rate not exceeding ε) behaves as [3], [9]

$$\log M^*(n, \varepsilon) = nC(\gamma) - \sqrt{nV(\gamma)}Q^{-1}(\varepsilon) + O(\log n), \quad (5)$$

where $C(\gamma) = \log(1+\gamma)$ is the channel capacity when the SNR is γ , $Q^{-1}(\cdot)$ is the inverse of the tail distribution function of the standard Gaussian random variable, and $V(\gamma) = \frac{\gamma(2+\gamma)}{(1+\gamma)^2}$ is the channel dispersion [3]. The first two terms in (6) are termed the *normal approximation* that will be denoted by

$$\kappa(n, \gamma, \varepsilon) := C(\gamma) - \sqrt{\frac{V(\gamma)}{n}}Q^{-1}(\varepsilon). \quad (6)$$

We shall refer to (first-order) achievable rate regions as $\mathcal{R}$, where the rate is taken as the limit as $n \rightarrow \infty$. For clarity when discussing second order results we shall use $\lambda_i := \log(M_i) = nR_i$. All second order achievable rate regions will be designated by $\mathcal{M}$.

III. TWO USER SECOND-ORDER ACHIEVABILITY

A. Concatenate and Code Protocol (CCP)

If we concatenate the bits of the two users in one single message and send one codeword as a common message, we obtain the following achievable rate region

$$\mathcal{R}^{(\text{ccp})} = \{ R_1 + R_2 \leq C(\min(\gamma_1, \gamma_2)) \} \quad (7)$$

When the error constraint is considered on a per-user basis, an achievable second-order rate region is given by

$$\mathcal{M}^{(\text{ccp})}(n, \varepsilon_1, \varepsilon_2) = \left\{ \begin{array}{l} \lambda_1 + \lambda_2 \leq n\kappa(n, \gamma_1, \varepsilon_1) \\ \lambda_1 + \lambda_2 \leq n\kappa(n, \gamma_2, \varepsilon_2) \end{array} \right\} \quad (8)$$

This is obtained through a direct application of finite block length point to point results from [3]. As we assume that at each time instant the noises on the two channels are independent, the probability of error for each user is independent.

B. NOMA - Superposition

The capacity region of the scalar static two-user AWGN BC is attained by superposition coding. The region, as derived in [6] for the BC with degraded message sets (which is capacity achieving for the more capable BC [6], and thus also for the stochastically degraded BC), is given by

$$\mathcal{R}^{(\text{sup})} = \bigcup_{\alpha \in [0,1]} \left\{ \begin{array}{l} R_1 \leq C(\alpha\gamma_1) \\ R_2 \leq C\left(\frac{(1-\alpha)\gamma_1}{1+\alpha\gamma_1}\right) \\ R_1 + R_2 \leq C(\gamma_1) \end{array} \right\} \quad (9)$$

where the last constraint in (9) is always redundant for less noisy BCs. The capacity achieving SUP coding scheme has a fixed ordering such that the less capable receiver (user 2) decodes its codeword while treating the codeword intended for the more capable receiver (user 1) as noise, while user 1 is required to recover both messages.

Due to the inclusion of error probabilities, it has not been shown that in the finite blocklength regime the same ordering always results in the largest achievable rate region. When user 2 has a lower reliability requirement, it might be capable

of decoding a larger combined message set within its error constraint.

When defining the second order achievable rate region we must then possibly enlarge the region by taking the union over both possible SUP code construction methods so that

$$\mathcal{M}^{(\text{sup})}(n, \varepsilon_1, \varepsilon_2) = \mathcal{M}^{(\text{sup-1})}(n, \varepsilon_1, \varepsilon_2) \cup \mathcal{M}^{(\text{sup-2})}(n, \varepsilon_1, \varepsilon_2) \quad (10)$$

where $\mathcal{M}^{(\text{sup-1})}(n, \varepsilon_1, \varepsilon_2)$ and $\mathcal{M}^{(\text{sup-2})}(n, \varepsilon_1, \varepsilon_2)$ are the regions attained by encoding in the cloud center user 1 and user 2 messages respectively. If we take α to be the power split between cloud center and satellite regardless of the ordering, the achievable rate regions are given by

$$\begin{aligned} & \mathcal{M}^{(\text{sup-1})}(n, \varepsilon_1, \varepsilon_2) \\ &= \bigcup_{\text{eq(13)}} \left\{ \begin{array}{l} \lambda_1 \leq n\kappa(n, \alpha\gamma_1, \varepsilon_{1,1}) \\ \lambda_2 \leq nC\left(\frac{(1-\alpha)\gamma_2}{1+\alpha\gamma_2}\right) - \sqrt{nV'(\alpha, \gamma_2)}Q^{-1}(\varepsilon_2) \\ \lambda_1 + \lambda_2 \leq n\kappa(n, \gamma_1, \varepsilon_{1,2}) \end{array} \right\} \end{aligned} \quad (11)$$

$$\begin{aligned} & \mathcal{M}^{(\text{sup-2})}(n, \varepsilon_1, \varepsilon_2) \\ &= \bigcup_{\text{eq(14)}} \left\{ \begin{array}{l} \lambda_1 \leq nC\left(\frac{(1-\alpha)\gamma_1}{1+\alpha\gamma_1}\right) - \sqrt{nV'(\alpha, \gamma_1)}Q^{-1}(\varepsilon_1) \\ \lambda_2 \leq n\kappa(n, \alpha\gamma_2, \varepsilon_{2,1}) \\ \lambda_1 + \lambda_2 \leq n\kappa(n, \gamma_2, \varepsilon_{2,2}) \end{array} \right\} \end{aligned} \quad (12)$$

where the union in (11) is over $(\alpha, \varepsilon_{1,1}, \varepsilon_{1,2}) \in \mathbb{R}_+^3$ such that

$$\alpha \in [0, 1], \quad (\text{power split}), \quad (13a)$$

$$1 - F(\varepsilon_{1,1}, \varepsilon_{1,2}) \leq \varepsilon_1 \quad (\text{user 1 error SUP}), \quad (13b)$$

and the union in (12) is over $(\alpha, \varepsilon_{2,1}, \varepsilon_{2,2}) \in \mathbb{R}_+^3$ such that

$$\alpha \in [0, 1], \quad (\text{power split}), \quad (14a)$$

$$1 - F(\varepsilon_{2,1}, \varepsilon_{2,2}) \leq \varepsilon_2 \quad (\text{user 2 error SUP}), \quad (14b)$$

and the functions $F(\cdot, \cdot)$, $V'(\cdot)$ and the quantity r are defined as

$$F(\varepsilon_0, \varepsilon_1) \quad (15a)$$

$$= \begin{cases} (1 - \varepsilon_0)(1 - \varepsilon_1) & r = 0 \\ 1 - \max(\varepsilon_0, \varepsilon_1) & r = 1 \\ 1 - \varepsilon_0 - \int_{-\infty}^{Q^{-1}(\varepsilon_0)} Q\left(\frac{Q^{-1}(\varepsilon_1) - rx}{\sqrt{1-r^2}}\right) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx & \end{cases}$$

$$r = \sqrt{\alpha \frac{2 + \gamma}{2 + \alpha\gamma}}, \quad (15b)$$

$$V'(\alpha, \gamma) = \frac{(1 - \alpha)\gamma(2\alpha\gamma^2 + \gamma + 3\alpha\gamma + 2)}{(\gamma + 1)^2(\alpha\gamma + 1)^2}. \quad (15c)$$

In (11) $\varepsilon_{1,1}$ is the probability of error of user 1 failing to decode the satellite, and $\varepsilon_{1,2}$ is the probability of error for user 1 failing to decode the combined code word given it decoded the satellite message. In (12) $\varepsilon_{2,1}$ and $\varepsilon_{2,2}$ are the probability of error for the same events when user 2 is the user encoded in the satellite.

Similar to [7] (11) and (12) can be proven using random coding on successive power shells, threshold decoding, a change of measure to a more convenient auxiliary product

distribution, and an application of the Berry-Esseen theorem. The major difference from [7] is the addition of a second decoder who is only required to recover the common message. A formal proof is beyond the scope of this paper, but is available in [10].

C. NOMA - Superposition with Rate Splitting (RS)

In the preceding NOMA analysis, one user decodes the message intended for the other user while simultaneously decoding its message encoded in the satellite. The message in the cloud center can be considered a common message in the sense that all users must decode it. This points to the possibility of the broadcaster allocating part of this common message to either user.

By splitting the message set for the user required to decode the satellite codeword (λ_1 for $\mathcal{M}^{(\text{sup-1+rs})}(n, \varepsilon_1, \varepsilon_2)$ and λ_2 for $\mathcal{M}^{(\text{sup-2+rs})}(n, \varepsilon_1, \varepsilon_2)$) as $\lambda_i = \lambda_{i,1} + \lambda_{i,2}$ with $i \in \{1, 2\}$ where $\lambda_{i,1}$ is an allocation of the cloud-center to user i and $\lambda_{i,2}$ is the portion of the message set in the overlay one can perform Fourier-Motzkin elimination on the split message achievable rate regions to arrive at

$$\begin{aligned} & \mathcal{M}^{(\text{sup+rs})}(n, \varepsilon_1, \varepsilon_2) \\ &= \mathcal{M}^{(\text{sup-1+rs})}(n, \varepsilon_1, \varepsilon_2) \cup \mathcal{M}^{(\text{sup-2+rs})}(n, \varepsilon_1, \varepsilon_2) \end{aligned} \quad (16)$$

where

$$\begin{aligned} & \mathcal{M}^{(\text{sup-1+rs})}(n, \varepsilon_1, \varepsilon_2) \\ &= \bigcup_{\text{eq(13)}} \left\{ \begin{array}{l} \lambda_2 \leq nC\left(\frac{(1-\alpha)\gamma_2}{1+\alpha\gamma_2}\right) - \sqrt{nV'(\alpha, \gamma_2)}Q^{-1}(\varepsilon_2) \\ \lambda_1 + \lambda_2 \leq n\kappa(n, \alpha\gamma_1, \varepsilon_{1,1}) \\ \quad + nC\left(\frac{(1-\alpha)\gamma_2}{1+\alpha\gamma_2}\right) - \sqrt{nV'(\alpha, \gamma_2)}Q^{-1}(\varepsilon_2) \\ \lambda_1 + \lambda_2 \leq n\kappa(n, \gamma_1, \varepsilon_{1,2}) \end{array} \right\} \end{aligned} \quad (17)$$

$$\begin{aligned} & \mathcal{M}^{(\text{sup-2+rs})}(n, \varepsilon_1, \varepsilon_2) \\ &= \bigcup_{\text{eq(14)}} \left\{ \begin{array}{l} \lambda_1 \leq nC\left(\frac{(1-\alpha)\gamma_1}{1+\alpha\gamma_1}\right) - \sqrt{nV'(\alpha, \gamma_1)}Q^{-1}(\varepsilon_1) \\ \lambda_1 + \lambda_2 \leq n\kappa(n, \alpha\gamma_2, \varepsilon_{2,1}) \\ \quad + nC\left(\frac{(1-\alpha)\gamma_1}{1+\alpha\gamma_1}\right) - \sqrt{nV'(\alpha, \gamma_1)}Q^{-1}(\varepsilon_1) \\ \lambda_1 + \lambda_2 \leq n\kappa(n, \gamma_2, \varepsilon_{2,2}) \end{array} \right\} \end{aligned} \quad (18)$$

which is at least as large as the union of $\mathcal{M}^{\text{cp}}(n, \varepsilon_1, \varepsilon_2)$ and $\mathcal{M}^{\text{sup}}(n, \varepsilon_1, \varepsilon_2)$ by construction. Note that $\mathcal{M}^{(\text{sup-1+rs})}(n, \varepsilon_1, \varepsilon_2)$ and $\mathcal{M}^{(\text{sup-2+rs})}(n, \varepsilon_1, \varepsilon_2)$ each have two sum-rate constraints and a single rate constraint for the message encoded in the cloud center.

IV. EVALUATION AND ANALYSIS

We first remark that, when evaluating second order regions, care must be taken to correctly interpret the results. In the “normal approximation” framework, there is a penalty term that can change independently of the capacity term through the reliability requirements. In addition, terms on the order of $\log(n)$ are omitted. This implies that the right-hand-side of the various bounds in the second-order regions we have stated may become negative for certain choices of the parameters.

As the λ 's are non-negative by definition, this implies that as evaluated through the second order, no positive rate can be achieved meeting the required reliability for some parameter combinations. This can occur for example when the channel SNR is very small or the required reliability is very high. This is in contrast to traditional information theory where some positive rate can always be achieved with zero error as $n \rightarrow \infty$.

In general, when the λ 's in the second-order regions approach zero for certain choices of the parameters, the “normal approximation” framework must be refined by either adding third-order terms [11], [12] or by evaluating the actual bounds based on tail behavior of the mutual information densities [13].

In light of these considerations in all plots we mark regions that are within $\frac{\log(n)}{n}$ of zero rate to indicate where results might be imprecise.

Remark 1. Superposition error splitting

For the user required to decode both the cloud center and satellite, the total probability of error is greater than the probability of error of either decoding step alone.

More formally, for a fixed ε all values $\varepsilon_0, \varepsilon_1$ satisfying $1 - F(\varepsilon_0, \varepsilon_1) < \varepsilon$ for $r > 0, \varepsilon_0 > 0, \varepsilon_1 > 0$ must have

$$\varepsilon_0 < \varepsilon, \quad \varepsilon_1 < \varepsilon \quad (19)$$

We can see from (15), $1 - F(\varepsilon_0, \varepsilon_1)$ (the total probability of error for the satellite user) is an increasing function of both ε_0 and ε_1 independent of r . Maintaining a constant $1 - F(\varepsilon_0, \varepsilon_1)$ while reducing ε_0 requires an increase in ε_1 and vice versa. As $\lim_{\varepsilon_i \rightarrow \varepsilon_j, \varepsilon_j \rightarrow 0} (1 - F(\varepsilon_i, \varepsilon_j)) = \varepsilon, \forall (i, j) \in \{0, 1\}, i \neq j$, (19) must then be true. Figure 1 shows this relationship for a fixed ε (0.5) for different values of r .

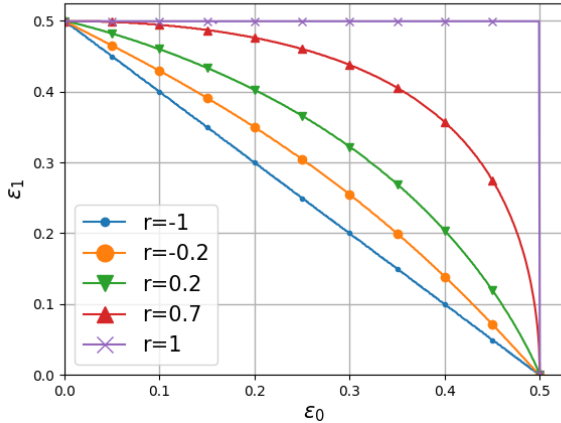


Fig. 1: ε_0 and ε_1 for a fixed $1 - F(\varepsilon_0, \varepsilon_1)$

Remark 2. Practical coding schemes.

We kick-off our analysis of achievable rate regions with a numerical evaluation of optimal coding schemes across a range of channel conditions for specific reliability requirements. In Figures 2 and 3 each point is colored based on the superposition ordering that generates any sum rates greater than the sum rate of $\mathcal{M}^{(ccp)}$.

We only consider points on the boundary where both user rates are greater than $\frac{\log n}{n}$. Excluding these bands close to the axis for each user means that for some channel conditions and reliability requirements no meaningful region enlargement over CCP is possible. This is the small red band clustered around the line corresponding to an equal Finite Block Length Point to Point (FBL P2P) code size.

Figure 2 is an evaluation with equal reliability constraints $\varepsilon_1 = \varepsilon_2 = 0.01$. Here it can be seen that only the capacity achieving ordering of superposition coding ever achieves the largest second order region. Figure 3 is an evaluation where $\varepsilon_2 = 10^{-6}$ and $\varepsilon_1 = 0.01$. The large area of blue underneath the equal SNR line and above the line representing an equal FBL P2P code size represents a large set of channel conditions where the non-standard superposition ordering achieves the largest rate region with this reliability requirements. The behavior seen in these numerical evaluations is examined in subsequent remarks.

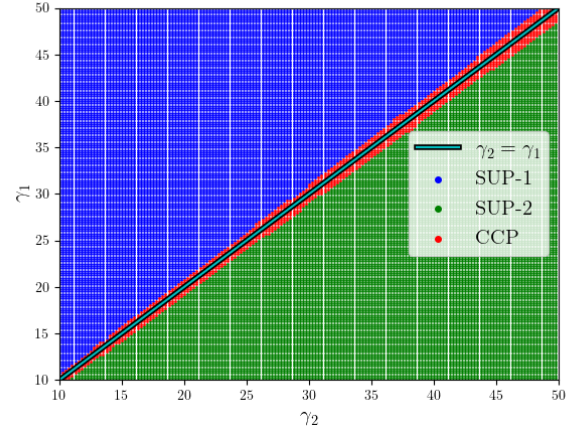


Fig. 2: Superior coding schemes for equal reliability

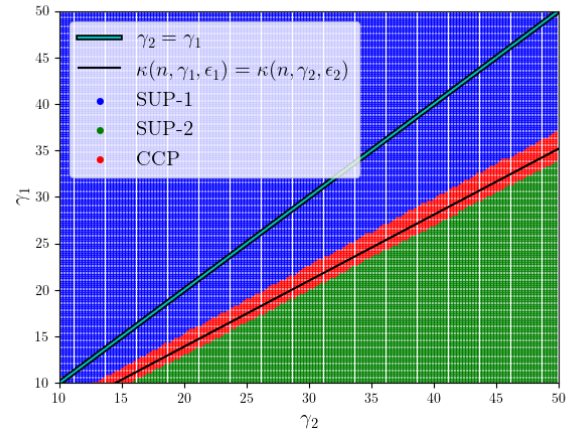


Fig. 3: Superior coding schemes for non-equal reliability

Remark 3. Fixed superposition ordering.

For fixed channel conditions to each user and fixed reliability requirements of each user, only one SUP ordering can ever outperform CCP.

A single ordering of cloud center and satellite determined by the largest (P2P FBL) capacity of each user will result in a $\mathcal{M}^{(\text{sup})}$ whose sum rate can be larger than $\mathcal{M}^{(\text{ccp})}$. The reversed ordering will result in an achievable rate region whose border has a sum rate uniformly less than $\mathcal{M}^{(\text{ccp})}$.

If $\kappa(n, \gamma_1, \varepsilon_1) \leq \kappa(n, \gamma_2, \varepsilon_2)$

$$\mathcal{M}^{(\text{sup-1})}(n, \varepsilon_1, \varepsilon_2) \subset \mathcal{M}^{(\text{ccp})}(n, \varepsilon_1, \varepsilon_2) \quad (20)$$

Alternatively, if $\kappa(n, \gamma_1, \varepsilon_1) \geq \kappa(n, \gamma_2, \varepsilon_2)$

$$\mathcal{M}^{(\text{sup-2})}(n, \varepsilon_1, \varepsilon_2) \subset \mathcal{M}^{(\text{ccp})}(n, \varepsilon_1, \varepsilon_2). \quad (21)$$

This is proven by contradiction, if Equation (20) is not true then this implies that there must be some SUP boundary point whose sum rate is greater than the sum rate of any point on the CCP boundary. IE

$$\kappa(n, \gamma_1, \varepsilon_{1,2}) > \min(\kappa(n, \gamma_1, \varepsilon_1), \kappa(n, \gamma_2, \varepsilon_2)), \quad (22)$$

which is impossible because $\kappa(\cdot)$ is a monotonically increasing function in ε , $\varepsilon_{1,2} < \varepsilon_1$ and $\min(\kappa(n, \gamma_1, \varepsilon_1), \kappa(n, \gamma_2, \varepsilon_2)) = \kappa(n, \gamma_1, \varepsilon_1)$. Proving (21) follows identically.

Figure 4 is a numerical evaluation for $\kappa(n, \gamma_1, \varepsilon_1) = \kappa(n, \gamma_2, \varepsilon_2)$ which shows an extension of Remark 3 where both FBL P2P conditions are true. As the P2P FBL capacity is equal, each achievable rate region for SUP is a subset of CCP. We note that the grey hatched area corresponds to rates for either user 1 or 2 less than or equal to $\frac{\log n}{n}$

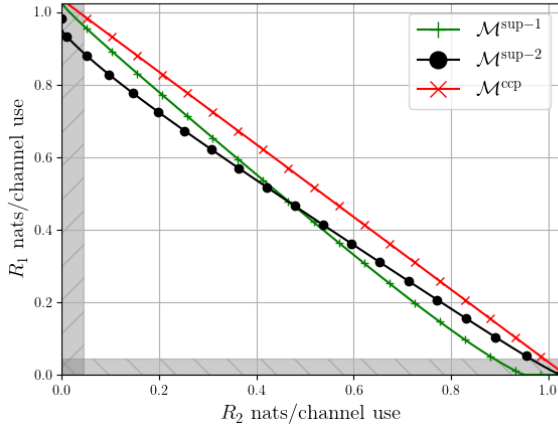


Fig. 4: Equal channel conditions and reliability reqs. ($\gamma_1 = \gamma_2 = 10, \varepsilon_1 = \varepsilon_2 = 0.01$)

When Rate Splitting is included the second sum-rate constraint implies that both superposition rate regions will exactly coincide with $\mathcal{M}^{(\text{ccp})}$ when $\alpha = 0$. This can be seen via direct substitution of $\alpha = 0$ in (17) and (18). Thus we find

If $\kappa(n, \gamma_1, \varepsilon_1) \leq \kappa(n, \gamma_2, \varepsilon_2)$

$$\mathcal{M}^{(\text{sup-1+rs})}(n, \varepsilon_1, \varepsilon_2) \subseteq \mathcal{M}^{(\text{ccp})}(n, \varepsilon_1, \varepsilon_2) \quad (23)$$

If $\kappa(n, \gamma_1, \varepsilon_1) \geq \kappa(n, \gamma_2, \varepsilon_2)$

$$\mathcal{M}^{(\text{sup-2+rs})}(n, \varepsilon_1, \varepsilon_2) \subseteq \mathcal{M}^{(\text{ccp})}(n, \varepsilon_1, \varepsilon_2). \quad (24)$$

Clearly, $\mathcal{M}^{(\text{sup+rs})}$ will only require the use of one ordering of superposition coding as determined by the FBL P2P capacity of each user. In contrast with classical network information theory results, the optimal ordering of the coding scheme does not depend solely on the properties of each channel. Rather it is the suitability of the channel based on the combination of channel conditions and the reliability requirements of the user as utilized in calculating the P2P FBL capacity that determines the optimal ordering. In the following remarks we will refer to the user with a larger FBL P2P capacity as the more suitable user.

Remark 4. Necessity of Superposition Coding

SUP can always generate points on the achievable rate region boundary whose sum-rate is greater than CCP. If $\kappa(n, \gamma_1, \varepsilon_1) > \kappa(n, \gamma_2, \varepsilon_2)$

$$\mathcal{M}^{(\text{sup-1})}(n, \varepsilon_1, \varepsilon_2) \not\subset \mathcal{M}^{(\text{ccp})}(n, \varepsilon_1, \varepsilon_2) \quad (25)$$

If $\kappa(n, \gamma_2, \varepsilon_2) > \kappa(n, \gamma_1, \varepsilon_1)$

$$\mathcal{M}^{(\text{sup-2})}(n, \varepsilon_1, \varepsilon_2) \not\subset \mathcal{M}^{(\text{ccp})}(n, \varepsilon_1, \varepsilon_2). \quad (26)$$

This can be seen by examining the sum rate constraint. $\kappa(n, \gamma_1, \varepsilon_1) > \kappa(n, \gamma_2, \varepsilon_2)$ implies it must be possible for

$$\kappa(n, \gamma_1, \varepsilon_{1,2}) > \kappa(n, \gamma_2, \varepsilon_2) \quad (27)$$

as $\varepsilon_{1,2}$ can be any positive value less than ε_1 . (26) can be shown in an identical fashion.

In Figure 5, $\mathcal{M}^{(\text{sup-1})}$, $\mathcal{M}^{(\text{sup-2})}$, $\mathcal{M}^{(\text{ccp})}$ are evaluated for a channel in which user 1 is the more suitable user ($n = 100, \gamma_1 = 34, \gamma_2 = 30, \varepsilon_1 = 0.01, \varepsilon_2 = 10^{-6}$). It can be seen a standard superposition ordering can achieve larger sum rates than CCP. (i.e. $\mathcal{M}^{(\text{sup-1})} \not\subset \mathcal{M}^{(\text{ccp})}$)

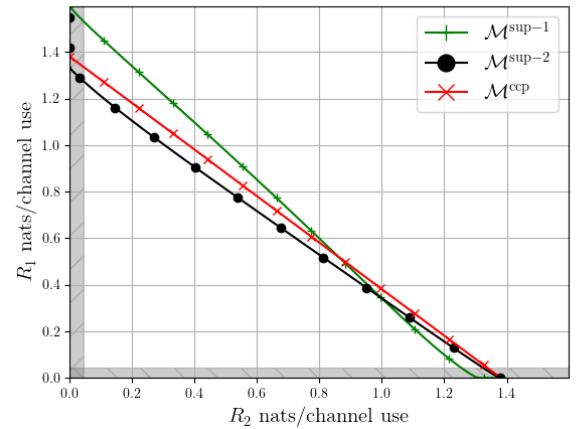


Fig. 5: Individual reliability constraints strong user more suitable

Figure 6 displays $\mathcal{M}^{(\text{sup-1})}$, $\mathcal{M}^{(\text{sup-2})}$, $\mathcal{M}^{(\text{ccp})}$ evaluated for a channel where user 2 is more suitable ($n = 100, \gamma_1 = 34, \gamma_2 = 30, \varepsilon_2 = 0.01, \varepsilon_1 = 10^{-6}$). This is a channel in which user 2 has a very relaxed reliability requirement compared to user 1, but such a network requirement is possible, and as shown a surprising construction of the superposition code is required to

achieve the greatest achievable rate region ($\mathcal{M}^{(\text{sup}-2)} \not\subset \mathcal{M}^{(\text{ccp})}$).

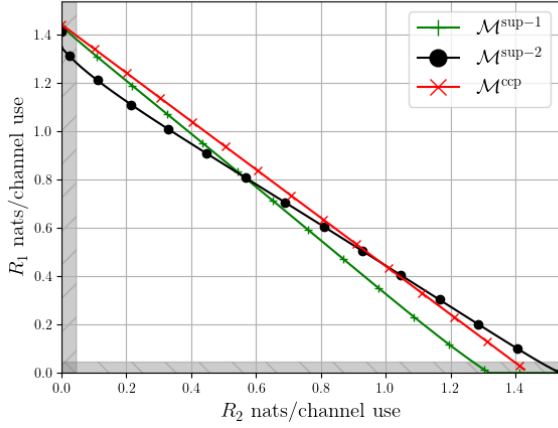


Fig. 6: Individual reliability constraints-non standard ordering

Remark 5. All or Nothing Superposition

As observed through extensive numerical simulation,

$$\mathcal{M}^{(\text{sup}+\text{rs})} = \mathcal{M}^{(\text{sup})} \cup \mathcal{M}^{(\text{ccp})} \quad (28)$$

That is, at the boundary of $\mathcal{M}^{(\text{sup}+\text{rs})}$, the cloud center message set will only be split between the two users in the case SUP coding is not being used ($\alpha = 0$).

Figure 7 shows an example for the same channel conditions and reliability requirements as Figure 5. The rate splitting region can be seen to be exactly equal to the union of the superposition and CCP regions.

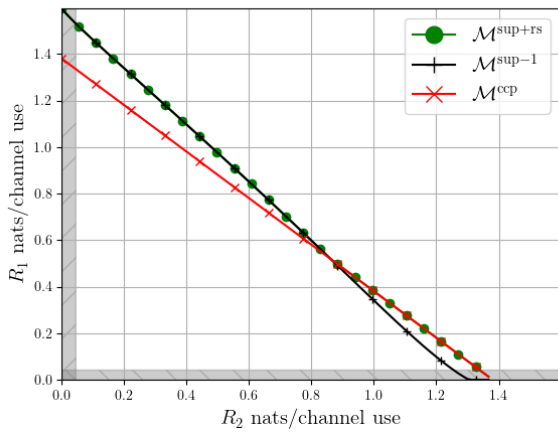


Fig. 7: Individual reliability constraints with rate splitting

V. CONCLUSION

This paper studies the two-user Gaussian broadcast channel at finite blocklength with per-user error constraints. The exploration of multi-user information theoretic problems from a finite blocklength perspective continues to show results

that surprise when compared to capacity (infinite blocklength) results. In our scheme, inspired by the optimal second order scheme for the multiple access channel with degraded message sets but with per-user error constraints, we showed that it is insufficient to consider only the capacity-achieving ordering when constructing a superposition code. In cases where the weak user's reliability requirement is looser than the strong user's, the weak user may be more suited to recovering both codewords. Along this line, it is shown that it is possible that placing the strong user's codeword in the cloud center will outperform the traditional ordering for a range of channel conditions. Superposition coding with rate splitting gives the largest second order region, as it also includes "concatenate and code" as a special case. No tight second order converse bound is yet known. It is therefore unknown if this scheme is second order optimal.

ACKNOWLEDGEMENT:

This work was supported in part by NSF Award 1900911. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the Authors and do not necessarily reflect the views of the NSF.

REFERENCES

- [1] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1711–1726, Sept 2016.
- [2] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai, "A survey on low latency towards 5g: Ran, core network and caching solutions," <https://arxiv.org/abs/1708.02562>, Jul. 2017.
- [3] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [4] P. Sheldon, D. Tuninetti, and B. Smida, "The gaussian broadcast channels with a hard deadline and a global reliability constraint," in *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1–6.
- [5] Y. Yuan, Z. Yuan, and L. Tian, "5g non-orthogonal multiple access study in 3gpp," *IEEE Communications Magazine*, vol. 58, no. 7, pp. 90–96, 2020.
- [6] A. El Gamal and Y.-H. Kim, *Network Information Theory*. UK: Cambridge University Press, 2011.
- [7] J. Scarlett and V. Y. F. Tan, "Second-order asymptotics for the gaussian mac with degraded message sets," *IEEE Transactions on Information Theory*, vol. 61, no. 12, pp. 6700–6718, Dec 2015.
- [8] V. Y. F. Tan and O. Kosut, "On the dispersions of three network information theory problems," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 881–903, Feb 2014.
- [9] M. Hayashi, "Information-spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947 – 4966, Nov. 2009.
- [10] D. Tuninetti, P. Sheldon, B. Smida, and N. Devroye, "On second order rate regions for the static scalar gaussian broadcast channel," 2022. [Online]. Available: <https://arxiv.org/abs/2209.15050>
- [11] M. V. Jamali and H. Mahdavi, "Massive coded-noma for low-capacity channels: A low-complexity recursive approach," *IEEE Transactions on Communications*, vol. 69, no. 6, pp. 3664–3681, 2021.
- [12] V. Y. F. Tan and M. Tomamichel, "The third-order term in the normal approximation for the awgn channel," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2430–2438, 2015.
- [13] S. Chen, A. Collins, G. Durisi, T. Erseghe, G. Carlo, V. Kostina, J. Östman, Y. Polyanskiy, I. Tal, and W. Yang, "Spectre," 2021. [Online]. Available: <https://github.com/yp-mit/spectre>