The Gaussian Broadcast Channels with a Hard Deadline and a Global Reliability Constraint

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

ICC 2021 - IEEE International Conference on Communications | 978-1-7281-7122-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICC42927.2021.9500658

***Abstract*—Recent push for low-latency and high-reliability systems requires derivation of fundamental trade-offs between reliability and rates in multi-user systems with hard deadlines and reliability constraints. Towards this goal, this paper provides a second-order analysis of superposition coding and other orthogonal access schemes for the two-user static Gaussian broadcast channel. Numerical evaluations show that, since reliability and rates are intertwined when a hard-deadline is imposed, the scheduling of resources among users, including the level of reliability for each user, must be done jointly at the physical layer in order to optimize the overall system performance.**

I. Introduction

Internet of Things (IoT) promises to bring wireless connectivity to various applications ranging from spectrum sharing to autonomous vehicles and drones. A successful implementation of this vision calls for wireless communication systems that are able to support the exchange of short data bursts at low-latency and high-reliability. The traditional information theoretic perspective characterizes the largest rate region one can support as the packet size (a proxy for delay) tends to infinity and the probability of error vanishes. These assumptions, inadequate for IoT, have motivated recent studies aimed at understanding 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 finite [1], [2]. Second-order is a sharp characterization of how fast the mutual information density concentrates around its mean, which equals the Shannon capacity, as a function of the block-length n, akin to the rate at which convergence takes place in the central limit theorem.

Incorporating hard deadlines (i.e., finite blocklength) requires re-thinking strategies commonly used in multiaccess techniques. In this paper, we focus on the downlink, or Broadcast Channel (BC), where one transmitter simultaneously sends independent messages to various users under a hard deadline and a global error probability constraint. Currently, many downlink systems orthogonalize the users, as in for example Time Division Multiplexing (TDM) or conceptually similar resource division techniques. These schemes are simple and practically relevant but are not in general capacity achieving. For the single antenna, static Gaussian BC, it is known that the superposition coding scheme is capacity achieving [3]. In this work, we study the behavior of these commonly used schemes, well understood in terms of achievable rates, in the non-asymptotic finite blocklength regime.

In this direction, we study the second-order region [4] of TDM, Concatenate-and-Code Protocol (CCP) [5], and SUPerposition coding (SUP) [3]. In TDM, the transmission time is partitioned into subintervals, each assigned to one user in order to avoid interference; the achieved rate for a user depends on the SNR on its channel and the duration of its subinterval. In CCP, the base station concatenates the users' message bits into a single data packet that is then broadcast to all users; since each user decodes the whole 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 [3, Problem 5.9.(c)]. In SUP, codewords are overlaid on top on one another, starting with the one for the user with the lowest SNR to finish with the one for the user with the largest SNR; here performance is determined by SINR as each user treats users with better channel qualities as noise. The high level question we ask is how to allocate resources, such as power and channel uses, and how to set users' reliability so as to attain the largest possible second-order region.

Breakthroughs have been made in understanding the finite blocklength performance for various point-to-point channels [4], [6]. In the important multi-user setting, only a handful of extensions exist for multiple-access [7], [8] and BCs [9], [10], where more open than settled problems remain. Such finite blocklength results often aim to derive approximations to the rate region in the spirit of the so-called "normal approximation," that is, a refined analysis of how the mutual information density concentrates to its mean which represents the achievable rate [4]. Some communication-theoretic papers investigated the reliability-rate trade-off for BC to be able to compare various orthogonal and non-orthogonal access schemes [11], [12]; we note that the approach of [11], [12], based on a repeated use of point-to-point finite blocklength rate expressions, is not precise for superposition coding. Because of this, here we follow the second-order approach initiated in [9], [10] and extend it to the Gaussian noise case.

Our main contributions are as follows.

- We provide a second-order analysis of SUP by combining elements of the asymmetric BC [9] and Gaussian Multiple Access Channel (MAC) with degraded message sets [10]. The work in [9] only considered the discrete memoryless BC. To the best of our knowledge, this paper is thus the first to investigate the problem in the continuous Gaussian BC setting.
- We show that, in terms of second-order regions, CCP

978-1-7281-7122-7/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: University of Illinois at Chicago Library. Downloaded on May 13,2022 at 20:19:47 UTC from IEEE Xplore. Restrictions apply.

not only beats TDM when the users have comparable channel qualities, but also SUP. This is rather surprising as SUP is capacity achieving, pointing out yet again that conclusions drawn based on first order regions do not necessarily hold for the second-order regions. A possible solution could be to consider the rate splitting approach for more capable BC as outlined in [3, Problem 6.18], which is not pursued here for sake of space.
- Through numerical evaluations, we show that scheduling of resources among users, including the level of reliability, must be done at the physical layer itself and can not be separated as in the capacity setting (where error rate is vanishing and is decoupled from the blocklength).

Compared to our past work [5], here we are not interested in deriving the probability of violating the hard deadline constraint. Instead, we derive the largest possible rate region for a fixed hard deadline such that the global (including all users) error rate remains below a pre-determined constant.

II. Channel Model and Problem Formulation

We consider the K-user complex-valued static Additive White Gaussian Noise (AWGN) Broadcast Channel (BC), where 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 $i \in [K]$. Here $N_i \sim \mathcal{N}(0,1)$ is the proper-complex Gaussian noise at receiver i (assumed to be independent of all other noises), and h_i is the static channel state at receiver i. The input X is subject to the power constraint $\mathbb{E}[|X|^2] \leq 1$. The SNR at receiver $i \in [K]$ is $\gamma_i := |h_i|^2$. The definitions of achievable rates and capacity region are as usual [3].

We are interested in the case where the base-station must convey information to the users within n channel uses, where n represents a hard deadline for the messages to be received, after which they become obsolete. We are thus in the realm of finite block length information theory, and in particular of the so-called second-order regime [9], [10], where the blocklength n is assumed to be large, but not infinite, and the global average probability of error ε is small but not vanishing in n. In this paper we fix the global probability of error to explore the interdependence of reliability to each user in maximizing the second-order region. For most memoryless point-to-point channels, $M^*(n,\varepsilon)$, defined as the largest number of messages that can be sent within n channel uses and with error rate not exceeding ε, behaves as [4], [13]

$$\log M^*(n,\varepsilon) = n\mathsf{C}(\gamma) - \sqrt{n\mathsf{V}(\gamma)}\mathsf{Q}^{-1}(\varepsilon) + O(\log n). \quad (1)$$

For the AWGN channel, $\mathsf{C}(\gamma) := \log(1+\gamma)$ is the capacity (infinite blocklength and vanishing error) when the SNR is γ, $\mathsf{Q}^{-1}(.)$ is the inverse of the tail distribution function of the standard normal random variable, and $\mathsf{V}(\gamma) := \frac{\gamma(2+\gamma)}{(1+\gamma)^2}$ is the *channel dispersion*, or variance of the information density [4]. The term $\sqrt{\frac{\mathsf{V}(\gamma)}{n}}\mathsf{Q}^{-1}(\varepsilon)$ corresponds to the approximate 'rate penalty' incurred by forcing decoding after n channel uses and allowing error $\varepsilon \in (0,1/2)$. The first two terms in (1) are termed the *normal approximation* that will be denoted by

$$\kappa(n,\gamma,\varepsilon) := \mathsf{C}(\gamma) - \sqrt{\frac{\mathsf{V}(\gamma)}{n}}\mathsf{Q}^{-1}(\varepsilon). \quad (2)$$

In the rest of this section we develop expressions akin to (1) for the two-user AWGN BC. Without loss of generality, in the following we let $0 < \gamma_{\rm w} \leq \gamma_{\rm s}$, and may refer to user with SNR $\gamma_{\rm w}$ as the "weak user" and to user with SNR $\gamma_{\rm s}$ as the "strong user." We focus here on the case $K=2$ for simplicity, but our analysis can be extended to any number of users.

A. Cut-set Outer Bound (CUT)

The cut-set bound [3] tells us that the capacity region of the Gaussian BC is within

$$\mathcal{R}^{(\text{cut-set})} = \left\{ \begin{array}{l} R_1 \leq \mathsf{C}(\gamma_{\rm w}) \\ R_1 + R_2 \leq \mathsf{C}(\gamma_{\rm s}) \end{array} \right\}, \quad (3)$$

since the weak user cannot receive at a higher rate than when it is the only user being served by the base-station, and the strong user can mimic the weak user (by adding extra noise to its received signal) and thus must be able to decode both messages. Following this same line of reasoning, one can show that the second-order coding region must satisfy

$$\mathcal{M}^{(\text{cut-set})}(n,\varepsilon) = \left\{ \begin{array}{l} \lambda_1 \leq n\ \kappa(n,\gamma_{\rm w},\varepsilon) \\ \lambda_2 + \lambda_1 \leq n\ \kappa(n,\gamma_{\rm s},\varepsilon) \end{array} \right\}. \quad (4)$$

where λ_i is used to indicate the number of bits that must be conveyed to user $i \in [2]$.

B. Time Division Multiplexing (TDM)

The achievable rate region for TDM with power control is

$$\mathcal{R}^{(\text{tdm})} = \bigcup_{\text{eq(6)}} \left\{ \begin{array}{l} R_1 \leq \tau\mathsf{C}(\alpha_1\gamma_{\rm w}) \\ R_2 \leq (1-\tau)\mathsf{C}(\alpha_2\gamma_{\rm s}) \end{array} \right\}, \quad (5)$$

where the union in (5) is over $(\alpha_1,\alpha_2,\tau) \in \mathbb{R}^3_+$ such that

$$\tau\alpha_1 + (1-\tau)\alpha_2 \leq 1, \qquad \text{(power constraint)}, \quad (6a)$$
$$\tau \in [0,1], \qquad \text{(time division)}. \quad (6b)$$

Since the channel is memoryless and communication for each user occurs on a separate time window, deriving the second-order region $\mathcal{M}^{(\text{tdm})}$ from $\mathcal{R}^{(\text{tdm})}$ in (5) is straightforward (i.e., two separate single-user decoding operations). We have

$$\mathcal{M}^{(\text{tdm})}(n,\varepsilon) = \bigcup_{\text{eq(6),eq(8)}} \left\{ \begin{array}{l} \lambda_1 \leq n\ \kappa(\tau n,\alpha_1\gamma_{\rm w},\epsilon_1) \\ \lambda_2 \leq n\ \kappa((1-\tau)n,\alpha_2\gamma_{\rm s},\epsilon_2) \end{array} \right\} \quad (7)$$

where the union in (7) is over $(\alpha_1,\alpha_2,\tau,\epsilon_1,\epsilon_2) \in \mathbb{R}^5_+$ such that they satisfy the constraints in (6) and in addition

$$1 - (1-\epsilon_1)(1-\epsilon_2) \leq \varepsilon, \qquad \text{(global error)}. \quad (8)$$

The meaning of the global error constraint in (8) is as follows: there are two decoding operations, one per user; user $i \in [2]$ is successful with probability $1-\epsilon_i$; since the noises are independent, the error events are independent, and thus the overall probability of success (equal to one minus the probability of error) is the product of the two individual probabilities of

Authorized licensed use limited to: University of Illinois at Chicago Library. Downloaded on May 13,2022 at 20:19:47 UTC from IEEE Xplore. Restrictions apply.

success. Optimizing over the individual probabilities of errors (ϵ_1, ϵ_2) subject to the constraint in (8) provides another degree of freedom to boost the finite blocklength performance.

C. Concatenate-and-Code (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

$$\mathcal{R}^{(\text{ccp})} = \left\{ \; R_1 + R_2 \le \mathsf{C}(\min(\gamma_\text{w}, \gamma_\text{s})) \; \right\}, \tag{9}$$

which at finite blocklength reads

$$\mathcal{M}^{(\text{ccp})}(n, \varepsilon) = \bigcup_{\text{eq(8)}} \left\{ \begin{array}{l} \lambda_1 + \lambda_2 \le n \; \kappa(n, \gamma_\text{w}, \epsilon_1) \\ \lambda_1 + \lambda_2 \le n \; \kappa(n, \gamma_\text{s}, \epsilon_2) \end{array} \right\}, \tag{10}$$

since there are two decoding operations, one per receiver, which must satisfy the global error constraint in (8). We would like to point out that the global error constraint for $\mathcal{M}^{(\text{ccp})}$ in (10) is because we assumed that at each time instant the noises on the two channels are *independent*.

D. Superposition Coding (SUP)

The capacity region of the two-user AWGN BC is attained by SUP. The SUP region, as derived in [3] for the BC with degraded message sets (which is capacity achieving for more capable BC [3], and thus also for the stochastically degraded AWGN BC considered here), is given by

$$\mathcal{R}^{(\text{sup})} = \bigcup_{\alpha \in [0,1]} \left\{ \begin{array}{l} R_1 \le \mathsf{C}\left(\frac{(1-\alpha)\gamma_\text{w}}{1+\alpha\gamma_\text{w}}\right) \\ R_2 \le \mathsf{C}(\alpha\gamma_\text{s}) \\ R_2 + R_1 \le \mathsf{C}(\gamma_\text{s}) \end{array} \right\}. \tag{11}$$

Note that the last constraint in (11) is always redundant as far as capacity is concerned. The second-order region with SUP is derived in Appendix and can be expressed as

$$\mathcal{M}^{(\text{sup})}(n, \varepsilon) = \bigcup_{\text{eq(13)}} \left\{ \begin{array}{l} \lambda_1 \le n\mathsf{C}\left(\frac{(1-\alpha)\gamma_\text{w}}{1+\alpha\gamma_\text{w}}\right) - \sqrt{nV_{11}}Q^{-1}(\epsilon_1) \\ \lambda_2 \le n \; \kappa(n, \alpha\gamma_\text{s}, \epsilon_2) \\ \lambda_2 + \lambda_1 \le n \; \kappa(n, \gamma_\text{s}, \epsilon_3) \end{array} \right\} \tag{12}$$

where the union in (12) is over $(\alpha, \epsilon_1, \epsilon_2, \epsilon_3) \in \mathbb{R}_+^4$ such that

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

$$1 - (1 - \epsilon_1)\mathsf{F}(\epsilon_2, \epsilon_3) \le \varepsilon \quad \text{(global error SUP)}, \tag{13b}$$

where (derivations can be found in Appendix) we have

$$\mathsf{F}(\epsilon_2, \epsilon_3) \tag{14}$$

$$= \begin{cases} (1 - \epsilon_2)(1 - \epsilon_3) & r = 0 \\ 1 - \max(\epsilon_2, \epsilon_3) & r = 1 \\ \int_{Q^{-1}(\epsilon_2)}^{\infty} Q\left(\frac{Q^{-1}(\epsilon_3) - rx}{\sqrt{1-r^2}}\right) \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}} \mathrm{d}x \; - \epsilon_2, & r \in (0, 1) \end{cases},$$

$$r = \sqrt{\alpha \frac{2 + \gamma_\text{s}}{2 + \alpha\gamma_\text{s}}}, \tag{15}$$

$$V_{11} = \frac{(1 - \alpha)\gamma_\text{w}(2\alpha\gamma_\text{w}^2 + \gamma_\text{w} + 3\alpha\gamma_\text{w} + 2)}{(\gamma_\text{w} + 1)^2(\alpha\gamma_\text{w} + 1)^2}. \tag{16}$$

Note that the first constraint in (12) is not $\kappa(n, \frac{(1-\alpha)\gamma_\text{w}}{1+\alpha\gamma_\text{w}}, \epsilon_1)$, as one would be tempted to guess based on a repeated use of the point-to-point normal approximation in (1); instead, the dispersion term is V_{11} in (16) and not $\mathsf{V}(\frac{(1-\alpha)\gamma_\text{w}}{1+\alpha\gamma_\text{w}})$. The term r in (15) represents the correlation coefficient between the equivalent noises in the stripping decoder at the strong receiver with resulting error rate $\mathsf{F}(\epsilon_2, \epsilon_3)$ in (14).

III. Numerical Evaluations

In this section, we numerically compare the performance of the schemes in Section II, as analytical closed-form solutions for the largest second-order regions remains elusive. The figures are for hard-deadline equal to $n = 100$ channels uses and global reliability $\varepsilon = 10^{-5}$. We also show the optimal values of some of the parameters in each second-order region.

Second-Order Regions: In terms of first-order rate regions we know that: (a) SUP attains the capacity region and coincides with TDM only for equal SNRs, (b) the cut-set bound is loose, and (c) CCP is always the worst among the three achievable strategies. We see that this 'ordering' does not hold in general for the second-order regions. From Fig. 1(left side) for equal SNRs $\gamma_\text{w} = \gamma_\text{s} = 10$, we surprisingly observe that CCP *uniformly*, i.e., over the whole range of rates, outperforms TDM and SUP. We observe an ordering as in the first order rate regions only when the SNRs are sufficiently different, as in Fig. 1(right side) for $\gamma_\text{w} = 10 \ll \gamma_\text{s} = 100$. This suggests that optimal second-order regions must include a *common message*, a codeword that is decoded by both users. Another way to look at this phenomena is to think of CCP as another form of 'multiple access' where the users are allocated distinct positions in the binary information codeword *before* being encoded by the *same* error correcting code; this is instead of occupying distinct resources in the same domain (as in TDM) *after* being encoded by *different* error correcting codes. The advantage of CCP can be easily understood as follows: the resulting codeword is longer (compared to each of the TDM codewords) and thus suffers less of the second-order penalty (quantified by the dispersion) that vanishes with the codeword length. It should be noted that for TDM the first user would receive its message earlier as compared to having to wait for the combined codeword as in CCP, but from the transmitter's point of view all users are served within the same global deadline.

TDM: Fig. 2 shows the optimal values of the parameters in (7) vs. the rate of the weak user, that is, the optimal power allocation (α_1, α_2) (left side) and reliability allocation (ϵ_1, ϵ_2) (right side) attaining the largest λ_2 in Fig 1(right side) vs $\lambda_1 \in [0, \kappa(n, \gamma_\text{w}, \varepsilon)]$. While the optimal power allocation does not appear to have a monotonic behavior as a function of rate in general, the optimal reliability allocation does. In particular, the error rate ϵ_u monotonically increases with rate λ_u, for each user $u \in [2]$, while the global error rate does not exceed $\varepsilon = 10^{-5}$. This can be understood from (1) since the second-order penalty decreases when ϵ_u increases.

CCP: The region $\mathcal{M}^{(\text{ccp})}(n, \epsilon)$ in (10) is equivalent to

$$\{\lambda_1 + \lambda_2 \le n \; \kappa(n, \gamma_\text{w}, \epsilon_1)\} \quad \text{for} \tag{17}$$

Authorized licensed use limited to: University of Illinois at Chicago Library. Downloaded on May 13,2022 at 20:19:47 UTC from IEEE Xplore. Restrictions apply.

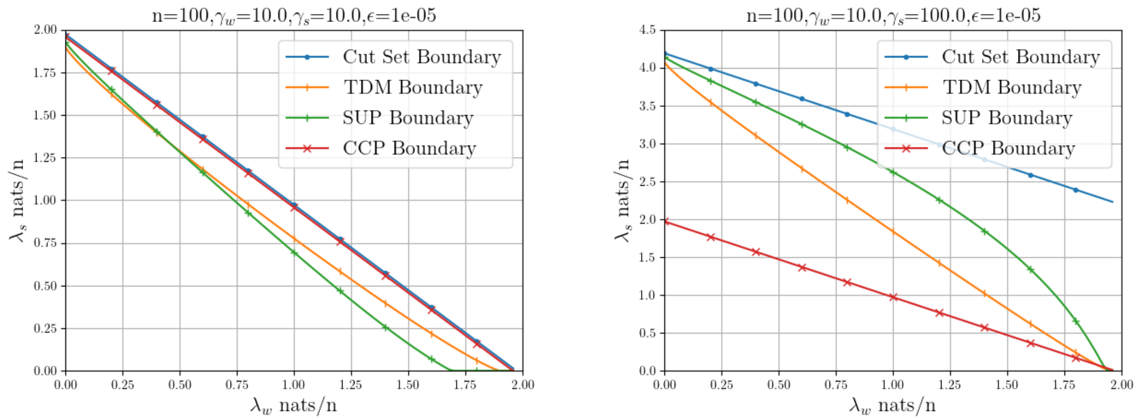

Fig. 1: Second-order regions. Left: same SNR $\gamma_{\mathrm{w}} = \gamma_{\mathrm{s}} = 10$. Right: different SNRs $\gamma_{\mathrm{w}} = 10 \ll \gamma_{\mathrm{s}} = 100$.

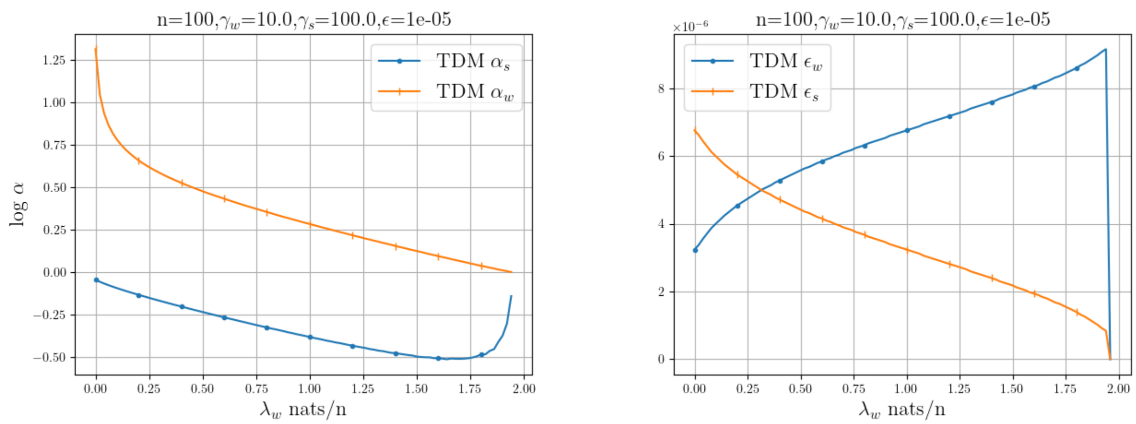

Fig. 2: TDM optimal parameters vs. λ_1 for $\gamma_{\mathrm{w}} = 10 \ll \gamma_{\mathrm{s}} = 100$. Left: power allocation. Right: reliability allocation.

$$\epsilon_1 \in [0, \varepsilon] : \kappa(n, \gamma_{\mathrm{w}}, \epsilon_1) = \kappa\left(n, \gamma_{\mathrm{s}}, \frac{\varepsilon - \epsilon_1}{1 - \epsilon_1}\right). \tag{18}$$

The optimal ϵ_1 and $\epsilon_2 = \frac{\varepsilon - \epsilon_1}{1 - \epsilon_1}$ in (18) only depend on $(n, \gamma_{\mathrm{s}}, \gamma_{\mathrm{w}})$ but not on the actual rate pair (λ_1, λ_2); this implies that the optimal reliability allocation is constant across the entire second-order region with $\epsilon_{\mathrm{w}} \geq \epsilon_{\mathrm{s}}$ for $\gamma_{\mathrm{s}} \geq \gamma_{\mathrm{w}}$. The overall CCP performance is determined by that of the bottleneck weak user, so it intuitively makes sense to relax the error rate constraint for the weak user as much as possible so as to make the second-order rate as large as possible.

SUP: Here the information for the strong user is overlaid as perturbations about the codeword intended for the weak user. The parameter α represents the fraction of the total available power allocated to the codeword for the strong user. In Fig. 3 we show the optimal values of the parameters in (12) vs. the rate of the weak user, that is, the optimal power split and resulting correlation coefficient (α, r) (left side) and reliability allocation $(\epsilon_1, 1 - \mathsf{F}(\epsilon_2, \epsilon_3), \epsilon_3)$ (right side) attaining the largest λ_2 in Fig 1(right side) vs $\lambda_1 \in [0, \kappa(n, \gamma_{\mathrm{w}}, \varepsilon)]$. As the correlation coefficient r in (15) is close to one for the whole range of rates (left side), from (14) we get $1 - \mathsf{F}(\epsilon_2, \epsilon_3) \approx \max(\epsilon_2, \epsilon_3) = \epsilon_2$ (right side).

We find that, in a manner similar to TDM in Fig. 3(right side), the error requirement for the weak user must be relaxed as the rate of the weak user increases. As expected, the value of ϵ_3 (the error rate for decoding the message of the weak user by the strong receiver) is the most stringent; this is so because a decoding error at this stage implies that the weak codeword cannot be stripped from the received signal and the intended message will be decoded in error with very high probability.

IV. Conclusion

In this paper we provided a second-order analysis of superposition coding for the two user Gaussian broadcast channel. We showed that, in terms of second-order regions, concatenate-and-code is superior to both superposition coding and time division when the users experience comparable channel qualities. Through numerical evaluation, we showed that resource scheduling must happen at the physical layer, including also the amount of reliability allocated to the users.

This work was supported in part by NSF Award 1900911.

Appendix

We aim to find the covariance matrix for the second-order approximation of SUP for the two-user Gaussian BC [9],

Authorized licensed use limited to: University of Illinois at Chicago Library. Downloaded on May 13,2022 at 20:19:47 UTC from IEEE Xplore. Restrictions apply.

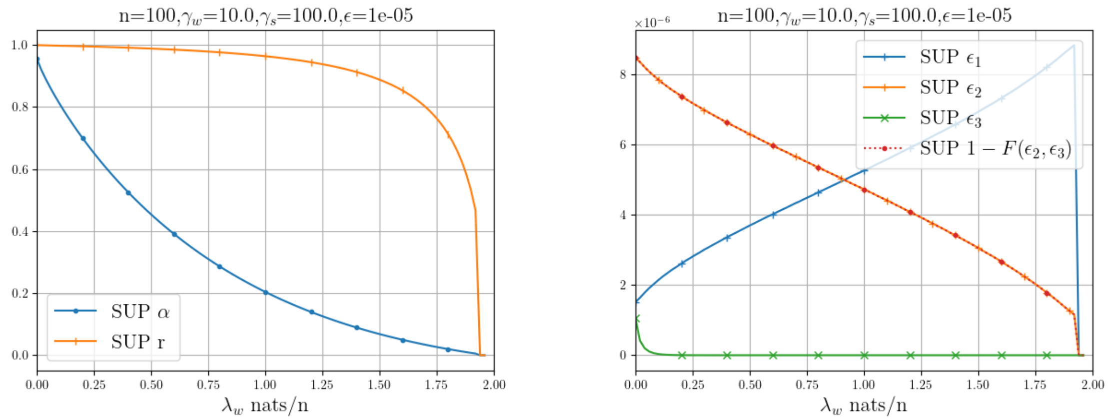

Fig. 3: SUP optimal parameters vs. λ_1 for $\gamma_{\rm w} = 10 \ll \gamma_{\rm s} = 100$. Left: power split. Right: reliability allocation.

[10]. The results are derived for the real-valued case; for the complex-valued case, the rates and the dispersions need to be multiplied by two. In order to introduce the reader to the notation, we first revisit the point-to-point case in [4].

A. Digression on the point-to-point AWGN channel

For the real-valued single-user AWGN $Y = X + Z$ with input $X \sim \mathcal{N}(0, P)$ independent of the noise $Z \sim \mathcal{N}(0, \sigma^2)$, the mutual information density random variable is

$$i(y; x) = \frac{1}{2}\ln\left(1 + \frac{P}{\sigma^2}\right) - \frac{(z)^2}{2\sigma^2} + \frac{(x+z)^2}{2(P+\sigma^2)}\Big|_{z:=y-x}; \quad (19)$$

define the following zero-mean random variable conditioned on $X = x$

$$T := i(Y; x) - \mathbb{E}\left[i(Y; x)|X = x\right] \quad (20)$$

$$= \frac{2N\text{sign}(x)\sqrt{t(x)s} + (1 - N^2)s}{2(s+1)}, \quad (21)$$

$$t(x) := \frac{x^2}{P},\ \ s := \frac{P}{\sigma^2},\ \ N := \frac{Z}{\sigma} \sim \mathcal{N}(0, 1), \quad (22)$$

whose conditional (on $X = x$) variance is

$$V(x) := \mathbb{E}\left[T^2|X = x\right] = \frac{(2t(x) + s)s}{2(s+1)^2}; \quad (23)$$

finally, by picking the codewords on the power shell [10], which implies $\Pr[t(X) = 1] = 1$ in (23), we get

$$V = \mathbb{E}\left[V(X)\right] = \frac{(2+s)s}{2(s+1)^2} = \mathsf{V}(s), \quad (24)$$

which is the channel dispersion in [4].

B. Two-user AWGN BC

Let $\sigma_1^2 > \sigma_2^2$ (i.e., user 1 is the weak receiver and user 2 the strong receiver). From the “ABC achievable region” in [9] with Gaussian inputs we have for $k \in [1:2]$

$$U = U_1 \sim \mathcal{N}(0, \xi_1^2),\ \text{(cloud center)}; \quad (25)$$

$$X = U_1 + U_2,\ U_2 \sim \mathcal{N}(0, \xi_2^2),\ \text{(superposition)}; \quad (26)$$

$$Y_k | U = u, X = x \sim \mathcal{N}(x, \sigma_k^2); \quad (27)$$

$$Y_k | U = u \sim \mathcal{N}(u, \xi_2^2 + \sigma_k^2); \quad (28)$$

$$Y_k \sim \mathcal{N}(0, \xi_1^2 + \xi_2^2 + \sigma_k^2),\ P := \xi_1^2 + \xi_2^2. \quad (29)$$

We aim to find the covariance matrix of mutual information density random vector $[i(Y_1; u), i(Y_2; x|u), i(Y_2; x, u)]$, whose mean values are the achievable rates given by

$$I_3^{(k)} = \frac{1}{2}\ln\left(1 + \frac{\xi_1^2 + \xi_2^2}{\sigma_k^2}\right)\ \text{(sum-rate, user } k = 2), \quad (30)$$

$$I_2^{(k)} = \frac{1}{2}\ln\left(1 + \frac{\xi_2^2}{\sigma_k^2}\right)\ (R_2,\ \text{user } k = 2), \quad (31)$$

$$I_1^{(k)} = I_3^{(k)} - I_2^{(k)},\ (R_1,\ \text{user } k = 1). \quad (32)$$

Similarly to the single-user case in Section A, we write

$$i_3^{(k)}(y; x, u) = I_3^{(k)} - \frac{(z)^2}{2(\sigma_k^2)} + \frac{(x+z)^2}{2(\xi_1^2 + \xi_2^2 + \sigma_k^2)}|_{z=y-x}; \quad (33)$$

$$i_2^{(k)}(y; x|u) = I_2^{(k)} - \frac{(z)^2}{2(\sigma_k^2)} + \frac{(x-u+z)^2}{2(\xi_2^2 + \sigma_k^2)}|_{z=y-x}, \quad (34)$$

$$i_1^{(k)}(y; u) = i_3^{(k)}(y; x, u) - i_2^{(k)}(y; x|u). \quad (35)$$

In order to characterize the second-order region we need to evaluate, conditioned on $(X = x, U = u)$, the covariance matrix of $[i_1^{(1)}(Y_1; u), i_2^{(2)}(Y_2; x|u), i_3^{(2)}(Y_2; x, u)]$; then we need to average over (X, U) which we pick as for the MAC with degraded message sets in [10]. Now, let

$$N_k := \frac{Z_k}{\sigma_k} \sim \mathcal{N}(0, 1), k \in [1:2]; \quad (36)$$

$$t_3 := \frac{x^2}{\xi_1^2 + \xi_2^2},\quad s_3^{(k)} := \frac{\xi_1^2 + \xi_2^2}{\sigma_k^2} = s_1^{(k)} + s_2^{(k)}, \quad (37)$$

$$t_2 := \frac{(x-u)^2}{\xi_2^2},\quad s_2^{(k)} := \frac{\xi_2^2}{\sigma_k^2}, \quad (38)$$

$$t_1 := \frac{u^2}{\xi_1^2},\quad s_1^{(k)} := \frac{\xi_1^2}{\sigma_k^2}, \quad (39)$$

then

$$T_3^{(k)} = \frac{2N_k\sqrt{s_3^{(k)}}\ \text{sign}(x)\sqrt{t_3} + (1 - N_k^2)s_3^{(k)}}{2(s_3^{(k)} + 1)}; \quad (40)$$

Authorized licensed use limited to: University of Illinois at Chicago Library. Downloaded on May 13,2022 at 20:19:47 UTC from IEEE Xplore. Restrictions apply.

$$T_2^{(k)} = \frac{2N_k\sqrt{s_2^{(k)}}\ \mathrm{sign}(x-u)\sqrt{t_2} + (1-N_k^2)s_2^{(k)}}{2(s_2^{(k)}+1)}\ ; \quad (41)$$

$$T_1^{(k)} = T_3^{(k)} - T_2^{(k)}. \quad (42)$$

Define the "noise correlation coefficient"

$$\rho = \begin{cases} \mathbb{E}\left[N_i N_j\right] & i \neq j \\ 1 & i = j \end{cases}. \quad (43)$$

Then, the expectations conditioned on $(X,U) = (x,u)$ (not explicitly indicated for sake of notation compactness) are

$$\mathbb{E}[T_3^{(i)}T_3^{(j)}] = \rho\frac{t_{33}(x,u)\sqrt{s_3^{(i)}s_3^{(j)}}}{(s_3^{(i)}+1)(s_3^{(j)}+1)} + \frac{\rho^2}{2}\frac{s_3^{(i)}s_3^{(j)}}{(s_3^{(i)}+1)(s_3^{(j)}+1)} \quad (44)$$

$$\mathbb{E}[T_2^{(i)}T_2^{(j)}] = \rho\frac{t_{22}(x,u)\sqrt{s_2^{(i)}s_2^{(j)}}}{(s_2^{(i)}+1)(s_2^{(j)}+1)} + \frac{\rho^2}{2}\frac{s_2^{(i)}s_2^{(j)}}{(s_2^{(i)}+1)(s_2^{(j)}+1)} \quad (45)$$

$$\mathbb{E}[T_3^{(i)}T_2^{(j)}] = \rho\frac{t_{23}(x,u)\sqrt{s_2^{(i)}s_2^{(j)}}}{(s_3^{(i)}+1)(s_2^{(j)}+1)} + \frac{\rho^2}{2}\frac{s_3^{(i)}s_2^{(j)}}{(s_3^{(i)}+1)(s_2^{(j)}+1)} \quad (46)$$

$$t_{33}(x,u) = \frac{x^2}{\xi_1^2+\xi_2^2}, \quad t_{22}(x,u) = \frac{(x-u)^2}{\xi_2^2}, \quad (47)$$

$$t_{23}(x,u) = \frac{x(x-u)}{\xi_2^2} \quad (48)$$

note $\mathrm{sign}(x)\mathrm{sign}(x-u)\ \sqrt{t_3 t_2 \frac{s_3^{(i)}}{s_2^{(i)}}} = \frac{x(x-u)}{\xi_2^2}$. Finally, by averaging over (X,U) and picking codes on the power shells as for the MAC with degraded message sets [10]. (i.e., the terms in (47) and (48) are equal to one, we get

$$\mathbf{V} = \mathrm{Cov}\begin{bmatrix} T_3^{(1)} - T_2^{(1)} \\ T_2^{(2)} \\ T_3^{(2)} \end{bmatrix} =: \begin{bmatrix} V_{11} & V_{12} & V_{13} \\ V_{12} & V_{22} & V_{23} \\ V_{12} & V_{23} & V_{33} \end{bmatrix}, \quad (49)$$

where, with $s_k = s_3^{(k)} = P/\sigma_k^2$, $s_2^{(k)} = \alpha s_k$, $s_1^{(k)} = (1-\alpha)s_k$, $\alpha \in [0:1],\ k \in [1:2]$, we get

$$V_{22} = \frac{(2+\alpha s_2)\alpha s_2}{2(\alpha s_2+1)^2} = \mathsf{V}(\alpha s_2), \quad (50)$$

$$V_{33} = \frac{(2+s_2)s_2}{2(s_2+1)^2} = \mathsf{V}(s_2), \quad (51)$$

$$V_{23} = \frac{(2+s_2)\ \alpha s_2}{2(s_2+1)(\alpha s_2+1)}, \quad (52)$$

for the terms relating to the strong user; for the weak user terms (that may include noises from both channels) we get

$$V_{11} = \frac{(1-\alpha)s_1(2\alpha s_1^2 + s_1 + 3\alpha s_1 + 2)}{2(s_1+1)^2(\alpha s_1+1)^2} \neq \mathsf{V}(\frac{(1-\alpha)s_1}{1+\alpha s_1}), \quad (53)$$

$$V_{12} = \left[-\rho\sqrt{\frac{s_1}{s_2}} + \frac{\rho^2}{2}\right]\frac{(1-\alpha)s_1}{(s_1+1)(\alpha s_1+1)}\ \frac{\alpha s_2}{\alpha s_2+1}\ , \quad (54)$$

$$V_{13} = \left[\rho\ \frac{1}{\sqrt{s_1 s_2}} + \frac{\rho^2}{2}\right]\frac{(1-\alpha)s_1}{(s_1+1)(\alpha s_1+1)}\ \frac{s_2}{s_2+1}\ . \quad (55)$$

The second-order region, akin to [10, eq(34)], is the set of (R_1, R_2) such that

$$\left\{\begin{bmatrix} R_1 \\ R_2 \\ R_1+R_2 \end{bmatrix} \in \begin{bmatrix} I_1^{(1)} \\ I_2^{(2)} \\ I_3^{(2)} \end{bmatrix} - \frac{1}{\sqrt{n}}\mathcal{S}(\mathbf{V},\epsilon) + O(\log n)\right\} \quad (56)$$

where the set $\mathcal{S}(\mathbf{V},\epsilon)$ is defined as

$$\mathcal{S}(\mathbf{V},\epsilon) := \left\{\mathbf{a}\in\mathbb{R}^3 : \Pr[\mathbf{Z}\leq\mathbf{a}] \geq 1-\varepsilon\ \mathbf{Z}\sim\mathcal{N}(0,\mathbf{V})\right\} \quad (57)$$

where $\mathbf{V}$ is given in (49).

In our model the noises on different channels are assumed independent, thus we use $\rho = 0$ in (54) and (55). We write the correlated noises $\mathbf{Z}\sim\mathcal{N}(0,\mathbf{V})$ in (57) as a the following linear combination of three iid standard noises $G_i, i\in[3]$

$$Z_1 = \sqrt{V_{11}}G_1, \quad Z_2 = \sqrt{V_{22}}G_2, \quad (58)$$

$$Z_3 = \sqrt{V_{33}}\left(rG_2 + \sqrt{1-r^2}G_3\right), \quad (59)$$

for $r := \frac{V_{23}}{\sqrt{V_{22}V_{33}}}$. We also parameterize $a_i := \sqrt{V_{ii}}Q^{-1}(\epsilon_i)$ in (57) to get the expression in (13b).

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] A. El Gamal and Y.-H. Kim, Network Information Theory. UK: Cambridge University Press, 2011.

[4] 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.

[5] D. Tuninetti, B. Smida, N. Devroye, and H. Seferoglu, "Scheduling on a gaussian broadcast channel with hard deadlines," International Conference on Communication (ICC)., May 2018.

[6] P. Popovski, J. J. Nielsen, C. Stefanovic, E. d. Carvalho, E. Strom, K. F. Trillingsgaard, A. Bana, D. M. Kim, R. Kotaba, J. Park, and R. B. Sorensen, "Wireless access for ultra-reliable low-latency communication: Principles and building blocks," IEEE Network, vol. 32, no. 2, pp. 16–23, March 2018.

[7] Y. W. Huang and P. Moulin, "Finite blocklength coding for multiple access channels," in 2012 IEEE International Symposium on Information Theory Proceedings, July 2012, pp. 831–835.

[8] E. MolavianJazi and J. N. Laneman, "A second-order achievable rate region for gaussian multi-access channels via a central limit theorem for functions," IEEE Transactions on Information Theory, vol. 61, no. 12, pp. 6719–6733, Dec 2015.

[9] 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.

[10] 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.

[11] Y. Yu, H. Chen, Y. Li, Z. Ding, and B. Vucetic, "On the performance of non-orthogonal multiple access in short-packet communications," IEEE Communications Letters, vol. 22, no. 3, pp. 590–593, 2018.

[12] X. Sun, S. Yan, N. Yang, Z. Ding, C. Shen, and Z. Zhong, "Short-packet downlink transmission with non-orthogonal multiple access," IEEE Transactions on Wireless Communications, vol. 17, no. 7, pp. 4550–4564, 2018.

[13] 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.

Authorized licensed use limited to: University of Illinois at Chicago Library. Downloaded on May 13,2022 at 20:19:47 UTC from IEEE Xplore. Restrictions apply.