

# On the Fundamental Limits of MIMO Massive Multiple Access Channels

Fan Wei, Yongpeng Wu, Wen Chen, Wei Yang, and Giuseppe Caire

**Abstract**—In this paper, we study multiple-antenna wireless communication networks, where a large number of devices simultaneously communicate with an access point. The capacity region of multiple-input multiple-output massive multiple access channels (MIMO mMAC) is investigated. While joint typicality decoding is utilized to establish the achievability of capacity region for conventional MAC with fixed number of the users, the technique is not directly applicable for the MIMO mMAC. Instead, an information-theoretic approach based on Gallager's error exponent analysis is exploited to characterize the finite dimension region of the MIMO mMAC. Theoretical results reveal that the region is dominated by the sum rate constraint only, and the individual user rates are dominated by specific factors that correspond to the allocation of the sum rate. The rate in conventional MAC is not achievable when the number of users is comparable with code length, which is due to the fact that successive interference cancellation cannot guarantee an arbitrary small error decoding probability for MIMO mMAC. The results further imply that, asymptotically, the individual user rate is independent of the number of transmit antennas, and channel hardening makes the individual user rate close to that when only statistic knowledge of channel is available at transmitter. The finite dimension region of MIMO mMAC is a generalization of the symmetric rate in Chen *et al.* (2017).

## I. INTRODUCTION

Massive connectivity and massive capacity are two central requirements for future cellular network. In contrast to the current human-centric networks, the expansion of connected machine-centric networks will potentially support millions of devices in each cellular network. As such, it is vital to understand the fundamental limits of such communication systems. The capacity region of conventional multiple access channel (MAC) where the number of simultaneous communicate devices is fixed has been established [1]–[3]. However, the proof of achievability by joint typicality may not be suitable when the number of devices in the network grows unbounded with code block length [4]. Moreover, with fixed transmit powers as well as fixed number of antennas, the individual

user rate would approach zero in the massive connectivity network, which makes the traditional notion of rate defined by  $R = \log M/n \rightarrow 0$  (where  $M$  and  $n$  denote the message set size and code length, respectively), i.e., bits per channel use a less meaningful performance metric.

In order to circumvent the above issues, Chen *et al.* [4], [5] proposed a symmetric code construction for the massive multiple access channels (mMAC), where the number of devices grows unbounded with the code block length. Meanwhile, the message-length capacity, which is a code length related notion of capacity was introduced. By the definition of message-length capacity, the individual user rate now is different from conventional vanishing rate in terms of bits per channel use, and is ensured to be bounded away from zero.

In this paper, the region of multiple-input multiple-output massive multiple access channels (MIMO mMAC) is investigated. We adopt the definition of message-length capacity from [4], [5] in order to avoid zero individual user rate. Asymmetric code construction, where different user groups may have codes with different rates is considered in this paper. Therefore the devices can transmit at diverse rates depending on their Quality of Service. The number of message set size is assumed to be finite and denumerable. Thereby, a finite dimension region of MIMO mMAC is formulated in this paper. Moreover, to meet the massive capacity requirement in the future cellular network, multiple-antenna system [6]–[9] is considered in our work.

In the massive connectivity regime, when the number of users grows to infinity with the code block length, the joint typicality method fails to achieve the region of MIMO mMAC due to the exponential number of error events. As such, we investigate the region of MIMO mMAC based on Gallager's error exponent analysis. In contrast to the conventional capacity region that is a polytope defined in general by  $2^K - 1$  rate constraint inequalities, where  $K$  is the number of users, the region of MIMO mMAC is shown to be subjected to the sum rate constraint only and each individual user rate can be expressed in terms of a specific factor multiplied by the sum rate. Theoretical analysis of the individual rate shows that asymptotically, there is no benefit of having multiple antenna at the transmitters and what matters is the number of antennas at the receiver. Further, we show that due to the channel hardening effect, the asymptotic individual user rate converges to that when transmitter has the statistical knowledge of channels only. Another interesting difference between the conventional MAC and the case studied in this paper is that, while successive interference cancellation (SIC) is able to achieve rate points on the capacity region boundary in conventional case, it fails in MIMO mMAC due to the arbitrary large interference from the other users.

For the related works of massive multiple access, the upper bound of sum rate in multiple antenna system is provided in [10] but with a sketch proof only. The key difference

The work of Y. Wu is supported in part by the National Science Foundation (NSFC) under Grant 61701301, Young Elite Scientist Sponsorship Program by CAST, and the open research project of State Key Laboratory of Integrated Services Networks (Xidian University) under Grant ISN20-03. The work of W. Chen is supported in part by NSFC under Grant 61671294, STCSM Project under Grant 16JC1402900 and 17510740700, National Science and Technology Major Project under Grant 2018ZX03001009 – 002. The corresponding author of this paper is Y. Wu.

F. Wei and W. Chen are with Shanghai Institute of Advanced Communications and Data Sciences, the Department of Electronic Engineering, Shanghai Jiao Tong University, Minhang 200240, China (E-mail: weifan89@sjtu.edu.cn; wenchen@sjtu.edu.cn).

Y. Wu is with the Department of Electronic Engineering, Shanghai Jiao Tong University, Minhang 200240, China (e-mail: yongpeng.wu@sjtu.edu.cn). Y. Wu is also with State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, China.

W. Yang is with Qualcomm Research, Qualcomm Technologies, Inc., San Diego, CA, 92121, USA (e-mail: weiyang@qti.qualcomm.com).

G. Caire is with Institute for Telecommunication Systems, Technical University Berlin, Einsteinufer 25, 10587 Berlin, Germany (Email: caire@tu-berlin.de).

between Yu's work in [10] and mMAC is that there the code block length goes to infinity before the number of devices goes to infinity. A compressed sensing based method is proposed for the active users detection in MIMO random users access [11]. It was shown that the probability of misdetection and false alarm vanish with an infinity number of antennas. In [12], the achievability bound for the random-access code is proposed and the tradeoff between the energy-per-bit and spectral efficiency for various MAC regimes in the massive multiple access is studied. Moreover, some recent works on the coding aspects of massive random access can be found in [13]–[18].

*Notation:* Lowercase letters  $x$ , bold lowercase letters  $\mathbf{x}$ , and bold uppercase letters  $\mathbf{X}$  denote scalars, column vectors, and matrices, respectively. We use  $(\cdot)^\dagger$  to denote complex conjugate and transpose for matrix and  $\text{Tr}\{\cdot\}$  to denote the trace operator of matrix. The notion  $\binom{n}{k}$  denotes the binomial coefficient of  $n$  choose  $k$ . The binary entropy function is denoted by  $H_2(p) = -p \log p - (1-p) \log(1-p)$ .

## II. SYSTEM MODEL

We consider the MIMO mMAC system where each user is equipped with  $N_T$  antennas while  $N_R$  antennas are deployed at the base station. In the massive multiple access, the number of users is comparable or even larger than the code block length, therefore we denote by  $k_n$  the number of users, which grows linearly with the code block length  $n$ , i.e.,  $k_n = O(n)$  [5], [12]. On the  $i_{th}$  channel use, the received signal can be written as

$$\mathbf{y}(i) = \sum_{k=1}^{k_n} \mathbf{H}_k(i) \mathbf{x}_k(i) + \mathbf{z}(i), \quad i = 1, 2, \dots, n, \quad (1)$$

where  $\mathbf{H}_k(i) \in \mathbb{C}^{N_R \times N_T}$  and  $\mathbf{x}_k(i) \in \mathbb{C}^{N_T \times 1}$  are the complex channel matrix and transmitted codewords vector of user  $k$ , respectively.  $\mathbf{z}(i) \in \mathbb{C}^{N_R \times 1}$  is the complex white Gaussian noise vector with normalized identity covariance matrix. Without loss of generality, we assume for every user  $k$  and every channel use  $i$ , the entries of channel matrix  $\mathbf{H}_k(i)$  equal to a fast fading coefficient multiplied by a large scale fading factor [7], i.e.,

$$h_{m,n}^{(k)} = \alpha_{m,n}^{(k)} \beta_k^{\frac{1}{2}}, \quad (2)$$

where  $\alpha_{m,n}^{(k)}$  denotes the fast fading coefficient, which is assumed to be complex Gaussian with zero mean and unit variance. The coefficient  $\beta_k = z_k / r_k^\eta$  is large scale fading factor where  $r_k$  is the distance between user  $k$  and the base station,  $\eta$  is the path loss exponent, and  $z_k$  is the log-normal random variable with a standard deviation of  $\sigma_{\text{shad}}$ . Both  $r_k$  and  $z_k$  are statistically independent of the antenna indexes but are statistically dependent on user index  $k$ . We also assume that the large scale fading factors remain constant during the transmission of codewords, and is known by both transceivers. On the other hand, for fast fading coefficient  $\alpha_{m,n}^{(k)}$ , the receiver may have a perfect knowledge while the transmitter has only the statistical knowledge. Meanwhile, the Gaussian noise  $\mathbf{z}(i)$  are supposed to be independent and identically distributed (i.i.d.) in each channel use  $i$ .

With the transmissions over  $n$  channel uses, the uplink received signal is given by

$$\mathbf{y}^n = \sum_{k=1}^{k_n} \mathbf{H}_k^n \mathbf{x}_k^n + \mathbf{z}^n, \quad (3)$$

where  $\mathbf{y}^n = [\mathbf{y}^T(1), \mathbf{y}^T(2), \dots, \mathbf{y}^T(n)]^T$  is the  $nN_R \times 1$  received signal. The block diagonal channel matrix and the transmitted codewords vector for each user  $k$  is given by  $\mathbf{H}_k^n = \text{diag}\{\mathbf{H}_k(1), \mathbf{H}_k(2), \dots, \mathbf{H}_k(n)\}$  and  $\mathbf{x}_k^n = [\mathbf{x}_k^T(1), \mathbf{x}_k^T(2), \dots, \mathbf{x}_k^T(n)]^T$ , respectively. The  $nN_R \times 1$  vector  $\mathbf{z}^n = [\mathbf{z}^T(1), \mathbf{z}^T(2), \dots, \mathbf{z}^T(n)]^T$  denotes the Gaussian noise. In the following of this paper, the variables are associated with superscript  $n$  when collections of channel uses are considered, whereas the superscript is removed to denote the per channel use variables.

The following definitions introduce the asymmetric codes construction for the MIMO mMAC as well as the notion of message-length capacity [5].

*Definition 1:* Let  $\mathcal{X}_k$  and  $\mathcal{Y}$  be the alphabet of input symbol for user  $k$  and output symbol of channel, respectively. An  $(M_1, M_2, \dots, M_{k_n}, n)$  code for MIMO mMAC  $(\mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_{k_n}, P_{Y|X_1, \dots, X_{k_n}}, \mathcal{Y})$  consists of  $k_n$  sets of integers  $\mathcal{W}_k \in \{1, 2, \dots, M_k\}$  called *message sets* and

- The set of *encoding functions*  $\mathcal{E}_k : \mathcal{W}_k \rightarrow \mathcal{X}_k^{nN_T}$  for every user  $k$ . Since all users are active, the transmitted codewords  $\mathbf{x}_k$  should satisfy the following power constraint:

$$P_k - \delta \leq \text{Tr}(\mathbf{Q}_k) \leq P_k,$$

where  $\delta > 0$  is an arbitrary small number and  $\beta_k P_k \in [P_{\min}, P_{\max}]$ . The positive semi-definite matrix  $\mathbf{Q}_k = \mathbb{E}\{\mathbf{x}_k \mathbf{x}_k^\dagger\}$  denotes the input covariance matrix for user  $k$ . For user  $k$ , the symbols  $\mathbf{x}_k$  are assumed to be generated i.i.d. according to the Gaussian distribution  $\mathcal{CN}(\mathbf{0}; \mathbf{Q}_k)$ .

- A *decoding function*  $\mathcal{D} : \mathcal{Y}^{nN_R} \rightarrow \mathcal{W}_1 \times \mathcal{W}_2 \times \dots \times \mathcal{W}_{k_n}$ , which is a deterministic rule that assigns a sequence of estimated messages  $(\mathcal{W}_1, \mathcal{W}_2, \dots, \mathcal{W}_{k_n})$  to each possible received vector.

The average error probability of the  $(M_1, M_2, \dots, M_{k_n}, n)$  code is given by

$$P_e^{(n)} = \Pr\{D(\mathbf{y}^n) \neq (\mathcal{W}_1, \mathcal{W}_2, \dots, \mathcal{W}_{k_n})\}, \quad (4)$$

where the messages  $(\mathcal{W}_1, \mathcal{W}_2, \dots, \mathcal{W}_{k_n})$  are generated independently over the message set.

*Definition 2 (Asymptotically Achievable Message-Length):* Given the set of functions  $R_k(\cdot)$  that map natural numbers to some positive value, we say the message-length  $R_k(n)$  are asymptotically achievable if there exists a sequence of  $(\lceil \exp(R_1(n)) \rceil, \lceil \exp(R_2(n)) \rceil, \dots, \lceil \exp(R_{k_n}(n)) \rceil, n)$  codes in the sense of Definition 1 such that the average error probability  $P_e^{(n)}$  vanishes as  $n \rightarrow \infty$ .

As stated in [3], [5], when the total number of senders is very large, the traditional notion of rate, measured in bits per channel use, goes to zero for each individual user even the total amount of information can be arbitrary large. Thus the message-length rate, as a function of the codelength  $n$ , is adopted here to avoid the deficiencies of conventional capacity.

The conventional definition for the capacity region of MIMO mMAC is the closure of the set of all achievable rate tuples  $(R_1(n), R_2(n), \dots, R_{k_n}(n))$ . However, as  $n$  becomes larger, the dimension of this rate tuple also increases since the number of users now grows linearly with  $n$ . Thus, to avoid an increasing dimension of capacity region and by noting that the message set size  $M_k$  may not necessarily different from each other, we define the finite dimension region of MIMO mMAC as follows.

*Definition 3 (Finite Dimension Message-Length Region):* Let  $K_j$  denote the number of users with message set size  $\exp\{V_j\}$ , i.e., with message-length rate  $R_k = V_j$  for  $k \in \mathcal{S}$ , where  $V_j$  are the set of “distinct” message-length rates in the system. Then by grouping together the rates  $R_k$  that are equal to each other, the  $J$ -dimensional region of MIMO mMAC is defined as the closure of the convex hull of all  $(K_1, K_2, \dots, K_J)$ , such that  $\sum_{j=1}^J K_j V_j$  is upper bounded by the sum rate of the system.

For the finite and denumerable message set size, the dimension  $J$  was expected to be independent of codelength  $n$ . Thus, in *Definition 3*, a finite dimension region is formulated through the number of sustainable users for some given rate.

For the presentation of the main results in next section, we further define the coefficients  $\mu_k^n = \frac{\log M_k}{\sum_{t \in \mathcal{S}} \log M_t} \in (0, 1)$ , given the message set size  $M_k$  of each user, and assume that the inverse of  $\mu_k^n$  grows linearly with codelength  $n$ , i.e.,  $\frac{1}{\mu_k^n} = O(n)$  so that  $\lim_{n \rightarrow \infty} n\mu_k^n = c_k$ , where  $c_k$  is some positive constant such that  $\sum_{k=1}^{k_n} c_k = n$ .

### III. THE FINITE REGION OF MIMO mMAC

In this section, we present the main results of this paper. We assume that perfect channel state information can be obtained at the receiver (CSIR) while only channel distribution information is available at the transmitter (CDIT). Besides, the users are supposed to be simultaneously active in the network<sup>1</sup>.

*Theorem 1:* For the MIMO mMAC described in (3), let  $\mathcal{S} = \{1, 2, \dots, k_n\}$  denote the user set in the network, where the total number of users scales as  $k_n = O(n)$ . Then message-length vector with components  $(R_1, \dots, R_{k_n})$  is asymptotically achievable if

$$R_k \leq c_k \mathbb{E}_{\mathbf{H}} \left\{ \log \det \left( \mathbf{I} + \sum_{t \in \mathcal{S}} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger \right) \right\}, \quad (5)$$

where  $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_{k_n})$  with  $\mathbf{H}_k$  denotes the channel matrix of user  $k$ , and  $c_k$  is defined as in Section II.  $\square$

*Corollary:* With massive accessed users, since  $k_n = O(n)$ , the asymptotic individual message-length rate grows as

$$R_k \xrightarrow{|S| \rightarrow \infty} c_k N_R \log \left( 1 + \sum_{t \in \mathcal{S}} \beta_t P_t \right) \quad (6)$$

$$= c_k N_R O(\log n) + O(1), \quad (7)$$

where the term  $O(\log n)$  is independent of any other constants but relates only to  $n$ , and the  $O(1)$  constant depends on the pathlosses and transmit powers.  $\square$

What is important here is the  $\log(n)$  dependence for the individual rate. This means that the number of bits that each user is able to deliver grows logarithmically with the block length and therefore also logarithmically with the number of users.

Notice that for each user  $k$ , rate (5) is determined by  $c_k$  (or  $M_k$ ). Since the message set sizes  $M_k$  may not be necessarily

different from each other, given the *Definition 3*, the finite dimension region of MIMO mMAC is formulated as

$$\mathcal{C}_{\text{MAC}} \triangleq \left\{ (K_1, \dots, K_J) : K_1 V_1 + K_2 V_2 + \dots + K_J V_J \leq n \mathbb{E}_{\mathbf{H}} \left\{ \log \det \left( \mathbf{I} + \sum_{t \in \mathcal{S}} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger \right) \right\} \right\}, \quad (8)$$

where  $J$  denotes dimension of the region, and  $V_j$  is on the order of  $O(\log n)$ . Eq. (8) states that any numbers of sustainable users are feasible, provided the total operational rate is constrained by the sum capacity of the system.

**Remark 1 :** The individual user rate in (5) is different from that of conventional MAC. In conventional MAC, the corner points of the region is achieved through SIC, i.e., the receiver decodes the first user's data by treating the signals from others as noises and then subtract it from the received signal. The SIC approach, however, may not be possible in the massive connectivity scenario since interferences plus noise become arbitrarily large as the number of users  $k_n$  grows without bound.<sup>2</sup> As shown in *Theorem 1*, the region of MIMO mMAC is only subjected to the sum rate constraint and the individual user rate is obtained by the sum rate multiplies a factor  $c_k$ , which is related to the size of message set  $M_k$  as shown in Section II. Note that the size  $M_k$  is determined when the codebook has been generated. This is in contrast with conventional MAC where the region is subjected to  $2^{k_n} - 1$  constraints.  $\blacksquare$

**Remark 2 :** From (6), the multiple antennas have provided the system with the degree of freedom gain [23], which is  $N_{\text{DoF}} = \min\{N_R, k_n N_T\} = N_R$ . One important implication of (6) is that, no matter how many transmitter antennas each user is equipped with, we will get the same asymptotical result. Thus there is no benefit of having multiple antenna at the transmitters and what matters is the number of antennas at the receiver. The result in (6) also indicates that due to the effect of channel hardening, the asymptotic rate is close to that when only the statistic knowledge of effective channel gain is available at transmitter.  $\blacksquare$

### IV. PROOF OF THE CONVERSE

The proof of the converse is to show that for any sequence of codes  $(M_1, M_2, \dots, M_{k_n}, n)$  with  $P_e^{(n)} \rightarrow 0$ , we must have  $R_k = \log M_k \leq c_k \mathbb{E}_{\mathbf{H}} \left\{ \log \det \left( \mathbf{I} + \sum_{t \in \mathcal{S}} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger \right) \right\}$  for some positive constant  $c_k$ .

Let  $\mathbf{W} = \{\mathcal{W}_1, \mathcal{W}_2, \dots, \mathcal{W}_{k_n}\}$  denote the message of  $k_n$  users. The entropy of messages are computed by

$$\begin{aligned} H(\mathbf{W}) &= H(\mathbf{W}|\mathbf{y}^n, \mathbf{H}^n) + I(\mathbf{W}; \mathbf{y}^n, \mathbf{H}^n) \\ &\leq H(\mathbf{W}|\mathbf{y}^n, \mathbf{H}^n) + I(\mathbf{X}^n; \mathbf{y}^n, \mathbf{H}^n), \end{aligned} \quad (9)$$

where  $\mathbf{X}^n = \{\mathbf{x}_1^n, \mathbf{x}_2^n, \dots, \mathbf{x}_{k_n}^n\}$  and (9) follows from the data processing inequality since  $\mathbf{W} \rightarrow \mathbf{X}^n \rightarrow (\mathbf{y}^n, \mathbf{H}^n)$  forms a Markov chain.

The mutual information in (9) can be written as,

$$\begin{aligned} I(\mathbf{X}^n; \mathbf{y}^n, \mathbf{H}^n) &= I(\mathbf{X}^n; \mathbf{y}^n | \mathbf{H}^n) \\ &= H(\mathbf{y}^n | \mathbf{H}^n) - H(\mathbf{y}^n | \mathbf{X}^n, \mathbf{H}^n) \end{aligned}$$

<sup>1</sup>The random multiple access where each user transmits with a certain probability (as the case in [11]) will be treated in the future work.

<sup>2</sup>A mathematical analysis for the single antenna AWGN case can be found in [5].

$$\leq \sum_{i=1}^n \left[ H(\mathbf{y}(i)|\mathbf{H}(i)) - H(\mathbf{z}(i)) \right] \\ \leq n\mathbb{E}_{\mathbf{H}} \left\{ \log \det \left( \mathbf{I} + \sum_{t \in \mathcal{S}} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger \right) \right\}, \quad (10)$$

where the second inequality follows because the entropy is maximized by normal distribution with the covariance  $\mathbf{K}_{\mathbf{y}|\mathbf{H}} = \mathbf{I} + \sum_{t \in \mathcal{S}} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger$ .

By Fano's inequality, the conditional entropy in (9) is upper bounded by

$$H(\mathbf{W}|\mathbf{y}^n, \mathbf{H}^n) \leq 1 + P_e^{(n)} \sum_{k \in \mathcal{S}} \log M_k. \quad (11)$$

For the uniformly distributed messages, the entropy is given by  $H(\mathbf{W}) = \sum_{k \in \mathcal{S}} \log M_k$ . Combining (9), (10), and (11), we obtain

$$\sum_{k \in \mathcal{S}} \log M_k \leq n\mathbb{E}_{\mathbf{H}} \left\{ \log \det \left( \mathbf{I} + \sum_{t \in \mathcal{S}} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger \right) \right\} \\ + 1 + P_e^{(n)} \sum_{k \in \mathcal{S}} \log M_k. \quad (12)$$

Multiplying both sides with  $\mu_k^n$ ,

$$(1 - P_e^{(n)}) \log M_k \leq \mu_k^n \\ + n\mu_k^n \mathbb{E}_{\mathbf{H}} \left\{ \log \det \left( \mathbf{I} + \sum_{t \in \mathcal{S}} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger \right) \right\}. \quad (13)$$

Now let  $n \rightarrow \infty$ , the first term on R.H.S. of (13) vanishes. As a consequence, the rate  $R_k$  is shown to be

$$\log M_k \leq c_k \mathbb{E}_{\mathbf{H}} \left\{ \log \det \left( \mathbf{I} + \sum_{t \in \mathcal{S}} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger \right) \right\}, \quad (14)$$

for some positive constant  $c_k$ .

## V. PROOF OF THE ACHIEVABILITY

In this section, we give the proof of achievability part for MIMO message-length capacity. The joint typicality decoding is shown to be achievable in conventional multiple access channel. This approach, however, fails in the case of MIMO mMAC as  $k_n$  now grows linearly with the codeword length  $n$ . As a consequence, the packing lemma [19] does not hold in this regime. An alternative approach for the achievability part is by Gallager's error exponent analysis [20], [21], which forms the basis of our proof in this section.

Assume that the decoder uses *maximum likelihood* detection for the data decoding, i.e.,

$$\hat{\mathcal{W}}_1, \hat{\mathcal{W}}_2, \dots, \hat{\mathcal{W}}_{k_n} = \arg \max_{\mathbf{W}} P\{\mathbf{y}^n | \mathbf{X}^n(\mathbf{W}), \mathbf{H}^n\}. \quad (15)$$

Consider the case when  $k_e$  out of  $k_n$  users incur a decoding error, where the number of such events equals  $\binom{k_n}{k_e}$ . Based on the Gallager's error exponent analysis, the decoding error probability for the  $l_{th}$  event is given by [20]

$$P_e^l(k_e) \leq \prod_{k \in A_l} (M_k - 1)^\rho \mathbb{E}_{\mathbf{H}} \left\{ \int Q\{\mathbf{X}_{A_l^c}\} \right. \\ \times \left[ \int Q\{\mathbf{X}_{A_l}\} P^{\frac{1}{1+\rho}}\{\mathbf{y}|\mathbf{X}, \mathbf{H}\} d\mathbf{X}_{A_l} \right]^{1+\rho} d\mathbf{X}_{A_l^c} d\mathbf{y} \Big\}^n \\ \leq \exp \left\{ - \left[ nE^l(\rho, Q) - \rho \sum_{k \in A_l} R_k \right] \right\}, \quad (16)$$

where  $\rho \in [0, 1]$ ,  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{k_n}\}$ , and  $A_l \subseteq \mathcal{S}$  is the  $l_{th}$  subset of  $\mathcal{S}$  with  $|A_l| = k_e$  while  $A_l^c$  is the complementary set of  $A_l$ . The error exponent  $E^l(\rho, Q)$  is given by

$$E^l(\rho, Q) = -\log \mathbb{E}_{\mathbf{H}} \left\{ \int Q\{\mathbf{X}_{A_l^c}\} \right. \\ \times \left[ \int Q\{\mathbf{X}_{A_l}\} P^{\frac{1}{1+\rho}}\{\mathbf{y}|\mathbf{X}, \mathbf{H}\} d\mathbf{X}_{A_l} \right]^{1+\rho} d\mathbf{X}_{A_l^c} d\mathbf{y} \Big\}, \quad (17)$$

where  $Q\{\cdot\}$  denotes a *priori* distribution of transmitted codewords and  $P\{\mathbf{y}|\mathbf{X}, \mathbf{H}\}$  is the likelihood function for each channel use. Assume the *priori* distribution of the transmitted codewords  $Q\{\cdot\}$  follows a Gaussian distribution  $\mathcal{CN}(\mathbf{0}; \mathbf{Q}_k)$ , after some algebra, Eq. (17) can be calculated as (analogous to the results in [6, p. 592])

$$E^l(\rho) = -\log \mathbb{E}_{\mathbf{H}} \left\{ \det \left( \mathbf{I} + \frac{1}{1+\rho} \sum_{t \in A_l} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger \right)^{-\rho} \right\}. \quad (18)$$

Since there are  $\binom{k_n}{k_e}$  error decoding events, using the inequality [3, Section 12.1]

$$\binom{k_n}{k_e} \leq \exp(k_n H_2(\gamma)), \quad (19)$$

where  $\gamma = k_e/k_n$ , and by union bound, we have for arbitrary  $l$ ,

$$P_e(k_e) \leq \binom{k_n}{k_e} P_e^l(k_e) \quad (20)$$

$$\leq \exp \left\{ - \left[ E_o(\rho) - k_n H_2(\gamma) \right] \right\}, \quad (21)$$

where  $E_o(\rho) = nE^l(\rho, Q) - \rho \sum_{k \in A_l} R_k$ .

Let  $E_r(\rho) = E^l(\rho) - \frac{\rho}{n} \sum_{k \in A_l} R_k - \frac{k_n}{n} H_2(\gamma)$  denote the per channel user error exponent, we have the following lemma.

**Lemma 1:** For  $\epsilon > 0$ , if the message-length rate  $R_k$  is given by

$$R_k = (1 - \epsilon) c_k \mathbb{E}_{\mathbf{H}} \left\{ \log \det \left( \mathbf{I} + \sum_{t \in \mathcal{S}} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger \right) \right\}, \quad (22)$$

for the constants  $c_k > 0$  such that  $\sum_{k \in \mathcal{S}} c_k = n$ , and assume the maximum transmit power  $P_{max}$  is upper bounded by some constant. Then there exists a positive constant  $c_0 > 0$  such that,

$$E_r(\rho) \geq c_0, \quad (23)$$

holds for all sufficient large codeword length  $n$ .

*Proof:* By Gallager's  $\rho$  trick [21],  $\rho$  can be an arbitrary real number within the interval  $[0, 1]$  in the error exponent analysis. Without loss of generality, we choose  $\rho = 1$  in the proof. Combining (18) and (22), the error exponent is given by

$$E_r(1, \gamma) = -\log \mathbb{E}_{\mathbf{H}} \left\{ \det \left( \mathbf{I} + \frac{1}{2} \sum_{t \in A_l} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger \right)^{-1} \right\} \\ - \frac{(1 - \epsilon)}{n} \sum_{k \in A_l} c_k \mathbb{E}_{\mathbf{H}} \left\{ \log \det \left( \mathbf{I} + \sum_{t \in \mathcal{S}} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger \right) \right\} \\ - \frac{k_n}{n} H_2(\gamma). \quad (24)$$

In the following, we consider two cases and show that  $E_r(1, \gamma) > c_0$  holds for sufficient large  $n$ .

*Case I:*  $k_e$  is a large number and grows linearly with the code length  $n$  so that  $\lim_{n \rightarrow \infty} \frac{k_e}{k_n} = \gamma > 0$  converges to some constant. Since the summation  $\frac{1}{n} \sum_{k \in A_l} c_k$  is less than 1 and  $H_2(\gamma) < 1$ , the error exponent is lower bounded by

$$\begin{aligned} E_r(1, \gamma) &\geq \epsilon \mathbb{E}_{\mathbf{H}} \left\{ \log \det \left( \mathbf{I} + \sum_{t \in \mathcal{S}} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger \right) \right\} - \frac{k_n}{n} \\ &\quad - \left\{ \log \mathbb{E}_{\mathbf{H}} \left[ \det \left( \mathbf{I} + \frac{1}{2} \sum_{t \in A_l} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger \right)^{-1} \right] \right. \\ &\quad \left. + \mathbb{E}_{\mathbf{H}} \left[ \log \det \left( \mathbf{I} + \sum_{t \in \mathcal{S}} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger \right) \right] \right\}. \end{aligned} \quad (25)$$

We now define  $\mathbf{G}_t = \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger$  where the  $(i, j)_{th}$  entry is given by

$$\begin{aligned} g_{i,j}^{(t)} &= \sum_n \sum_m h_{i,m}^{(k)} q_{m,n}^{(t)} (h_{j,n}^{(t)})^* \\ &= \sum_n \sum_m \alpha_{i,m}^{(k)} q_{m,n}^{(t)} (\alpha_{j,n}^{(t)})^* \beta_t \end{aligned} \quad (26)$$

where  $h_{m,n}^{(t)} = \alpha_{m,n}^{(t)} \beta_t^{\frac{1}{2}}$  and  $q_{m,n}^{(t)}$  denote the  $(m, n)_{th}$  entry of matrices  $\mathbf{H}_t$  and  $\mathbf{Q}_t$ , respectively. Since the fast fading coefficients  $\alpha_{m,n}^{(t)}$  are *i.i.d.* random variables with normal distribution  $\mathcal{CN}(0, 1)$ , by Kolmogorov's strong law of large number [22],

$$\lim_{k_n \rightarrow \infty} \sum_{t \in \mathcal{S}} g_{i,j}^{(t)} = \begin{cases} \sum_{t \in \mathcal{S}} \beta_t \text{Tr}(\mathbf{Q}_t), & i = j; \\ 0, & i \neq j. \end{cases} \quad (27)$$

and

$$\lim_{k_e \rightarrow \infty} \sum_{t \in A_l} g_{i,j}^{(t)} = \begin{cases} \sum_{t \in A_l} \beta_t \text{Tr}(\mathbf{Q}_t), & i = j; \\ 0, & i \neq j. \end{cases} \quad (28)$$

Therefore for larger number  $k_e$  and  $k_n$ ,

$$\lim_{k_n \rightarrow \infty} \det \left( \mathbf{I} + \sum_{t \in \mathcal{S}} \mathbf{G}_t \right) = \left( 1 + \sum_{t \in \mathcal{S}} \beta_t \text{Tr}(\mathbf{Q}_t) \right)^{N_R}, \quad (29)$$

and

$$\lim_{k_e \rightarrow \infty} \det \left( \mathbf{I} + \frac{1}{2} \sum_{t \in A_l} \mathbf{G}_t \right) = \left( 1 + \frac{1}{2} \sum_{t \in A_l} \beta_t \text{Tr}(\mathbf{Q}_t) \right)^{N_R}. \quad (30)$$

Consequently,

$$\begin{aligned} E_r(1, \gamma) &\geq \epsilon N_R \log \left( 1 + \sum_{t \in \mathcal{S}} \beta_t \text{Tr}(\mathbf{Q}_t) \right) - \frac{k_n}{n} \\ &\quad + N_R \log \left\{ \left( 1 + \frac{1}{2} \sum_{t \in A_l} \beta_t \text{Tr}(\mathbf{Q}_t) \right) / \left( 1 + \sum_{t \in \mathcal{S}} \beta_t \text{Tr}(\mathbf{Q}_t) \right) \right\} \\ &\geq \epsilon N_R \log \left( 1 + \sum_{t \in \mathcal{S}} \beta_t \text{Tr}(\mathbf{Q}_t) \right) - \frac{k_n}{n} \\ &\quad + N_R \log \left\{ \left( 1 + \frac{1}{2} \gamma k_n (P_{\min} - \delta) \right) / \left( 1 + k_n P_{\max} \right) \right\}, \end{aligned} \quad (31)$$

where the last inequality in (31) follows because  $P_{\min} - \delta = \min_{t \in A_l} \beta_t \text{Tr}(\mathbf{Q}_t)$  and  $P_{\max} = \max_{t \in \mathcal{S}} \beta_t \text{Tr}(\mathbf{Q}_t)$ . Since the last two terms in (31) converge to some constant while the first term grows unbounded as  $n$  increases, the error exponent is shown to be large than some positive constant.

*Case II:*  $k_e$  scales sublinearly with code length  $n$  so that  $\lim_{n \rightarrow \infty} \gamma = \lim_{n \rightarrow \infty} \frac{k_e}{k_n} = 0$ . By choosing  $\rho = 1$ , we have

$$\begin{aligned} E_r(1, \gamma) &= -\log \mathbb{E}_{\mathbf{H}} \left\{ \det \left( \mathbf{I} + \frac{1}{2} \sum_{t \in A_l} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger \right)^{-1} \right\} \\ &\quad - \frac{(1 - \epsilon)}{n} \sum_{k \in A_l} c_k \mathbb{E}_{\mathbf{H}} \left\{ \log \det \left( \mathbf{I} + \sum_{t \in \mathcal{S}} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger \right) \right\} \\ &\quad - \frac{k_n}{n} H_2(\gamma). \end{aligned} \quad (32)$$

Note that since matrix  $\mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger$  is positive semi-definite, its largest eigenvalue is strictly positive, and hence the largest eigenvalue of  $\mathbf{I} + \frac{1}{2} \sum_{t \in A_l} \mathbf{H}_t \mathbf{Q}_t \mathbf{H}_t^\dagger$  is strictly larger than 1, whereas the rest eigenvalues can be greater than or equal to 1. As a consequence, the first term in (32) is strictly positive.

For the second term, we have  $\frac{1}{n} \sum_{k \in A_l} c_k \leq k_e c_{\max}/n$ . Since  $c_k$  is a positive constant for all  $k$  while the expectation in the second term is on the order  $O(\log n)$ . In the following, we consider two cases for the sublinearly growing  $k_e$ :

- 1)  $k_e = O(\frac{n}{\log n})$ , i.e.,  $k_e$  is on the order of  $\frac{n}{\log n}$ ,
- 2)  $k_e = o(\frac{n}{\log n})$ , i.e.,  $k_e$  grows slower than the speed  $\frac{n}{\log n}$ .

In case 1), the first term grows unbounded with the code length, the second term is lower bounded by some constant and the last term vanishes. In case 2), the last two terms vanishes while the first term is positive. Therefore, in both cases the error exponent is shown to be large than zero.

Since  $l$  is arbitrary, by (20), Lemma 1 and combining the Case I and Case II, the error probability  $P_e(k_e)$  decays exponentially with code length  $n$ . Hence,

$$P_e \leq \sum_{k=1}^{k_n} P_e(k_e) \leq k_n e^{-n c_0}, \quad (33)$$

vanishes as  $n$  increases. As a consequence, the achievability of Theorem 1 is established. The corollary can be obtained by using the same metric in the proof of Case I.

## VI. NUMERICAL RESULTS

This section presents the numerical results for the MIMO mMAC channels. We consider a single cell network with  $k_n$  users randomly deployed within the cellular. Each user has a transmit power  $P_k$  generated uniformly within the interval [5, 15]. The distance between an individual user and the base station ranges from 100m to 1000m. For large scale fading, the decay exponent is  $\eta = 3.8$  whereas the shadow-fading standard deviation is set to be  $\sigma_{\text{shad}} = 8.0$  dB.

In Fig. 1, we plot the sum rate of MIMO mMAC as a function of code length  $n$  for  $2 \times 2$ ,  $4 \times 4$  and  $8 \times 8$  MIMO, respectively. The number of users is set  $k_n = n/2$ , which grows linearly with the code length  $n$ . Note that since  $k_n$  grows unbounded with  $n$  and the message-length capacity considered in this paper itself is a function of  $n$ , the capacity of MIMO mMAC system thereby grows with an increasing of code length as shown in Fig. 1. Moreover, it can be also observed that the asymptotic rate (6) matches well with the operational rate (5). Hence, with *i.i.d.* Rayleigh coefficients, the performance of MIMO mMAC is only affected by large scale fading.

We also demonstrate the impact of receiving antenna numbers in Fig. 2. From the figure, the sum rates grow almost linearly with  $N_R$  for different code length  $n$ . This is due to the degree of freedom gain of MIMO system [23]. The similar

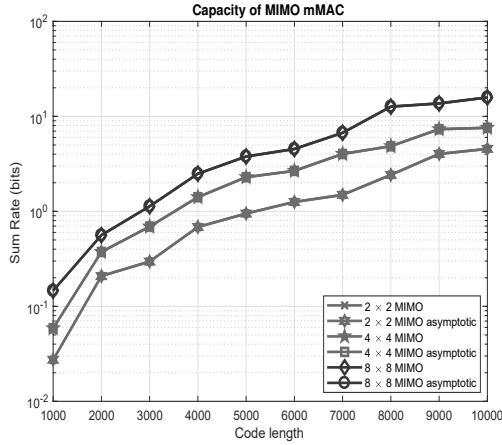


Fig. 1: The sum rate of MIMO mMAC versus growing code length  $n$ .

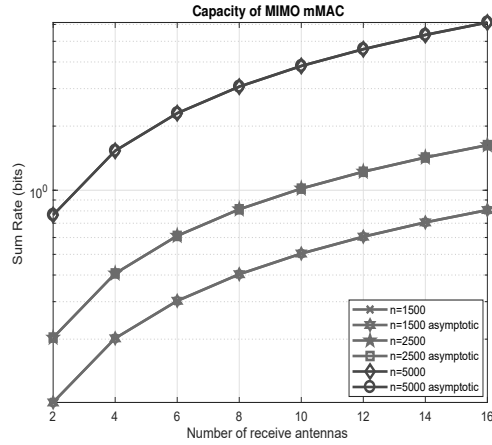


Fig. 2: The sum rate of MIMO mMAC versus growing receive antennas  $N_R$ ,  $N_T = 2$ .

phenomenon has been well observed in [24], where they showed that the number of degree of freedom of a MIMO system is limited by the minimum of the number of transmit antenna and the number of receive antenna in high SNR regime. With massive users, the uplink SNR is high enough. Therefore, we have  $N_{\text{DoF}} = \min\{N_R, k_n N_T\} = N_R$ . As a result, the capacity of MIMO mMAC grows linearly with  $N_R$  as is shown in Fig. 2.

## VII. CONCLUSION

In this work, we have investigated the capacity of mMAC with multiple antennas. The capacity of conventional MAC with fixed number of users is achieved by a joint typicality decoder at the receiver. However, as the number of users grows unbounded, the joint typicality approach does not work due to the exponential number of error events. As such, we derived the finite dimension region of MIMO mMAC channels by using the Gallager's error exponent analysis. Theoretical analysis indicates that rather than the interference plus noise, the individual user rate in MIMO mMAC is determined by a specific factor that corresponds to the allocation of sum capacity.

In this paper, we assumed that the set of active users is known to the receiver. In a random access scenario, each user has an independent transmit probability and the receiver is lack of the active users knowledge. This generalized model will be addressed in the future work.

## REFERENCES

- [1] R. Ahlswede, "Multi-way communication channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Sep. 1971, pp. 23–52.
- [2] H. Liao, "A coding theorem for multiple access communications," in *Proc. IEEE Int. Symp. Inf. Theory*, Pacific Grove, CA, USA, 1972.
- [3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.
- [4] X. Chen and D. Guo, "Gaussian many-access channels: Definition and symmetric capacity," in *Proc. IEEE Inf. Theory Workshop*, Seville, Spain, Sep. 2013, pp. 1–5.
- [5] X. Chen, T.-Y. Chen, and D. Guo, "Capacity of Gaussian many-access channels," *IEEE Trans. Inform. Theory*, vol. 63, no. 6, pp. 3516–3539, Jun. 2017.
- [6] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–596, Nov. 1999.
- [7] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [8] Y. Wu, R. Schober, D. W. K. Ng, C. Xiao, and G. Caire, "Secure massive MIMO transmission with an active eavesdropper," *IEEE Trans. Inf. Theory*, vol. 62, pp. 3880–3900, Jul. 2016.
- [9] Y. Wu, C. Xiao, Z. Ding, X. Gao, and S. Jin, "A survey on MIMO transmission with finite input signals: Technical challenges, advances, and future trends," *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1779–1833, Oct. 2018.
- [10] W. Yu, "On the fundamental limits of massive connectivity," in *Proc. Inf. Theory Appl. Workshop*, pp. 1–6, Feb. 2017.
- [11] L. Liu, W. Yu, "Massive device connectivity with massive MIMO," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1–6, Jun. 2017.
- [12] Y. Polyanskiy, "A perspective on massive random-access," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2523–2527, Jun. 2017.
- [13] O. Ordentlich and Y. Polyanskiy, "Low complexity schemes for the random access Gaussian channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2528–2532.
- [14] M. Effros, V. Kostina, and R. C. Yavas, "Random access channel coding in the finite blocklength regime," arXiv:1801.09018, Feb. 2018.
- [15] A. Hasanzadeh, J. Chamberland, and K. Narayanan, "Rate selection and power adaptation using maximal ratio combining for the random access Gaussian channel," arXiv:1801.03557, Jan. 2018.
- [16] V. K. Amalladinne, A. Vem, D. K. Soma, K. R. Narayanan, J.-F. Chamberland, "A coupled compressive sensing scheme for uncoordinated multiple access," arXiv:1809.04745v1 [eess.SP], Sep. 2018.
- [17] F. Wei, W. Chen, Y. Wu, J. Li, and Y. Luo, "Toward 5G wireless interface technology: Enabling nonorthogonal multiple access in the sparse code domain," *IEEE Veh. Technol. Mag.*, vol. 13, no. 4, pp. 18–27, Dec. 2018.
- [18] F. Wei, W. Chen, Y. Wu, J. Ma, and T. A. Tsiftsis, "Message-passing receiver design for joint channel estimation and data decoding in uplink grant-free SCMA systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 167–181, Jan. 2019.
- [19] A. El Gamal and Y. -H. Kim, *Network Information Theory*. Cambridge University Press, 2012.
- [20] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inf. Theory*, vol. 31, no. 2, pp. 124–142, Mar. 1985.
- [21] R. G. Gallager, *Information Theory and Reliable Communication*, New York, NY, USA: Wiley, 1968.
- [22] P. K. Sen and J. M. Singer, *Large Sample Methods in Statistics: An Introduction with Applications*, New York, Chapman & Hall, Inc. New York 1993.
- [23] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [24] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering, and keyhole," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2636–2647, Oct. 2003.