Energy-Efficient and Low-Latency Massive SIMO using Noncoherent ML Detection for Industrial IoT Communications

Xiang-Chuan Gao, Jian-Kang Zhang, He Chen, Zheng Dong, and Branka Vucetic

***Abstract*—To enable ultra-reliable low-latency wireless communications required in the industrial Internet of Things, in this paper we develop an energy-based modulation (i.e., nonnegative pulse amplitude modulation (PAM)) constellation design framework for noncoherent detection in massive single-input multiple-output (SIMO) systems. We consider that one single-antenna transmitter communicates to a receiver with a large number of antennas over a Rayleigh fading channel, and the receiver decodes the transmitted information at the end of every symbol. For such a SIMO system with nonnegative PAM modulation, we first propose a fast noncoherent maximum likelihood (ML) decoding algorithm and derive a closed-form expression of its symbol error probability (SEP). We then enhance the system energy efficiency by finding the optimal PAM constellation that minimizes the exact SEP subject to a total signal power constraint for such a system with an arbitrary number of receiver antennas, signal-to-noise ratio (SNR), and constellation size. Furthermore, the closed-form upper and lower bounds on the optimal SEP are derived. Based on these bounds, the exact expression for coding gain of the dominant term of the SEP is presented for such an optimal massive SIMO system. We also present an asymptotic SEP expression at a high signal-to-noise ratio (SNR) regime and the approximate diversity gain of the system. Simulation results for the proposed optimal PAM constellation validate the theoretical analysis, and show that our presented optimal constellation attains significant performance gains over the currently available minimum-distance based constellation systems.**

***Index Terms*—Ultra-reliable low-latency communications, massive SIMO, noncoherent ML detection, energy-based modulation, industrial Internet of Things.**

I. Introduction

The Internet of Things (IoT) is changing our daily life, and transforming businesses and industries. As the most important part of the IoT, industrial IoT (IIoT) has been attracting tremendous attention from governments, researchers, and business leaders due to the unprecedented opportunities it can potentially bring to business and society. Generally speaking, the IIoT is "a new industrial ecosystem that combines intelligent and autonomous machines, advanced predictive analytics, and machine-human collaboration to improve productivity, efficiency and reliability" [1]. While the conventional IoT is mostly targeted for low-power consumer usage scenarios [2], [3], IIoT mainly focuses on the mission-critical industrial use cases, such as manufacturing, supply chain, transportation, and medical sectors. The intelligent industrial operations largely count on the information collected by various sensors and the analysis of the gathered information [4]. In this sense, connectivity is critical to the successful implementation of the IIoT [5].

To provide network connectivity for IIoT devices such as sensors, actuators, and controllers, wireless communications have a number of merits over the currently-used wired communications: low deployment and maintenance cost, easier implementation in scenarios where cables are difficult to deploy, and high long-term reliability by avoiding the wear and tear issues [6]. Wireless technologies have made remarkable advances and have been deployed widely to provide ubiquitous human-to-human connectivity to the people during the past decades. However, until now, wireless networks have not been widely applied in industrial environments: wireless technologies have been mainly used in noncritical applications (e.g., condition monitoring), where a failure of communication does not result in serious economic losses and safety problems; the adoption of wireless in more crucial scenarios, e.g., closed-loop control where an unsuccessful delivery of data may lead to serious accidents, has been rare [5]. The main reason behind this is that current standardized industrial wireless protocols are unable to meet the stringent performance requirements of mission-critical industrial use cases, with high reliability of packet error rate down to 10^{-9} and ultra-low latency at the level of sub-microsecond [7], [8]. More specifically, current industrial wireless technologies have been reusing the physical layers designed for human-oriented communications and modifying the upper layers [7]. These physical layers have been designed to achieve higher and higher data rate with the reliability and latency suited to human perception; and critical industrial applications need reliability and latencies several orders of magnitude better than what is achievable in today's networks. As such, a dedicated physical layer design with superior reliability and latency performance is crucial to fully unlock

X.-C. Gao is with the School of Information and Engineering, Zhengzhou University, Zhengzhou, China (e-mail:iexcgao@zzu.edu.cn). J.-K. Zhang is with the Department of Electrical and Computer Engineering, McMaster University, 1280 Main Street West, L8S 4K1, Hamilton, Ontario, Canada (e-mail: jkzhang@mail.ece.mcmaster.ca). H. Chen, Z. Dong and B. Vucetic are with the School of Electrical and Information Engineering, The University of Sydney, NSW 2006, Australia (emails:{he.chen, zheng.dong, branka.vucetic}@sydney.edu.au).

The work of X.-C. Gao was supported in part by the China National Natural Science Foundation of China (61640003, U1204607, 61501404), the Major Projects of Henan Province (161100210200), National 863 Project (2014AA01A705), the Outstanding Young Talent Research Fund of ZZU (1521318003) and the Startup Research Fund of Zhengzhou University under Grant 1512318005. The work of J.-K. Zhang was supported in part by Natural Sciences and Engineering Research Council of Canada (NSERC). The work of H. Chen, Z. Dong, and B. Vucetic was supported in part by the Australian Research Council Grant FL160100032.

the potential of the IIoT.

Achieving ultra-reliable low-latency communications (URLLC) over wireless links is highly challenging, since wireless channels are unstable and susceptible to fading, path-loss, shadowing, and interferences. It has been shown that a high-order diversity is critical to realize ultra-reliable communication within very low latency budget [9], [10]. In theory, diversity in wireless communications can be achieved in the time, frequency and/or space domains [11]. In particular, time diversity adds redundant symbols to the original data to increase its successful decoding probability at the receiver side. However, time diversity is not preferable in URLLC, since it enhances the communication reliability at the cost of latency. On the other hand, to harness the frequency diversity, the information symbols should be transmitted over uncorrelated frequency channels. This can be costly since IIoT devices typically operate in the frequency band between 100 MHz to 6 GHz due to cost, power, and size constraints, where the spectrum resource is scarce [12], [13]. The frequency band above 6 GHz is mainly reserved for fixed links or satellite communications [14], where the propagation loss is too high for IIoT applications. Therefore, spatial diversity, also known as space/antenna diversity, has been regarded as the most desirable diversity scheme to enable the URLLC [10]. To obtain spatial diversity gains, multiple antennas need to be equipped at the transmitters and/or the receivers of wireless systems. In fact, multiple-antenna technologies have been extensively investigated during the last two decades and applied to various wireless standards, for example LTE-Advanced [15]. According to the number of antennas at the transmitter and receiver, multiple-antenna technologies consist of three formats: single-input multiple-output (SIMO), multiple-input single-output (MISO) and multiple-input multiple-output (MIMO). The use of multiple antennas yields additional spatial degrees of freedom that enhance system reliability for a given throughput through spatial diversity, or increase the throughput for a given reliability requirement through the exploitation of spatial multiplexing. These two effects cannot be maximized concurrently, but can be used simultaneously at some sub-optimum levels and there exists a fundamental tradeoff between diversity and multiplexing [16]. In current wireless systems offering human-oriented mobile broadband services, multiple-antenna technologies have typically been designed and optimized so as to maximize the system throughput [17]. These designs and optimizations are no longer suitable for URLLC applications, in which the amount of data to be transmitted (e.g., status updates and control commands) is normally small, and the reliability and latency performance has a higher priority than the throughput.

To achieve more dramatic gains as well as to simplify the required signal processing, large-scale multiple-antenna systems, also termed massive MIMO, have been proposed, in which a large number of antennas are used at both the transmitters and receivers [18]–[22]. It has been shown that massive MIMO has several benefits over its conventional version: the effects of uncorrelated noise and small-scale fading are eliminated, and even simple linear signal processing approaches perform well in the asymptotic limit. With the capability to create a large number of spatial diversity paths, massive MIMO has very recently been treated as one of the most promising techniques for enabling URLLC [23], [24]. However, the aforementioned benefits of massive MIMO are conditioned on the acquisition of the instantaneous channel state information (CSI) and the implementation of coherent detection at the receiver. The estimation of instantaneous CSI may introduce a "long" training symbols (delay) with a size comparable to that of short data payload in the IIoT use cases [25], especially in low signal-to-noise ratio (SNR) regime that is the case for most battery-powered industrial sensors and devices. On the other hand, for those mobile IIoT devices (e.g., automated guided vehicle), estimating the CSI can be ineffective since the channel changes quickly due to their mobility. Considering these facts, the authors in [24] envisioned that "in high-mobility or low-SNR scenarios, fulfilling the requirements of URLLC might require shifting to a basic time-division multiple access (TDMA) system with noncoherent receivers based on energy detection (ED)". The initial simulation performed in [24] has validated this vision. Specifically, for a SIMO system with 128 antenna with user mobility, the symbol-error probability (SEP) of the ED-based noncoherent receiver can be orders of magnitude better than that of its coherent maximum ratio combining counterpart.

There have been some recent efforts on designing massive SIMO systems using ED-based noncoherent receivers, in which no instantaneous CSI is needed at either the transmitter or the receiver, and the receiver decodes transmitted information after each received symbol by measuring only the average signal energy across all receiver antennas [26]–[29]. Among them, the authors in [26] characterized the scaling law of the noncoherent massive SIMO systems with energy-based detectors over independent and identically distributed channel fading when the number of receiver antenna goes to infinity. The results in [26] showed that energy-based noncoherent schemes are able to achieve the same scaling behavior in terms of achievable rates as optimal coherent schemes with an increasing number of antennas. A simple constellation design scheme was also proposed for one-user and two-user cases, which was based on maximizing the minimum Euclidean distance of the sum constellation at the energy-based detector [26]. However, for the two-user case, the closed-form solution to the constellation design problem was only obtained for some specific channels and low-order modulations, and numerical search is needed otherwise. Recently, Zhang *et al.* have generalized the results attained in [26] to the scenarios with arbitrary numbers of users and channel fading models in [27]. A more comprehensive analysis of the constellation design of noncoherent energy-based massive SIMO systems was conducted in [28], where the constellation design problem was formulated as the maximization of the error exponent of the SEP with respect to the number of receiver antennas, and the formulated problem was resolved under different assumptions on the availability and imperfection of CSI statistics. A similar constellation design problem was also considered in [29], which applied the central limit theorem, instead of the law of large numbers used in the previous papers, to attain a Gaussian

approximation of the instantaneous received signal energy, which further leads to a closed-form expression of the detection threshold for the adopted energy-based detector. However, all the aforementioned designs applied certain approximation to the distribution of the received signal by resorting to the asymptotic channel orthogonality and hardening properties of massive MIMO systems. In practice, due to a finite number of receiver antennas, these properties are not strictly satisfied. As such, the approximations will result in error in the performance analysis and the associated constellation design. Furthermore, to the best knowledge of the authors, the optimal diversity and multiplexing tradeoff for noncoherent SIMO systems has not been well investigated. Besides, the condition for achieving the full diversity in noncoherent SIMO systems needs to be better understood.

In this paper, as in [28], [29], we consider an energy-based massive SIMO system having a single-antenna user communicating to an access point (AP) equipped with a large number of antennas over Rayleigh fading channels. The user transmitter modulates its information on the amplitude of transmitted symbols and the AP adopts the noncoherent maximum-likelihood (ML) detector for the symbol-by-symbol estimation of the user information symbols from the received signal. This model can represent typical IIoT use cases, wherein multiple sensors transmit their measurements to a central controller in a TDMA manner. Different from the existing approaches, we will derive the *exact* decision region of the noncoherent ML detector as well as the corresponding *exact* SEP, which will be subsequently used for the optimal constellation design. The main contributions of this paper can be summarized as follows:

- A fast closed-form noncoherent ML detector for any energy-based modulation (i.e., nonnegative PAM) constellation is developed and its average SEP is derived. This enables us to enhance the system energy efficiency by finding the SEP-optimal constellation by solving a convex optimization problem.
- For the case when SNR goes to infinity, we prove that the optimal received constellation that maximizes the noncoherent diversity gain is the constellation consisting of the geometrical sequence with the noise variance as the initial term. In addition, an asymptotic series of the resulting optimal SEP is derived. Thus, the optimal tradeoff between the multiplexing gain and the diversity gain is also found for the noncoherent SIMO system.
- For a massive SIMO system with a large number of antennas, a new concept of *receiver diversity gain* and *geometrical coding gain* is introduced. When the number of the received antennas goes to infinity, we establish lower and upper bounds on SEP and then, prove that any nonnegative PAM signalling enables full receiver diversity with the noncoherent ML detector. Again and very interestingly, we find that the optimal received constellation that maximizes the geometrical coding gain is also the same geometrical constellation. Furthermore, we quantitatively characterize two important asymptotic behaviours of the resulting optimal geometrical coding gains.
  1) The first is the asymptotic behaviour when SNR goes to infinity and the transmission data rate is fixed, showing that the optimal geometrical coding gain tends to infinity with a certain diversity gain. Therefore, unlike currently available constellation designs resulting from directly using the central limit theorem for massive MIMO systems in which there is error floor when SNR goes to infinity, there is no longer error floor in our optimal design.
  2) The second is the asymptotic behaviour when the transmission data rate varies with the number of the received antennas and goes to infinity, but SNR is fixed. Hence, a novel error performance metric on multiplexing-receiver diversity gain tradeoff, following Zheng and Tse's concept on multiplexing-diversity gain tradeoff for the coherent MIMO systems [16], is proposed for the noncoherent massive SIMO system.

II. System Model and Noncoherent ML Detection

We consider a point-to-point massive single-input multiple-output (SIMO) system for IIoT communications, where a single-antenna user[1] (e.g., sensor) transmits to the receiver (e.g., controller) equipped with N antennas. For such a system, the input and output relationship can be written as

$$\mathbf{y} = \mathbf{h}s + \mathbf{n}, \tag{1}$$

where $\mathbf{y} = [y_1, \cdots, y_N]^T$ denotes the $N \times 1$ received signal vector, and $\mathbf{h} = [h_1, \cdots, h_N]^T$, $\mathbf{n} = [n_1, \cdots, n_N]^T$ represent $N \times 1$ channel coefficient and noise vectors, respectively. We also assume that the transmitted symbol s is a scalar which is randomly and equally likely drawn from a finite-alphabet constellation to be designed later. Throughout this paper, we make the following two assumptions:

1) The instantaneous channel coefficient vector $\mathbf{h}$, with all elements independently and identically circularly-symmetric complex Gaussian distributed with zero mean and unit variance (i.e., Rayleigh fading), is *not available* at either the transmitter or the receiver and may change to other independent values in the next time slot [30]–[32]. This assumption is to eliminate the channel estimation/feedback overhead for achieving ultra-low latency required by the IIoT communications, especially for fast fading channels. Nevertheless, the mean and variance of the channel are assumed to be known for the optimal constellation design.
2) The noise vector $\mathbf{n}$ is circularly-symmetric complex Gaussian distributed with zero mean and covariance matrix being $\sigma^2\mathbf{I}_N$, and σ^2 is assumed to be known.

Under the above assumptions, for given transmitted signal s, the output $\mathbf{y}$ is also circularly-symmetric complex Gaussian distributed with zero mean, whose covariance matrix is determined by

$$\mathbb{E}[\mathbf{y}\mathbf{y}^H] = \mathbb{E}[(\mathbf{h}s + \mathbf{n})(\mathbf{h}s + \mathbf{n})^H] = (|s|^2 + \sigma^2)\mathbf{I}_N.$$

[1]Note that the considered system can have multiple users working in a TDMA manner.

Then, the probability density function (PDF) of $\mathbf{y}$ conditioned on s, $f(\mathbf{y}|s)$ is given by

$$f(\mathbf{y}|s) = \frac{1}{\pi^N(|s|^2+\sigma^2)^N}\ \exp\Big(-\frac{\|\mathbf{y}\|^2}{|s|^2+\sigma^2}\Big), \tag{2}$$

from which we note that the phase information of the transmitted signal is lost. Therefore, we will focus on an energy-based transceiver design: the information is modulated on the energy of transmitted symbols, and the receiver estimates the transmitted information bits based on the energy of received signals. In this paper, we consider a normalized symbol duration and we thus use "energy" and "power" interchangeably hereafter.

We consider that the power of transmitted information symbol $|s|^2$ is chosen from one of the L power levels from PAM constellation $\mathcal{E} = \big\{E_i : E_i > 0\big\}_{i=1}^{L}$ subject to an instantaneous average power constraint such that

$$\frac{1}{L}\sum_{i=1}^{L} E_i \le E_s.$$

Without loss of generality, the power levels are taken to be in an ascending order with $E_i < E_{i+1}$, for $i = 1, \ldots, L-1$.

For such a noncoherent SIMO system, it is known that the optimal decoder is a noncoherent maximum-likelihood (ML) decoder, which is to find s such that $f(\mathbf{y}|s)$ is maximized [28], [29]. That is to solve the following optimization problem:

$$\hat{s} = \underset{|s|^2\in\mathcal{E}}{\arg\max}\ \ \ln\big(f(\mathbf{y}|s)\big). \tag{3}$$

Combing (2) and (3), the original optimization problem for a noncoherent ML decoder is simplified as

$$\hat{a} = \underset{a\in\mathcal{A}}{\arg\min}\ \frac{\|\mathbf{y}\|^2}{a} + N\ln a, \tag{4}$$

where $\mathcal{A}$ denotes the resulting received constellation given by

$$\mathcal{A} = \{a_i\}_{i=1}^{L} = \{E_i + \sigma^2\}_{i=1}^{L}. \tag{5}$$

From (4), we notice that the noncoherent ML decoder corresponds to specifying the decision regions for the sufficient statistic $\|\mathbf{y}\|^2$. In order to further simplify the ML detector (4) and to analyze its error performance, we need to establish the following lemma, which can help us to simplify the decision region of the noncoherent ML receiver.

Lemma 1: For $r > 1$, the two functions defined by

$$u(r) = \frac{\ln r}{r-1}, \quad v(r) = \frac{r\ln r}{r-1}, \tag{6}$$

have the following properties:

1) $0 < u(r) < 1$ and $v(r) > 1$;
2) $u(r)$ and $v(r)$ are monotonically decreasing and increasing functions of r, respectively.

■

The proof is provided in Appendix A.

With the preceding lemma, we are now in a position to propose the following *fast* noncoherent ML decoding algorithm for the considered SIMO system, which can be implemented by a quantization operation[2].

Algorithm 1: Fast Noncoherent ML Detector: Consider constellation set $\mathcal{A} = \{a_i\}_{i=1}^{L}$ defined in (5), and decision threshold set $\mathcal{B} = \{b_i\}_{i=1}^{L-1}$ with b_i given by

$$b_i = \frac{a_i a_{i+1}\ln\big(\frac{a_{i+1}}{a_i}\big)}{a_{i+1}-a_i}, \quad i = 1, \cdots, L-1. \tag{7}$$

Then, the optimal estimates of the transmitted signals using the noncoherent ML detector are determined as follows:

$$\hat{a} = \begin{cases} a_1, & \text{if } \frac{\|\mathbf{y}\|^2}{N} \le b_1; \\ a_i, & \text{if } b_{i-1} < \frac{\|\mathbf{y}\|^2}{N} \le b_i,\ i = 2, \ldots, L-1; \\ a_L, & \text{if } \frac{\|\mathbf{y}\|^2}{N} > b_{L-1}. \end{cases} \tag{8}$$

■

The proof is provided in Appendix B.

III. Symbol Error Probability Analysis and Optimal Constellation Design

In this section, we first analyze the *exact* average SEP of the noncoherent ML detector for the considered SIMO system, and then find the optimal constellation that minimizes the average SEP.

A. Exact Average SEP Analysis

We notice that the random variable $X = \frac{\|\mathbf{y}\|^2}{a_i}$ follows a Chi-squared distribution [33] with PDF and cumulative distribution function (CDF) determined respectively by

$$f_X(x) = \frac{1}{\Gamma(N)}x^{N-1}e^{-x}, \tag{9a}$$

$$G(x) = 1 - e^{-x}\sum_{k=0}^{N-1}\frac{x^k}{k!}, \quad x > 0. \tag{9b}$$

For notation simplicity, we define

$$u_i = u(r_i) = \frac{\ln r_i}{r_i - 1}, \quad v_i = v(r_i) = \frac{r_i\ln r_i}{r_i - 1}, \tag{10}$$

where $r_i = \frac{a_{i+1}}{a_i}$ with a_i defined in (5), and then by the definition of b_i in (7), we have $b_{i-1} = a_i u_{i-1}$ and $b_i = a_i v_i$. As a consequence, the decision rule given in (8) can be reformulated in terms of $X = \frac{\|\mathbf{y}\|^2}{a_i}$ by

$$\hat{a} = \begin{cases} a_1, & \text{if } \frac{\|\mathbf{y}\|^2}{a_1} \le Nv_1; \\ a_i, & \text{if } Nu_{i-1} < \frac{\|\mathbf{y}\|^2}{a_i} \le Nv_i, i = 2, \ldots, L-1; \\ a_L, & \text{if } \frac{\|\mathbf{y}\|^2}{a_L} > Nu_{L-1}. \end{cases} \tag{11}$$

Now, using the CDF in (9b) and decision region in terms of $X = \frac{\|\mathbf{y}\|^2}{a_i}$ given in (11), we can obtain the expression of the

[2]Since the quantization levels are not uniform, the quantization algorithm can be performed by using a bisection search with a complexity $\mathcal{O}\big(\ln(L)\big)$ while the original ML method has a complexity of $\mathcal{O}(L)$.

probability of correct decision on the i-th symbol a_i, denoted by $\mathrm{P}_{\mathrm{c},i}$, as follows:

$$\mathrm{P}_{\mathrm{c},i} = \begin{cases} G(Nv_1), & i = 1; \\ G(Nv_i) - G(Nu_{i-1}), & i = 2, \cdots, L-1; \\ 1 - G(Nu_{L-1}), & i = L. \end{cases} \quad (12)$$

As each symbol is drawn with equal probability, the average SEP denoted by P_e can be represented by

$$\begin{aligned} \mathrm{P}_e &= 1 - \frac{1}{L}\sum_{i=1}^{L} \mathrm{P}_{c,i} \\ &= \frac{1}{L}\sum_{i=1}^{L-1} \big(G(Nu_i) + 1 - G(Nv_i)\big). \end{aligned} \quad (13)$$

All the above discussions can be summarized as the following theorem:

Theorem 1: The average SEP for Rayleigh fading SIMO channels using an L-levels constellation $\mathcal{E} = \{E_1, \ldots, E_L\}$ and the noncoherent ML detector is given by

$$\mathrm{P}_e = \frac{1}{L}\sum_{i=1}^{L-1} \Big(G(Nu_i) + 1 - G(Nv_i)\Big), \quad (14)$$

where $u_i = u(r_i)$ and $v_i = v(r_i)$ are defined in (10) with $r_i = \frac{a_{i+1}}{a_i} = \frac{E_{i+1}+\sigma^2}{E_i+\sigma^2}$. ■

B. Optimal Constellation Design

Our primary purpose in this subsection is to propose an optimal constellation design that minimizes the average SEP subject to an average power constraint. Mathematically, the optimization problem can be formally stated as below:

$$\min_{\mathcal{E}} \ \mathrm{P_e} \qquad \text{s.t.} \begin{cases} \frac{1}{L}\sum_{i=1}^{L} E_i \le E_s, \\ 0 \le E_1 < \cdots < E_L. \end{cases} \quad (15)$$

In order to solve this optimization problem, we now replace the original design variables of the transmitted signal points by the ratios between any two adjacent received constellation signal points, i.e., $r_i = \frac{E_{i+1}+\sigma^2}{E_i+\sigma^2} = \frac{a_{i+1}}{a_i}$, $i = 1, \cdots, L-1$. In this case, the optimization problem (15) can be rewritten as

$$\min_{a_1, \{r_i\}_{i=1}^{L-1}} \mathrm{P_e} \qquad \text{s.t.} \begin{cases} 1 + \sum_{j=1}^{L-1}\prod_{i=1}^{j} r_i \le \frac{L(E_s+\sigma^2)}{a_1}, \\ r_i > 1, \ i = 1, \cdots, L-1. \end{cases} \quad (16)$$

We note that (16) is a constrained multivariate optimization problem and in order to efficiently solve this problem, we first establish the following lemma, which helps us to identify the convexity of the objective function.

Lemma 2: Consider functions $u(r)$ and $v(r)$ defined in Lemma 1. If we let function $F(t)$ be defined by $F(t) = G(Nu(e^t)) + 1 - G(Nv(e^t))$, $t > 0$, where $G(t)$ is defined in (9b), then $F(t)$ is convex for $t > 0$. ■

The proof has been provided in Appendix C.

We restate the optimization problem (16) in terms of variables $t_i = \ln r_i$ as follows:

$$\min_{a_1, \{t_i\}_{i=1}^{L-1}} \mathrm{P}_e = \frac{1}{L}\sum_{i=1}^{L-1} F(t_i) \quad (17a)$$

$$\text{s.t.} \begin{cases} 1 + \sum_{j=1}^{L-1} \exp(\sum_{i=1}^{j} t_i) \le \frac{L(E_s+\sigma^2)}{a_1}, \\ t_i > 0, \ i = 1, 2, \ldots, L-1. \end{cases} \quad (17b)$$

We note that a_1 is only included in the constraint (17b). As such, we should let $\widetilde{a}_1 = \sigma^2$ to maximize the feasible region of t_i, $i = 1, \ldots, L-1$, and remove it from the optimization problem (17). In this case, we denote the system SNR as

$$\mathrm{SNR} = \frac{E_s}{\sigma^2}. \quad (18)$$

As a consequence, (17) can be stated formally in the following problem:

Problem 1: Given constellation size $L \ge 2$, find $L-1$ positive numbers t_i for $i = 1, \ldots, L-1$ such that

$$\min_{\{t_i\}_{i=1}^{L-1}, \{w_j\}_{j=1}^{L-1}} \quad \mathrm{P}_e = \frac{1}{L}\sum_{i=1}^{L-1} F(t_i) \quad (19a)$$

$$\text{s.t. } 1 + \sum_{j=1}^{L-1} \exp(w_j) \le L(1 + \mathrm{SNR}), \quad (19b)$$

$$\sum_{i=1}^{j} t_i = w_j, \ t_i > 0, \ i, j = 1, \ldots, L-1. \quad (19c)$$

■

By Lemma 2, we know that the objective function in (19a) is a sum of $L-1$ convex functions $F(t_i)$, $i = 1, \ldots, L-1$, and hence is convex for $\{t_i\}_{i=1}^{L-1}$. Moreover, we note that the left hand side of (19b) is a sum of $L-1$ convex functions $\exp(w_j)$, $j = 1, \ldots, L-1$ plus 1, which is convex for $\{w_j\}_{j=1}^{L-1}$. Also, the constraint in (19c) is a linear constraint, which is also convex. In other words, Problem 1 has now been reformulated as a convex optimization problem, and therefore can be efficiently solved by using the interior-point method [34].

Though the problem is convex, a closed-form optimal solution to Problem 1 is hard to obtain since the expression of the objective function contains high-order polynomials, where we cannot have a closed-form expression using the Karush-Kuhn-Tucker (KKT) conditions. However, when $L = 2$, we can immediately attain that the optimal solution $\widetilde{t}_1$ is $\widetilde{t}_1 = \ln \mathrm{SNR}$ and thus, the corresponding optimal ratio $\widetilde{r}_1 = \mathrm{SNR}$. We have $\widetilde{E}_1 = \widetilde{a}_1 - \sigma^2 = 0$, and $\widetilde{E}_2 = \widetilde{a}_1\widetilde{r}_1 - \sigma^2 = \sigma^2(\mathrm{SNR} - 1)$. Therefore, in this case, the optimal transmitted binary constellation that minimizes the exact bit error probability is $\widetilde{\mathcal{E}} = \{0, \sigma^2(\mathrm{SNR} - 1)\}$.

C. Optimal Constellation in Large SNR Regime

In this subsection, we propose an optimal constellation design that maximizes the diversity gain for the considered noncoherent SIMO system when the value of SNR goes asymptomatically large. The analysis can provide us with useful insight of the constellation design when the transmitter

is allowed to transmit at a relatively high power. Furthermore, it also serves as a benchmark for the subsequent asymptotic study for the case when the number of receiver antennas goes to infinity. Essentially, the diversity gain quantitatively characterizes how quickly SEP decays when SNR goes to infinity and its formal definition is given as follows [16]:

$$D_{\rm snr} = -\lim_{\rm SNR\to\infty} \frac{\log \rm P_e}{\log \rm SNR}. \tag{20}$$

Recall that $\rm P_e = \frac{1}{L}\sum_{i=1}^{L-1} F(r_i)$, where $F(r) = G\big(Nu(r)\big) + 1 - G\big(Nv(r)\big),\ r > 1$. We denote $r_{\min} = \min\ \{r_i\}_{i=1}^{L-1}$. By Lemma 1, we know that $F(r)$ is a monotonically decreasing function of $r > 1$, and hence P_e is lower and upper bounded by

$$\frac{1}{L}F(r_{\min}) \le \rm P_e \le \frac{L-1}{L}F(r_{\min}). \tag{21}$$

By the definition of the diversity gain given in (20), we have $D_{\rm snr} = -\lim_{\rm SNR\to\infty}\frac{\log F(r_{\min})}{\log \rm SNR}$. Therefore, maximizing the diversity gain is reduced to maximizing $r_{\min} = \min\ \{r_i\}_{i=1}^{L-1}$ as $F(r)$ is a decreasing function for $r > 1$. That is, Problem described in (17) can be re-written as

Problem 2: Given constellation size $L \ge 2$, find $\widetilde{r}$ such that

$$\widetilde{r} = \max_{\{r_i\}_{i=1}^{L-1}} \min\ \{r_i\}_{i=1}^{L-1} \tag{22a}$$

$$\text{s.t.} \begin{cases} 1 + \sum_{j=1}^{L-1}\prod_{i=1}^{j} r_i \le L(1+\rm SNR), \\ r_i > 1,\ i = 1, \cdots, L-1. \end{cases} \tag{22b}$$

■

As $E_i < E_{i+1}$ for $i = 1, \ldots, L-1$, we always have $r_i = \frac{E_{i+1}+\sigma^2}{E_i+\sigma^2} > 1$ and hence $r_{\min} = \min\ \{r_i\}_{i=1}^{L-1} > 1$. By (22b) and denoting $r_{\min}^0 = 1$, we can conclude that

$$\sum_{j=0}^{L-1} r_{\min}^j \overset{(a)}{\le} 1 + \sum_{j=1}^{L-1}\prod_{i=1}^{j} r_i \overset{(b)}{\le} L(\rm SNR+1), \tag{23}$$

where $r_{\min}^j$ denotes the jth power of $r_{\min}$. As $\sum_{j=0}^{L-1} r_{\min}^j$ is an increasing function of $r_{\min}$, then we have $\widetilde{r} = \max\ r_{\min}$ when inequalities (a) and (b) in (23) are achieved simultaneously. In other words, the maximum of $r_{\min}$ is achieved if and only if the following conditions are satisfied simultaneously:

$$\widetilde{a}_1 = \sigma^2,\ r_1 = \cdots = r_{L-1} = \widetilde{r}, \tag{24}$$

where $\widetilde{r}$ can be obtained by solving the following equation

$$\sum_{j=0}^{L-1} \widetilde{r}^j = L(\rm SNR+1). \tag{25}$$

As $\rm SNR > 0$, then from (25), we always have $\widetilde{r} > 1$.

Remark 1: The solution given in (24) shows that the optimal received constellation maximizing the diversity gain is the constellation consisting of *the geometrical sequence* with the noise variance as the initial term and $\widetilde{r}$ as a ratio implicitly determined by (25). ■

We proceed to characterize the asymptotic behavior of the average SEP more accurately for the constellation determined by (24) when SNR goes to infinity, and the results can be summarized as the following theorem:

Theorem 2: Given constellation size L and the noise variance σ^2, the optimal L-level constellation $\widetilde{\mathcal{E}}$ maximizing the diversity gain for the Rayleigh fading SIMO channels using the noncoherent ML detection is given by

$$\widetilde{\mathcal{E}} = \Big\{0, \sigma^2(\widetilde{r}-1), \sigma^2(\widetilde{r}^2-1), \cdots, \sigma^2(\widetilde{r}^{L-1}-1)\Big\},$$

where $\widetilde{r} > 1$ is attained by numerically solving (25). Correspondingly, the resulting SEP is determined by

$$\widetilde{\rm P}_e = \frac{L-1}{L}\big\{G\big(N\widetilde{u}\big) + 1 - G\big(N\widetilde{v}\big)\big\}, \tag{26}$$

where $\widetilde{u} = u(\widetilde{r})$ and $\widetilde{v} = v(\widetilde{r})$. Moreover, when SNR goes to infinity, the SEP for such optimal constellation has the following asymptotic formula

$$\widetilde{P}_e = \widetilde{P}_e^{\infty} + \mathcal{O}\Big((\rm SNR)^{-\frac{N+1}{L-1}}(\ln \rm SNR)^{N+1}\Big), \tag{27}$$

where $\widetilde{\rm P}_e^{\infty}$ is given by

$$\widetilde{\rm P}_e^{\infty} = \frac{L-1}{L}(L\,\rm SNR)^{-\frac{N}{L-1}}\sum_{k=0}^{N}\frac{\big(N\ln(L\rm SNR)\big)^k}{(L-1)^k k!}. \tag{28}$$

■

The proof is given in Appendix D.

From (28), we can observe that, if we let $L = (\ln \rm SNR)^{\alpha}$ with $0 < \alpha < 1$, then we have

$$\begin{aligned}
&\lim_{\rm SNR\to\infty}\frac{\ln\ln \widetilde{\rm P}_e^{-1}}{\ln\ln \rm SNR} \\
&= \lim_{\rm SNR\to\infty}\frac{\ln\ln\Big(\frac{L-1}{L}(L\,\rm SNR)^{-\frac{N}{L-1}}\sum_{k=0}^{N}\frac{\big(N\ln(L\rm SNR)\big)^k}{(L-1)^k k!}\Big)^{-1}}{\ln\ln \rm SNR} \\
&= \lim_{\rm SNR\to\infty}\frac{\ln\ln(L\,\rm SNR)^{\frac{N}{L-1}}}{\ln\ln \rm SNR} + \lim_{\rm SNR\to\infty}\frac{\ln\ln\Big(\frac{L-1}{\ln(L\rm SNR)}\Big)^N}{\ln\ln \rm SNR} \\
&= \lim_{\rm SNR\to\infty}\frac{\ln\ln\big((\ln \rm SNR)^{\alpha}\,\rm SNR\big)}{\ln\ln \rm SNR} - \lim_{\rm SNR\to\infty}\frac{\ln(\ln \rm SNR)^{\alpha}}{\ln\ln \rm SNR} \\
&\quad + \lim_{\rm SNR\to\infty}\frac{\ln\ln\Big(\frac{(\ln \rm SNR)^{\alpha}}{\ln\big((\ln \rm SNR)^{\alpha}\rm SNR\big)}\Big)^N}{\ln\ln \rm SNR} \\
&= 1-\alpha.
\end{aligned} \tag{29}$$

Remark 2: Three important remarks on Theorem 2 can be made as follows:

1) *Diversity Gain*. It can be observed from Theorem 2 that for a fixed transmission rate $\rm R = \ln L$ nats/channel use, the diversity gain for the optimal system is $\frac{N}{L-1}$. In particular, the optimal binary modulation (i.e. $L = 2$) enables full diversity gain N for the noncoherent ML detector. This result is fundamentally different from that for a noncoherent generalized likelihood ratio test (GLRT) receiver analyzed in [35], where in order to achieve full diversity, the number of channel uses is at least twice as large as the number of transmit antennas (i.e., 2 in our considered SIMO system).

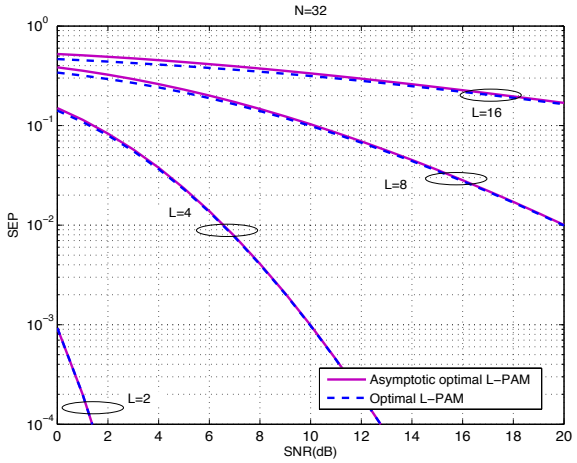


Fig. 1. Comparison of SEP of the optimal constellation (i.e., the solution to Problem 1) and the asymptotically optimal constellation for high SNR given in Theorem 2 versus SNR.

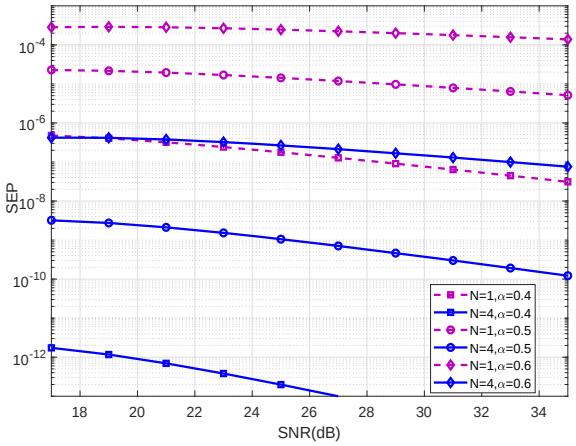


Fig. 2. The asymptotic SEP against SNR with different number of receiver antennas N and α.

2) *Multiplexing-Diversity Gain Tradeoff for Large SNR*. If we let the transmission rate R varies with SNR such that $\mathrm{R} = \ln L = \ln(\ln \mathrm{SNR})^{\alpha} = \alpha \ln \ln \mathrm{SNR}$, where $0 < \alpha < 1$, then by Theorem 2 we can attain $\lim_{\mathrm{SNR}\to\infty} \frac{\ln \ln \widetilde{\mathrm{P}}_e^{-1}}{\ln \ln \mathrm{SNR}} = 1 - \alpha$, as derived in (29). Following the concept of diversity-multiplexing tradeoff for coherent MIMO communications developed in [16], we can term α and $1-\alpha$ as *multiplexing gain* and *diversity gain* for the considered noncoherent SIMO system, respectively. As far as we know, this is for the first time that such a diversity-multiplexing tradeoff is characterized for non-coherent SIMO systems with the CSI not available at both the transmitter and receiver.
3) *Understanding of the Diversity-Multiplexing Tradeoff.* From (29), we can observe that, we should keep $\alpha < 1$ in order to make sure that $\widetilde{\mathrm{P}}_e$ can be arbitrarily small when SNR goes arbitrarily large. That in turn means the maximum transmission rate when SNR goes to infinity is dominated by $\mathrm{R} = \alpha \ln \ln \mathrm{SNR} < \ln \ln \mathrm{SNR}$, which coincide with the information-theoretical study in [32] on the memoryless (i.e., channel coherence time $T_c = 1$ as assumed in our channel model) noncoherent SIMO systems.

In Fig. 1, we illustrate the SEP curves of the optimal constellation (i.e., the solution to Problem 1) and the asymptotically optimal constellation for high SNR, given in Theorem 2. We can see that for various values of L, the latter quickly matches the former, which verifies the correctness of our asymptotic analysis in this subsection. In Fig. 2, we also plot the asymptotic SEP against SNR with different number of receiver antennas and α to show the multiplexing-diversity tradeoff given in Remark 2. It can be observed that extreme low SEPs can be achieved by reducing the transmission rate, i.e., decrease α in a high SNR regime.

D. Optimal Constellation Design for Massive SIMO Systems

In this subsection, we first quantitatively characterize how quickly the SEP decays when the number of the receiver antennas goes to infinity, and then design an optimal constellation that maximizes the geometrical coding gain. To that end, we first establish the following Lemma, which plays an important role in simplifying the expression of the SEP bounds.

Lemma 3: Let function $\rho(t)$ be defined by $\rho(t) = \frac{t}{e^{t-1}}$, and then $\rho(t)$ monotonically increases for $t < 1$, and monotonically decreases for $t > 1$. Moreover, $\rho\big(u(r)\big) = \rho\big(v(r)\big)$ holds for $r > 0$ and $r \neq 1$, where the functions $u(r)$ and $v(r)$ are defined in Lemma 1. ■

The proof is provided in Appendix E.

We are now ready to establish the following bounds on the average SEP when the number of receiver antennas N goes to infinity.

Theorem 3: The SEP derived in Theorem 1 is lower and upper bounded by $\mathrm{P}_e^L < \mathrm{P}_e < \mathrm{P}_e^U$, where

$$\mathrm{P}_e^U = \frac{4(L-1)\sqrt{N}}{L\sqrt{2\pi}}\, \rho_{\min}^N, \tag{30a}$$

$$\mathrm{P}_e^L = \frac{2\sqrt{N-1/6}}{L\sqrt{2\pi N}\, v(r_{\min})}\, \rho_{\min}^N, \tag{30b}$$

with $\rho_{\min}$ defined by

$$\rho_{\min} = \rho\left(v(r_{\min})\right) = \frac{v(r_{\min})}{e^{v(r_{\min})-1}}. \tag{31}$$

■

The proof is provided in Appendix F.

Theorem 3 naturally allow us to unveil the significant asymptotic behavior of the noncoherent massive SIMO systems. Before formally stating it, let us first introduce the following new concept on full-receiver diversity, which is in parallel to [16].

Definition 1: A transmission scheme is said to enable *full-receiver diversity* N for a detector D if, for any given positive numbers ϵ_L and ϵ_U, there exist positive constants C_1, C_2 and $\mathcal{G}_{\mathrm{D}}(\mathrm{SNR}, R) > 1$ independent of N such that the probability of detection error, $\mathrm{P_D}$ satisfies the following condition:

$$C_1 \mathcal{G}_{\mathrm{D}}^{-N(1+\epsilon_L)}(\mathrm{SNR}, \mathrm{R}) \leq \mathrm{P_D} \leq C_2 \mathcal{G}_{\mathrm{D}}^{-N(1-\epsilon_U)}(\mathrm{SNR}, \mathrm{R}).$$

Here, the constant $\mathcal{G}_{\mathrm{D}}(\mathrm{SNR},\mathrm{R})$, for presentation convenience, is called *geometrical coding gain* and relies on both SNR and the transmission data rate R. ■

From Definition 1 and Theorem 3, we conclude that any energy-based signalling enables full receiver diversity N for the noncoherent ML detector. In addition, we can obtain $\lim_{N\to\infty} \mathrm{P}_e^{\frac{1}{N}} = \rho_{\min}$ and thus, the corresponding geometrical coding gain is $\mathcal{G}_{\mathrm{ML}}(\mathrm{SNR},\mathrm{R}) = \rho_{\min}^{-1}$ with $\rho_{\min}$ defined in (31), which essentially characterizes how rapidly SEP decays when the number of the received antennas goes to infinity. Therefore, we should maximize it. Following the same argument as that of maximizing diversity gain in Sec. III-C when SNR tends to infinity, we can arrive at the following theorem.

Theorem 4: Any energy-based constellation $\mathcal{E}$ enables the full-receiver diversity gain for the massive SIMO systems using non-coherent ML detector, and the geometrical coding gain is determined by

$$\mathcal{G}_{\mathrm{ML}}(\mathrm{SNR},\mathrm{R}) = \frac{u(r_{\min})}{e^{u(r_{\min})-1}} = \frac{v(r_{\min})}{e^{v(r_{\min})-1}}.$$

Moreover, the optimal constellation that optimizes the geometrical coding gain is the same as one given in Theorem 2, i.e.,

$$\widetilde{\mathcal{E}} = \big\{0, \sigma^2(\widetilde{r}-1), \cdots, \sigma^2(\widetilde{r}^{L-1}-1)\big\},$$

where $\widetilde{r} > 1$ is obtained by numerically solving (25). Correspondingly, the optimal SEP is determined by

$$\widetilde{\mathrm{P}}_e = \frac{L-1}{L}\Big\{G\big(N\widetilde{u}\big) + 1 - G\big(N\widetilde{v}\big)\Big\}, \tag{32}$$

with $\widetilde{u} = u(\widetilde{r})$ and $\widetilde{v} = v(\widetilde{r})$, and it is lower and upper bounded by

$$\frac{2(L-1)\sqrt{N-1/6}}{L\sqrt{2\pi}Nv(\widetilde{r})}\ \widetilde{\rho}^N \le \widetilde{\mathrm{P}}_e \le \frac{4(L-1)\sqrt{N}}{L\sqrt{2\pi}}\ \widetilde{\rho}^N \tag{33}$$

with the optimal geometrical coding gain expressed by

$$\widetilde{\mathcal{G}}(\mathrm{SNR},\mathrm{R}) = \widetilde{\rho}^{-1} = \frac{e^{u(\widetilde{r})-1}}{u(\widetilde{r})} = \frac{e^{v(\widetilde{r})-1}}{v(\widetilde{r})}. \tag{34}$$

■

As shown in Fig. 3, the SEP curve of the asymptotically optimal constellation for high N, given in Theorem 4, quickly approaches that of the optimal constellation (i.e., the solution to Problem 1) as N increases.

We now see clearly that the optimal SEP decays geometrically when the number of the receiver antennas goes to infinity. However, it not clear yet that for a fixed N, how fast the decaying speed is when either SNR or $\mathrm{R} = \ln L$ goes to infinity, since the geometrical coding gain is a function of $\widetilde{r}$, which is implicitly determined by (25), and thus an implicit function of SNR and R. To make it more clear, let us now investigate the asymptotic behaviour of the geometrical coding gain when either SNR or R becomes large.

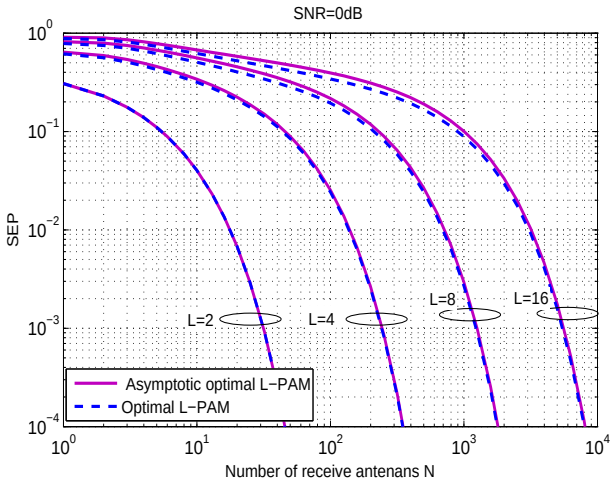


Fig. 3. Comparison of SEP of the optimal constellation (i.e., the solution to Problem 1) and the asymptotically optimal constellation for high SNR given in Theorem 4 versus N.

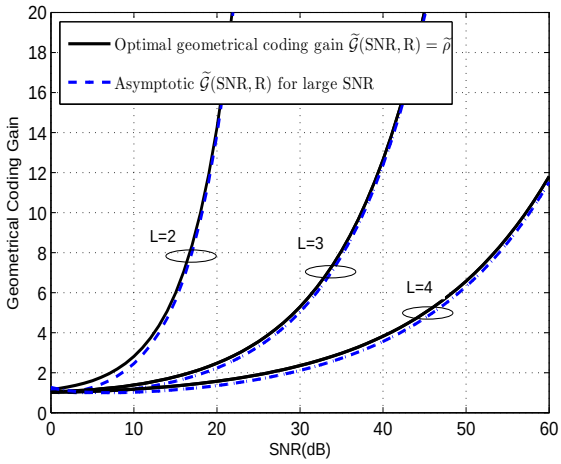


Fig. 4. The optimal geometrical coding gain (34) and its asymptotic expression (35) for high SNR versus SNR, with different values of fixed L.

1) Asymptotic Behaviour of $\widetilde{\mathcal{G}}_{\mathrm{ML}}(\mathrm{SNR},\mathrm{R})$ for Large SNR: For the case that R (i.e., L) is fixed and SNR goes to infinity, using (45) given in Appendix D, we have

$$\begin{aligned}
\widetilde{\mathcal{G}}_{\mathrm{ML}}^{-1}(\mathrm{SNR},\mathrm{R}) &= e\widetilde{u}\,e^{-\widetilde{u}} \\
&= e\Big(\frac{\ln(L\mathrm{SNR})}{L-1}(L\mathrm{SNR})^{-\frac{1}{L-1}} + \mathcal{O}\big((L\mathrm{SNR})^{-\frac{2}{L-1}}\ln\mathrm{SNR}\big)\Big) \\
&\quad \times \big(1 + \mathcal{O}((L\mathrm{SNR})^{-\frac{1}{L-1}}\ln\mathrm{SNR})\big) \\
&= \frac{e\ln(L\mathrm{SNR})}{L-1}(L\mathrm{SNR})^{-\frac{1}{L-1}} + \mathcal{O}\big((L\mathrm{SNR})^{-\frac{2}{L-1}}\ln^2\mathrm{SNR}\big) \\
&= \frac{e\ln(L\mathrm{SNR})}{L-1}(L\mathrm{SNR})^{-\frac{1}{L-1}}\Big(1 + \mathcal{O}\big((L\mathrm{SNR})^{-\frac{1}{L-1}}\ln\mathrm{SNR}\big)\Big)
\end{aligned}$$

which is equivalent to giving us:

Proposition 1: The optimal geometrical coding gain has the following asymptotic formula:

$$\widetilde{\mathcal{G}}_{\mathrm{ML}}(\mathrm{SNR},\mathrm{R}) = \frac{(L-1)(L\mathrm{SNR})^{\frac{1}{L-1}}}{e\ln(L\mathrm{SNR})} + \mathcal{O}\big((L\mathrm{SNR})^{-\frac{1}{L-1}}\ln\mathrm{SNR}\big). \tag{35}$$

■

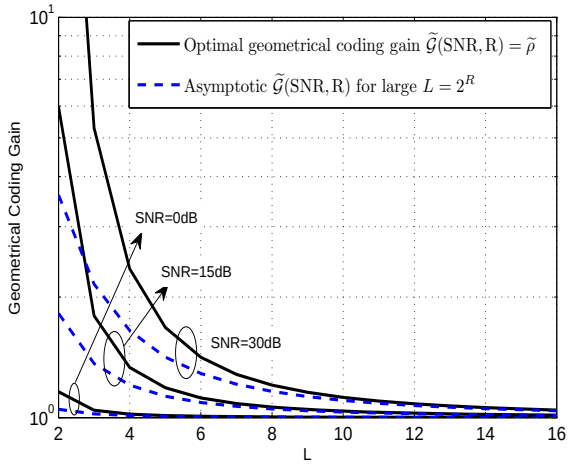


Fig. 5. The optimal geometrical coding gain (34) and its asymptotic expression (38) for large R (i.e., L) versus R, with different values of fixed SNR.

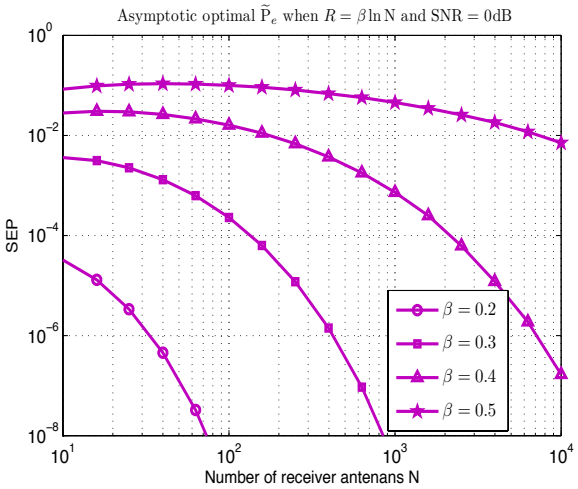


Fig. 6. The array multiplexing gain and diversity gain tradeoff for large N with different β.

We can see from Proposition 1 that the geometrical coding gain goes to infinity when SNR tends to infinity. As a consequence, it is expected to see that there is no error floor in the SEP curve when SNR tends to infinity. Our asymptotic analysis of the geometrical coding gain $\widetilde{\mathcal{G}}_{\mathrm{ML}}(\mathrm{SNR},\mathrm{R})$ when SNR goes large is validated in Fig. 4.

2) Asymptotic Behavior of $\widetilde{\mathcal{G}}_{\mathrm{ML}}(\mathrm{SNR},\mathrm{R})$ for Large R*:* On the other hand, in order to investigate an asymptotic behaviour of the optimal geometrical coding gain when the transmission rate R or equivalently the constellation size L tends to infinity, and SNR is fixed, we need to establish the following two lemmas.

Lemma 4: For any given value of SNR, the equation

$$t=\ln\big(1+t(1+\mathrm{SNR})\big) \tag{36}$$

always has a unique positive solution for $t>\ln(1+\mathrm{SNR})$, and the resulting unique solution is denoted by $\tau(\mathrm{SNR})$. ■

The proof is provided in Appendix G.

Lemma 5: Let $\widetilde{r}$ be determined by (25). Then, if L goes to infinity, $\widetilde{r}$ has the following asymptotic formula

$$\widetilde{r}=1+\tau(\mathrm{SNR})L^{-1}+\frac{\big(1+\tau(\mathrm{SNR})(1+\mathrm{SNR})\big)\ln^2\big(1+\tau(\mathrm{SNR})(1+\mathrm{SNR})\big)}{\tau(\mathrm{SNR})(1+\mathrm{SNR})}L^{-2}+\mathcal{O}\big(L^{-3}\big), \tag{37}$$

where $\tau(\mathrm{SNR})$ is determined by (36). ■

The proof is provided in Appendix H.

Proposition 2: Let $\widetilde{r}$ be determined by (25). Then, the optimal geometrical coding gain $\widetilde{\mathcal{G}}(\mathrm{SNR},\mathrm{R})$ defined in Theorem 4 has the following asymptotic formula,

$$\widetilde{\mathcal{G}}(\mathrm{SNR},\mathrm{R})=1+\frac{\tau^2(\mathrm{SNR})}{8L^2}+\mathcal{O}\big(L^{-3}\big). \tag{38}$$

when L goes to infinity, where $\tau(\mathrm{SNR})$ is determined by (36). ■

The proof is provided in Appendix I.

In Figure 5, we plot the curves of the geometrical coding gain and its asymptotic expression when R approaches infinity. We can see from the figure that the curves of the asymptotic results quickly coincide with that of the exact results, which verifies the effectiveness of our asymptotic analysis.

3) Diversity-Multiplexing Gain Tradeoff for Large N: To characterize the new diversity-multiplexing gain tradeoff for large N, we let the rate $R=\ln L$ varies with N. To model this relationship, we introduce a new concept of *array-multiplexing gain* defined as follows:

Definition 2: The *array-multiplexing gain* $0<r_{\mathrm{array}}<1$ describes how the rate R increases as the number of antenna N approaches to infinity. Mathematically,

$$r_{\mathrm{array}}=\lim_{N\to\infty}\frac{\mathrm{R}(N)}{\frac{1}{2}\ln N},$$

in which the factor $\frac{1}{2}$ is due to the usage of real signalling. ■

We now derive the diversity gain by investigating the asymptotic behavior of the error performance $\widetilde{\mathrm{P}}_e$ given in (32) as the transmission rate R tends to infinity together with N for any fixed SNR. We then attain the following important result on the tradeoff between the array-multiplexing gain and the diversity gain.

Theorem 5: For any fixed SNR, if we set the *array multiplexing gain* as β, then the corresponding geometrical coding gain is $\exp\big(\frac{\tau^2(\mathrm{SNR})}{8}\big)$ and the diversity gain is $N^{1-\beta}$. ■

The proof is given in Appendix J.

Theorem 5 reveals such a significant fact that if about N^{β} receiving antennas are used for enhancing the transmission data rate, then there are about $N^{1-\beta}$ receiving antennas left for reliably recovering the information symbols. The asymptotic SEP performance against the number of receiver antennas is plotted in Fig. 6 with different β to show the array multiplexing and diversity gain tradeoff given in Theorem 5. As can be found that, with the increase of the number of receiver antennas, we can achieve very low SEP as targeted in IIoT.

IV. Simulation Results

In this section, we carry out computer simulations to verify our theoretical results. We also compare our proposed opti-

mal system with the minimum-distance constellation design scheme in [26], i.e., a unipolar L-level PAM constellation (L-PAM) with equal power distance, in which the average received signal energy is approximated as Gaussian random variable with constant variance in the large-antenna limit by the central limit theorem. Therefore, it leads to a suboptimal detector. For fair comparisons and simplicity, we estimate the transmitted signals using the proposed fast ML detector in Algorithm 1 for both constellations.

Figs. 7(a) and 7(b) show the SEP in the noncoherent SIMO system over Rayleigh fading channels with the number of receiver antennas increasing and various constellation sizes at $\mathrm{SNR} = 0\mathrm{dB}$ and $\mathrm{SNR} = 10\mathrm{dB}$, respectively. In order to examine the derived closed-form expression on SEP in (13), we carry out Monte Carlo simulations to compare the simulated average SEP results with the theoretical expression in these two figures. The results show that they match well with each other, which verifies the correctness of our analysis. Therefore, in the following, we only use the theoretical result to show some properties. We can also see that the SEP of both schemes approximately exhibits exponential decay as the number of receiver antenna increases, and the proposed optimal constellation yields much better error performance gain than the minimum-distance PAM constellation. Specifically, we can observe from Figs. 7(a) and 7(b) that the performance gain also depends on the number of receiver antennas and SNR. Furthermore, the performance gain enlarges as the number of receiver antennas or SNR increases.

The SEP of the both constellations when SNR increases is depicted in Figs. 8(a) and 8(b), where the number of receiver antennas is fixed at $N = 32$ and $N = 64$, respectively. We can observe from these two figures that the SEP of the proposed optimal constellation approximately exhibits a polynomial decay as SNR increases, whereas the minimum-distance PAM constellation yields error performance floors at the medium and high SNR regime. The reason for this phenomenon is that the proposed optimal constellation adapts to SNR, while the PAM constellation is fixed for any SNR.

V. Conclusions

In this paper, we have devised an optimal energy-based modulation (i.e., nonnegative PAM) constellation for noncoherent SIMO systems over Rayleigh fading, which is designed to enable the ultra-reliable low-latency communications required by critical industrial IoT use cases. Specifically, we first derive a fast noncoherent ML detection algorithm and the closed-form expression of its SEP, and then find an optimal geometrical received nonnegative PAM constellation that can maximize the diversity gain for large SNR. In addition, the lower and upper bounds of SEP have been derived to quantitatively characterize how fast SEP decays as the number of receiver antennas tends to large, which show that any nonnegative PAM constellation enables full receiver diversity for the large number of antennas. Again, the optimal received constellation that maximizes the geometrical coding gain has been found to be geometrical for the considered noncoherent massive SIMO system. Moreover, for this optimal constellation, two kinds of multiplexing and diversity gain tradeoffs in terms of SNR and the number of the received antennas have been characterized. Simulation results have corroborated our analytical analysis and showed that the efficient utilization of the statistics of the channels at the transceiver leads to significant performance enhancement, particularly for the massive SIMO system or the high SNR regime. As a future work, we will extend the design framework developed in this paper to the multi-user scenario with more than one users accessing the channel at the same time. More importantly, we will also implement the developed algorithms on a real-world software-defined radio (SDR) platform.

Appendix

A. Proof of Lemma 1

First, we note that the fundamental inequality in information theory [36, Lemma 2.29] is

$$t - 1 - \ln t \geq 0, \quad t > 0, \tag{39}$$

where the equality holds if and only if $t = 1$.

1) For $r > 1$, we note that $\ln r > 0$, $r \ln r > 0$, and $r - 1 > 0$. Hence, we have $u(r) > 0$ and $v(r) > 0$. Now, to show $u(r) < 1$ and $v(r) > 1$ for $r > 1$ is equivalent to show that: $(r-1) - \ln r > 0$, $r \ln r - (r-1) = r\Big(\frac{1}{r} - 1 - \ln\big(\frac{1}{r}\big)\Big) > 0$, which are both true for $r > 1$ by the fundamental inequality (39).
2) By using the fundamental inequality (39) again, we can show that, for $r > 1$, the first-order derivatives of $u(r)$ and $v(r)$ have the following properties:

$$u'(r) = -\frac{\frac{1}{r} - 1 - \ln\frac{1}{r}}{(r-1)^2} < 0, \; v'(r) = \frac{r - 1 - \ln r}{(r-1)^2} > 0,$$

Hence, $u(r)$ and $v(r)$ are monotonically decreasing and increasing functions of r, for $r > 1$, respectively.

This completes the proof of Lemma 1. □

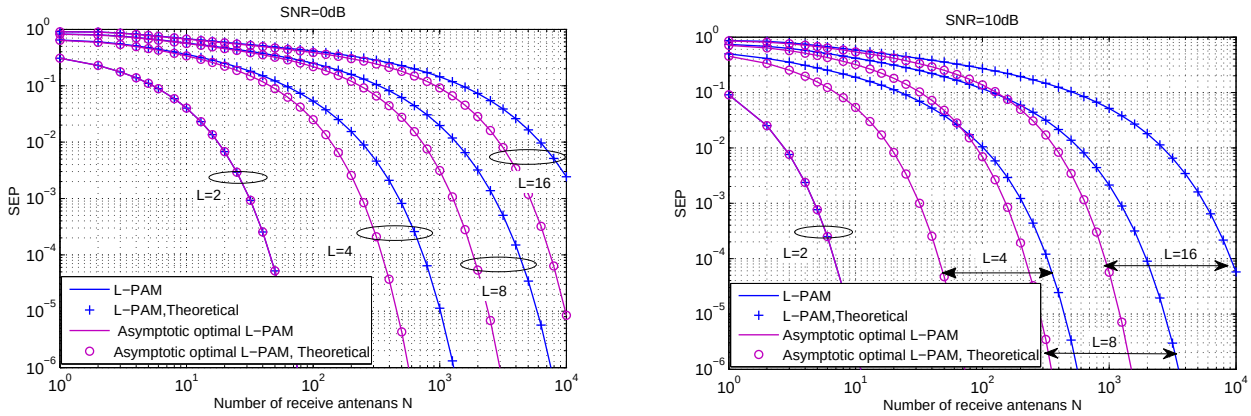
B. Proof of Algorithm 1

We first consider then case when $b_{i-1} < \frac{\|\mathbf{y}\|^2}{N} \leq b_i$, for $i = 2, \ldots, L-1$. By definition (7), for $k = 1, \ldots, i-1$, we have:

$$\begin{aligned}\frac{\|\mathbf{y}\|^2}{N} > b_{i-1} &= \frac{a_{i-1}a_i \ln\big(\frac{a_i}{a_{i-1}}\big)}{a_i - a_{i-1}} = \frac{a_i \ln\big(\frac{a_i}{a_{i-1}}\big)}{\frac{a_i}{a_{i-1}} - 1} \\ &= a_i u\Big(\frac{a_i}{a_{i-1}}\Big) \overset{(a)}{\geq} a_i u\Big(\frac{a_i}{a_k}\Big) = \frac{a_k a_i \ln\big(\frac{a_i}{a_k}\big)}{a_i - a_k}.\end{aligned} \tag{40}$$

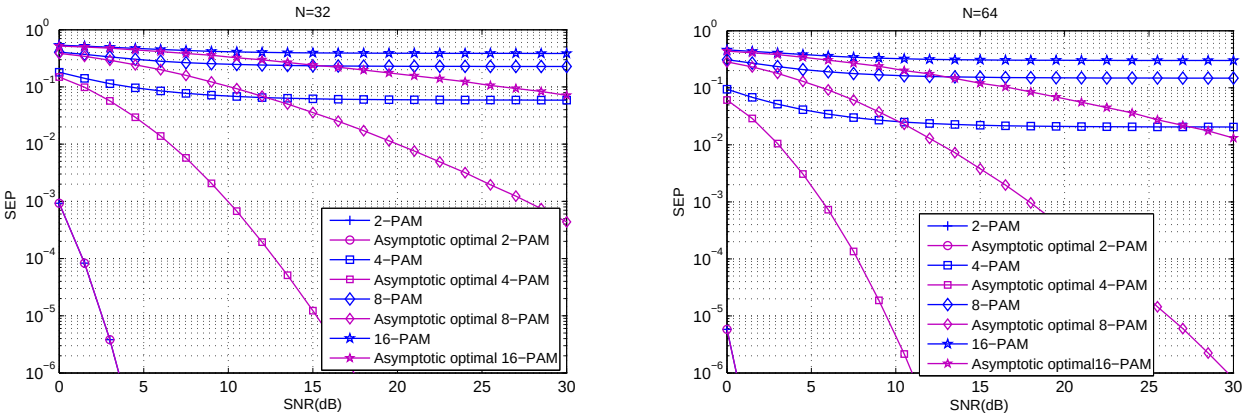
where inequality (a) holds since $u\big(\frac{a_i}{a_{i-1}}\big) \geq u\big(\frac{a_i}{a_k}\big)$ for $a_k \leq a_{i-1}$, as shown by Lemma 1. Likewise, for $\ell = i+1, \ldots, L$, we have that:

$$\begin{aligned}\frac{\|\mathbf{y}\|^2}{N} \leq b_i &= \frac{a_i a_{i+1} \ln\big(\frac{a_{i+1}}{a_i}\big)}{a_{i+1} - a_i} = \frac{a_i \frac{a_{i+1}}{a_i} \ln\big(\frac{a_{i+1}}{a_i}\big)}{\frac{a_{i+1}}{a_i} - 1} \\ &= a_i v\Big(\frac{a_{i+1}}{a_i}\Big) \overset{(b)}{\leq} a_i v\Big(\frac{a_\ell}{a_i}\Big) = \frac{a_i a_\ell \ln\big(\frac{a_\ell}{a_i}\big)}{a_\ell - a_i}.\end{aligned} \tag{41}$$



(a) SNR = 0 dB (b) SNR = 10 dB

Fig. 7. Performance comparison for 1-bit ($L = 2$), 2-bit ($L = 4$), 3-bit ($L = 8$) and 4-bit ($L = 16$) against the number of receiver antennas N. (a) SNR = 0dB, (b) SNR = 10dB.



(a) $N = 32$ (b) $N = 64$

Fig. 8. Performance comparison for 1-bit ($L = 2$), 2-bit ($L = 4$), 3-bit ($L = 8$) and 4-bit ($L = 16$) versus SNR. (a) $N = 32$, (b) $N = 64$.

where (b) is true since $v\left(\frac{a_\ell}{a_i}\right) \geq v\left(\frac{a_{i+1}}{a_i}\right)$ for $a_\ell \geq a_{i+1}$ as indicated in Lemma 1. Then, by Eqs. (40) and (41), we have

$$\frac{\|\mathbf{y}\|^2}{a_i} + N\ln a_i < \frac{\|\mathbf{y}\|^2}{a_k} + N\ln a_k,\ k = 1,\ldots,i-1,$$
$$\frac{\|\mathbf{y}\|^2}{a_i} + N\ln a_i \leq \frac{\|\mathbf{y}\|^2}{a_\ell} + N\ln a_\ell,\ \ell = i+1,\ldots,L.$$

Now, according to the ML decision rule in (4), we conclude that for $b_{i-1} < \frac{\|\mathbf{y}\|^2}{N} \leq b_i$, we have $\hat{a} = a_i$, $i = 2,\ldots,L-1$. The cases when $\frac{\|\mathbf{y}\|^2}{N} \leq b_1$ and $\frac{\|\mathbf{y}\|^2}{N} > b_{L-1}$ can be proved in a similar fashion, and hence they are omitted for brevity. This completes the proof of Algorithm 1. □

C. Proof of Lemma 2

It is sufficient to prove that the second-order derivative of $F(t)$ is always positive for $t > 0$. To that end, we denote $u_1(t) = u(e^t)$ and $v_1(t) = v(e^t)$. Now, using the PDF in (9a) and its CDF in (9b), we have

$$\begin{aligned} G'(Nu_1(t)) &= f_X(Nu_1(t))Nu_1'(t) \\ &= \frac{N^N u_1^{N-1}(t)e^{-Nu_1(t)}u_1'(t)}{(N-1)!}, \end{aligned} \tag{42a}$$
$$\begin{aligned} G'(Nv_1(t)) &= f_X(Nv_1(t))Nv_1'(t) \\ &= \frac{N^N v_1^{N-1}(t)e^{-Nv_1(t)}v_1'(t)}{(N-1)!}. \end{aligned} \tag{42b}$$

Since $v_1(t) = t + u_1(t) = e^t u_1(t)$, where $u_1(t)$ is determined by $u_1(t) = \frac{t}{e^t-1}$, we can obtain $v_1^{N-1}(t) = e^{-t}e^{Nt}u_1^{N-1}(t)$, $v_1'(t) = e^t(u_1(t) + u_1'(t))$, and $e^{-Nv_1(t)} = e^{-Nt}e^{-Nu_1(t)}$. Thus, (42a) and (42b) have the following relationship:

$$\begin{aligned} G'(Nv_1(t)) &= \frac{N^N u_1^{N-1}(t)e^{-Nu_1(t)}(u_1(t) + u_1'(t))}{(N-1)!} \\ &= G'(Nu_1(t)) + \frac{N^N u_1^N(t)e^{-Nu_1(t)}}{(N-1)!}. \end{aligned}$$

Now, the first-order derivative of $F(t)$ can be reformulated by

$$\begin{aligned} F'(t) &= G'(Nu_1(t)) - G'(Nv_1(t)) \\ &= -\frac{N^N u_1^N(t) e^{-Nu_1(t)}}{(N-1)!}. \end{aligned}$$

As a consequence, the second-order derivative of $F(t)$ can be calculated by

$$F''(t) = -\frac{N^{N+1} u_1^{N-1}(t) u_1'(t)(1-u_1(t)) e^{-Nu_1(t)}}{(N-1)!}.$$

By Lemma 1 and the definition of $u_1(t)$, we have $0 < u_1(t) < 1$. Also, we let $r = e^t$, then we have $u_1'(t) = u'(r)e^t < 0$. Now, we can conclude that $F''(t) > 0$ for $t > 0$.

This completes the proof of Lemma 2. □

D. Proof of Theorem 2

First, starting from the optimal ratio $\widetilde{r}$ determined by (25), we can attain that $\widetilde{r}^{L-1}$ is lower and upper bounded by

$$\mathrm{SNR} + 1 < \widetilde{r}^{L-1} < L(\mathrm{SNR} + 1). \tag{43}$$

Hence, we can obtain the following asymptotic formula of $\widetilde{r}$ when SNR tends to infinity:

$$\begin{aligned} \widetilde{r} &= \big(L(\mathrm{SNR}+1) + \mathcal{O}(\widetilde{r}^{L-2})\big)^{\frac{1}{L-1}} \\ &= (L\,\mathrm{SNR})^{\frac{1}{L-1}} \big(1 + \mathcal{O}(\mathrm{SNR}^{-\frac{1}{L-1}})\big)^{\frac{1}{L-1}} \\ &= (L\,\mathrm{SNR})^{\frac{1}{L-1}} + \mathcal{O}(1). \end{aligned} \tag{44}$$

Now, substituting (44) into $\widetilde{u}$ and $\widetilde{v}$ produces

$$\begin{aligned} \widetilde{u} &= \ln \widetilde{r}\Big(\frac{1}{\widetilde{r}} + \frac{1}{\widetilde{r}^2} + \cdots\Big) \\ &= \frac{\ln(L\mathrm{SNR})}{L-1}(L\mathrm{SNR})^{-\frac{1}{L-1}} + \mathcal{O}\big((L\mathrm{SNR})^{-\frac{2}{L-1}} \ln \mathrm{SNR}\big), \end{aligned} \tag{45}$$

$$\widetilde{v} = \widetilde{r}\widetilde{u} = \frac{\ln(L\mathrm{SNR})}{L-1} + \mathcal{O}\big((L\mathrm{SNR})^{-\frac{1}{L-1}} \ln \mathrm{SNR}\big). \tag{46}$$

Then, by expanding e^t into the Taylor series at $t = 0$ with the Lagrange remainder term, we obtain

$$e^{N\widetilde{u}} = \sum_{k=0}^{N} \frac{(N\widetilde{u})^k}{k!} + \frac{(N\widetilde{u})^{N+1} e^{\xi}}{(N+1)!}, \tag{47}$$

where $0 \le \xi \le N\widetilde{u}$. Hence, we have

$$\begin{aligned} G\big(N\widetilde{u}\big) &= 1 - e^{-N\widetilde{u}} \sum_{k=0}^{N-1} \frac{(N\widetilde{u})^k}{k!} \\ &= e^{-N\widetilde{u}}\Big(e^{N\widetilde{u}} - \sum_{k=0}^{N-1} \frac{(N\widetilde{u})^k}{k!}\Big) \\ &\overset{(a)}{=} \frac{(N\widetilde{u})^N e^{-N\widetilde{u}}}{N!} + \frac{(N\widetilde{u})^{N+1} e^{-(N\widetilde{u}-\xi)}}{(N+1)!} \\ &\overset{(b)}{=} \frac{\big(\frac{N\ln(L\mathrm{SNR})}{L-1}\big)^N}{N!}(L\mathrm{SNR})^{-\frac{N}{L-1}} \\ &\quad + \mathcal{O}\big((\mathrm{SNR})^{-\frac{N+1}{L-1}} (\ln \mathrm{SNR})^{N+1}\big), \end{aligned}$$

where we have used (47) in (a) and (45) in (b).

Likewise, utilizing (44), and (46), we can have

$$\begin{aligned} &1 - G\big(N\widetilde{v}\big) = e^{-N\widetilde{v}} \sum_{k=0}^{N-1} \frac{(N\widetilde{v})^k}{k!} \\ &= e^{-N\widetilde{v}}\Big(\sum_{k=0}^{N-1} \frac{(\frac{N\ln(L\mathrm{SNR})}{L-1})^k}{k!} + \mathcal{O}\big((\mathrm{SNR})^{-\frac{1}{L-1}} (\ln \mathrm{SNR})^{N-1}\big)\Big) \\ &= \Big((L\mathrm{SNR})^{-\frac{N}{L-1}} + \mathcal{O}\big(\mathrm{SNR}^{-\frac{N+1}{L-1}}\big)\Big) \\ &\quad \times \Big(\sum_{k=0}^{N-1} \frac{(\frac{N\ln(L\mathrm{SNR})}{L-1})^k}{k!} + \mathcal{O}\big((\mathrm{SNR})^{-\frac{1}{L-1}} (\ln \mathrm{SNR})^{N-1}\big)\Big). \end{aligned}$$

Therefore, we have

$$\begin{aligned} \widetilde{\mathrm{P}}_e &= \frac{L-1}{L}\big\{G\big(N\widetilde{u}\big) + 1 - G\big(N\widetilde{v}\big)\big\}, \\ &= \widetilde{P}_e^{\infty} + \mathcal{O}\big((\mathrm{SNR})^{-\frac{N+1}{L-1}} (\ln \mathrm{SNR})^{N+1}\big), \end{aligned}$$

where notation $\widetilde{\mathrm{P}}_e^{\infty}$ denotes the dominant term of $\widetilde{\mathrm{P}}_e$ at high SNR regime given by:

$$\widetilde{\mathrm{P}}_e^{\infty} = \frac{L-1}{L}(L\,\mathrm{SNR})^{-\frac{N}{L-1}} \sum_{k=0}^{N} \frac{(N\ln(L\mathrm{SNR}))^k}{(L-1)^k k!}.$$

We completes the proof of Theorem 2. □

E. Proof of Lemma 3

Since $\rho'(t) = \frac{1-t}{e^{t-1}}$, we have $\rho'(t) > 0$ for $t < 1$ and $\rho'(t) < 0$ for $t > 1$, i.e., with the increases of t, $\rho(t)$ monotonically increases for $t < 1$ and monotonically decreases for $t > 1$. In addition, for $r > 0$ and $r \neq 1$, it can be verified directly that $\frac{\rho(v(r))}{\rho(u(r))} = \frac{v(r)}{u(r)} \exp\big(u(r) - v(r)\big) = r \exp\big(\frac{\ln r}{r-1} - \frac{r \ln r}{r-1}\big) = r \exp(-\ln r) = 1$, and hence we have $\rho\big(u(r)\big) = \rho\big(v(r)\big)$.

This completes the proof of Lemma 3. □

F. Proof of Theroem 3

Let us prove the upper bound first. On the one hand, similar to (47), by using Taylor series at $t = 0$, we know that e^{Nu} can be expanded as

$$e^{Nu} = \sum_{k=0}^{3N-1} \frac{(Nu)^k}{k!} + \frac{(Nu)^{3N} e^{\xi}}{(3N)!}. \tag{48}$$

where $0 \le \xi \le 3Nu$. Therefore, according to (9b), we can rewrite $G\big(Nu\big)$ as

$$\begin{aligned} G\big(Nu\big) &= 1 - \Big(\sum_{k=0}^{3N-1} \frac{(Nu)^k}{k!} - \sum_{k=N}^{3N-1} \frac{(Nu)^k}{k!}\Big) e^{-Nu} \\ &\overset{(a)}{=} 1 - \Big(e^{Nu} - \frac{(Nu)^{3N} e^{\xi}}{(3N)!} - \sum_{k=N}^{3N-1} \frac{(Nu)^k}{k!}\Big) e^{-Nu} \\ &= \frac{(Nu)^{3N} e^{-(Nu-\xi)}}{(3N)!} + e^{-Nu} \sum_{k=N}^{3N-1} \frac{(Nu)^k}{k!}, \end{aligned} \tag{49}$$

where we have used (48) in step (a). On the other hand, for $0 < u < 1$, we consider sequence $c_k = \frac{(Nu)^k}{k!}$ where $\frac{c_{k+1}}{c_k} = \frac{Nu}{k} \le u < 1$ for $k = N, \ldots, 3N-1$. Hence, we have

$$\frac{(Nu)^k}{k!} \le \frac{(Nu)^N}{N!}, \quad k = N, \cdots, 3N-1. \tag{50}$$

Now, combining (49) with (50) yields

$$G\big(Nu\big) \leq \frac{(Nu)^{3N}}{(3N)!} + \frac{2N(Nu)^N}{N!}e^{-Nu} \tag{51}$$

Then, using the upper bound of the Stirling's inequality in [37], i.e.,

$$\frac{N^N}{N!} \leq \frac{e^N}{\sqrt{2\pi N}}. \tag{52}$$

We can further give the upper bound of $G\big(Nu\big)$ in (51) by

$$\begin{aligned} G\big(Nu\big) &< \frac{1}{\sqrt{6\pi N}}\Big(\frac{ue}{3}\Big)^{3N} + \frac{\sqrt{2N}}{\sqrt{\pi}}\Big(\frac{u}{e^{u-1}}\Big)^N \\ &< \frac{3\sqrt{N}}{\sqrt{2\pi}}\Big(\frac{u}{e^{u-1}}\Big)^N, \end{aligned} \tag{53}$$

where we have used the fact that $\frac{ue}{3} < \frac{u}{e^{u-1}}$ for $0 < u < 1$. Regarding the second term of P_e, we notice that

$$1 - G\big(Nv\big) = e^{-Nv}\sum_{k=0}^{N-1}\frac{(Nv)^k}{k!}. \tag{54}$$

From Lemma 1 we know that $v > 1$. In this case, we consider sequence $d_k = \frac{(Nv)^k}{k!}$, where $\frac{d_{k+1}}{d_k} = \frac{Nv}{k+1} \geq v > 1$ for $k = 0, 1, \cdots, N-1$, we have that

$$\frac{(Nv)^k}{k!} \leq \frac{(Nv)^{N-1}}{(N-1)!}, \quad k = 0, \ldots, N-1. \tag{55}$$

Now, combining (54) with (55), we can upper bound $1 - G\big(Nv\big)$ by

$$1 - G\big(Nv\big) \leq e^{-Nv}\frac{N(Nv)^{N-1}}{(N-1)!}. \tag{56}$$

Using the upper bound of the Stirling inequality (52) again, (56) can be further upper bounded by

$$1 - G\big(Nv\big) \quad \leq \frac{\sqrt{N}}{v\sqrt{2\pi}}\big(\tfrac{v}{e^{v-1}}\big)^N < \frac{\sqrt{N}}{\sqrt{2\pi}}\big(\tfrac{v}{e^{v-1}}\big)^N, \tag{57}$$

By Lemma 3, we know

$$\rho = \frac{u}{e^{u-1}} = \frac{v}{e^{v-1}}. \tag{58}$$

Combing (53), (57), (58) with (21), we conclude that

$$\begin{aligned} \mathrm{P}_e &= \frac{L-1}{L}G\big(Nu(r_{\min})\big) + 1 - G\big(Nv(r_{\min})\big) \\ &\leq \frac{4(L-1)\sqrt{N}}{L\sqrt{2\pi}}\ \rho_{\min}^N \end{aligned}$$

This completes the proof of Proposition 3 on the upper bound.

Now, we consider to prove the lower bound of Lemma (3) and by (21) again, we have

$$\begin{aligned} \mathrm{P}_e &= \frac{1}{L}\sum_{i=1}^{L-1}\Big(G(Nu_i) + 1 - G(Nv_i)\Big) \\ &\geq \frac{1}{L}\Big(e^{-Nu_{\min}}\frac{(Nu_{\min})^N}{N!} + e^{-Nv_{\min}}\frac{(Nv_{\min})^{N-1}}{(N-1)!}\Big) \\ &\overset{(a)}{\geq} \frac{1}{L}\Big(\frac{\sqrt{N-1/6}}{\sqrt{2\pi}Nv_{\min}}\Big(\frac{u_{\min}}{e^{u_{\min}-1}}\Big)^N + \frac{\sqrt{N-1/6}}{\sqrt{2\pi}Nv_{\min}}\Big(\frac{v_{\min}}{e^{v_{\min}-1}}\Big)^N\Big) \\ &\overset{(b)}{=} \frac{2\sqrt{N-1/6}}{L\sqrt{2\pi}Nv_{\min}}\Big(\frac{v_{\min}}{e^{v_{\min}-1}}\Big)^N. \end{aligned}$$

where (a) is true by using the lower bound of Stirling's inequality in [38], i.e.,

$$\frac{N^N}{N!} \geq \frac{e^N\sqrt{N-1/6}}{\sqrt{2\pi}N}. \tag{59}$$

Also, in (b), we have used (58). Now, we complete the proof of the lower bound, and thus Theorem 3. □

G. Proof of Lemma 4

Consider a function $f(t) = e^t - \big(1 + t(1+\mathrm{SNR})\big)$ for $t > \ln(1+\mathrm{SNR})$. Since its first-order derivative $f'(t) = e^t - (1+\mathrm{SNR}) > 0$ when $t > \ln(1+\mathrm{SNR})$, it monotonically increases. In order to prove $f(\ln(1+\mathrm{SNR})) < 0$, we consider a function $g(t) = t - (1 + t\ln t)$ for $t > 1$. Notice its first-order derivative $g'(t) = -\ln t < 0$ for $t > 1$. Hence, $g(t)$ monotonically decreases. Since $g(1) = 0$, we have $g(t) < 0$ for $t > 1$. Therefore, we attain $f(\ln(1+\mathrm{SNR})) = (1+\mathrm{SNR}) - \big(1 + (1+\mathrm{SNR})\ln((1+\mathrm{SNR}))\big) < 0$. In addition, it is not difficult to see that $f(\infty) = \infty$. Hence, there exists a unique positive number $\tau(\mathrm{SNR})$ dependent of SNR satisfying $e^{\tau(\mathrm{SNR})} = 1 + \tau(\mathrm{SNR})(1+\mathrm{SNR})$, i.e., $\tau(\mathrm{SNR}) = \ln\big(1 + \tau(\mathrm{SNR})(1+\mathrm{SNR})\big)$. This completes the proof of Lemma 4. □

H. Proof of Lemma 5

Substituting the identity $\sum_{j=0}^{L-1}\widetilde{r}^j = \frac{\widetilde{r}^L - 1}{\widetilde{r} - 1}$ into (25), we obtain $\widetilde{r}^L = L(1+\mathrm{SNR})(\widetilde{r}-1) + 1$. Hence, we have

$$\begin{aligned} \widetilde{r} &= \big(L(1+\mathrm{SNR})(\widetilde{r}-1) + 1\big)^{\frac{1}{L}} \\ &= \exp\Big(\frac{\ln\big(L(1+\mathrm{SNR})(\widetilde{r}-1)+1\big)}{L}\Big) \\ &= 1 + \frac{\ln\big(L(1+\mathrm{SNR})(\widetilde{r}-1)+1\big)}{L} \\ &\quad + \frac{\ln^2\big(L(1+\mathrm{SNR})(\widetilde{r}-1)+1\big)}{2L^2} + \mathcal{O}(L^{-3}). \end{aligned} \tag{60}$$

On the other hand, from (25) we know $\widetilde{r}^{L-1} > 1 + \mathrm{SNR}$. Therefore, we have $\widetilde{r} > (1+\mathrm{SNR})^{1/(L-1)} > 1 + \frac{\ln(1+\mathrm{SNR})}{L}$ and thus, $\lim_{L\to\infty} L(\widetilde{r}-1) \geq \ln(1+\mathrm{SNR})$. Combining this with Lemma 4 and (60) yields $\tau(\mathrm{SNR}) = \lim_{L\to\infty} L(\widetilde{r}-1)$. Hence, we can write

$$\widetilde{r} = 1 + \tau(\mathrm{SNR})L^{-1} + AL^{-2} + \mathcal{O}(L^{-3}), \tag{61}$$

so that $\ln\big(L(1+\mathrm{SNR})(\widetilde{r}-1)+1\big)$ can be represented using its Taylor expansion by

$$\begin{aligned} &\ln\big(L(1+\mathrm{SNR})(\widetilde{r}-1)+1\big) = \ln\tau(\mathrm{SNR}) \\ &\quad + \ln\big(1 + \frac{A}{(1+\tau(\mathrm{SNR})(1+\mathrm{SNR}))L} + \mathcal{O}(L^{-2})\big) \\ &= \ln\tau(\mathrm{SNR}) + \frac{A}{(1+\tau(\mathrm{SNR})(1+\mathrm{SNR}))L} + \mathcal{O}(L^{-2}) \end{aligned} \tag{62}$$

Now, substituting (61) and (62) into the respective left side and the right side of (60) produces

$$A = \frac{\big(1+\tau(\mathrm{SNR})(1+\mathrm{SNR})\big)\ln^2\big(1+\tau(\mathrm{SNR})(1+\mathrm{SNR})\big)}{\tau(\mathrm{SNR})(1+\mathrm{SNR})}.$$

This completes the proof of Lemma 5. □

I. Proof of Proposition 2

Using the Taylor expansion of function $\ln(1+t)$ and Lemma 5 gives us

$$u(\widetilde{r}) = \frac{\ln \widetilde{r}}{\widetilde{r}-1} = \frac{\ln\big(1+(\widetilde{r}-1)\big)}{\widetilde{r}-1} = 1 - \frac{\widetilde{r}-1}{2} + \frac{(\widetilde{r}-1)^2}{3} + \mathcal{O}(L^{-3}). \tag{63}$$

Then, using the Taylor expansion of function e^t and (63), we have

$$e^{1-u(\widetilde{r})} = 1 + (1-u(\widetilde{r})) + \frac{(1-u(\widetilde{r}))^2}{2} + \mathcal{O}(L^{-3}) = 1 + \frac{\widetilde{r}-1}{2} - \frac{5(\widetilde{r}-1)^2}{24} + \mathcal{O}(L^{-3}). \tag{64}$$

Now, combining (60) and (64) with the definition of $\widetilde{\rho}$ in Theorem 4 results in

$$\widetilde{\rho} = u(\widetilde{r})e^{1-u(\widetilde{r})} = 1 - \frac{(\widetilde{r}-1)^2}{8} + \mathcal{O}(L^{-3}) = 1 - \frac{\tau^2(\mathrm{SNR})}{8L^2} + \mathcal{O}(L^{-3}). \tag{65}$$

Therefore, we obtain

$$\widetilde{\mathcal{G}}(\mathrm{SNR},\mathrm{R}) = \widetilde{\rho}^{-1} = 1 + \frac{\tau^2(\mathrm{SNR})}{8L^2} + \mathcal{O}(L^{-3}).$$

This completes the proof of Proposition 2. □

J. Proof of Theorem 5

First of all, by Definition 2, we have $\beta = \lim_{N\to\infty} \frac{R}{\frac{1}{2}\ln N}$. Recall that $\mathrm{R} = \ln L$, then we have $L = N^{\frac{\beta}{2}} + o(N^{\frac{\beta}{2}})$. As $\beta > 0$ by definition, L increases without bound when N goes to infinity.

We now consider how the error performance changes when N (as well as L) goes to infinity. We first study the geometrical coding gain defined in (34). By (38) in Proposition 2, we notice that

$$\begin{aligned}
\lim_{N\to\infty} \widetilde{\rho}^{N^\beta} &\overset{(a)}{=} \lim_{N\to\infty} \widetilde{\mathcal{G}}(\mathrm{SNR},\mathrm{R})^{-N^\beta} \\
&\overset{(a)}{=} \lim_{N\to\infty} \Big(1 + \frac{\tau^2(\mathrm{SNR})}{8N^\beta}\Big)^{-N^\beta} \\
&\overset{(b)}{=} \lim_{N\to\infty} \Big(1 + \frac{\tau^2(\mathrm{SNR})}{8N^\beta}\Big)^{\frac{8N^\beta}{\tau^2(\mathrm{SNR})}\frac{-\tau^2(\mathrm{SNR})}{8}} \\
&\overset{(c)}{=} \exp\big(-\frac{\tau^2(\mathrm{SNR})}{8}\big).
\end{aligned}$$

where (a) is due to (34), while we used (38) and the fact that $L = N^{\frac{\beta}{2}} + o(N^{\frac{\beta}{2}})$ in (b), and (c) results from the definition of Euler's number.

Then, with the help of (37) in Lemma 5, we know $\lim_{N\to\infty}\widetilde{r} = 1$. In addition, from (63) we have $\lim_{N\to\infty} u(\widetilde{r}) = 1$. Furthermore, as $v(\widetilde{r}) = \widetilde{r}u(\widetilde{r})$, we obtain $\lim_{N\to\infty} v(\widetilde{r}) = 1$. Now, by (33) in Theorem 4, we have $\exp\big(-\frac{\tau^2(\mathrm{SNR})}{8}\big) \le \lim_{N\to\infty} \widetilde{\mathrm{P}}_e^{\frac{1}{N^{1-\beta}}} \le \exp\big(-\frac{\tau^2(\mathrm{SNR})}{8}\big)$. Therefore, by Definition 1, we conclude that the geometrical coding gain is $\exp\big(\frac{\tau^2(\mathrm{SNR})}{8}\big)$ and the diversity gain is $N^{1-\beta}$.

This completes the proof of Theorem 5. □

REFERENCES

[1] M. N. Sadiku, Y. Wang, S. Cui, and S. M. Musa, "Industrial internet of things," *Int. J. Adv. Sci. Res. Eng.*, vol. 3, 2017.

[2] S. H. Alsamhi, O. Ma, M. S. Ansari, and Q. Meng, "Greening internet of things for smart everythings with a green-environment life: A survey and future prospects," *eprint arXiv:1805.00844*, May 2018.

[3] C. Zhu, V. C. M. Leung, L. Shu, and E. C. . Ngai, "Green internet of things for smart world," *IEEE Access*, vol. 3, pp. 2151–2162, 2015.

[4] T. Sauter, S. Soucek, W. Kastner, and D. Dietrich, "The evolution of factory and building automation," *IEEE Ind. Electron. Mag.*, vol. 5, pp. 35–48, Sept. 2011.

[5] V. K. L. Huang, Z. Pang, C. J. A. Chen, and K. F. Tsang, "New trends in the practical deployment of industrial wireless: From noncritical to critical use cases," *IEEE Ind. Electron. Mag.*, vol. 12, pp. 50–58, June 2018.

[6] M. Luvisotto, Z. Pang, and D. Dzung, "Ultra high performance wireless control for critical applications: Challenges and directions," *IEEE Trans. Ind. Informat.*, vol. 13, pp. 1448–1459, June 2017.

[7] M. Luvisotto, Z. Pang, D. Dzung, M. Zhan, and X. Jiang, "Physical layer design of high-performance wireless transmission for critical control applications," *IEEE Trans. Ind. Informat.*, vol. 13, pp. 2844–2854, Dec. 2017.

[8] H. Chen, R. Abbas, P. Cheng, M. Shirvanimoghaddam, W. Hardjawana, W. Bao, Y. Li, and B. Vucetic, "Ultra-reliable low latency cellular networks: Use cases, challenges and approaches," *to appear in IEEE Commun. Mag.*, 2018.

[9] N. A. Johansson, Y. P. E. Wang, E. Eriksson, and M. Hessler, "Radio access for ultra-reliable and low-latency 5g communications," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, pp. 1184–1189, June 2015.

[10] N. Brahmi, O. N. C. Yilmaz, K. W. Helmersson, S. A. Ashraf, and J. Torsner, "Deployment strategies for ultra-reliable and low-latency communication in factory automation," in *2015 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, Dec. 2015.

[11] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[12] ITU-R RSG1SG5-IOT-16 Information Document 8, "Brief summary of the ITU-R study group 1 related studies (incl. RA-15 and WRC-15 related outcomes," Nov. 2016.

[13] ACMA, "The internet of things and the ACMA's areas of focus emerging issues in media and communications occasional paper," Nov. 2015.

[14] Ofcom, "Spectrum above 6 GHz for future mobile communications," Feb. 2015.

[15] H. Holma and A. Toskala, *LTE advanced: 3GPP solution for IMT advanced*. New York, NY, USA:Wiley, 1st ed., 2012.

[16] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 1073–1096, May 2003.

[17] A. Lozano and N. Jindal, "Transmit diversity vs. spatial multiplexing in modern MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 186–197, Jan. 2010.

[18] 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.

[19] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[20] R. Baldemair, E. Dahlman, G. Fodor, G. Mildh, S. Parkvall, Y. Selen, H. Tullberg, and K. Balachandran, "Evolving wireless communications: Addressing the challenges and expectations of the future," *IEEE Veh. Technol. Mag.*, vol. 8, no. 1, pp. 24–30, Mar. 2013.

[21] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[22] E. Larsson, F. Tufvesson, O. Edfors, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[23] T. K. Vu, C. F. Liu, M. Bennis, M. Debbah, M. Latva-aho, and C. S. Hong, "Ultra-reliable and low latency communication in mmwave-enabled massive MIMO networks," *IEEE Commun. Lett.*, vol. 21, pp. 2041–2044, Sept. 2017.

[24] P. Popovski, J. J. Nielsen, C. Stefanovic, E. d. Carvalho, E. Strom, K. F. Trillingsgaard, A. S. 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 Netw.*, vol. 32, pp. 16–23, Mar. 2018.

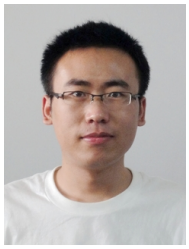
[25] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. of the IEEE*, vol. 104, pp. 1711–1726, Sept. 2016.
[26] M. Chowdhury, A. Manolakos, and A. Goldsmith, "Scaling laws for noncoherent energy-based communications in the SIMO MAC," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1980–1992, Apr. 2016.
[27] Y.-Y. Zhang, J.-K. Zhang, and H.-Y. Yu, "Physically securing energy-based massive MIMO MAC via joint alignment of multi-user constellations and artificial noise," *IEEE J. Sel. Areas Commun.*, vol. 36, Apr. 2018.
[28] A. Manolakos, M. Chowdhury, and A. Goldsmith, "Energy-based modulation for noncoherent massive SIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 7831–7846, Nov. 2016.
[29] L. Jing, E. De Carvalho, P. Popovski, and A. O. Martinez, "Design and performance analysis of noncoherent detection systems with massive receiver arrays," *IEEE Trans. Signal Processing*, vol. 64, pp. 5000–5010, Oct. 2016.
[30] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 47, pp. 1290–1301, May 2001.
[31] M. C. Gursoy, H. V. Poor, and S. Verdu, "The noncoherent Rician fading channel-part I: Structure of the capacity-achieving input," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2193–2206, Sept. 2005.
[32] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2426–2467, Oct 2003.
[33] R. J. Muirhead, *Aspects of multivariate statistical theory*. New York: John Wiley & Sons, INC, 1982.
[34] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge Univ. Press, 2004.
[35] M. Brehler and M. K. Varanasi, "Asymptotic error probability analysis of quadratic receiver in Rayleigh-fading channels with applications to a unified analysis of coherent and noncoherent space-time receivers," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2383–2399, Sept. 2001.
[36] R. W. Yeung, *Information Theory and Network Coding*. Springer, 2008.
[37] S. Jozsef and L. Debnath, "On certain inequalities involving the constant e and their applications," *J. Math. Anal. Appl.*, vol. 249, no. 2, pp. 569–582, 2000.
[38] N. Batir, "Sharp inequalities for factorial n," *Proyecciones*, vol. 27, no. 1, pp. 97–102, 2008.

Xiang-Chuan Gao received the B.Sc. and M.Eng. degrees from Zhengzhou University, Zhengzhou, China, in 2005 and 2008, respectively, and the Ph.D. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2011. His research interests include massive MIMO, cooperative communications, and visible light communication.

Jian-Kang Zhang (SM'09) received the B.S. degree in information science (math.) from Shaanxi Normal University, Xian, China, the M.S. degree in information and computational science (math.) from Northwest University, Xian, China, and the Ph.D. degree in electrical engineering from Xidian University, Xian, China, in 1983, 1988, and 1999, respectively. He is currently Associate Professor in the Department of Electrical and Computer Engineering at McMaster University, Hamilton, ON, Canada. He has hold research positions in McMaster University and Harvard University.

His research interests are in the general area of signal processing, digital communications, signal detection and estimation. Dr. Zhang is the coauthor of the paper that received the IEEE Signal Processing Society Best Young Author Award in 2008. He has served as Associate Editors for the IEEE SIGNAL PROCESSING LETTERS and IEEE TRANSACTIONS ON SIGNAL PROCESSING. He is currently serving as an Associate Editor for the Journal of Electrical and Computer Engineering.

He (Henry) Chen (S'10-M'16) received the Ph.D. degree in Electrical Engineering from the University of Sydney, Sydney, Australia, in 2015. He is currently a Research Fellow at the School of Electrical and Information Engineering, University of Sydney. His current research interests are in the field of industrial Internet of Things, with a particular focus on ultra-reliable low latency wireless, real-time industrial Ethernet, lightweight authentication and encryption, and anomaly detection. He received the Outstanding Bachelor Thesis of Shandong University, the Outstanding Master Thesis of Shandong Province, and the Chinese Government Award for Outstanding Self-Financed Students Abroad.



Zheng Dong (M'18) received his B.Sc. and M.Eng. degrees from the School of Information Science and Engineering, Shandong University, Jinan, China, in 2009 and 2012, respectively. He received the Ph.D. degree from the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada, in 2016. He is currently a Research Fellow at the School of Electrical and Information Engineering, The University of Sydney. His research interests include Industrial Internet of Things (IIoT), Ultra-Reliable Low-Latency Communications (URLLC), Non-Orthogonal Multiple Access (NOMA), and Information Theory and Coding. He received the Outstanding Bachelor Thesis of Shandong University, the Outstanding Master Thesis of Shandong Province.

Branka Vucetic (SM'00-F'03) is an ARC Laureate Fellow and Professor of Telecommunications, Director of the Centre of Excellence in Telecommunications at the University of Sydney. During her career she has held research and academic positions in Yugoslavia, Australia, UK and China. Her research interests include coding, communication theory and signal processing and their applications in wireless networks and industrial internet of things.

Prof Vucetic co-authored four books and more than four hundred papers in telecommunications journals and conference proceedings. She is a Fellow of the Australian Academy of Technological Sciences and Engineering and a Fellow of the IEEE.