Detection of Jamming Attack in Non-Coherent Massive SIMO Systems

Shengbo Xu, Weiyang Xu, *Member, IEEE*, Cunhua Pan, *Member, IEEE*, and Maged Elkashlan, *Member, IEEE*

***Abstract*— In recent studies, a simple non-coherent communication scheme based on energy detection is proposed in massive single-input multiple-output (SIMO) systems. Before data transmission, the transmitter sends pilots to the receiver for the purpose of estimating the channel statistics. However, this training phase unintentionally provides opportunity for a malicious jammer to attack legitimate communication. In order to secure the legitimate communication, this paper proposes a jamming detection method in non-coherent SIMO systems, in which the information of channel statistics is not required. First, the transmitter sends pilots to the receiver, then the receiver sends the conjugate of its received signal (which may contain jammer signal) back to the transmitter, where the final decision on jamming detection is made. According to the likelihood ratio test principle, two detectors based on variance and standard variance normalization are proposed. The performance analysis indicates that these two detectors are of similar detection performance but of different complexity. Furthermore, it is revealed that the probability of detection initially grows with the number of receive antennas but converges quickly then, whereas the channel statistics from the jammer to the receiver always greatly influences the performance. Finally, the numerical simulations are carried out to validate the proposed detection method.**

***Index Terms*— Non-coherent communications, massive single-input multiple-output (SIMO), energy detection, physical layer security, likelihood ratio test.**

I. Introduction

EMPLOYING a large number of antennas at base stations (BSs) while sharing the same time-frequency resources, which is known as massive multiple-input multiple-output (MIMO), has recently received a great deal of interest due to its huge potential gains [1], [2]. For example, massive MIMO is energy efficient as the transmit power scales down with the number of antennas at BS. Besides, channel vectors associated with different users are asymptotically orthogonal, thus both intra- and inter-cell interference can be eliminated with simple detection and precoding algorithms [3].

Meanwhile, physical layer security has become one of the research hotspots recently in wireless communications [4]. Rather than high level cryptographic methods, it is possible to secure physical layer transmission by applying the principles of information-theoretic security and signal processing techniques [5]. For example, it has been shown that the security in cognitive radio networks and non-orthogonal multiple access (NOMA) can be enhanced by employing game theory and power allocation [6], [7]. In this study, we focus on the security issue in systems with large antenna array. Specifically, passive and active attacks are the two major concerns in physical layer security. In particular, massive MIMO provides improvements of the physical layer security against passive attack, due to its capability to focus the transmission energy in the desired direction [8].

However, if the eavesdropper can actively attack the legitimate communication, the achievable secrecy capacity will be dramatically reduced. Hence, great efforts have been made to address the detection of active attack, and valuable algorithms are obtained. To improve the reliability of data transmission, Cumanan et al. [9] propose to tackle active attack via exploiting artificial noise. A jamming detection scheme based on random matrix theory is introduced in [10], where the final decision is made by analyzing the maximum eigenvalue of the sample covariance matrix of the received signal. Based on the generalized likelihood ratio test (GLRT), an algorithm of detection is designed for the uplink of massive MIMO systems, where unused orthogonal pilots are employed [11], [12]. Moreover, the unused pilots are employed to estimate the legitimate channel and jamming channel simultaneously, and then the estimate of jamming channel is used to construct linear receiver filters that reject the impact of the jamming signal [13]. With the intention of detecting the pilot spoofing attack, which is carried out during the channel training phase, a two-way training-based scheme is presented in [14]. Besides, [15] proposes a detector that leverages the asymmetry of received signal power levels at the transmitter and legitimate receiver when there exists a pilot spoofing attack. More recently, a pilot retransmission scheme is presented to detect the jammer by examining the pilot contamination in the uplink and downlink [16].

The above studies concentrate on coherent massive MIMO systems, where accurate channel state information (CSI)

Manuscript received September 10, 2018; revised January 3, 2019; accepted February 7, 2019. Date of publication February 14, 2019; date of current version June 5, 2019. This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 2018CDXYTX0011, and in part by the Key Program of Natural Science Foundation of Chongqing under Grant CSTC2017JCYJBX0047. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Walid Saad. *(Corresponding author: Weiyang Xu.)*

S. Xu and W. Xu are with the School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China (e-mail: shengboxu@cqu.edu.cn; weiyangxu@cqu.edu.cn).

C. Pan and M. Elkashlan are with the School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K. (e-mail: c.pan@qmul.ac.uk; maged.elkashlan@qmul.ac.uk).

Digital Object Identifier 10.1109/TIFS.2019.2899484

associated with all users is required at BS. However, a massive antenna array would make acquiring CSI in a timely manner much more challenging than before. Furthermore, pilot contamination, attributed to reusing pilots among adjacent cells, makes the problem even worse. Meanwhile, receivers that use simple, robust and energy efficient designs are thus attractive to realizing the benefits of large antenna systems, especially when it comes to applications of mmWave carrier frequency [17]. In recent studies, a simple non-coherent transmission scheme based on energy detection (ED) is proposed in massive single-input multiple-output (SIMO) systems [18]. Utilizing non-negative pulse amplitude modulation (PAM), the transmit symbols can be decoded through averaging the received signal power across all antennas. Moreover, given that the number of receive antennas is asymptotically infinite, the ED-based non-coherent system can provide the same error performance as that of the coherent counterparts. Inspired by the pioneering work in [18], non-coherent massive SIMO systems have drawn a lot of attention from the research community [19]–[21].

Existing studies clearly demonstrate that in coherent massive MIMO systems, the active attack could reduce the secrecy capacity remarkably, and seriously degrade the reliability of legitimate communications. However, how the active attack influences the non-coherent SIMO systems and the way to detect the jamming sources efficiently is still unclear. For this purpose, this paper presents a jamming detection algorithm for ED-based non-coherent massive SIMO systems. To the best of the authors' knowledge, this is the first study concerning physical layer security in non-coherent massive SIMO systems. The main contributions of this paper are summarized as follows.

- It has been shown that the jamming attack in the training phase can deteriorate the error performance notably. Even worse, this performance loss cannot be compensated for by increasing transmit power or deploying more receive antennas.
- We propose to let the receiver retransmit the conjugate of its received pilots back to the transmitter, where the final decision on jamming detection is made. This can be regarded as some extent of cooperation. Via preprocessing the received signal at the transmitter, two kinds of decision metrics are designed according to the likelihood ratio test (LRT).
- We derive the closed-form expression of probability density function (PDF) of decision metrics. Different from existing detection algorithms where the channel statistics are assumed as *a prior*, the proposed algorithm estimates the channel statistics, thus making it more applicable in real situations. Performance analysis regarding probabilities of false alarm (P_{FA}) and detection (P_D) is carried out. Results indicate P_D initially grows with the number of receive antennas but converges quickly then, while the channel statistics from the jammer to the receiver always influences the performance greatly.

The remainder of this paper is organized as follows. The considered system model is illustrated in Section II. The design of decision metrics based on LRT are detailed in Section III. Section IV presents the proposed jamming detection algorithm. Numerical simulations are conducted to validate our detection scheme in Section V. Finally, concluding remarks are drawn in Section VI.

Notation: $\mathbb{C}^{n\times m}$ indicates a complex matrix of size $n \times m$. Bold variables represent matrices and vectors. For a random variable x, $x \sim \mathcal{CN}(\mu, \sigma^2)$ and $x \sim \mathcal{N}(\mu, \sigma^2)$ indicate complex and real Gaussian distributions with mean μ and variance σ^2, respectively. $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^*$ and $\|\cdot\|_2^2$ denote the transpose, conjugate transpose, complex conjugate and $\mathcal{L}_2$ norm operators. $\Re\{\cdot\}$ and $\Im\{\cdot\}$ refer to the real and imaginary parts of complex numbers. erf$(\cdot)$ and erfc$(\cdot)$ separately represents the error and complementary error functions. Finally, $\mathbb{E}[\cdot]$, var$[\cdot]$ and cov$[\cdot]$ indicate the expectation, variance, and covariance operators, respectively.

II. System Model and Problem Formulation

First, the ED-based non-coherent massive SIMO system is briefly reviewed. Afterwards, the influence of jamming attack in the training phase is analyzed, of which result clearly shows the necessity of jamming detection.

A. ED-Based Non-Coherent Massive SIMO Systems

Consider a SIMO network consisting of a single-antenna transmitter (`Alice`) and a receiver (`Bob`) with M ($M \gg 1$) antennas. Hence, the $M \times 1$ received signal vector by the massive antennas array is

$$\mathbf{y} = \sqrt{\mathcal{P}_u}\mathbf{h}w + \mathbf{n} \tag{1}$$

where $\mathcal{P}_u$ is the transmit power and assumed to be unity, w indicates data symbol drawn from a non-negative constellation $\left\{\sqrt{p_1}, \sqrt{p_2}, \ldots, \sqrt{p_N}\right\}$ of size N, $\mathbf{h} = [h_1, h_2, \ldots, h_M]^T$ denotes the channel vector with the m-th item $h_m \sim \mathcal{CN}(0, \sigma_h^2)$ and $\mathbf{n} = [n_1, n_2, \ldots, n_M]^T$ is the noise vector with its m-th component $n_m \sim \mathcal{CN}(0, \sigma_n^2)$. In this study, channel and noise vectors are mutually independent. In addition, the signal-to-noise ratio (SNR) at `Bob` is defined as $\mathrm{SNR}_{\mathtt{Bob}} = \mathbb{E}[w^2]\sigma_h^2/\sigma_n^2$. If $M \to \infty$, after the received signal having been filtered, squared and integrated, the average power across all antennas can be represented by [18]

$$\begin{aligned}\Omega &= \frac{1}{M}\mathbf{y}^H\mathbf{y} \\ &= \frac{1}{M}\mathbf{h}^H\mathbf{h}w^2 + \frac{2}{M}\Re\left(\mathbf{h}^H\mathbf{n}\right)w + \frac{1}{M}\mathbf{n}^H\mathbf{n} \\ &\overset{M\to\infty}{=} w^2\sigma_h^2 + \sigma_n^2.\end{aligned} \tag{2}$$

Given the knowledge of σ_h^2 and σ_n^2, the estimate of w can be readily obtained according to (2).

Since M is always finite, Ω can be approximated to one of the N Gaussian variables depending on *a prior* information of transmit symbols. For example, with a non-negative PAM of $N = 4$, the PDF of Ω over an additive white Gaussian noise (AWGN) channel is shown in Fig. 1, where $M = 100$ and $\mathrm{SNR}_{\mathtt{Bob}} = 4$ dB. Four distinct Gaussian-like curves are observed, corresponding to four constellation points. Accordingly, the positive line is partitioned into multiple decoding

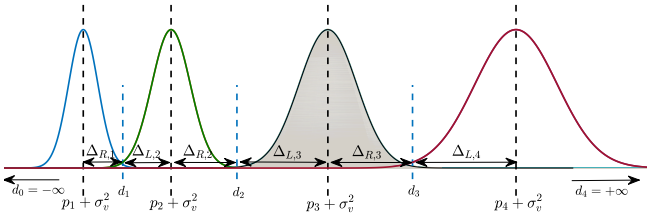


Fig. 1. Decoding regions of a non-negative PAM of size $N = 4$, where $M = 100$ and $\mathrm{SNR}_{\mathtt{Bob}} = 4$ dB.

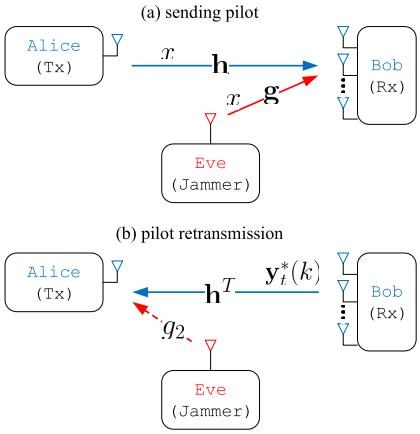


Fig. 2. The proposed jamming detection algorithm. (a) Alice sends pilots to Bob in the presence of Eve. (b) Bob sends the conjugate of its received signal to Alice, and Eve could continuously send jamming signal.

regions $\{d_n\}_{n=0}^N$ to decide which symbol was transmitted based on the observation of $\mathbf{\Omega}$, i.e.

$$\widehat{w} = \sqrt{p_n}, \quad \text{if } d_{n-1} \le \mathbf{\Omega} < d_n. \tag{3}$$

Concretely, d_0 is $-\infty$ for $\sqrt{p_1}$, d_N is $+\infty$ for $\sqrt{p_N}$ and [19]

$$d_n = \frac{(p_n + p_{n+1})\,\sigma_h^2}{2} + \sigma_n^2, \quad \text{for } 1 \le n \le N-1. \tag{4}$$

This scheme relies on the information of σ_h^2 and σ_n^2, which can be estimated by sending a sequence of pilots before data transmission. However, it provides opportunity for a malicious jammer to attack the legitimate transmission.

B. The Influence of Jamming Attack in the Training Phase

As shown in Fig. 2 (a), a jammer (Eve) transmits the same pilots to the receiver in the training phase if pilots are publicly known [15]. Note that this signifies the worst case of jamming detection. Hence, depending on whether the jammer is present or not, the received signal at Bob is denoted by

$$\begin{aligned} \mathcal{H}_0: &\quad \mathbf{y} = \sqrt{\mathcal{P}_u}\mathbf{h}x + \mathbf{n} \\ \mathcal{H}_1: &\quad \mathbf{y} = \sqrt{\mathcal{P}_u}\mathbf{h}x + \sqrt{\mathcal{P}_j}\mathbf{g}x + \mathbf{n} \end{aligned} \tag{5}$$

where $\mathcal{H}_0$ and $\mathcal{H}_1$ separately indicate hypotheses of absence and presence of the jammer, x denotes the pilot symbol, $\mathcal{P}_j$ is the transmit power of Eve and also assumed to be unity. $\mathbf{g} = [g_1, g_2, \ldots, g_M]^T$, which is independent of $\mathbf{h}$, models the channel vector from Eve to Bob with its m-th element being $g_m \sim \mathcal{CN}(0, \sigma_g^2)$.

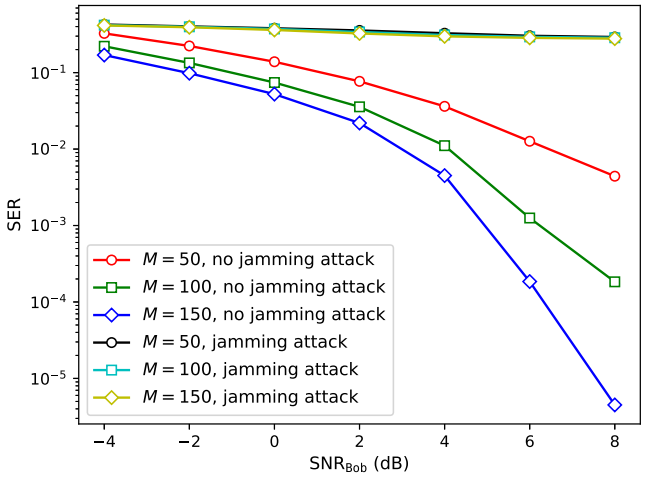


Fig. 3. The SER comparison under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$. It is demonstrated that the error performance is severely deteriorated.

According to (2), in the limit of $M \to \infty$, the channel statistics from Alice to Bob is estimated by

$$\begin{aligned} \mathcal{H}_0: &\quad \widehat{\sigma}_h^2 = \left(\|\mathbf{y}\|_2^2/M - \sigma_n^2\right)/x^2 = \sigma_h^2, \\ \mathcal{H}_1: &\quad \widehat{\sigma}_h^2 = \left(\|\mathbf{y}\|_2^2/M - \sigma_n^2\right)/x^2 = \sigma_h^2 + \sigma_g^2. \end{aligned} \tag{6}$$

After the training phase, data symbols are transmitted and then decoded by using $\widehat{\sigma}_h^2$. In the presence of jamming attack, the estimate of channel statistics is $\sigma_h^2 + \sigma_g^2$ instead of σ_h^2. This could modify the decoding regions and make the final decision prone to errors. As an example, Fig. 3 compares the symbol-error rate (SER) in two hypotheses. Clearly enough, the SER is severely deteriorated and cannot be lowered by increasing the transmit power or the number of receive antennas.

III. Decision Metrics for the Proposed Jamming Detection Algorithm

Different from existing literatures, our study doesn't assume a prior knowledge of σ_h^2 and σ_g^2, thus making the jamming detection realistic but also more challenging. Towards this end, we propose a new detection scheme that Bob retransmits the conjugate of its received signal back to Alice (shown in Fig. 2 (b)), where the final decision of whether Eve is present or not is made.

A. Signal Retransmission

The exact estimate of channel statistics based on (2) requires $M \to \infty$. However, the estimation accuracy by this method in real scenarios is not sufficiently high since M is always finite. Accordingly, Alice sends pilots in K channel coherence intervals and then Bob carries out estimation through averaging over both space (M antennas) and time (K intervals). In our proposed scheme, each channel coherence interval consists of training, retransmission and data phases. Specifically, during the k-th coherence interval, Alice sends τ consecutive pilot symbols to Bob. Let $\mathbf{y}_t(k) \in \mathbb{C}^{M\times 1}$ denote the received signal by Bob at time index t during the k-th interval, it then

comes to the following observations

$$\begin{aligned}\mathcal{H}_0:&\quad \mathbf{y}_t(k)=\mathbf{h}(k)x_t(k)+\mathbf{n}_t(k)\\ \mathcal{H}_1:&\quad \mathbf{y}_t(k)=\mathbf{h}(k)x_t(k)+\mathbf{g}(k)x_t(k)+\mathbf{n}_t(k)\\ &\qquad k=1,2,\ldots,K\end{aligned}\tag{7}$$

where $t\in[1,\tau]$, $\mathbf{h}(k)$ and $\mathbf{g}(k)$ remain unchanged during the k-th interval and vary independently among intervals, $x_t(k)$ denotes pilot symbol sent at time t during the k-th interval and $\mathbf{n}_t(k)\in\mathbb{C}^{M\times 1}$ is the AWGN. Afterwards, Bob retransmits the conjugate of $\mathbf{y}_t(k)$ back to Alice in the retransmission phase. Suppose the received signal by Alice at time $t+\tau$ is $z_{t+\tau}(k)$, then one can obtain the average $z(k)=\sum_{t=1}^{\tau}z_{t+\tau}(k)/\tau$. Let $Q(k)$ indicate the real part of $z(k)$, then we have

$$\begin{aligned}\mathcal{H}_0:&\quad Q(k)=\mathbf{h}^T(k)\mathbf{h}^*(k)\bar{x}(k)+\Re\left\{\mathbf{h}^T(k)\bar{\mathbf{n}}^*(k)\right\}\\ &\qquad +\Re\left\{\bar{v}(k)\right\}\\ \mathcal{H}_1:&\quad Q(k)=\mathbf{h}^T(k)\mathbf{h}^*(k)\bar{x}(k)+\Re\left\{\mathbf{h}^T(k)\mathbf{g}^*(k)\right\}\bar{x}(k)\\ &\qquad +\Re\left\{\mathbf{h}^T(k)\bar{\mathbf{n}}^*(k)\right\}+\Re\left\{\bar{v}(k)\right\}\end{aligned}\tag{8}$$

where $\bar{x}(k)=\sum_{t=1}^{\tau}x_t(k)/\tau$ denotes the average of pilots sent in the k-th channel coherence interval, thus $\bar{x}(k)=1$ since all pilots are assumed to be unity, while $\bar{\mathbf{n}}^*(k)=\sum_{t=1}^{\tau}\mathbf{n}_t^*(k)/\tau$ follows the distribution of $\mathcal{CN}(\mathbf{0},\sigma_n^2\mathbf{I}_M/\tau)$. $\bar{v}(k)=\sum_{t=1}^{\tau}v_{t+\tau}(k)/\tau$, and $v_{t+\tau}(k)\sim\mathcal{CN}(0,\sigma_v^2)$ denotes the AWGN generated by Alice at time $t+\tau$ during the k-th coherence interval. Without loss of generality, the noise power at both Alice and Bob is the same, namely $\sigma_v^2=\sigma_n^2$, hence it comes to $\bar{v}(k)\sim\mathcal{CN}(0,\sigma_n^2/\tau)$.

According to the central limit theorem (CLT), $Q(k)$ can be approximated as a real Gaussian variable if M is sufficiently large, the proof can be found in Appendix A. Therefore, in the absence of Eve, it is shown that $Q(k)\sim\mathcal{N}(\mu_0,\sigma_0^2)$ with

$$\begin{aligned}\mu_0&=M\sigma_h^2,\\ \sigma_0^2&=M\sigma_h^4+M\sigma_h^2\frac{\sigma_n^2}{2\tau}+\frac{\sigma_n^2}{2\tau}.\end{aligned}\tag{9}$$

In the presence of Eve, it is derived that $Q(k)\sim\mathcal{N}(\mu_1,\sigma_1^2)$ with

$$\begin{aligned}\mu_1&=M\sigma_h^2,\\ \sigma_1^2&=M\sigma_h^4+M\sigma_h^2\frac{\sigma_g^2}{2}+M\sigma_h^2\frac{\sigma_n^2}{2\tau}+\frac{\sigma_n^2}{2\tau}.\end{aligned}\tag{10}$$

The derivation is included in Appendix B. From (9) and (10), it is observed that $\mu_0=\mu_1$, which reveals the mean of $Q(k)$ is identical under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$. Fortunately, since $\mu_0=\mu_1=M\sigma_h^2$, the channel statistics from Alice to Bob can be estimated no matter Eve is present or not. On the other hand, the variance of $Q(k)$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ are different from each other. In the following analysis, it will be shown how to take advantage of these two features to design robust jammer detection scheme.

Before that, *it is worth noting that there is a possibility that* Eve *sends pilots continuously in the retransmission phase*, as shown in Fig. 2 (b). Thus observations by Alice under $\mathcal{H}_1$ change to

$$\begin{aligned}Q'(k)&=\mathbf{h}^T(k)\mathbf{h}^*(k)\bar{x}(k)+\Re\left\{\mathbf{h}^T(k)\mathbf{g}^*(k)\right\}\bar{x}(k)\\ &\quad+\Re\left\{\mathbf{h}^T(k)\bar{\mathbf{n}}^*(k)\right\}+\Re\left\{\bar{v}(k)\right\}+\Re\left\{g_2(k)\bar{x}(k)\right\}\\ &=Q(k)+\Re\left\{g_2(k)\bar{x}(k)\right\}\end{aligned}\tag{11}$$

where $g_2(k)\sim\mathcal{CN}(0,\sigma_{g_2}^2)$ models the channel from Eve to Alice. Given $Q(k)$ approximates to a Gaussian random variable, it can be derived that $Q'(k)\sim\mathcal{N}(\mu_1,\sigma_1^2+\sigma_{g_2}^2/(2\tau))$. As $\sigma_{g_2}^2/(2\tau)\ll\sigma_1^2$ when M is a large number, the distribution of $Q'(k)$ is very close to that of $Q(k)$. Hence, we ignore the situation in (11), since jamming signal in the retransmission phase has little effect on the detection performance.

B. The Design of Decision Metric

In our scheme, Alice sends pilots to Bob in each training phase, while Bob retransmits the conjugate of its received signal back to Alice. After K channel coherence intervals, one can obtain K observations, namely $\{Q(1),Q(2),\ldots,Q(K)\}$.

It is not straightforward to come up with decision metric from $Q(k)$, thus we consider a general situation first. Suppose there are K observations $\mathbf{S}=\{S_1,S_2,\ldots,S_K\}$, where S_k are independent and identically distributed (i.i.d.). Moreover, the distributions of S_k under hypotheses $\mathcal{H}_A$ and $\mathcal{H}_B$ are different, i.e.,

$$\begin{aligned}\mathcal{H}_A:&\quad S_k\sim\mathcal{N}\left(\mu_A,\sigma_A^2\right),\quad k=1,2,\ldots,K\\ \mathcal{H}_B:&\quad S_k\sim\mathcal{N}\left(\mu_B,\sigma_B^2\right),\quad k=1,2,\ldots,K\end{aligned}\tag{12}$$

According to the Neyman-Pearson theorem, the LRT principle is exploited to decide which hypothesis is true. Concretely, the LRT can be provided as follows [22]

$$\begin{aligned}L(\mathbf{S})&=\frac{p(\mathbf{S};\mathcal{H}_B)}{p(\mathbf{S};\mathcal{H}_A)}\underset{\mathcal{H}_B}{\overset{\mathcal{H}_A}{\lessgtr}}\Upsilon\\ &=\frac{\frac{1}{\left(2\pi\sigma_B^2\right)^{\frac{K}{2}}}\exp\left(-\frac{1}{2\sigma_B^2}\sum_{k=1}^{K}(S_k-\mu_B)^2\right)}{\frac{1}{\left(2\pi\sigma_A^2\right)^{\frac{K}{2}}}\exp\left(-\frac{1}{2\sigma_A^2}\sum_{k=1}^{K}(S_k-\mu_A)^2\right)}\underset{\mathcal{H}_B}{\overset{\mathcal{H}_A}{\lessgtr}}\Upsilon\end{aligned}\tag{13}$$

where $p(\mathbf{S};\mathcal{H})$ is the joint distribution of $\mathbf{S}$ under hypothesis $\mathcal{H}$, and Υ is the threshold. When the logarithmic operation is applied on both sides of (13), it comes to

$$\begin{aligned}&\frac{\sigma_B^2-\sigma_A^2}{2\sigma_A^2\sigma_B^2}\sum_{k=1}^{K}S_k^2+\frac{\mu_B\sigma_A^2-\mu_A\sigma_B^2}{\sigma_A^2\sigma_B^2}\sum_{k=1}^{K}S_k\underset{\mathcal{H}_B}{\overset{\mathcal{H}_A}{\lessgtr}}\\ &\qquad\times\ln\left(\left(\frac{\sigma_B}{\sigma_A}\right)^K\Upsilon\right)-\frac{K\left(\mu_A^2\sigma_B^2-\mu_B^2\sigma_A^2\right)}{2\sigma_A^2\sigma_B^2}.\end{aligned}\tag{14}$$

To construct a simple decision metric, one can eliminate the first or second component on the left side of (14). Then let's

return to the original problem, in which we have the following hypotheses

$$\begin{aligned}\mathcal{H}_0:&\quad Q(k)\sim\mathcal{N}\left(\mu_0,\sigma_0^2\right),\quad k=1,2,\ldots,K\\ \mathcal{H}_1:&\quad Q(k)\sim\mathcal{N}\left(\mu_1,\sigma_1^2\right),\quad k=1,2,\ldots,K\end{aligned}\tag{15}$$

where $\mu_0=\mu_1$. Still, the decision metric cannot be obtained by directly substituting (15) into (14). Therefore, preprocessing of K observations is required.

1) Standard Variance Normalization: The first method of preprocessing normalizes $Q(k)$ using its standard variance, namely $Q_{\mathsf{std}}(k)=Q(k)/\sqrt{\mathrm{var}[Q(k)]}$. Consequently, the following relationship is obtained

$$\begin{aligned}\mathcal{H}_0:&\quad Q_{\mathsf{std}}(k)\sim\mathcal{N}\left(\frac{\mu_0}{\sigma_0},1\right),\quad k=1,2,\ldots,K\\ \mathcal{H}_1:&\quad Q_{\mathsf{std}}(k)\sim\mathcal{N}\left(\frac{\mu_1}{\sigma_1},1\right),\quad k=1,2,\ldots,K.\end{aligned}\tag{16}$$

And then substitute (16) into (14) and keep in mind that $\mu_0=\mu_1$, one can get the relation

$$\left(\frac{\mu_0}{\sigma_1}-\frac{\mu_0}{\sigma_0}\right)\sum_{k=1}^{K}Q_{\mathsf{std}}(k)\underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}}\ln\Upsilon-\frac{K}{2}\left(\frac{\mu_0^2}{\sigma_0^2}-\frac{\mu_0^2}{\sigma_1^2}\right).\tag{17}$$

With some mathematical manipulations, the final decision of jamming detection is made by

$$\sum_{k=1}^{K}Q_{\mathsf{std}}(k)\underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\lessgtr}}\frac{\sigma_0\sigma_1}{\mu_0\left(\sigma_0-\sigma_1\right)}\left(\ln\Upsilon-\frac{K}{2}\left(\frac{\mu_0^2}{\sigma_0^2}-\frac{\mu_0^2}{\sigma_1^2}\right)\right)\tag{18}$$

where the decision metric is $\sum_{k=1}^{K}Q_{\mathsf{std}}(k)$. It is observed that the first kind of preprocessing eliminates the first component on the left side of (14).

2) Variance Normalization: The second decision metric is similar to the first one, but this time the second component on the left side of (14) is canceled. With operation $Q_{\mathsf{var}}(K)=Q(k)/\mathrm{var}[Q(k)]$, the following distributions are obtained

$$\begin{aligned}\mathcal{H}_0:&\quad Q_{\mathsf{var}}(k)\sim\mathcal{N}\left(\frac{\mu_0}{\sigma_0^2},\frac{1}{\sigma_0^2}\right),\quad k=1,2,\ldots,K\\ \mathcal{H}_1:&\quad Q_{\mathsf{var}}(k)\sim\mathcal{N}\left(\frac{\mu_1}{\sigma_1^2},\frac{1}{\sigma_1^2}\right),\quad k=1,2,\ldots,K.\end{aligned}\tag{19}$$

As before, one can get the following results by substituting (19) into (14)

$$\frac{\sigma_0^2-\sigma_1^2}{2}\sum_{k=1}^{K}Q_{\mathsf{var}}^2(k)\underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}}\ln\left(\left(\frac{\sigma_0}{\sigma_1}\right)^K\Upsilon\right)-\frac{K\mu_0^2}{2}\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)\tag{20}$$

which is equivalent to

$$\sum_{k=1}^{K}Q_{\mathsf{var}}^2(k)\underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\lessgtr}}\frac{2}{\sigma_0^2-\sigma_1^2}\times\left\{\ln\left(\left(\frac{\sigma_0}{\sigma_1}\right)^K\Upsilon\right)-\frac{K\mu_0^2}{2}\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)\right\}\tag{21}$$

where the decision metric is $\sum_{k=1}^{K}Q_{\mathsf{var}}^2(k)$.

C. Decision Metrics in Real Situation

The proposed decision metrics are derived based on *a prior* knowledge of variance of $Q(k)$, which depends on σ_h^2 and σ_g^2. However, the accurate variance is unavailable for two reasons. First, it is unreasonable to assume in this situation that σ_h^2 and σ_g^2 are known as *a prior*; Second, infinite observations of $Q(k)$ are required to obtain the accurate estimation. Therefore, the sample variance, which is calculated based on K observations of $Q(k)$, is chosen instead of accurate variance. Let $\widehat{\sigma}$ and $\widehat{\mu}$ indicate the sample variance and sample mean of $Q(k)$, the first decision metric changes to

$$\phi_{\mathsf{std}}=\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\frac{Q(k)}{\widehat{\sigma}}\tag{22}$$

and the second decision metric

$$\phi_{\mathsf{var}}=\frac{1}{K}\sum_{k=1}^{K}\frac{Q^2(k)}{\widehat{\sigma}^2}\tag{23}$$

where $\widehat{\sigma}$ and $\widehat{\mu}$ are obtained by

$$\begin{aligned}\widehat{\mu}&=\frac{1}{K}\sum_{k=1}^{K}Q(k),\\ \widehat{\sigma}&=\sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left(Q(k)-\widehat{\mu}\right)^2}.\end{aligned}\tag{24}$$

The $1/\sqrt{K}$ and $1/K$ in (22) and (23) facilitate the derivation of PDF of decision metrics, as will be shown in Section IV.

The proposed method could raise concern that the time it takes to collect enough observations can be too long. Non-coherent systems require the channel statistics σ_h^2 rather than the instantaneous CSI. Generally, the duration of a large-scale fading coefficient lasts for multiple seconds or minutes, while a typical value for channel coherence interval is 2.5 milliseconds [23]. As a result, the accuracy of estimation of σ_h^2 can be improved by averaging the training results of multiple coherence intervals over which σ_h^2 is constant. This finding enlightens us to store received signals of different intervals and decode them using the averaged estimate. Hence in non-coherent systems, there is enough time to collect the required number of observations.

Besides, the impact of sporadic jammer needs to be paid attention due to the possible long time of observations [24]. In the considered systems, the estimation of σ_h^2 is carried out in each coherence interval. In the absence of sporadic jammer, the training result in each interval is approximately the same, namely about σ_h^2. On the other hand, the presence of sporadic jammer can result in abnormal training outcomes of about $\sigma_h^2+\sigma_g^2$. Since these outcomes are larger than σ_h^2, they can be detected by setting a threshold. After that, these results are discarded and the estimation of σ_h^2 will not be impacted by sporadic jammer.

IV. Jamming Detection Algorithm in Non-Coherent Massive SIMO Systems

A. Preliminaries

Lemma 1: Suppose $x_1, x_2, \ldots, x_i, \ldots, x_n$ are n independent samples from a Gaussian distribution with variance σ^2, the sample mean is $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Accordingly, the sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ follows a χ^2 distribution with $(n-1)$ degrees of freedom according to the Cochran's Theorem [25]

$$(n-1)\,\frac{s^2}{\sigma^2} \sim \chi^2_{n-1}. \tag{25}$$

Lemma 2: Suppose Y is a normally distributed random variable with zero mean and unit variance, V is a χ^2 random variable with υ degrees of freedom that is independent of Y, then

$$T = \frac{Y + \delta}{\sqrt{V/\upsilon}}$$

follows a non-central t distribution with υ degrees of freedom and non-centrality parameter δ, i.e., $T \sim t(\upsilon, \delta)$.

Lemma 3: Suppose $x_1, x_2, \ldots, x_i, \ldots, x_k$ are k independent, normally distributed random variables with means μ_i and unit variance, then the random variable

$$Z = \sum_{i=1}^{k} x_i^2$$

follows the non-central χ^2 distribution with k degrees of freedom and non-centrality parameter $\lambda = \sum_{i=1}^{k} \mu_i^2$, namely $Z \sim \chi^2_k(\lambda)$.

Lemma 4: Suppose Q is a non-central χ^2 random variable with q degrees of freedom and non-centrality parameter $\varsigma \neq 0$, P is χ^2-distributed with p degrees of freedom that is statistically independent of Q, then

$$F = \frac{Q/q}{P/p}$$

is a non-central f random variable with p and q degrees of freedom respectively, and non-centrality parameter ς, namely, $F \sim f(q, p, \varsigma)$.

With the above lemmas, the analysis of the proposed jamming detection algorithm can be carried out.

B. Distributions of Decision Metrics

1) Sample Standard Variance Normalization: The decision metric in (22) can be rewritten as

$$\phi_{\mathsf{std}} = \frac{\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\frac{Q(k)}{\sigma}}{\sqrt{\frac{(K-1)\,\widehat{\sigma}^2}{(K-1)\,\sigma^2}}} \tag{26}$$

where σ^2 and $\widehat{\sigma}^2$ stand for the variance and sample variance of $Q(k)$, respectively. According to our analysis, $Q(k)$ approximates to a Gaussian variable with mean μ and variance σ^2. Hence, the distribution of numerator in (26) is denoted by

$$\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\frac{Q(k)}{\sigma} \sim \sqrt{K}\frac{\mu}{\sigma} + \mathcal{N}(0,1)\,. \tag{27}$$

Since $\widehat{\sigma}^2$ is the sample variance of $Q(k)$, thus according to Lemma 1, we have

$$(K-1)\,\frac{\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{K-1}. \tag{28}$$

Then based on the results in (27), (28) and Lemma 2, the PDF of decision metric ϕ_{std} under two hypotheses are

$$\begin{aligned}\mathcal{H}_0\colon\quad & \phi_{\mathsf{std}} \sim t\left(K-1, \sqrt{K}\frac{\mu_0}{\sigma_0}\right)\\ \mathcal{H}_1\colon\quad & \phi_{\mathsf{std}} \sim t\left(K-1, \sqrt{K}\frac{\mu_1}{\sigma_1}\right)\end{aligned} \tag{29}$$

where μ_0, σ_0, μ_1 and σ_1 can be found in (9) and (10). It is worth noting that (29) relies on the independence between the numerator and denominator in (26). First, the sample mean and sample variance of a Gaussian random variable are mutually independent according to Basu's theorem. Second, the numerator and denominator in (26) are functions of sample mean and sample variance of an approximated Gaussian variable $Q(k)$, respectively. Third, functions of independent random variables are independent. Hence, the independence is assured. Besides, K observations of $Q(k)$ can be divided into two sets, one is used to calculate the numerator and the other the denominator. As data in two sets are strictly independent, thus the required independence in (29) is well founded.

2) Sample Variance Normalization: Similarly, the decision metric in (23) can be rewritten as

$$\phi_{\mathsf{var}} = \frac{\frac{1}{K}\sum_{k=1}^{K}\frac{Q^2(k)}{\sigma^2}}{\frac{(K-1)\,\widehat{\sigma}^2}{(K-1)\,\sigma^2}}. \tag{30}$$

Because $Q(k)$ approximates to a Gaussian variable with mean μ and variance σ^2, then according to Lemma 3, one can have

$$\sum_{k=1}^{K}\frac{Q^2(k)}{\sigma^2} \sim \chi^2_K\left(K\frac{\mu^2}{\sigma^2}\right). \tag{31}$$

In the light of (28), (31) and Lemma 4, the PDF of decision metric ϕ_{var} under two hypotheses are

$$\begin{aligned}\mathcal{H}_0\colon\quad & \phi_{\mathsf{var}} \sim f\left(K, K-1, K\frac{\mu_0^2}{\sigma_0^2}\right)\\ \mathcal{H}_1\colon\quad & \phi_{\mathsf{var}} \sim f\left(K, K-1, K\frac{\mu_1^2}{\sigma_1^2}\right).\end{aligned} \tag{32}$$

As before, the result in (32) builds upon the independence between the numerator and denominator in (30).

To validate the above results, Fig. 4 compares the PDFs in (29) and numerical histogram of ϕ_{std} using (22) and (24) under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, where $M = 100$, $K = 200$ in Fig. 4 (a) and $K = 400$ in Fig. 4 (b). It can be observed that analytical results match numerical ones quite well. More

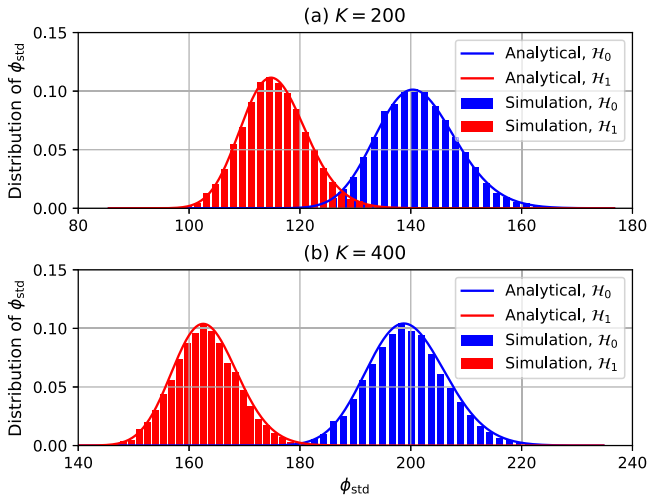


Fig. 4. Histogram and PDF of ϕ_{std} under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$.

importantly, this figure indicates ϕ_{std} under $\mathcal{H}_0$ is on the right-hand side of ϕ_{std} under $\mathcal{H}_1$. In addition, it is displayed that increasing K could reduce the overlap region, thus helps to improve the probability of detection.

C. Probabilities of False Alarm and of Detection

This section will discuss the approach to compute theoretical P_{FA} and P_D. To this end, analytical results in (29) and (32) are employed.

1) Sample Standard Variance Normalization: When ϕ_{std} is chosen as the decision metric, the probability of false alarm actually indicates the CDF of ϕ_{std} under hypothesis $\mathcal{H}_0$, i.e.,

$$P_{FA} = P\left\{\phi_{\text{std}} \leq \eta_{\text{std}}; \mathcal{H}_0\right\} \tag{33}$$

where η_{std} is the decision threshold. In general, η_{std} is computed by setting a predefined P_{FA}

$$\eta_{\text{std}} = P^{-1}_{\upsilon=K-1,\delta=\sqrt{K}\frac{\mu_0}{\sigma_0}}\left(P_{FA}\right) \tag{34}$$

where $P^{-1}_{\upsilon,\delta}(y)$ represents the inverse function of CDF of a non-central t random variable with υ degrees of freedom and non-centrality parameter δ. According to (9) and (10), the expectation of $Q(k)$ is identical under both $\mathcal{H}_0$ and $\mathcal{H}_1$. Besides, the sample mean $\widehat{\mu}_0$ is a good approximation of μ_0 when the number of observations K is large. Meanwhile, σ_0^2 is a function of μ_0, thus the PDF of ϕ_{std} under $\mathcal{H}_0$ is always available. After that, the threshold η_{std} is computed by (34), and then one can make the final decision through comparing the simulated decision metric with η_{std}.

On the other hand, analytical P_D can be considered as the CDF of ϕ_{std} under $\mathcal{H}_1$ conditioned on η_{std}, i.e.,

$$P_D = P\left\{\phi_{\text{std}} < \eta_{\text{std}}; \mathcal{H}_1\right\}. \tag{35}$$

Although a closed-form expression of P_D doesn't exist (or at least very difficult to obtain), the numerical results are attainable by exploiting commercial softwares, e.g., MATLAB. Note that P_D in (35) is actually unavailable during detection since it relies on the channel statistics from `Eve` to `Bob`, which is obviously unknown to `Alice`. However, through comparing (35) with the simulated P_D, it provides a benchmark to verify our proposed algorithm.

Algorithm 1 Proposed Jamming Attack Detection Algorithm

1: `Alice` sends pilot sequences over K independent channel coherence intervals;
2: `Bob` retransmits the conjugate of its received pilots back to `Alice`;
3: By averaging, `Alice` collects K independent observations $Q(k)$ for jamming detection;
4: Compute the decision metric either by (22) or (23);
5: Calculate the threshold η_{std} numerically with (34) given a required P_{FA}, or η_{var} with (37);
6: **if** $\phi_{\text{std}} > \eta_{\text{std}}$ or $\phi_{\text{var}} > \eta_{\text{var}}$ **then**
7: there is no jamming attack, and declares $\mathcal{H}_0$;
8: **else**
9: there is jamming attack, and declares $\mathcal{H}_1$;
10: **end if**

2) Sample Variance Normalization: When ϕ_{var} is chosen as the decision metric, the probability of false alarm can be represented the same way

$$P_{FA} = P\left\{\phi_{\text{var}} \leq \eta_{\text{var}}; \mathcal{H}_0\right\} \tag{36}$$

where η_{var} is the decision threshold in this scenario. Then the threshold is calculated with a predefined P_{FA}

$$\eta_{\text{var}} = P^{-1}_{q=K,p=K-1,\varsigma=K\mu_0^2/\sigma_0^2}\left(P_{FA}\right) \tag{37}$$

where $P^{-1}_{q,p,\varsigma}(y)$ represents the inverse function of CDF of a non-central f random variable with q and p degrees of freedom, and non-centrality parameter ς. Similarly, the PDF of ϕ_{var} under $\mathcal{H}_0$ is attainable, and then η_{var} is calculated by using (37) for the final decision.

As for P_D, it is represented by

$$P_D = P\left\{\phi_{\text{var}} < \eta_{\text{var}}; \mathcal{H}_1\right\}. \tag{38}$$

As before, although a closed-form expression of P_D may not exist, its numerical result is available. In conclusion, the proposed detection scheme can be summarized on the top right of this page.

D. The Impact of M on the Detection Performance

It is found that the number of antennas M poses little effect on the detection performance if M is larger than a certain number. In this section, we will analyze the cause behind this observation. First, the following lemma is introduced.

Lemma 5: Suppose T is a t random variable with υ degrees of freedom and non-centrality parameter δ, the mean and variance of T are [26]

$$\begin{aligned}\mathbb{E}\left[T\right] &= \delta\sqrt{\frac{\upsilon}{2}}\frac{\Gamma\left(\frac{\upsilon-1}{2}\right)}{\Gamma\left(\frac{\upsilon}{2}\right)}, \quad \upsilon > 1\\ \text{var}\left[T\right] &= \frac{\upsilon\left(1+\delta^2\right)}{\upsilon-2} - \mathbb{E}^2\left[T\right], \quad \upsilon > 2\end{aligned} \tag{39}$$

where $\Gamma(\cdot)$ denotes the gamma function.

Here we choose decision metric ϕ_{std} to analyze the impact of M on the detection performance, the detection using ϕ_{var}

is managed the same way. In general, K is on the order of several hundreds, thus ϕ_{std} is well approximated as a Gaussian variable. According to Lemma 5, the mean and variance of ϕ_{std} under hypothesis $\mathcal{H}_0$ are

$$\begin{aligned}&\mathbb{E}\left[\phi_{\mathsf{std}};\mathcal{H}_0\right]\\&=\sqrt{K}\frac{\mu_0}{\sigma_0}\sqrt{\frac{K-1}{2}}\frac{\Gamma\left(\frac{K-2}{2}\right)}{\Gamma\left(\frac{K-1}{2}\right)},\\&\mathrm{var}\left[\phi_{\mathsf{std}};\mathcal{H}_0\right]\\&=\frac{(K-1)\left(1+K\frac{\mu_0^2}{\sigma_0^2}\right)}{(K-3)}-\frac{K\,(K-1)}{2}\frac{\mu_0^2}{\sigma_0^2}\left(\frac{\Gamma\left(\frac{K-2}{2}\right)}{\Gamma\left(\frac{K-1}{2}\right)}\right)^2.\end{aligned}\tag{40}$$

Given (9) and (10), the following relationship is attainable

$$\frac{\mu_0}{\sigma_0}=\frac{M\sigma_h^2}{\sqrt{M\sigma_h^4+M\sigma_h^2\frac{\sigma_n^2}{2\tau}+\frac{\sigma_n^2}{2\tau}}}\approx\frac{M\sigma_h^2}{\sqrt{M\sigma_h^4+M\sigma_h^2\frac{\sigma_n^2}{2\tau}}}\tag{41}$$

if M is a large number. From (40), it is derived that $\mathbb{E}[\phi_{\mathsf{std}};\mathcal{H}_0]$ is proportional to μ_0/σ_0, and $\mathrm{var}[\phi_{\mathsf{std}};\mathcal{H}_0]$ is in proportion to μ_0^2/σ_0^2 when $K\mu_0^2/\sigma_0^2\gg 1$. Given that ϕ_{std} is Gaussian-distributed, the probability of false alarm is represented as the CDF of ϕ_{std} under hypothesis $\mathcal{H}_0$, namely

$$P_{FA}=\frac{1}{2}\left(1+\mathrm{erf}\left(\frac{\eta_{\mathsf{std}}-\mathbb{E}\left[\phi_{\mathsf{std}};\mathcal{H}_0\right]}{\sqrt{2\mathrm{var}\left[\phi_{\mathsf{std}};\mathcal{H}_0\right]}}\right)\right).\tag{42}$$

Since P_{FA} is predefined, then the threshold η_{std} is given by

$$\eta_{\mathsf{std}}=\mathbb{E}\left[\phi_{\mathsf{std}};\mathcal{H}_0\right]+\sqrt{2\mathrm{var}\left[\phi_{\mathsf{std}};\mathcal{H}_0\right]}\mathrm{erf}^{-1}\left(2P_{FA}-1\right)\tag{43}$$

where $\mathrm{erf}^{-1}(\cdot)$ indicates the inverse function of $\mathrm{erf}(\cdot)$.

Similarly, the mean and variance of ϕ_{std} under $\mathcal{H}_1$ are

$$\begin{aligned}&\mathbb{E}\left[\phi_{\mathsf{std}};\mathcal{H}_1\right]\\&=\sqrt{K}\frac{u_1}{\sigma_1}\sqrt{\frac{K-1}{2}}\frac{\Gamma\left(\frac{K-2}{2}\right)}{\Gamma\left(\frac{K-1}{2}\right)},\\&\mathrm{var}\left[\phi_{\mathsf{std}};\mathcal{H}_1\right]\\&=\frac{(K-1)\left(1+K\frac{\mu_1^2}{\sigma_1^2}\right)}{(K-3)}-\frac{K\,(K-1)}{2}\frac{\mu_1^2}{\sigma_1^2}\left(\frac{\Gamma\left(\frac{K-2}{2}\right)}{\Gamma\left(\frac{K-1}{2}\right)}\right)^2.\end{aligned}\tag{44}$$

Accordingly, P_D is written as the CDF of ϕ_{std} under hypothesis $\mathcal{H}_1$, namely

$$P_D=\frac{1}{2}\left(1+\mathrm{erf}\left(\frac{\eta_{\mathsf{std}}-\mathbb{E}\left[\phi_{\mathsf{std}};\mathcal{H}_1\right]}{\sqrt{2\mathrm{var}\left[\phi_{\mathsf{std}};\mathcal{H}_1\right]}}\right)\right).\tag{45}$$

As before, $\mathbb{E}[\phi_{\mathsf{std}};\mathcal{H}_1]$ is proportional to μ_1/σ_1. In addition, $\mathrm{var}[\phi_{\mathsf{std}};\mathcal{H}_1]$ is proportional to μ_1^2/σ_1^2 when $K\mu_1^2/\sigma_1^2\gg 1$. Then according to (9) and (10), if $M\gg 1$, one can arrive at

$$\begin{aligned}\frac{\mu_1}{\sigma_1}&=\frac{M\sigma_h^2}{\sqrt{M\sigma_h^4+M\sigma_h^2\frac{\sigma_g^2}{2}+M\sigma_h^2\frac{\sigma_n^2}{2\tau}+\frac{\sigma_n^2}{2\tau}}}\\&\approx\frac{M\sigma_h^2}{\sqrt{M\sigma_h^4+M\sigma_h^2\frac{\sigma_g^2}{2}+M\sigma_h^2\frac{\sigma_n^2}{2\tau}}}.\end{aligned}\tag{46}$$

According to the analysis above, if the number of antennas increases from M to nM, both μ_0/σ_0 and $\mathbb{E}[\phi_{\mathsf{std}};\mathcal{H}_0]$ will be $\sqrt{n}$ times larger, and $\mathrm{var}[\phi_{\mathsf{std}};\mathcal{H}_1]$ will be n times larger. As a result, the new threshold changes to $\sqrt{n}\eta_{\mathsf{std}}$ in the light of (43). Similarly, the mean and variance of ϕ_{std} under $\mathcal{H}_1$ change to $\sqrt{n}\mathbb{E}[\phi_{\mathsf{std}};\mathcal{H}_1]$ and $n\mathrm{var}[\phi_{\mathsf{std}};\mathcal{H}_1]$, respectively. Therefore, the new probability of detection approximates to

$$P_D'\approx\frac{1}{2}\left(1+\mathrm{erf}\left(\frac{\sqrt{n}\eta_{\mathsf{std}}-\sqrt{n}\mathbb{E}\left[\phi_{\mathsf{std}};\mathcal{H}_1\right]}{\sqrt{2n\mathrm{var}\left[\phi_{\mathsf{std}};\mathcal{H}_1\right]}}\right)\right).\tag{47}$$

By comparing (45) with (47), one can find that the detection performance keeps unchanged. It is also worth noting that the aforementioned result is based on the assumption that M is a large number, which is reasonable in massive antenna systems. Actually, if M is small, P_D grows initially with the increase of M, as will be shown by numerical simulation. Therefore, it comes to the conclusion that deploying more antennas doesn't pose any impact on P_D if M is larger than a certain threshold. In addition, this conclusion is also valid in the case of ϕ_{var}, the derivation is omitted here for the purpose of briefness.

E. The Impact of K on the Detection Performance

In this subsection, the impact of K on P_D is investigated. Decision metric ϕ_{std} is selected as reference, with its mean and variance under hypothesis $\mathcal{H}_0$ given in (40). Then if the number of observations increases from K to nK, the mean and variance of ϕ_{std} become to

$$\begin{aligned}&\mathbb{E}'\left[\phi_{\mathsf{std}};\mathcal{H}_0\right]\approx n\mathbb{E}\left[\phi_{\mathsf{std}};\mathcal{H}_0\right],\\&\mathrm{var}'\left[\phi_{\mathsf{std}};\mathcal{H}_0\right]\approx n^2\mathrm{var}\left[\phi_{\mathsf{std}};\mathcal{H}_0\right]-K\left(n^2-n\right)\frac{\mu_0^2}{\sigma_0^2}.\end{aligned}\tag{48}$$

According to (48) and (43), the updated threshold η_{std}' can be obtained. In particular, we have $\mathrm{erf}^{-1}(2P_{FA}-1)<0$ since normally $0<P_{FA}<0.5$, and $K(n^2-n)\mu_0^2/\sigma_0^2>0$ because it is assumed $n>1$. Therefore, compared the updated threshold with the original one, we have $\eta_{\mathsf{std}}'>n\eta_{\mathsf{std}}$.

In addition, the distribution of decision metric under $\mathcal{H}_1$ also changes, with its mean and variance derived as follows

$$\begin{aligned}&\mathbb{E}'\left[\phi_{\mathsf{std}};\mathcal{H}_1\right]\approx n\mathbb{E}\left[\phi_{\mathsf{std}};\mathcal{H}_1\right],\\&\mathrm{var}'\left[\phi_{\mathsf{std}};\mathcal{H}_1\right]\approx n^2\mathrm{var}\left[\phi_{\mathsf{std}};\mathcal{H}_1\right]-K\left(n^2-n\right)\frac{\mu_1^2}{\sigma_1^2}.\end{aligned}\tag{49}$$

where $K(n^2-n)\mu_1^2/\sigma_1^2>0$. Therefore, the updated probability of detection is denoted by

$$P_D'=\frac{1}{2}\left(1+\mathrm{erf}\left(\frac{\eta_{\mathsf{std}}'-\mathbb{E}'\left[\phi_{\mathsf{std}};\mathcal{H}_1\right]}{\sqrt{2\mathrm{var}'\left[\phi_{\mathsf{std}};\mathcal{H}_1\right]}}\right)\right).\tag{50}$$

By substituting $\eta_{\mathsf{std}}'>n\eta_{\mathsf{std}}$ and (49) into (50), it comes to the following result

$$\begin{aligned}P_D'&>\frac{1}{2}\left(1+\mathrm{erf}\left(\frac{n\eta_{\mathsf{std}}-n\mathbb{E}[\phi_{\mathsf{std}};\mathcal{H}_1]}{\sqrt{2n^2\mathrm{var}[\phi_{\mathsf{std}};\mathcal{H}_1]-2K(n^2-n)\frac{\mu_1^2}{\sigma_1^2}}}\right)\right)\\&>\frac{1}{2}\left(1+\mathrm{erf}\left(\frac{n\eta_{\mathsf{std}}-n\mathbb{E}\left[\phi_{\mathsf{std}};\mathcal{H}_1\right]}{2n^2\mathrm{var}\left[\phi_{\mathsf{std}};\mathcal{H}_1\right]}\right)\right)\\&>\frac{1}{2}\left(1+\mathrm{erf}\left(\frac{\eta_{\mathsf{std}}-\mathbb{E}\left[\phi_{\mathsf{std}};\mathcal{H}_1\right]}{2\mathrm{var}\left[\phi_{\mathsf{std}};\mathcal{H}_1\right]}\right)\right)=P_D.\end{aligned}\tag{51}$$

TABLE I
COMPUTATIONAL COMPLEXITY OF ALGORITHMS USING ϕ_{std} AND ϕ_{var}

Algorithms	Complexity analysis
Detection using ϕ_{std}	$4KC_{add} + KC_{mul} + 3C_{div} + C_{sqrt}$
Detection using ϕ_{var}	$4KC_{add} + 2KC_{mul} + 3C_{div}$

Evidently, increasing the number of observations always helps to improve the detection performance. However, a larger K indicates a larger processing delay, thus careful attention needs to paid during system design.

F. Computational Complexity

This subsection discusses the computational complexity of jamming detection algorithms using ϕ_{std} and ϕ_{var}. The complexity of the proposed schemes is mainly attributed to estimating the mean and variance of $Q(k)$, and computing decision metrics. The results of comparison are listed in Table I, where C_{add}, C_{div}, C_{sqrt} and C_{mul} denotes a single real addition, real division, square root and real multiplication operations. From this table, it is shown that although algorithm using ϕ_{std} needs an extra square root calculation than that using ϕ_{var}, it saves K times real multiplication operations. Therefore, the jamming detection scheme using the first decision metric enjoys a lower complexity. Moreover, since the complexity of both algorithms using ϕ_{std} and ϕ_{var} grows with K, a trade-off between performance and complexity exists in the system design.

V. NUMERICAL RESULTS

The considered model contains a legitimate transmit-receive pair (`Alice` and `Bob`), and a malicious jammer `Eve`. Channels of different transmit-receive pairs are modeled by independent Rayleigh fading. The SNR considered in all simulations is the SNR at `Bob`. Both analytical and numerical results are included for the purpose of comparison. Analytical results are obtained by (35) and (38). As for numerical results, random signals are required to be generated at `Alice` and sent to `Bob`, and retransmitted back to `Alice`. After that, the final decision is made through comparing the simulated decision metric with the threshold.

Fig. 5 illustrates the receiver operating characteristic (ROC) curves of the proposed scheme, where $M = 100$, $\sigma_h^2 = 1$, $\sigma_g^2 = 0.5$, $\sigma_n^2 = 1$ and $\tau = 10$. All results are averaged over 10,000 Monte Carlo realizations. First, the well match between the simulated and analytical results validate the effectiveness of our analysis. As expected, K is critical to the success of jammer detection. For example, the ROC curve has the steepest slope when $K = 400$, which means the increment of P_D is the largest as P_{FA} increases. Moreover, Fig. 5 reveals that algorithms using ϕ_{std} and ϕ_{var} possess almost the same performance, thus only ϕ_{std} is taken into account in the following simulations.

Among the existing literatures, the one that is most related to our method is the energy ratio detector (ERD) presented in [15]. Fig. 6 draws the ROC curves of our scheme and ERD.

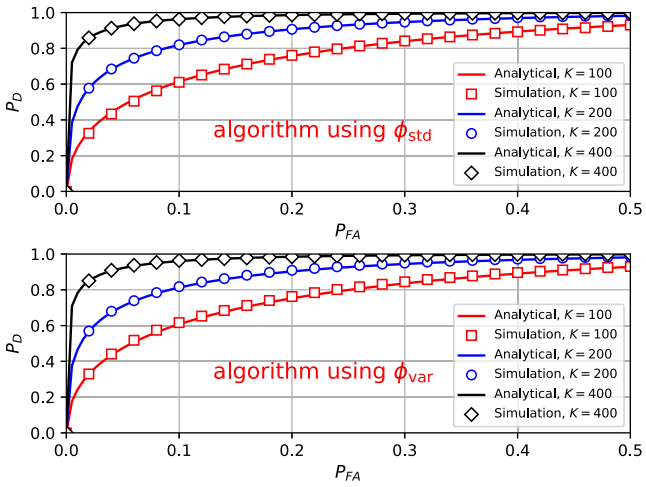


Fig. 5. The ROC curves of the proposed jamming attack detection algorithm, where K varies.

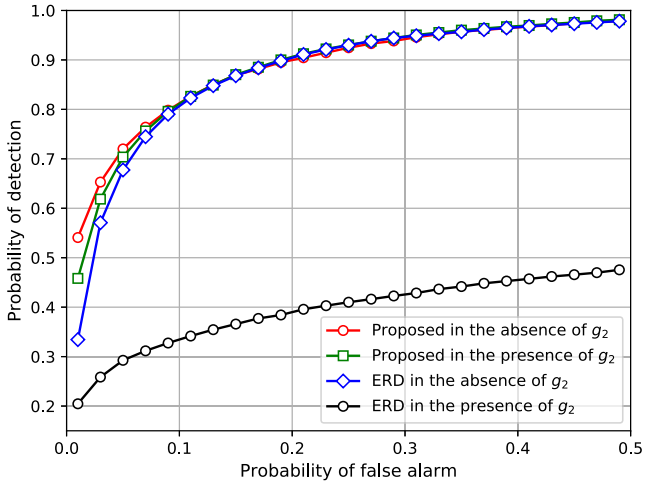


Fig. 6. Probability of detection versus probability of false alarm, where the proposed scheme and ERD are included.

The common parameters are $M = 100$, $\sigma_h^2 = 1$, $\sigma_g^2 = 0.5$ and $\sigma_n^2 = 1$. For our proposed method, it is assumed that $\tau = 10$ and $K = 200$, while $N_1 = N_2 = 50$ in ERD where N_1 and N_2 denote the length of pilot sequence and time of retransmissions, respectively. g_2 is an indicator of jamming attack during retransmission. Concretely, two methods achieve almost the same P_D in the absence of g_2. However, compared with our algorithm, ERD is quite sensitive to the attack in the retransmission phase. This is attributed to the fact that jamming attack during retransmission poses little impact on the distribution of $Q(k)$, as denoted in (11). Therefore, the proposed detection method is more robust than ERD. Since our scheme and ERD are separately designed for non-coherent and coherent communications, we focus on the performance of our scheme in non-coherent massive SIMO systems in the following discussion.

Alternatively, Fig. 7 draws the ROC curves of the proposed detection scheme using ϕ_{std} with various σ_g^2, where $K = 100$ and other parameters can be referred to in Fig. 5. The key observation lies in that P_D increases with σ_g^2, namely the channel statistics from `Eve` to `Bob`. This is reasonable since

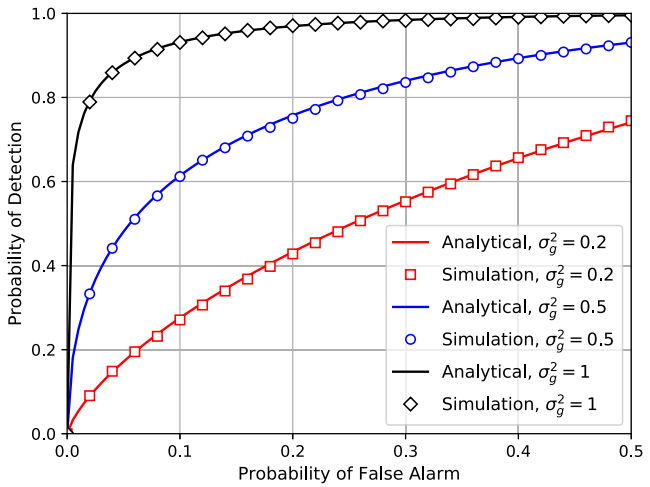


Fig. 7. The ROC curves of the proposed jamming attack detection algorithm, where σ_g^2 varies.

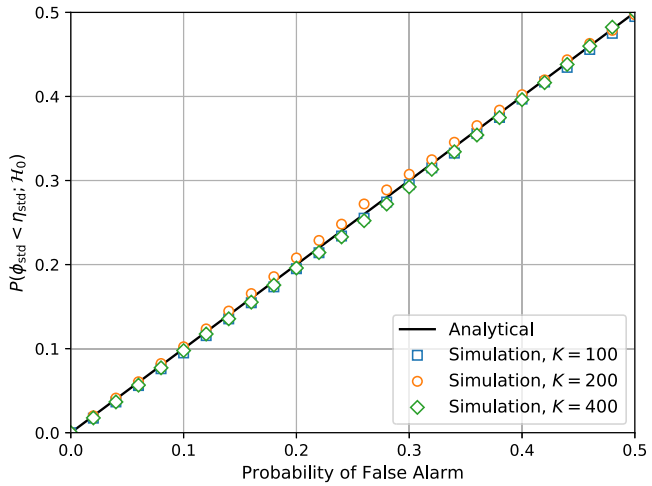


Fig. 8. The comparison between simulated P_{FA} and predefined P_{FA}.

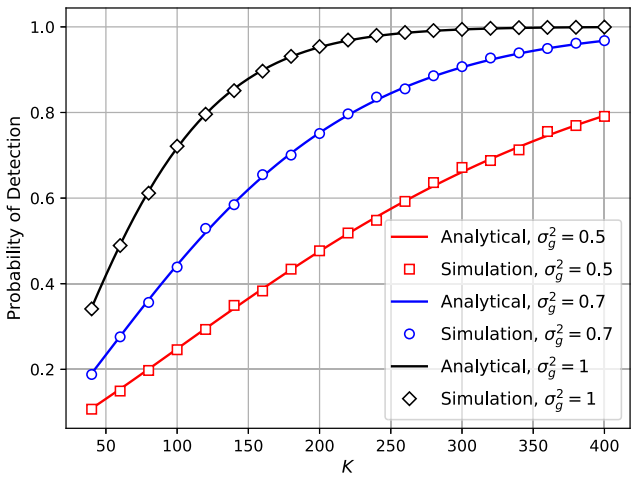


Fig. 9. The relationship between P_D and K, where σ_g^2 varies.

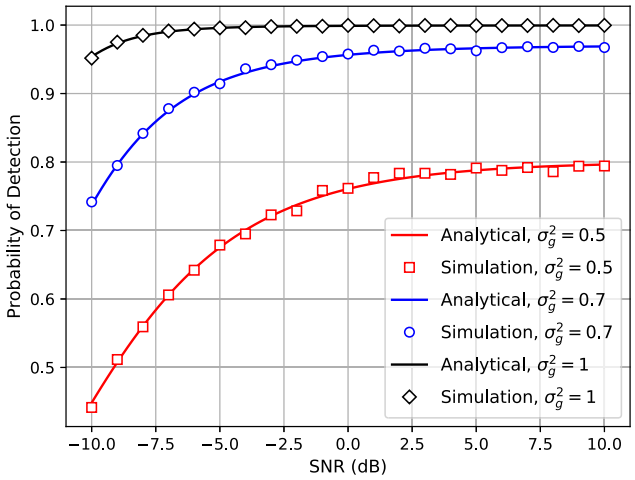


Fig. 10. The relationship between P_D and SNR, where σ_g^2 varies.

a larger σ_g^2 indicates a greater portion of jamming signal in the composite received $\mathbf{y}_t(k)$. In real application, a larger σ_g^2 often represents scenarios that Eve is close to Bob.

Fig. 8 shows the comparison results between a predefined P_{FA} and simulated $P\{\phi_{\text{std}} < \eta_{\text{std}}; \mathcal{H}_0\}$, where parameters can be referred to in Fig. 5. Since approximations are employed to derive our scheme, thus some extent of mismatch could exist between a predefined P_{FA} and $P\{\phi_{\text{std}} < \eta_{\text{std}}; \mathcal{H}_0\}$. A large mismatch often signifies the simulation results are of low degree of confidence, which should be paid close attention to. Fortunate enough, the close match between a predefined P_{FA} and simulated $P\{\phi_{\text{std}} < \eta_{\text{std}}; \mathcal{H}_0\}$ proves the proposed algorithm works well.

Fig. 9 demonstrates how P_D varies with K, where $P_{FA} = 0.01$ and other parameters can be referred to in Fig. 5. Results in this figure once again proves increasing the number of observations $Q(k)$ is an effective way to improve P_D. However, this comes at the price of increased computational complexity and processing delay. Hence, from a system prospective point of view, K is required to be carefully chosen to strike a balance between complexity and performance. As before, Fig. 9 indicates σ_g^2 has a great impact on P_D.

In Fig. 10, the relationship between P_D and SNR (SNR_{Bob}) is illustrated, where σ_g^2 varies, $K = 400$ and other parameters can be referred to in Fig. 5. It can be revealed that although P_D is in direct proportion to SNR, it would reach an error floor that no matter how large the SNR is, P_D cannot be improved. In addition, σ_g^2 presents itself as a key parameter to guarantee a desirable performance at low SNR. For example, P_D increases from 0.25 to about 0.8 as σ_g^2 changes from 0.5 to 1, when SNR $= -10$ dB. Therefore, high SNRs would not be a necessity in this situation.

Fig. 11 shows how P_D varies with M, where $P_{FA} = 0.01$ and $K = 400$. As expected, P_D quickly converges to constant as M grows, which is in accordance to (47). Moreover, it can be observed that P_D increases initially when M is small, e.g. $M \leq 20$. To reveal the reason behind, let's revisit (40), where both $\mathbb{E}[\phi_{\text{std}}; \mathcal{H}_0]$ and $\sqrt{\text{var}[\phi_{\text{std}}; \mathcal{H}_0]}$ are $\sqrt{n}$ times larger if the number of antennas increases from M to nM. Beware of the former result is built upon $K\mu_0^2/\sigma_0^2 \gg 1$, which is not the case if M is relatively small. According to the property of non-central t distributions that $\sqrt{\frac{\upsilon}{2}}\Gamma(\frac{\upsilon-1}{2})/\Gamma(\frac{\upsilon}{2}) \approx \frac{4\upsilon-1}{4\upsilon-4}$ [27],

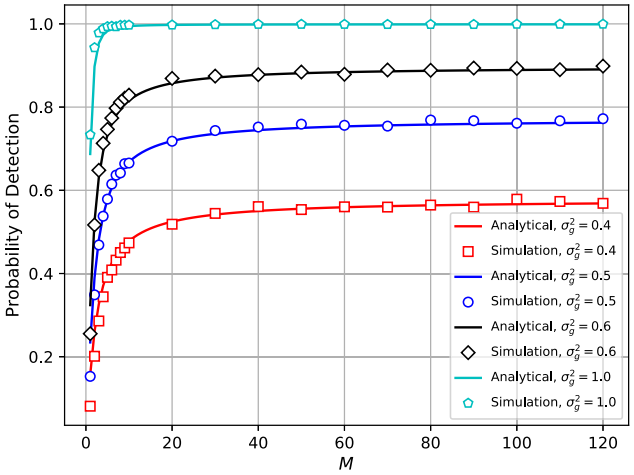


Fig. 11. The relationship between P_D and M, where σ_g^2 varies.

$\text{var}[\phi_{\text{std}}; \mathcal{H}_0]$ can be approximated to

$$K\frac{u_0^2}{\sigma_0^2}\left(\frac{(K-1)}{(K-3)}\left(1+\frac{\sigma_0^2}{Ku_0^2}\right)-\left(1-\frac{3}{4K-5}\right)^{-2}\right).$$

Due to the presence of $\left(Ku_0^2/\sigma_0^2\right)^{-1}$, if the number of antennas increases from M to nM, the updated mean and variance are

$$\begin{aligned}\mathbb{E}'[\phi_{\text{std}}; \mathcal{H}_0] &= \sqrt{n}\mathbb{E}[\phi_{\text{std}}; \mathcal{H}_0]\\ \text{var}'[\phi_{\text{std}}; \mathcal{H}_0] &< n\text{var}[\phi_{\text{std}}; \mathcal{H}_0]\end{aligned} \tag{52}$$

which indicates the growth rate of mean is larger than that of variance. Note that the same conclusion holds in the scenario of $\mathcal{H}_1$. Intuitively, the growth of mean indicates the separation of two PDFs becomes larger, while the growth of variance means two PDFs get closer. Consider the growth rates of mean and variance, the overlap between two PDFs will reduce, thus a successful detection are more likely to happen.

VI. Conclusion

Non-coherent receivers are promising in systems with massive antenna array, primarily due to their low complexity and costs. However, non-coherent systems are vulnerable to active attack, just as their coherent counterparts.

Different from existing literatures, this paper proposes to detect the jamming attack by `Alice` instead of `Bob`. To be more realistic, our proposed algorithm doesn't assume any knowledge of channel statistics. Concretely, `Bob` retransmits the conjugate of its received signal back to `Alice`, where the final decision is made. According to the LRT principle, two decision metrics are introduced based on the normalization of K independent observations. Simulation results indicate the proposed detection scheme is effective. Interestingly, the number of antennas at `Bob` has a marginal impact on P_D if M is greater than a certain threshold. On the other hand, two key parameters, K and σ_g^2, pose a great influence on the detection performance. Since σ_g^2 is an environmental parameter and increasing M cannot continuously improve P_D, K is regarded as a key parameter in system design.

Appendix A
The Proof of $Q(k)$ Being a Real Gaussian Variable

Suppose $\{X_1, X_2, \ldots, X_n\}$ is a sequence of i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2 < \infty$. Then according to the Lindeberg-Lévy CLT, as n approaches infinity, the random variables $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)$ converge in distribution to $\mathcal{N}(0, \sigma^2)$, i.e.

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right). \tag{53}$$

The Gaussian distributions are closed with linear transformations. That is, if X is normally distributed with mean μ and variance σ^2, then a linear transform $aX + b$ (for some real numbers a and b) is also normally distributed [28]

$$aX + b \sim \mathcal{N}\left(a\mu + b,\, a^2\sigma^2\right). \tag{54}$$

Therefore, (53) changes to

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{d} \mu + \frac{1}{\sqrt{n}}\mathcal{N}\left(0, \sigma^2\right) = \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right). \tag{55}$$

Then let's return back to our problem. First, $z_{t+\tau}(k)$, which denotes the received signal by `Alice` at time $t+\tau$ in the k-th channel coherence interval, is expanded as follows

$$\begin{aligned}z_{t+\tau}(k) &= \mathbf{h}^T(k)\mathbf{y}_t^*(k) + v_{t+\tau}(k)\\ &= \sum_{m=1}^{M} h_m(k)y_{t,m}^*(k) + v_{t+\tau}(k)\end{aligned} \tag{56}$$

where $h_m(k)$ is the channel with respect to the m-th antenna in the k-th interval, and $y_{t,m}(k)$ denotes the m-th item of $\mathbf{y}_t(k)$. If `Eve` is absent, it comes to

$$\sum_{m=1}^{M} h_m(k)y_{t,m}^*(k) = \sum_{m=1}^{M} h_m(k)\left(h_m^*(k)x_t^*(k) + n_{t,m}^*(k)\right) \tag{57}$$

where $n_{t,m}^*(k)$ is the m-th item of $\mathbf{n}_t(k)$. According to (55), $\sum_{m=1}^{M} h_m(k)y_{t,m}^*(k)$ converges to complex Gaussian distribution if M is large. Besides, $z_{t+\tau}(k)$ is complex Gaussian-distributed as $v_{t+\tau}(k)$ is AWGN. As $z(k) = \sum_{t=1}^{\tau} z_{t+\tau}(k)/\tau$, its real part $Q(k)$ is Gaussian. In the presence of `Eve`, the analysis is carried out similarly. Thus, the proof is concluded.

Appendix B
The Derivation of (9) and (10)

Under hypothesis $\mathcal{H}_0$, the mean of $Q(k)$ is computed by

$$\begin{aligned}\mu_0 &= \mathbb{E}\left[\mathbf{h}^T(k)\mathbf{h}^*(k)\right]\bar{x} + \mathbb{E}\left[\Re\left\{\mathbf{h}^T(k)\bar{\mathbf{n}}^*(k)\right\}\right]\\ &\quad + \mathbb{E}\left[\Re\left\{\bar{v}(k)\right\}\right]\\ &= \mathbb{E}\left[\sum_{i=1}^{M}|h_i(k)|^2\right] + \mathbb{E}\left[\Re\left\{\sum_{i=1}^{M} h_i(k)\bar{n}_i^*(k)\right\}\right]\\ &= M\sigma_h^2.\end{aligned} \tag{58}$$

While the variance of $Q(k)$ is expanded as follows

$$\begin{aligned}\sigma_0^2 &= \mathrm{var}\left[\mathbf{h}^T(k)\mathbf{h}^*(k)\bar{x} + \Re\left\{\mathbf{h}^T(k)\bar{\mathbf{n}}^*(k)\right\} + \Re\{\bar{v}(k)\}\right] \\ &= \mathrm{var}\left[\mathbf{h}^T(k)\mathbf{h}^*(k)\right] + \mathrm{var}\left[\Re\left\{\mathbf{h}^T(k)\bar{\mathbf{n}}^*(k)\right\}\right] \\ &\quad + 2\mathrm{cov}\left[\mathbf{h}^T(k)\mathbf{h}^*(k), \Re\left\{\mathbf{h}^T(k)\bar{\mathbf{n}}^*(k)\right\}\right] + \frac{\sigma_n^2}{2\tau}. \quad (59)\end{aligned}$$

The first element on the right hand side of (59) is

$$\begin{aligned}\mathrm{var}\left[\mathbf{h}^T(k)\mathbf{h}^*(k)\right] &= \mathrm{var}\left[\sum_{i=1}^{M}|h_i(k)|^2\right] \\ &= M\mathrm{var}\left[(\Re\{h_i(k)\})^2 + (\Im\{h_i(k)\})^2\right] \\ &= \frac{M\sigma_h^4}{4}\mathrm{var}\left[\begin{array}{l}\left(\frac{\sqrt{2}}{\sigma_h}\Re\{h_i(k)\}\right)^2 \\ +\left(\frac{\sqrt{2}}{\sigma_h}\Im\{h_i(k)\}\right)^2\end{array}\right] \\ &= M\sigma_h^4 \quad (60)\end{aligned}$$

where $\frac{\sqrt{2}}{\sigma_h}\Re\{h_i(k)\} \sim \mathcal{N}(0,1)$ and $\frac{\sqrt{2}}{\sigma_h}\Im\{h_i(k)\} \sim \mathcal{N}(0,1)$, thus $(\Re\{h_i(k)\})^2 + (\Im\{h_i(k)\})^2$ is a χ^2 random variable with its variance being twice of its degree of freedom.

The second component on the right side of (59) can be simplified as

$$\begin{aligned}\mathrm{var}\left[\Re\left\{\mathbf{h}^T(k)\bar{\mathbf{n}}^*(k)\right\}\right] &= \frac{1}{2}\mathrm{var}\left[\sum_{i=1}^{M}h_i(k)\bar{n}_i^*(k)\right] \\ &= \frac{1}{2}\sum_{i=1}^{M}\mathrm{var}\left[h_i(k)\bar{n}_i^*(k)\right] \\ &= \frac{M}{2}\mathrm{var}\left[h_i(k)\right]\mathrm{var}\left[\bar{n}_i^*(k)\right] \\ &= M\sigma_h^2\frac{\sigma_n^2}{2\tau}. \quad (61)\end{aligned}$$

While the third component is simplified as

$$\begin{aligned}&2\mathrm{cov}\left[\mathbf{h}^T(k)\mathbf{h}^*(k), \Re\left\{\mathbf{h}^T(k)\bar{\mathbf{n}}^*(k)\right\}\right] \\ &= 2\mathbb{E}\left[\left(\sum_{i=1}^{M}h_i(k)h_i^*(k)\right)\Re\left\{\sum_{j=1}^{M}h_j(k)\bar{n}_j^*(k)\right\}\right] \\ &\quad - 2\mathbb{E}\left[\sum_{i=1}^{M}|h_i(k)|^2\right]\mathbb{E}\left[\Re\left\{\sum_{i=1}^{M}h_i(k)\bar{n}_i^*(k)\right\}\right] \\ &= 2\mathbb{E}\left[\left(\sum_{i=1}^{M}h_i(k)h_i^*(k)\right)\Re\left\{\sum_{j=1}^{M}h_j(k)\bar{n}_j^*(k)\right\}\right]. \quad (62)\end{aligned}$$

The former equation is the sum of M^2 different monomials, namely, $\mathbb{E}[h_i(k)h_i^*(k)\Re\{h_j(k)\bar{n}_j^*(k)\}]$. Depending on $i = j$ or $i \neq j$, we discuss the following two situations.

1) $i = j$: In this scenario, $h_i(k)h_i^*(k)$ relates to $\Re\left\{h_j(k)\bar{n}_j^*(k)\right\}$. Therefore

$$\begin{aligned}&\mathbb{E}\left[h_i(k)h_i^*(k)\Re\left\{h_j(k)\bar{n}_j^*(k)\right\}\right] \\ &= \mathbb{E}\left[h_i(k)h_i^*(k)\Re\left\{h_i(k)\bar{n}_i^*(k)\right\}\right] \\ &= \mathbb{E}\left[h_i(k)h_i^*(k)\left(\begin{array}{l}\Re\{h_i(k)\}\Re\left\{\bar{n}_i^*(k)\right\} \\ +\Im\{h_i(k)\}\Im\left\{\bar{n}_i^*(k)\right\}\end{array}\right)\right] \\ &= \mathbb{E}\left[h_i(k)h_i^*(k)\Re\{h_i(k)\}\right]\mathbb{E}\left[\Re\left\{\bar{n}_i^*(k)\right\}\right] \\ &\quad + \mathbb{E}\left[h_i(k)h_i^*(k)\Im\{h_i(k)\}\right]\mathbb{E}\left[\Im\left\{\bar{n}_i^*(k)\right\}\right] \\ &= 0 \quad (63)\end{aligned}$$

2) $i \neq j$: In this scenario, $h_i(k)h_i^*(k)$ is independent of $\Re\{h_j(k)\bar{n}_j^*(k)\}$. Therefore

$$\begin{aligned}&\mathbb{E}\left[h_i(k)h_i^*(k)\Re\left\{h_j(k)\bar{n}_j^*(k)\right\}\right] \\ &= \mathbb{E}\left[h_i(k)h_i^*(k)\right]\mathbb{E}\left[\Re\left\{h_j(k)\bar{n}_j^*(k)\right\}\right] \\ &= 0 \quad (64)\end{aligned}$$

Then combine the results in (63) and (64), one can have $2\mathrm{cov}[\mathbf{h}^T(k)\mathbf{h}^*(k), \Re\{\mathbf{h}^T(k)\bar{\mathbf{n}}^*(k)\}] = 0$. Therefore, it arrives at

$$\sigma_0^2 = M\sigma_h^4 + M\sigma_h^2\frac{\sigma_n^2}{2\tau} + \frac{\sigma_n^2}{2\tau}. \quad (65)$$

As before, the derivation of statistics of $Q(k)$ under hypothesis $\mathcal{H}_1$ can be carried out, i.e.,

$$\begin{aligned}\mu_1 &= \mathbb{E}\left[\mathbf{h}^T(k)\mathbf{h}^*(k)\right]\bar{x} + \mathbb{E}\left[\Re\left\{\mathbf{h}^T(k)\mathbf{g}^*(k)\right\}\right]\bar{x} \\ &\quad + \mathbb{E}\left[\Re\left\{\mathbf{h}^T(k)\bar{\mathbf{n}}^*(k)\right\}\right] + \mathbb{E}\left[\Re\{\bar{v}(k)\}\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{M}|h_i(k)|^2\right] + \mathbb{E}\left[\Re\left\{\sum_{i=1}^{M}h_i(k)g_i^*(k)\right\}\right] \\ &\quad + \mathbb{E}\left[\Re\left\{\sum_{i=1}^{M}h_i(k)\bar{n}_i^*(k)\right\}\right] \\ &= M\sigma_h^2 \quad (66) \\ \sigma_1^2 &= \mathrm{var}\left[\begin{array}{l}\mathbf{h}^T(k)\mathbf{h}^*(k)\bar{x} + \Re\left\{\mathbf{h}^T(k)\mathbf{g}^*(k)\right\}\bar{x} \\ +\Re\left\{\mathbf{h}^T(k)\bar{\mathbf{n}}^*(k)\right\} + \Re\{\bar{v}(k)\}\end{array}\right] \\ &= \mathrm{var}\left[\mathbf{h}^T(k)\mathbf{h}^*(k)\right] + \mathrm{var}\left[\Re\left\{\mathbf{h}^T(k)\mathbf{g}^*(k)\right\}\right] \\ &\quad + 2\mathrm{cov}\left[\mathbf{h}^T(k)\mathbf{h}^*(k), \Re\left\{\mathbf{h}^T(k)\mathbf{g}^*(k)\right\}\right] \\ &\quad + 2\mathrm{cov}\left[\begin{array}{l}\mathbf{h}^T(k)\mathbf{h}^*(k) + \Re\left\{\mathbf{h}^T(k)\mathbf{g}^*(k)\right\}, \\ \Re\left\{\mathbf{h}^T(k)\bar{\mathbf{n}}^*(k)\right\}\end{array}\right] \\ &\quad + \mathrm{var}\left[\Re\left\{\mathbf{h}^T(k)\bar{\mathbf{n}}^*(k)\right\}\right] + \frac{\sigma_n^2}{2\tau} \quad (67)\end{aligned}$$

where the third and fourth components in (67) are zero, of which the derivation can be referred to in (62). Hence, it comes to

$$\sigma_1^2 = M\sigma_h^4 + M\sigma_h^2\frac{\sigma_g^2}{2} + M\sigma_h^2\frac{\sigma_n^2}{2\tau} + \frac{\sigma_n^2}{2\tau}. \quad (68)$$

Finally, the derivation of (9) and (10) is concluded.

REFERENCES

[1] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[4] W. Trappe, "The challenges facing physical layer security," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 16–20, Jun. 2015.

[5] N. Yang, L. Wang, G. Geraci, M. Elkashlan, J. Yuan, and M. Di Renzo, "Safeguarding 5G wireless communication networks using physical layer security," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 20–27, Apr. 2015.

[6] L. Xiao, J. Liu, Q. Li, N. B. Mandayam, and H. V. Poor, "User-centric view of jamming games in cognitive radio networks," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 12, pp. 2578–2590, Dec. 2015.

[7] L. Xiao, Y. Li, C. Dai, H. Dai, and H. V. Poor, "Reinforcement learning-based NOMA power allocation in the presence of smart jamming," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3377–3389, Apr. 2017.

[8] D. Kapetanovic, G. Zheng, and F. Rusek, "Physical layer security for massive MIMO: An overview on passive eavesdropping and active attacks," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 21–27, Jun. 2015.

[9] K. Cumanan *et al.*, "Physical layer security jamming: Theoretical limits and practical designs in wireless networks," *IEEE Access*, vol. 5, pp. 3603–3611, 2016.

[10] J. Vinogradova, E. Björnson, and E. G. Larsson, "Detection and mitigation of jamming attacks in massive MIMO systems using random matrix theory," in *Proc. IEEE 17th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2016, pp. 1–5.

[11] J. Vinogradova, E. Björnson, and E. G. Larsson, "Jamming massive MIMO using massive MIMO: Asymptotic separability results," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 3454–3458.

[12] H. Akhlaghpasand, S. M. Razavizadeh, E. Björnson, and T. T. Do, "Jamming detection in massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 2, pp. 242–245, Apr. 2018.

[13] T. T. Do, E. Björnson, E. G. Larsson, and S. M. Razavizadeh, "Jamming-resistant receivers for the massive MIMO uplink," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 1, pp. 210–223, Jan. 2018.

[14] Q. Xiong, Y.-C. Liang, K. H. Li, Y. Gong, and S. Han, "Secure transmission against pilot spoofing attack: A two-way training-based scheme," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 5, pp. 1017–1026, May 2016.

[15] Q. Xiong, Y.-C. Liang, K. H. Li, and Y. Gong, "An energy-ratio-based approach for detecting pilot spoofing attack in multiple-antenna systems," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 5, pp. 932–940, May 2015.

[16] T. T. Do, H. Q. Ngo, T. Q. Duong, T. J. Oechtering, and M. Skoglund, "Massive MIMO pilot retransmission strategies for robustification against jamming," *IEEE Wireless Commun. Lett.*, vol. 6, no. 1, pp. 58–61, Feb. 2017.

[17] R. W. Heath, Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2017.

[18] A. Manolakos, M. Chowdhury, and A. J. Goldsmith, "Constellation design in noncoherent massive SIMO systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 3690–3695.

[19] L. Jing, E. De Carvalho, P. Popovski, and À. O. Martínez, "Design and performance analysis of noncoherent detection systems with massive receiver arrays," *IEEE Trans. Signal Process.*, vol. 64, no. 19, pp. 5000–5010, Oct. 2016.

[20] A. Manolakos, M. Chowdhury, and A. Goldsmith, "Energy-based modulation for noncoherent massive SIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7831–7846, Nov. 2016.

[21] 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.

[22] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.

[23] D. N. C. Tse and P. Viswanath, *Fundamentals Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

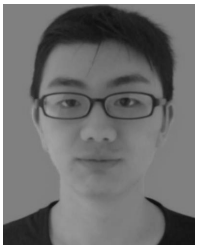
[24] H. Rahbari, M. Krunz, and L. Lazos, "Swift jamming attack on frequency offset estimation: The Achilles' heel of OFDM systems," *IEEE Trans. Mobile Comput.*, vol. 15, no. 5, pp. 1264–1278, May 2016.

[25] N. Mukhopadhyay, *Probability and Statistical Inference*. New York, NY, USA: Marcel Dekker, 2000.

[26] A. Leon-Garcia, *Probability, Statistics, and Random Processes for Electrical Engineering*. Upper Saddle River, NJ, USA: Pearson Prentice Hall, 2017.

[27] G. J. Resnikoff and G. J. Lieberman, *Tables of the Non-Central T-Distribution*. Stanford, CA, USA: Stanford Univ. Press, 1957.

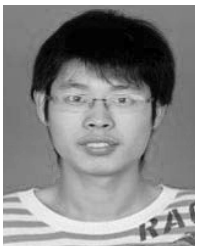
[28] *Operations on Normal Deviates*. Accessed: Aug. 5, 2018. [Online]. Available: https://en.wikipedia.org/wiki/Normal_distribution#Operations_on_normal_deviates



Shengbo Xu received the B.S. degree in electronic and information engineering from Chongqing University, China, in 2017, where he is currently pursuing the master's degree. His research interests include massive MIMO, machine learning, and physical layer security.

Weiyang Xu (M'16) received the B.S.E. and M.S.E. degrees from Xi'an Jiaotong University, Xi'an, China, in 2004 and 2007, respectively, and the Ph.D. degree from Fudan University, Shanghai, China, in 2010.

In 2014, he was a Visiting Scholar at the University of Southern Queensland, Australia. He is currently an Associate Professor with the School of Microelectronics and Communication Engineering, Chongqing University, China. His research interests include massive MIMO and cognitive radio techniques.



Cunhua Pan received the B.S. and Ph.D. degrees from the School of Information Science and Engineering, Southeast University, Nanjing, China, in 2010 and 2015, respectively.

From 2015 to 2016, he was a Research Associate at the University of Kent, U.K. He held a post-doctoral position at Queen Mary University of London, U.K., from 2016 and 2019, where he is currently a Lecturer. His research interests mainly include ultra-dense C-RAN, machine learning, UAV, Internet of Things, and mobile edge computing. He serves as a TPC member for numerous conferences, such as ICC and GLOBECOM, and the Student Travel Grant Chair for ICC 2019. He also serves as an Editor of IEEE ACCESS.

Maged Elkashlan received the Ph.D. degree in electrical engineering from The University of British Columbia, Canada, in 2006.

From 2007 to 2011, he was with the Commonwealth Scientific and Industrial Research Organization, Australia. During this time, he has held visiting appointments at the University of New South Wales and the University of Technology Sydney. In 2011, he joined the School of Electronic Engineering and Computer Science, Queen Mary University of London, U.K. His research interests include the broad areas of communication theory and statistical signal processing. He received the Best Paper Awards at the IEEE International Conference on Communications (ICC) in 2016 and 2014, the International Conference on Communications and Networking in China (CHINACOM), in 2014, and the IEEE Vehicular Technology Conference (VTC-Spring), in 2013. He currently serves as an Editor of IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.