

Analysis and Design of Secure Massive MIMO Systems in the Presence of Hardware Impairments

Jun Zhu, Derrick Wing Kwan Ng, Ning Wang, Robert Schober, and
Vijay K. Bhargava,
The University of British Columbia

Abstract

To keep the hardware costs of future communications systems manageable, the use of low-cost hardware components is desirable. This is particularly true for the emerging massive multiple-input multiple-output (MIMO) systems which equip base stations (BSs) with a large number of antenna elements. However, low-cost transceiver designs will further accentuate the hardware impairments which are present in any practical communication system. In this paper, we investigate the impact of hardware impairments on the secrecy performance of downlink massive MIMO systems in the presence of a passive multiple-antenna eavesdropper. Thereby, for the BS and the legitimate users, the joint effects of multiplicative phase noise, additive distortion noise, and amplified receiver noise are taken into account, whereas the eavesdropper is assumed to employ ideal hardware. We derive a lower bound for the ergodic secrecy rate of a given user when matched filter (MF) data precoding and artificial noise (AN) transmission are employed at the BS. Based on the derived analytical expression, we investigate the impact of the various system parameters on the secrecy rate and optimize both the pilot sets used for uplink training and the AN precoding. Our analytical and simulation results reveal that 1) the additive distortion noise at the BS may be beneficial for the secrecy performance, especially if the power assigned for AN emission is not sufficient; 2) all other hardware impairments have a negative impact on the secrecy performance; 3) despite their susceptibility to pilot interference in the presence of phase noise, so-called spatially orthogonal pilot sequences are preferable unless the phase noise is very strong; 4) the proposed generalized null-space (NS) AN precoding method can efficiently mitigate the negative effects of phase noise.

This work has been in part presented at The 17th IEEE International Workshop on Signal Processing Advances in Wireless Communications 2016 (SPAWC 2016) [1].

I. INTRODUCTION

The emerging massive multiple-input multiple-output (MIMO) architecture promises tremendous performance gains in terms of network throughput and energy efficiency by employing simple coherent processing across arrays of hundreds or even thousands of base station (BS) antennas, serving tens or hundreds of mobile terminals [2], [3]. Thereby, physical layer channel impairments such as fading, additive Gaussian noise, and interference are averaged out in the limit of an infinite number of antennas [2]- [5]. As an additional benefit, massive MIMO is inherently more secure than conventional MIMO systems, as the large-scale antenna array equipped at the transmitter (Alice) can accurately focus a narrow and directional information beam on the intended terminal (Bob), such that the received signal power at Bob is several orders of magnitude higher than that at any incoherent passive eavesdropper (Eve) [6]. Unfortunately, this benefit may vanish if Eve also employs a massive antenna array for eavesdropping. In this case, unless additional measures to secure the communication are taken by Alice, even a single passive Eve may be able to intercept the signal intended for Bob [7], [8].

Since security is a critical concern for future communication systems, facilitating secrecy at the physical layer of (massive) MIMO systems has received significant attention recently. Physical layer security for conventional (non-massive) MIMO transmission has been extensively studied in the literature, e.g. [9]–[11]. A large system secrecy analysis of MIMO systems was provided in [12], [13]. Thereby, availability of Eve’s channel state information (CSI) at Alice was assumed, which is an optimistic assumption in practice. Artificial noise (AN) generation [14] was employed to provide physical layer security in a multi-cell massive MIMO system with pilot contamination in [7] for the case when Eve’s CSI is not known. Thereby, it was shown that secure communication can be achieved even with simple matched filter (MF) precoding of the data and null-space (NS) precoding of the AN. Nevertheless, it was revealed in [15] that significant additional performance gains are possible with more sophisticated data and AN precoders, including polynomial precoders. Furthermore, AN-aided jamming of Rician fading massive MIMO channels was investigated in [16]. In the context of massive MIMO relaying, the work presented in [17] compared two classic relaying schemes, i.e., amplify-and forward (AF) and decode-and-forward (DF), for physical layer security with imperfect CSI at the massive MIMO relay. While [7]- [17] assumed that Eve is passive, the so-called pilot contamination

attack, a form of active eavesdropping, was also considered in the literature. In particular, several techniques for detection of the pilot contamination attack were proposed in [6]. Moreover, the authors in [18] developed a secret key agreement protocol under the pilot contamination attack, and the authors in [19] proposed to encrypt the pilot sequence in order to hide it from the attacker. Several techniques for combating the pilot contamination attack at the physical layer of a multi-cell massive MIMO system were proposed in [20].

All aforementioned works on secure massive MIMO are based on the assumption that the transceivers of the legitimate users are equipped with perfect hardware components, i.e., the effects of hardware impairments (HWIs) were not taken into account. Nevertheless, all practical implementations do suffer from HWIs such as phase noise, quantization errors, amplification noise, and nonlinearities [21]. These impairments are expected to be particularly pronounced in massive MIMO systems as the excessive number of BS antennas makes the use of low-cost components desirable to keep the overall capital expenditures for operators manageable. Although HWIs can be mitigated by analog and digital signal processing techniques [22], they cannot be removed completely, due to the randomness introduced by the different sources of imperfection. The remaining residual HWIs can be modelled by a combination of phase noise and additive distortion noises at the transmitter and the receiver [22]. Several works have investigated the impact of HWIs on massive MIMO systems [21], [23]- [26]. The impact of phase noise originating from free-running oscillators on the downlink performance of massive MIMO systems was studied in [23] for different linear precoder designs. Constant envelope precoding for massive MIMO was studied in [24] and [25] with the objective of avoiding distortions caused by power amplifier nonlinearities at the transmitter. The impact of the aggregate effects of several HWIs originating from different sources on massive MIMO systems was studied in [21] by modelling the residual impairments remaining after compensation as additive distortion noises [22]. The authors in [26] presented closed-form expressions for the achievable user rates in uplink massive MIMO systems for a general residual HWI model including both multiplicative phase noise and additive distortion noise. The aforementioned works demonstrated that HWIs can severely limit the performance of massive MIMO systems. Thereby, a crucial role is played by the degradation caused by phase noise to the quality of the CSI estimates needed for precoder design. On the one hand, phase noise causes the CSI estimates to become outdated more quickly. On the other hand, it may cause a loss of orthogonality of the pilot sequences employed by the different users

in a cell for uplink training. To overcome the latter effect, so-called temporally orthogonal (TO) and spatially orthogonal (SO) pilot sequences were investigated in [26]. Furthermore, the impact of the number of local oscillators (LOs) employed at the massive MIMO BS on performance in the presence of phase noise was studied in [23], [26], while the effect of HWIs on full-duplex massive MIMO relaying was considered in [27].

Communication secrecy aspects are not adequately considered in existing works studying the impact of HWIs in the context of massive MIMO system design [21], [23]- [27]. However, if communication secrecy is considered, an additional challenge arises: Whereas the legitimate user of the system will likely employ low-cost equipment giving rise to HWIs, the eavesdropper is expected to employ high-quality HWI-free equipment. This disparity in equipment quality was not considered in the related work on physical layer security [7]- [20] nor in the related work on HWIs [21], [23]- [26] and necessitates the development of a new analysis and design framework. For example, NS AN precoding, which was widely used to enhance the achievable secrecy rate of massive MIMO systems [7], [15], [29], becomes ineffective in the presence of phase noise.

Motivated by the above considerations, in this paper, we present the first study of physical layer security in hardware constrained massive MIMO systems. Thereby, we focus on the downlink and adopt for the legitimate links the generic residual HWI model from [22], [26], which includes the effects of multiplicative phase noise and additive distortion noise at the BS and the users. As a worst-case scenario, the eavesdropper is assumed to employ ideal hardware. Our main contributions are summarized as follows.

- For the adopted generic residual HWI model, we derive a tight lower bound for the ergodic secrecy rate achieved by a downlink user when MF data precoding is employed at the massive MIMO BS. The derived bound provides insight into the impact of various system and channel parameters, such as the phase noise variance, the additive distortion noise parameters, the AN precoder design, the amount of power allocated to the AN, the pilot sequence design, the number of deployed LOs, and the number of users, on the ergodic secrecy rate.
- As conventional NS AN precoding is sensitive to phase noise, we propose a novel generalized NS (G-NS) AN precoding design, which mitigates the AN leakage caused to the legitimate user in the presence of phase noise at the expense of a reduction of the available spatial degrees of freedom. The proposed method leads to significant performance gains,

especially in systems with large numbers of antennas at the BS.

- We generalize the SO and TO pilot sequence designs from [26] to orthogonal pilot sequences with arbitrary numbers of non-zero elements. Although SO sequences, which have no zero elements, are preferable for small phase noise variances, sequence designs with zero elements become beneficial in the presence of strong phase noise.
- Our analytical and numerical results reveal that while HWIs in general degrade the achievable secrecy rate, the proposed countermeasures are effective in limiting this degradation. Furthermore, surprisingly, there are cases when the additive distortion noise at the BS is beneficial for the secrecy performance as it can have a similar effect as AN.

The remainder of this paper is organized as follows. In Section II, the models for uplink training and downlink data transmission in the considered massive MIMO system with imperfect hardware are presented. In Section III, we derive a lower bound on the achievable ergodic secrecy rate and introduce the proposed G-NS AN precoder design. In Section IV, the impact of the various system and channel parameters on the secrecy performance is investigated based on the derived lower bound. In Section V, the achievable secrecy rate is studied via simulation and numerical evaluation of the derived analytical expressions. Conclusions are drawn in Section VI.

Notation: Superscripts T and H stand for the transpose and conjugate transpose, respectively. \mathbf{I}_N is the N -dimensional identity matrix. The expectation operation and the variance of a random variable are denoted by $\mathbb{E}[\cdot]$ and $\text{var}[\cdot]$, respectively. $\text{diag}\{\mathbf{x}\}$ denotes a diagonal matrix with the elements of vector \mathbf{x} on the main diagonal. $\text{tr}\{\cdot\}$ denotes the trace of a matrix. $\mathbb{C}^{m \times n}$ represents the space of all $m \times n$ matrices with complex-valued elements. $\mathbf{x} \sim \mathbb{CN}(\mathbf{0}_N, \mathbf{\Sigma})$ denotes a circularly symmetric complex Gaussian vector $\mathbf{x} \in \mathbb{C}^{N \times 1}$ with zero mean and covariance matrix $\mathbf{\Sigma}$. $[\mathbf{A}]_{kl}$ denotes the element in the k^{th} row and l^{th} column of matrix \mathbf{A} . $[x]^+ = \max\{x, 0\}$ and $\lfloor x \rfloor$ stands for the largest integer no greater than x . Finally, $|\mathcal{S}|$ represents the cardinality of set \mathcal{S} .

II. SYSTEM AND CHANNEL MODELS

The considered massive MIMO system model comprises an N -antenna BS, K single-antenna mobile terminals (MTs), and an N_E -antenna eavesdropper. The eavesdropper is passive in order to hide its existence from the BS and the MTs. Similar to [21], [26], we assume that after proper compensation the residual HWIs manifest themselves at the BS and the MTs in the

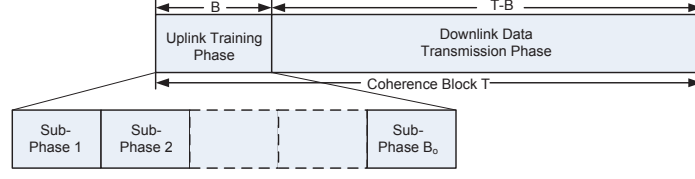


Fig. 1. Uplink training and downlink transmission phase.

form of 1) multiplicative phase noises at transmitter and receiver, 2) transmit and receive power dependent distortion noises at transmitter and receiver, respectively, and 3) amplified thermal noise at the receiver. The impact of this general HWI model on uplink training and downlink data transmission is investigated in Sections II-A and II-B, respectively, and the signal model for the eavesdropper is presented in Section II-C. In this paper, we consider a single-cell system. This allows us to concentrate on the main focus of our work, i.e., studying the impact of HWIs on physical layer security in massive MIMO systems. Naturally, the obtained results can serve as a benchmark for multi-cell massive MIMO systems with HWIs.

A. Uplink Pilot Training under HWIs

In massive MIMO systems, the CSI is usually acquired via uplink training by exploiting the channel reciprocity between uplink and downlink in time-division duplex (TDD) mode [3], [5]. Here, we assume that the first B symbol intervals of the coherence time, which comprises T symbol intervals, are used for uplink training. Thereby, we split the training phase into B_o sub-phases of lengths B_b , $1 \leq b \leq B_o$, where $\sum_{b=1}^{B_o} B_b = B$, cf. Fig. 1. Furthermore, the K MTs are assigned to B_o disjunct sets \mathcal{S}_b , $1 \leq b \leq B_o$, with $|\mathcal{S}_b| \leq B_b$ and $\sum_{b=1}^{B_o} |\mathcal{S}_b| = K$. In training sub-phase b , the MTs in set \mathcal{S}_b emit mutually orthogonal pilot sequences $\omega_k = [\omega_k(1), \omega_k(2), \dots, \omega_k(B_b)]^T \in \mathbb{C}^{B_b \times 1}$, $k \in \mathcal{S}_b$, for which we assume a per-pilot power constraint $|\omega_k(t)|^2 = p_\tau, \forall k, t$, whereas all MTs $k \notin \mathcal{S}_b$ are silent¹. For larger values of B_b , the total energy of the pilot sequences is larger but, as will be shown later, the loss of orthogonality caused by phase noise becomes also more pronounced. Hence, B_b or equivalently B_o (assuming a fixed B) should be optimized for maximization of the secrecy rate. We note that the proposed pilot

¹We adopt a per-pilot power constraint, as in practice, systems are peak power limited, e.g., [4], [5], [21]. We note that some of our results and conclusions may change if an average power constraint for the pilot sequence was employed.

design is a generalization of the SO and TO pilot designs proposed in [23], [26] which result as special cases for $B_o = 1$ and $B_o = B$, respectively.

In symbol interval $t \in \mathcal{T}_b$, where \mathcal{T}_b denotes the set of symbol intervals in training sub-phase b , $1 \leq b \leq B_o$, the received uplink vector $\mathbf{y}^{\text{UL}}(t) \in \mathbb{C}^{N \times 1}$ at the BS is given by

$$\mathbf{y}^{\text{UL}}(t) = \sum_{k \in \mathcal{S}_b} \mathbf{\Theta}_k(t) \mathbf{g}_k(\omega_k(t) + \eta_{t,k}^{\text{MT}}(t)) + \boldsymbol{\eta}_r^{\text{BS}}(t) + \boldsymbol{\xi}^{\text{UL}}(t). \quad (1)$$

Here, the channel vector of the k^{th} MT, $\mathbf{g}_k \sim \mathbb{CN}(\mathbf{0}_N, \beta_k \mathbf{I}_N)$, is modelled as block Rayleigh fading, where β_k denotes the path-loss. Thereby, \mathbf{g}_k is assumed to be constant during coherence time T and change independently afterwards. In (1), the terms $\mathbf{\Theta}_k(t)$, $\eta_{t,k}^{\text{MT}}(t)$, $\boldsymbol{\eta}_r^{\text{BS}}(t)$, and $\boldsymbol{\xi}^{\text{UL}}(t)$ characterize the HWIs affecting the uplink training phase and are explained in detail in the following:

1) Phase noise: Matrix

$$\mathbf{\Theta}_k(t) = \text{diag} \left(e^{j\theta_k^1(t)} \mathbf{1}_{1 \times N/N_o}, \dots, e^{j\theta_k^{N_o}(t)} \mathbf{1}_{1 \times N/N_o} \right) \in \mathbb{C}^{N \times N} \quad (2)$$

models the phase noise originating from the free-running LOs equipped at the BS and the MTs [23]. Thereby, we assume that at the BS each group of $N/N_o \in \mathbb{Z}$ antennas is connected to one free-running LO. $\theta_k^l(t) = \psi_l(t) + \phi_k(t)$ is the phase noise that distorts the link between the l^{th} LO at the BS and the k^{th} MT. Adopting the discrete-time Wiener phase noise model [23], in time interval t , the phase noises at the l^{th} LO of the BS and the k^{th} MT are modelled as $\psi_l(t) \sim \mathbb{CN}(\psi_l(t-1), \sigma_\psi^2)$, $1 \leq l \leq N_o$, and $\phi_k(t) \sim \mathbb{CN}(\phi_k(t-1), \sigma_\phi^2)$, $1 \leq k \leq K$, where σ_ψ^2 and σ_ϕ^2 are the phase noise (increment) variances at the BS and the MTs, respectively.

2) Distortion noise: $\eta_{t,k}^{\text{MT}}(t) \in \mathbb{C}$ and $\boldsymbol{\eta}_r^{\text{BS}}(t) \in \mathbb{C}^{N \times 1}$ model the additive distortion noises at the k^{th} MT and the BS, respectively, which originate from residual effects after compensation of HWIs such as power amplifier non-linearities at the transmitter, quantization noise in the analog-to-digital converters (ADCs) at the receiver, etc. [21]. Distortion noise is modeled as a Gaussian distributed random process in the literature [21], [22]. This model has been experimentally verified in [28]. Furthermore, at each antenna, the distortion noise power is proportional to the corresponding signal power, i.e., $\eta_{t,k}^{\text{MT}}(t) \sim \mathbb{CN}(0, v_{t,k}^{\text{MT}})$ and $\boldsymbol{\eta}_r^{\text{BS}}(t) \sim \mathbb{CN}(\mathbf{0}_N, \boldsymbol{\Upsilon}_r^{\text{BS}})$, where

$$v_{t,k}^{\text{MT}} = \kappa_t^{\text{MT}} \mathbb{E}[|\omega_k(t)|^2] \quad \text{and} \quad \boldsymbol{\Upsilon}_r^{\text{BS}} = \kappa_r^{\text{BS}} \sum_{k=1}^K \mathbb{E}[|\omega_k(t)|^2] \mathbf{R}_k^{\text{diag}}. \quad (3)$$

Here, $\mathbf{R}_k^{\text{diag}} = \text{diag}(|g_k^1|^2, \dots, |g_k^N|^2)$, where g_k^i denotes the i^{th} element of \mathbf{g}_k , and parameters $\kappa_t^{\text{MT}}, \kappa_r^{\text{BS}} > 0$ denote the ratio between the additive distortion noise variance and the signal power and are measures for the severity of the residual HWIs.

3) Amplified thermal noise: $\xi^{\text{UL}}(t) \sim \mathbb{CN}(\mathbf{0}_N, \xi^{\text{UL}} \mathbf{I}_N)$ models the thermal noise amplified by the low noise amplifier and other components such as mixers at the receiver [26]. Therefore, the variance of this noise is generally larger than that of the actual thermal noise σ_n^2 , i.e., $\xi^{\text{UL}} > \sigma_n^2$.

For channel estimation, we collect the signal vectors received during the b^{th} training phase in vector $\psi_b = [(\mathbf{y}^{\text{UL}}(\overline{B}_{b-1} + 1))^T, \dots, (\mathbf{y}^{\text{UL}}(\overline{B}_b))^T]^T \in \mathbb{C}^{B_b N \times 1}$, $b = 1, \dots, B_o$, where $\overline{B}_b \triangleq \sum_{i=1}^b B_i$ and $\overline{B}_0 = 0$, and define the effective channel vector at time t as $\mathbf{g}_k(t) = \Theta_k(t) \mathbf{g}_k$. With these definitions, the linear minimum mean-square error (LMMSE) estimate of the channel of MT $k \in \mathcal{S}_b$ at time $t \in \{B + 1, \dots, T\}$ (i.e., during the data transmission phase) can be written as [26]

$$\hat{\mathbf{g}}_k(t) = \mathbb{E}[\mathbf{g}_k(t) \psi_b^H] (\mathbb{E}[\psi_b \psi_b^H])^{-1} \psi_b = (\beta_k \boldsymbol{\omega}_k^H \Theta_{\sigma(t)}^b \boldsymbol{\Sigma}_b^{-1} \otimes \mathbf{I}_N) \psi_b, \quad (4)$$

where

$$\Theta_{\sigma(t)}^b = \text{diag} \left(e^{-\frac{\sigma_\psi^2 + \sigma_\phi^2}{2} |t - \overline{B}_{b-1} - 1|}, \dots, e^{-\frac{\sigma_\psi^2 + \sigma_\phi^2}{2} |t - \overline{B}_b|} \right) \text{ and } \boldsymbol{\Sigma}_b = \sum_{k \in \mathcal{S}_b} \beta_k (\mathbf{W}_k^b + \mathbf{U}_k^b) + \xi^{\text{UL}} \mathbf{I}_{B_b}. \quad (5)$$

Here, we adopted the definitions $[\mathbf{W}_k^b]_{i,j} = \omega_k(i) \omega_k^*(j) e^{-\frac{\sigma_\psi^2 + \sigma_\phi^2}{2} |i-j|}$, $i, j \in \{1, \dots, B_b\}$, and $\mathbf{U}_k^b = (\kappa_t^{\text{MT}} + \kappa_r^{\text{BS}}) p_\tau \mathbf{I}_{B_b}$.

Considering the properties of LMMSE estimation, the channel can be decomposed as $\mathbf{g}_k(t) = \hat{\mathbf{g}}_k(t) + \mathbf{e}_k(t)$, $t = 1, \dots, B$, where $\hat{\mathbf{g}}_k(t)$ denotes the LMMSE channel estimate given in (4) and $\mathbf{e}_k(t)$ represents the estimation error. $\hat{\mathbf{g}}_k(t)$ and $\mathbf{e}_k(t)$ are mutually uncorrelated and have zero mean [21, Theorem 1]. The error covariance matrix is given by

$$\mathbb{E}[\mathbf{e}_k(t) \mathbf{e}_k^H(t)] = \beta_k (1 - \beta_k \boldsymbol{\omega}_k^H \Theta_{\sigma(t)}^b \boldsymbol{\Sigma}_b^{-1} \Theta_{\sigma(t)}^b \boldsymbol{\omega}_k) \mathbf{I}_N. \quad (6)$$

Eqs. (4)-(6) reveal that for $|\mathcal{S}_b| > 1$ and $\sigma_\psi^2, \sigma_\phi^2 > 0$, the channel estimate of the k^{th} MT contains contributions from channels of other MTs emitting their pilots in the same training sub-phase, i.e., the pilots interfere with each other although the emitted pilot sequences are orthogonal. This loss of orthogonality at the receiver is introduced by the phase noise via matrices $\Theta_{\sigma(t)}^b$ and \mathbf{W}_k^b , and can be avoided only by enforcing that in any sub-phase only one MT emits its pilots, i.e., $|\mathcal{S}_b| = 1$, $1 \leq b \leq B_o$. In particular, for the case $|\mathcal{S}_b| = B_b = 1$, $1 \leq b \leq B_o = B$, for symbol

interval $t \in \{B+1, \dots, T\}$, the LMMSE channel estimate of MT $k \in \mathcal{S}_b$ can be simplified to

$$\hat{\mathbf{g}}_k(t) = \frac{p_\tau \beta_k e^{-\frac{\sigma_\psi^2 + \sigma_\phi^2}{2}|t-b|}}{p_\tau \beta_k (1 + \kappa_t^{\text{MT}} + \kappa_r^{\text{BS}}) + \xi^{\text{UL}}} \mathbf{y}^{\text{UL}}(b), \quad (7)$$

with $\mathbf{y}^{\text{UL}}(t)$ given in (1), i.e., $\hat{\mathbf{g}}_k(t)$ is not affected by the channels of other MTs despite the phase noise. The corresponding error covariance matrix simplifies to

$$\mathbb{E}[\mathbf{e}_k(t) \mathbf{e}_k^H(t)] = \beta_k \left(1 - \frac{p_\tau \beta_k e^{-(\sigma_\psi^2 + \sigma_\phi^2)|t-b|}}{p_\tau \beta_k (1 + \kappa_t^{\text{MT}} + \kappa_r^{\text{BS}}) + \xi^{\text{UL}}} \right) \mathbf{I}_N. \quad (8)$$

Eqs. (4) and (7) reveal that the channel estimate depends on time t . As a consequence, ideally, the channel-dependent data and AN precoders employed for downlink transmission should be recomputed in every symbol interval of the data transmission phase, in accordance with the corresponding channel estimate, which entails a high computational complexity. Therefore, in the following, we assume that data and AN precoders are computed based on the channel estimate for one symbol interval t_0 (e.g., $t_0 = B+1$) and are then employed for precoding during the entire data transmission phase, i.e., for $t \in \{B+1, \dots, T\}$. For notational conciseness, we denote the corresponding channel estimate by $\hat{\mathbf{g}}_k = \hat{\mathbf{g}}_k(t_0)$, $k = \{1, \dots, K\}$.

B. Downlink Data Transmission and Linear Precoding

Assuming channel reciprocity, during the downlink data transmission phase, the received signal at the k^{th} MT in time interval $t \in \{B+1, \dots, T\}$ is given by

$$y_k^{\text{DL}}(t) = \mathbf{g}_k^H \boldsymbol{\Theta}_k^H(t) (\mathbf{x} + \boldsymbol{\eta}_t^{\text{BS}}(t)) + \eta_{r,k}^{\text{MT}}(t) + \xi_k^{\text{DL}}(t). \quad (9)$$

In (9), similar to the uplink, $\boldsymbol{\eta}_t^{\text{BS}}(t) \sim \mathbb{CN}(\mathbf{0}_N, \boldsymbol{\Upsilon}_t^{\text{BS}})$ and $\eta_{r,k}^{\text{MT}}(t) \sim \mathbb{CN}(0, v_{r,k}^{\text{MT}}(t))$ denote the downlink distortion noise [21] at the BS and the k^{th} MT, respectively, where

$$\boldsymbol{\Upsilon}_t^{\text{BS}} = \kappa_t^{\text{BS}} \text{diag}(X_{11}, \dots, X_{NN}) \quad \text{and} \quad v_{r,k}^{\text{MT}}(t) = \kappa_r^{\text{MT}} \mathbf{g}_k^H(t) \mathbf{X} \mathbf{g}_k(t) \quad (10)$$

with $\mathbf{X} = \mathbb{E}[\mathbf{x} \mathbf{x}^H]$ and $X_{ii} = [\mathbf{X}]_{ii}$, $i = 1, \dots, N$. Furthermore, $\xi_k^{\text{DL}}(t) \sim \mathbb{CN}(0, \xi^{\text{DL}})$ represents the amplified thermal noise at the k^{th} MT. For simplicity of presentation, we assume that parameters κ_t^{BS} , κ_r^{MT} , and ξ^{DL} are identical for all MTs.

The downlink transmit signal $\mathbf{x} \in \mathbb{C}^{N \times 1}$ in (9) is modeled as

$$\mathbf{x} = \sqrt{p} \mathbf{F} \mathbf{s} + \sqrt{q} \mathbf{A} \mathbf{z} \in \mathbb{C}^{N \times 1}, \quad (11)$$

where the data symbol vector $\mathbf{s} \in \mathbb{C}^{K \times 1}$ and the AN vector $\mathbf{z} \in \mathbb{C}^{L \times 1}$, $L \leq N$, are multiplied by data precoder $\mathbf{F} \in \mathbb{C}^{N \times K}$ and AN precoder $\mathbf{A} \in \mathbb{C}^{N \times L}$, respectively. As we assume that the eavesdropper's CSI is not available at the BS, AN is injected to degrade the eavesdropper's ability to decode the data intended for the MTs [7], [15], [29]. Thereby, it is assumed that the components of \mathbf{s} and \mathbf{z} are independent and identically distributed (i.i.d.) circularly symmetric complex Gaussian (CSCG) random variables, i.e., $\mathbf{s} \sim \mathbb{CN}(\mathbf{0}_K, \mathbf{I}_K)$ and $\mathbf{z} \sim \mathbb{CN}(\mathbf{0}_L, \mathbf{I}_L)$. In (11), $p = \phi P_T/K$ and $q = (1 - \phi)P_T/L$ denote the power assigned to each MT and each column of the AN, where P_T is the total power budget and $\phi \in (0, 1]$ can be used to strike a balance between data transmission and AN emission. Combining (11) and (9) we obtain

$$y_k^{\text{DL}}(t) = \sqrt{p} \mathbf{g}_k^H(t) \mathbf{f}_k s_k + \sum_{l \neq k}^K \sqrt{p} \mathbf{g}_k^H(t) \mathbf{f}_l s_l + \sqrt{q} \mathbf{g}_k^H(t) \mathbf{A} \mathbf{z} + \mathbf{g}_k^H(t) \boldsymbol{\eta}_t^{\text{BS}}(t) + \eta_{r,k}^{\text{MT}}(t) + \xi_k^{\text{DL}}(t), \quad (12)$$

where s_k and \mathbf{f}_k denote the k^{th} element of \mathbf{s} and the k^{th} column of matrix \mathbf{F} , respectively.

C. Signal Model of the Eavesdropper

We assume that the eavesdropper is silent during the training phase, i.e., for $t \in \{1, \dots, B\}$, and eavesdrops the signal intended for MT k during the data transmission phase, i.e., for $t \in \{B + 1, \dots, T\}$. Let \mathbf{G}_E denote the channel matrix between the BS and the eavesdropper with i.i.d. zero-mean complex Gaussian elements having variance β_E , where β_E is the path-loss between the BS and the eavesdropper. Since the capabilities of the eavesdropper are not known at the BS, we make worst-case assumptions regarding the hardware and signal processing capabilities of the eavesdropper with respect to communication secrecy. In particular, we assume the received signal at the eavesdropper at time $t \in \{B + 1, \dots, T\}$ can be modelled as

$$\mathbf{y}_E(t) = \mathbf{G}_E^H \boldsymbol{\Psi}^H(t) (\mathbf{x} + \boldsymbol{\eta}_t^{\text{BS}}(t)) \in \mathbb{C}^{N_E \times 1}, \quad (13)$$

where $\boldsymbol{\Psi}(t) = \text{diag} \left(e^{j\psi_1(t)} \mathbf{1}_{1 \times N/N_o}^T, \dots, e^{j\psi_{N_o}(t)} \mathbf{1}_{1 \times N/N_o}^T \right)$. Thereby, we assumed that the eavesdropper employs high-quality hardware such that the only HWIs are the phase noise and the additive distortion noise at the BS. Eq. (13) also implies that the thermal noise at the eavesdropper is negligibly small [7], [15], [29]. Furthermore, we assume that the eavesdropper has perfect CSI, i.e., it can perfectly estimate the effective eavesdropper channel matrix $\mathbf{G}_E^H \boldsymbol{\Psi}^H(t)$, and can perfectly decode and cancel the interference caused by all MTs except for the MT of interest

[7], [15], [29]. These worst-case assumptions lead to an upper bound on the ergodic capacity of the eavesdropper in time interval t given by

$$C_E(t) = \mathbb{E}[\log_2(1 + \gamma_E(t))] \quad (14)$$

where

$$\gamma_E(t) = p\mathbf{g}_E^k(t) (\mathbf{G}_E^H \mathbf{\Psi}^H(t)(q\mathbf{A}\mathbf{A}^H + \mathbf{\Upsilon}_t^{\text{BS}})\mathbf{\Psi}(t)\mathbf{G}_E)^{-1} (\mathbf{g}_E^k(t))^H \quad (15)$$

and $\mathbf{g}_E^k(t) = \mathbf{f}_k^H \mathbf{\Psi}(t)\mathbf{G}_E$. We note that since we assumed that the thermal noise at the receiver of the eavesdropper is negligible, $\gamma_E(t)$, and consequently $C_E(t)$, are independent of the path-loss of the eavesdropper, β_E . Furthermore, we observe from (15) that the additive distortion noise at the BS, $\boldsymbol{\eta}_t^{\text{BS}}(t)$, with co-variance matrix $\mathbf{\Upsilon}_t^{\text{BS}}$, affects the ergodic capacity of the eavesdropper in a similar manner as the injected AN.

III. ACHIEVABLE ERGODIC SECRECY RATE IN THE PRESENCE OF HWIs

In this section, we analyze the achievable ergodic secrecy rate of a massive MIMO system employing non-ideal hardware. To this end, we derive a lower bound on the achievable ergodic secrecy rate in Section III-A, and present an asymptotic analysis for the downlink data rate of the legitimate MTs when MF data precoding is adopted by the BS in Section III-B. In Section III-C, a generalized NS AN precoder is proposed to avoid the AN leakage caused by phase noise for conventional NS AN precoding. Finally, in Section III-D, a simple closed-form upper bound for the eavesdropper's capacity for the new AN precoder is presented.

A. Lower Bound on Achievable Ergodic Secrecy Rate

In this paper, we assume that communication delay is tolerable and coding over many independent channel realizations is possible. Hence, we adopt the ergodic secrecy rate achieved by a given MT as performance metric [29].

Before analyzing the secrecy rate, we first employ [26, Lemma 1] to obtain a lower bound on the achievable rate for the multiple-input single-output (MISO) phase noise channel given by (9). In particular, the achievable rate of the k^{th} MT, $1 \leq k \leq K$, in symbol interval $t \in \{B+1, \dots, T\}$ is lower bounded by

$$R_k(t) \geq \underline{R}_k(t) = \log_2(1 + \gamma_k(t)), \quad (16)$$

with SINR $\gamma_k(t) =$

$$\frac{p \left| \mathbb{E} [\mathbf{g}_k^H(t) \mathbf{f}_k] \right|^2}{\sum_{l=1}^K p \mathbb{E} \left[\left| \mathbf{g}_k^H(t) \mathbf{f}_l \right|^2 \right] - p \left| \mathbb{E} [\mathbf{g}_k^H(t) \mathbf{f}_k] \right|^2 + \mathbb{E} [\mathbf{g}_k^H(t) (q \mathbf{A} \mathbf{A}^H + \mathbf{\Upsilon}_t^{\text{BS}}) \mathbf{g}_k(t)] + \mathbb{E} [v_{k,r}^{\text{MT}}(t)] + \xi^{\text{DL}}}. \quad (17)$$

The expectation operator in (17) is taken with respect to channel vectors, \mathbf{g}_k , as well as the phase noise processes, $\psi_l(t)$ and $\phi_k(t)$. The rate given in (16) is achievable because: 1) The SINR in (17) is underestimated by assuming that only the BS has channel estimates, while the MTs only know the mean of the effective channel gain $\left| \mathbb{E} [\mathbf{g}_k^H(t) \mathbf{f}_k] \right|$ and employ it for signal detection. The deviation from the average effective channel gain is treated as Gaussian noise having variance $\mathbb{E} \left[\left| \mathbf{g}_k^H(t) \mathbf{f}_k \right|^2 \right] - \left| \mathbb{E} [\mathbf{g}_k^H(t) \mathbf{f}_k] \right|^2$, cf. [5], [7], [15], [20]; 2) Following [26, Lemma 1], we treat the multiuser interference and distortion noises as independent Gaussian noises, which is a worst-case assumption for the calculation of the mutual information. The tightness of the bound will be confirmed in Section V. Based on (16), we provide a lower bound on the achievable ergodic secrecy rate of the k^{th} MT, $1 \leq k \leq K$, in the following Lemma.

Lemma 1: The achievable ergodic secrecy rate of the k^{th} MT, $1 \leq k \leq K$, is bounded below by

$$R_k^{\text{sec}} \geq \underline{R}_k^{\text{sec}} = \frac{1}{T} \sum_{t \in \{B+1, \dots, T\}} [\underline{R}_k(t) - C_E(t)]^+, \quad (18)$$

where $\underline{R}_k(t)$, $1 \leq k \leq K$, is the lower bound of the achievable ergodic rate of the k^{th} MT given in (16) and $C_E(t)$ is the ergodic capacity between the BS and the eavesdropper given in (14).

Proof: Please refer to Appendix B. ■

The sum in (18) is over the $T - B$ time slots used for data transmission. Motivated by the coding scheme for the non-secrecy case in [30], a similar coding scheme that supports the secrecy rate given in (18) is described as follows. For a given $t \in \{B + 1, \dots, T\}$, the statistics of $\mathbf{g}_k(t)$ in (17) given the estimate $\hat{\mathbf{g}}_k$ are identical across all coherence intervals and the corresponding channel realizations are i.i.d. Hence, we employ $T - B$ parallel channel codes for each MT; one code for each time $t \in \{B + 1, \dots, T\}$, i.e., the t^{th} channel code is employed across the t^{th} time slots of multiple coherence intervals. Then, at each MT, the t^{th} received symbols across the multiple coherence intervals are jointly decoded [30]. With this coding strategy the ergodic secrecy rate given in (18) is achieved provided the parallel codes span sufficiently many (ideally

an infinite number) of independent channel realizations \mathbf{g}_k and phase noise samples $\psi_l(t)$ and $\phi_k(t)$.

B. Asymptotic Analysis of Achievable Rate for MF Precoding

In this subsection, we analyze the lower bound on the achievable ergodic rate of the k^{th} MT, $1 \leq k \leq K$, in (16) in the asymptotic limit $N, K \rightarrow \infty$ for fixed ratio $\beta = K/N$. Thereby, we adopt MF precoding at the BS, i.e., $\mathbf{f}_k = \hat{\mathbf{g}}_k / \|\hat{\mathbf{g}}_k\|$, as is commonly done for massive MIMO systems because of complexity concerns for more sophisticated precoder designs. In the following Lemma, we provide a closed-form expression for the gain of the desired signal.

Lemma 2: For MF precoding at the BS, the numerator of (17) reflecting the gain of the desired signal at MT $k \in \mathcal{S}_b$, $1 \leq b \leq B_o$, can be expressed as

$$\mathbb{E} [\mathbf{g}_k^H \mathbf{\Theta}_k^H(t) \mathbf{f}_k] = \sqrt{\beta_k N \lambda_k} \cdot e^{-\frac{\sigma_\psi^2 + \sigma_\phi^2}{2} |t - t_0|}, \text{ where } \lambda_k = \beta_k \boldsymbol{\omega}_k^H \mathbf{\Theta}_{\sigma(t_0)}^b \boldsymbol{\Sigma}_b^{-1} \mathbf{\Theta}_{\sigma(t_0)}^b \boldsymbol{\omega}_k. \quad (19)$$

Proof: Please refer to Appendix C. ■

The term $e^{-\frac{\sigma_\psi^2 + \sigma_\phi^2}{2} |t - t_0|}$ in (19) reveals the impact of the accumulated phase noise from the time of channel estimation, t_0 , to the time of data transmission, t , on the received signal strength at MT k . On the other hand, the phase noise within the training phase affects λ_k , and consequently the received signal strength, via $\mathbf{\Theta}_{\sigma(t_0)}^b$ and $\boldsymbol{\Sigma}_b$, cf. (5), when multiple pilot sequences are simultaneously emitted in a given training sub-phase. In contrast, when TO pilots are adopted, i.e., only a single user emits pilots in each training sub-phase and $B_b = 1$, $1 \leq b \leq B$, λ_k in (19) reduces to $\lambda_k = \frac{p_\tau \beta_k}{p_\tau \beta_k (1 + \kappa_t^{\text{MT}} + \kappa_r^{\text{BS}}) + \xi^{\text{UL}}}$ and is not affected by the phase noise.

Next, an expression for the multiuser interference power in the first term of the denominator of (17) is derived.

Lemma 3: When MF precoding is adopted at the BS, the power of the multiuser interference caused by the signal intended for the l^{th} MT, $l \neq k$, at MT $k \in \mathcal{S}_b$, $1 \leq b \leq B_o$, is given by

$$\mathbb{E} [|\mathbf{g}_k^H \mathbf{\Theta}_k^H(t) \mathbf{f}_l|^2] = \left(\beta_k + \left(X_{k,l}^{(1)} + X_{k,l}^{(2)} + X_{k,l}^{(3)} \right) \left(\frac{1 - \epsilon}{N_o} + \epsilon \right) \right), \text{ if } l \in \mathcal{S}_b \quad (20)$$

and by β_k otherwise. Here, $\epsilon = e^{-\sigma_\psi^2 |t - t_0|}$, $X_{k,l}^{(1)} = \frac{\beta_k^2 \boldsymbol{\omega}_l^H \mathbf{\Theta}_{\sigma(t_0)}^b \boldsymbol{\Sigma}_b^{-1} \mathbf{U}_k^b \boldsymbol{\Sigma}_b^{-1} \mathbf{\Theta}_{\sigma(t_0)}^b \boldsymbol{\omega}_l}{\boldsymbol{\omega}_l^H \mathbf{\Theta}_{\sigma(t_0)}^b \boldsymbol{\Sigma}_b^{-1} \mathbf{\Theta}_{\sigma(t_0)}^b \boldsymbol{\omega}_l}$, $X_{k,l}^{(2)} = \frac{N}{N_o} \cdot \frac{\beta_k^2 \boldsymbol{\omega}_l^H \mathbf{\Theta}_{\sigma(t_0)}^b \boldsymbol{\Sigma}_b^{-1} \mathbf{W}_k^b \boldsymbol{\Sigma}_b^{-1} \mathbf{\Theta}_{\sigma(t_0)}^b \boldsymbol{\omega}_l}{\boldsymbol{\omega}_l^H \mathbf{\Theta}_{\sigma(t_0)}^b \boldsymbol{\Sigma}_b^{-1} \mathbf{\Theta}_{\sigma(t_0)}^b \boldsymbol{\omega}_l}$, and $X_{k,l}^{(3)} = N \left(1 - \frac{1}{N_o} \right) \cdot \frac{\left| \beta_k \boldsymbol{\omega}_k^H \mathbf{\Theta}_{\sigma(t_0)}^b \boldsymbol{\Sigma}_b^{-1} \mathbf{\Theta}_{\sigma(t_0)}^b \boldsymbol{\omega}_l \right|^2}{\boldsymbol{\omega}_l^H \mathbf{\Theta}_{\sigma(t_0)}^b \boldsymbol{\Sigma}_b^{-1} \mathbf{\Theta}_{\sigma(t_0)}^b \boldsymbol{\omega}_l}$.

Proof: Please refer to Appendix D. ■

Lemma 3 confirms that when the number of BS antennas is sufficiently large, i.e., $N \rightarrow \infty$, as long as $l \notin \mathcal{S}_b$, the impact of the multiuser interference from the l^{th} MT vanishes, as is commonly assumed in the massive MIMO literature, e.g. [3], [4]. However, the same is not true for MTs that emit pilots in the same training sub-phase as MT k , i.e., MTs $l \in \mathcal{S}_b$. Because of the impairment incurred by the phase noise during the training phase, the interference power of these MTs grows linearly with N and does not vanish compared to the strength of the desired signal in (19) in the limit of $N \rightarrow \infty$.

Furthermore, for the summand with $l = k$ in the sum in the first term of the denominator of (17), we obtain $\mathbb{E} \left[\left| \mathbf{g}_k^H \mathbf{\Theta}_k^H(t) \mathbf{f}_k \right|^2 \right] =$

$$\mathbb{E} \left[\text{tr} \left(\mathbf{g}_k(t_0) \mathbf{g}_k^H(t_0) \mathbf{\Psi}_{t_0}^H(t) \frac{\hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H}{\|\hat{\mathbf{g}}_k\|^2} \mathbf{\Psi}_{t_0}(t) \right) \right] = \beta_k + \beta_k(N-1)\lambda_k \left(\frac{1-\epsilon}{N_o} + \epsilon \right), \quad (21)$$

where $k \in \mathcal{S}_b$ and $\mathbf{\Psi}_{t_0}(t)$ is defined in Appendix C. The last equality in (21) is obtained by applying Theorem 1 in Appendix A [34]. The variance of the gain of the desired signal, $\mathbf{g}_k^H \mathbf{\Theta}_k^H(t) \mathbf{f}_k$, is obtained by subtracting the right hand side of (21) from the square of the right hand side of (19).

The two terms in the denominator of (17) originating from the HWIs at the BS and the k^{th} MT, i.e., $\eta_t^{\text{BS}}(t)$ and $\eta_{r,k}^{\text{MT}}(t)$, respectively, can be calculated as

$$\mathbb{E} \left[\left| \mathbf{g}_k^H \mathbf{\Theta}_k^H(t) \mathbf{\Upsilon}_t^{\text{BS}} \mathbf{\Theta}_k(t) \mathbf{g}_k \right| \right] = \beta_k \kappa_t^{\text{BS}} P_T \quad \text{and} \quad \mathbb{E} \left[v_{r,k}^{\text{MT}}(t) \right] = \beta_k \kappa_r^{\text{MT}} P_T. \quad (22)$$

Substituting the results in (19)-(22) into (17), we obtain the received SINR at MT $k \in \mathcal{S}_b$ in symbol interval t as

$$\gamma_k(t) = \frac{pN\beta_k\bar{\lambda}_k}{p\beta_k(a_k + c_k) + qL_{\text{AN}}^k + \beta_k(\kappa_t^{\text{BS}} + \kappa_r^{\text{MT}})P_T + \xi^{\text{DL}}}, \quad (23)$$

with

$$a_k = \sum_{l \in \mathcal{S}_b} \left(1 + \left(X_{k,l}^{(1)} + X_{k,l}^{(2)} + X_{k,l}^{(3)} \right) \left(\frac{1-\epsilon}{N_o} + \epsilon \right) / \beta_k \right) + (K - |\mathcal{S}_b|), \quad (24)$$

$$c_k = \left(1 - \frac{1}{N_o} \right) (1 - \epsilon) + [(N-1)\lambda_k + 1] \left(\frac{1-\epsilon}{N_o} + \epsilon \right) - N\bar{\lambda}_k, \quad (25)$$

where $\bar{\lambda}_k = \lambda_k e^{-(\sigma_\psi^2 + \sigma_\phi^2)|t-t_0|}$. Furthermore, a_k and c_k represent the multiuser interference received at the k^{th} MT and the variance of the gain of the desired signal, respectively. Moreover, the term $L_{\text{AN}}^k = \mathbb{E} \left[\mathbf{g}_k^H \mathbf{\Theta}_k^H(t) \mathbf{A} \mathbf{A}^H \mathbf{\Theta}_k(t) \mathbf{g}_k \right]$ in (23) represents the AN leakage in the received signal of the k^{th} MT in time slot t . This term will be characterized in detail for the considered AN precoders in Section III-C.

C. Generalized NS AN Precoding

The AN leakage term L_{AN}^k in (23) depends on the particular AN precoder used. Therefore, in this subsection, we first evaluate L_{AN}^k for the conventional NS precoder, where \mathbf{A} is designed to lie in the NS of the estimated channel vectors of all MTs, $\hat{\mathbf{g}}_k$, $1 \leq k \leq K$, which is the most common design used in the literature [7], [15], [29]. Subsequently, we propose and analyze the G-NS AN precoder design which is less sensitive to HWIs than the conventional NS design.

The AN leakage incurred by the conventional NS AN precoder is given in the following Lemma.

Lemma 4: For the conventional NS AN precoder, where $L = N - K$ [7], [15], [29], the AN leakage power received at MT $k \in \mathcal{S}_b$ in time interval t is given by

$$L_{\text{AN}}^k = \beta_k(N - K) \left(\left(1 - \frac{1}{N_o}\right) (1 - \epsilon) + 1 - \lambda_k \right). \quad (26)$$

Proof: Please refer to Appendix E. ■

In Lemma 4, the terms ϵ and λ_k reflect the negative impact of the HWIs on the AN power leakage. If only one LO is employed, i.e., $N_o = 1$, the impact of ϵ is eliminated. However, the negative effect of ϵ increases as the number of LOs, N_o , increases since the phase noise processes of different LOs are independent destroying the orthogonality of the columns of \mathbf{A} and $\mathbf{g}_k(t)$, $1 \leq k \leq K$.

This problem can be mitigated by employing M_o NS AN precoders where each precoder encodes the data signals intended for the antennas connected to N_o/M_o LOs. Thereby, N_o is assumed to be a multiple of M_o , i.e., $N_o/M_o \in \mathbb{Z}$. The resulting AN preorder is referred to as G-NS AN precoder. More in detail, for the G-NS AN precoder, we divide each channel estimation vector, $\hat{\mathbf{g}}_k$, $1 \leq k \leq K$, into M_o sub-vectors

$$\hat{\mathbf{g}}_k = \left[\left(\hat{\mathbf{g}}_k^{(1)} \right)^T, \left(\hat{\mathbf{g}}_k^{(2)} \right)^T, \dots, \left(\hat{\mathbf{g}}_k^{(M_o)} \right)^T \right]^T, \quad (27)$$

where $\hat{\mathbf{g}}_k^{(m)} \in \mathbb{C}^{N/M_o \times 1}$, which contains the $((m-1)N/M_o + 1)^{\text{th}}$ to the $(mN/M_o)^{\text{th}}$ elements of $\hat{\mathbf{g}}_k$ for $1 \leq m \leq M_o$. Correspondingly, we split matrix \mathbf{A} into M_o sub-matrices as follows

$$\mathbf{A} = \left[\mathbf{A}_{(1)}^T, \mathbf{A}_{(2)}^T, \dots, \mathbf{A}_{(M_o)}^T \right]^T, \quad (28)$$

with $\mathbf{A}_{(m)} \in \mathbb{C}^{N/M_o \times (N/M_o - K)}$, $1 \leq m \leq M_o$, i.e., we have $L = N/M_o - K$. Now, matrix $\mathbf{A}_{(m)}$ is designed to lie in the null-space of $\hat{\mathbf{g}}_k^{(m)}$, $1 \leq k \leq K$, i.e., $\mathbf{A}_{(m)} \hat{\mathbf{g}}_k^{(m)} = \mathbf{0}$, $1 \leq k \leq K$,

$1 \leq m \leq M_o$. For $M_o = 1$, the G-NS precoder simplifies to the conventional NS precoder. On the other hand, for $M_o = N_o$, the antennas connected to each LO have their own NS AN precoder.

The AN leakage of the G-NS precoder is analyzed in the following Lemma.

Lemma 5: For the G-NS AN precoder, where $L = N/M_o - K$ and $1 \leq M_o \leq N_o$, the AN leakage power received at MT $k \in \mathcal{S}_b$ in time interval t is given by

$$L_{\text{AN}}^k = \beta_k \left(\frac{N}{M_o} - K \right) \left(\left(1 - \frac{M_o}{N_o} \right) (1 - \epsilon) + 1 - \lambda_k \right). \quad (29)$$

Proof: Please refer to Appendix F. ■

Several observations can be made from (29). First, we note that, as expected, for $M_o = 1$, (29) reduces to (26). Second, the negative impact of the phase noise via ϵ on the AN leakage can be completely eliminated by choosing $M_o = N_o$. Third, the G-NS precoder requires the calculation of M_o null spaces of dimension $N/M_o \times K$. Hence, the computational complexity increases with M_o . We will elaborate on the optimal choice of M_o in Sections IV and V.

Remark 1: We note that the proposed G-NS AN precoder is not optimal for the maximization of the achievable secrecy rate. Nevertheless, the G-NS AN precoder achieves high performance and facilitates the derivation of closed-form expressions for the achievable secrecy rate, which provide substantial technical insight for system design. Optimizing the AN precoder for maximization of the secrecy rate [35] is an interesting topic for future work.

The achievable rates of MT $k \in \mathcal{S}_b$ in time slot t with conventional NS and G-NS precoding are obtained by inserting (26) and (29) into (23), respectively. Hence, for the proposed G-NS precoder, we obtain

$$\underline{R}_k(t) = \log_2 \left(1 + \frac{\bar{\lambda}_k \phi N}{(a_k + c_k - \beta \mu_k) \phi + \beta \mu_k + \xi_k} \right), \quad (30)$$

where $\mu_k = (\frac{N}{M_o} - K) \left(\left(1 - \frac{M_o}{N_o} \right) (1 - \epsilon) + 1 - \lambda_k \right)$, $\xi_k = \beta(\kappa_r^{\text{MT}} + \kappa_t^{\text{BS}} + \xi^{\text{DL}}/(\beta_k P_T))$, and $\beta = K/N > 0$.

D. Upper Bound on the Eavesdropper's Capacity

In the following Proposition, we provide a tight and tractable upper bound on eavesdropper's capacity.

Proposition 1: For $N \rightarrow \infty$ and (G-)NS AN precoding, the eavesdropper's capacity in (14) can be upper bounded as

$$C_E(t) \leq \overline{C}_E = \log_2 \left(1 + \frac{pN_E}{qL + \kappa_t^{\text{BS}}P_T - \chi N_E} \right), \text{ with } \chi = \frac{(1 + \kappa_t^{\text{BS}})^2 q^2 L + (\kappa_t^{\text{BS}})^2 p^2 K}{(1 + \kappa_t^{\text{BS}})qL + \kappa_t^{\text{BS}}pK}, \quad (31)$$

for $qL + \kappa_t^{\text{BS}}P_T > \chi N_E$, and where $L = N - K$ and $L = N/M_o - K$ for the conventional NS and the G-NS precoders, respectively.

Proof: Please refer to Appendix G. ■

We observe from (31) that, as expected, the capacity of the eavesdropper is increasing in the number of its equipped antennas, N_E . Another non-trivial observation is that the bound provided in Proposition 1 is no longer a function of time slot index t , due to the worst-case assumption that the eavesdropper has perfect instantaneous CSI, cf. Section II-C. Interestingly, when no AN is injected, i.e., $q = 0$, (31) reduces to

$$\overline{C}_E \Big|_{q=0} = \log_2 \left(1 + \frac{N_E}{\kappa_t^{\text{BS}}(K - N_E)} \right), \quad (32)$$

for $K > N_E$. For perfect BS hardware, we have $\kappa_t^{\text{BS}} \rightarrow 0$ and $\overline{C}_E \rightarrow \infty$ making secure communication impossible. Hence, if AN is not injected, HWIs may in fact be beneficial for secure communication as the distortion noise at the BS acts like AN and may facilitate secrecy. This surprising insight will be studied more carefully in the next section. Furthermore, the number of independent distortion noise processes at the BS is equal to the number of users, K . Hence, $K > N_E$ is needed to prevent the eavesdropper from nulling out the distortion noise and for achieving secrecy.

The worst-case ergodic secrecy rate achieved by MT k in time slot t is lower bounded by $\left[\underline{R}_k(t) - \overline{C}_E \right]^+$, where $\underline{R}_k(t)$ and \overline{C}_E are given in (30) and (31), respectively. Hence, although the CSI and location information of the eavesdropper assumed to be not available at the BS, with the proposed transmission strategy, the BS can still guarantee the derived worst-case ergodic secrecy rate in the presence of HWIs. In non-worst-case scenarios, higher ergodic secrecy rates are expected.

IV. GUIDELINES FOR SYSTEM DESIGN

In this section, we exploit the analytical results derived in the previous section to gain some insight into the impact of the various system and HWI parameters on system design. To this end,

we carefully study the closed-form lower bound on the achievable ergodic secrecy rate obtained by combining (18), (30), and (31).

A. Design of the Pilot Sequences

Assuming that we assign the maximum number of users to each training sub-phase, i.e., $|\mathcal{S}_b| = B_b$, the relevant design parameter for the pilot sequences is the number of training sub-phases B_o , or equivalently, the size of the training sub-phases B_b as $\sum_{b=1}^{B_o} B_b = B$. In particular, B_b affects the lower bound on the achievable ergodic rate of MT k in (30) via λ_k , a_k , and c_k , where c_k becomes proportional to λ_k for $N \rightarrow \infty$, cf. (25). Thereby, close inspection of (19) reveals that λ_k , which reflects the power of the received useful signal, is not monotonic in B_b . This can be explained as follows. On the one hand, since the power of each pilot symbol is constrained, i.e., $|\omega_k(t)|^2 = p_\tau$, $\forall k, t$, the sum power of the pilot sequence per MT increases with B_b . On the other hand, for larger B_b , more MTs are allowed to emit pilots in training sub-phase b introducing more mutual pilot interference due to phase noise. This has an adverse effect on the quality of the channel estimate and consequently on the power of the received useful signal. Similarly, close inspection of (24) reveals that a_k , which reflects the multiuser interference incurred to the k^{th} MT, is a monotonically increasing function of B_b , as a lower channel estimation accuracy caused by more mutual pilot interference, gives rise to stronger multiuser interference. Considering the behaviour of λ_k , a_k , and c_k and their impact on the achievable ergodic rate of MT k in (30), we conclude that B_b , $1 \leq b \leq B_o$, should be optimized and the optimal value depends on the channel and HWI parameters. Thereby, the optimal B_b is decreasing in the phase noise variances, σ_ψ^2 and σ_ϕ^2 , as the degradation introduced by concurrent pilot emission by multiple MTs is increasing in these parameters. This conclusion will be verified in Section V-B by numerically evaluating (18).

B. Selection of M_o for G-NS AN Precoding

The number of G-NS AN precoding sub-matrices, M_o , $1 \leq M_o \leq N_o$, employed affects the achievable ergodic secrecy rate via the AN leakage L_{AN}^k in (29) and via the (bound on the) eavesdropper capacity \bar{C}_E in (31). The AN leakage is a decreasing function with respect to M_o , i.e., as far as the AN leakage is concerned, $M_o = N_o$ is preferable. On the other hand, since the dimensionality of the G-NS AN precoder is given by $L = N/M_o - K$, the eavesdropper capacity

is an increasing function of M_o , cf. (31), which has a negative effect on the ergodic secrecy rate. Hence, M_o has to be optimized. Since the eavesdropper capacity does not depend on the phase noise, we expect that the optimal M_o increases with increasing BS phase noise variance, σ_ψ^2 , as σ_ψ^2 affects the AN leakage via ϵ in (29). This conjecture will be numerically verified in Section V-D.

C. Secrecy in the Absence of AN

In [7], [15] it was shown that if perfect hardware is employed, injection of AN is necessary to achieve secrecy. In particular, without AN generation, under worst-case assumptions regarding the noise at the eavesdropper, the eavesdropper capacity is unbounded. On the other hand, we showed in Section III-D that in the presence of HWIs the eavesdropper capacity is bounded since the distortion noise generated at the BS has a similar effect as AN. Motivated by this observation, in this section, we calculate the maximum number of eavesdropper antennas N_E that can be tolerated if a positive secrecy rate is desired without AN emission.

If AN is not emitted, we have $\phi = 1$ or $q = 0$. In this case, the proposed lower bound on the ergodic secrecy rate of the k^{th} MT in time interval t simplifies to

$$\left. \underline{R}_k^{\text{sec}}(t) \right|_{q=0} = \left[\log_2 \left(1 + \frac{\bar{\lambda}_k N}{a_k + c_k + \xi_k} \right) - \log_2 \left(1 + \frac{\alpha}{\kappa_t^{\text{BS}}(\beta - \alpha)} \right) \right]^+. \quad (33)$$

where $\alpha = N_E/N$ denotes the normalized number of eavesdropper antennas. In the following Proposition, we provide a condition for the number of eavesdropper antennas that has to be met for secure communication to be possible.

Proposition 2: If AN is not generated, the maximum number of eavesdropper antennas that the k^{th} MT can tolerate while achieving a positive ergodic secrecy rate is $N_E = \lfloor \alpha_{\text{AN}} N \rfloor$, where

$$\alpha_{\text{AN}} = \left. \frac{\bar{\lambda}_k N \kappa_t^{\text{BS}} \beta}{\bar{\lambda}_k N \kappa_t^{\text{BS}} + a_k + c_k + \xi_k} \right|_{t=B+1}. \quad (34)$$

Proof: First, we note that $\underline{R}_k(t)$ is a decreasing function of t . Hence, considering (18), it is sufficient to ensure $\underline{R}_k(B+1) > \bar{\mathcal{C}}_E$ for achieving a positive ergodic secrecy rate. Eq. (34) is obtained by setting (33) to zero and observing that $\left. \underline{R}_k^{\text{sec}}(t) \right|_{q=0}$ is a decreasing function of α . ■

Eq. (34) clearly shows that the additive distortion noise at the BS is essential for achieving a positive secrecy rate if AN is not injected as $\alpha_{\text{AN}} = 0$ results if $\kappa_t^{\text{BS}} = 0$. On the other hand, α_{AN} is a decreasing function of all other HWI parameters, i.e., κ_r^{BS} , κ_t^{MT} , κ_r^{MT} , ζ^{DL} , σ_ψ^2 , and

σ_ψ^2 , as the corresponding HWIs affect only the achievable ergodic rate of the MT but not the ergodic capacity of the eavesdropper. We note that α_{AN} is an increasing function of β since the dimensionality of the additive distortion noise at the BS is proportional to β .

D. Maximum Number of Eavesdropper Antennas

Now, we consider the maximum number of eavesdropper antennas that can be tolerated if a positive ergodic secrecy rate is desired and AN injection is possible. Combining (18), (30), and (31), the lower bound on the ergodic secrecy rate in time interval t can be expressed as

$$\underline{R}_k^{\text{sec}}(t) = \left[\log_2 \left(1 + \frac{\bar{\lambda}_k \phi N}{(a_k + c_k)\phi + \beta \mu_k (1 - \phi) + \xi_k} \right) - \log_2 \left(1 + \frac{\alpha \phi}{\beta(1 - \phi + \kappa_t^{\text{BS}} - \chi' \alpha)} \right) \right]^+, \quad (35)$$

where $\chi' = \frac{(1 + \kappa_t^{\text{BS}})^2 (1 - \phi)^2 N / L + (\kappa_t^{\text{BS}})^2 \phi^2 / \beta}{1 - \phi + \kappa_t^{\text{BS}}}$.

Proposition 3: If AN injection is possible, a positive secrecy rate can be achieved by the k^{th} MT if the number of eavesdropper antennas does not exceed $N_E = \lfloor \alpha_{\text{sec}} N \rfloor$, where

$$\alpha_{\text{sec}} = \frac{(1 + \kappa_t^{\text{BS}}) \bar{\lambda}_k L}{L / N (\mu_k + \kappa_r^{\text{MT}} + \kappa_t^{\text{BS}} + \xi^{\text{DL}} / (\beta_k P_T)) + \bar{\lambda}_k N (1 + \kappa_t^{\text{BS}})} \Big|_{t=B+1} \quad (36)$$

and $\phi \rightarrow 0$, i.e., almost all transmit power is employed for AN generation.

Proof: Exploiting again that $\underline{R}_k(t)$ is a decreasing function of t it suffices to consider the ergodic secrecy rate for $t = B + 1$. Then, an expression for α_{sec} is obtained by setting $\underline{R}_k^{\text{sec}}(t)$ in (35) to zero. This expression is monotonically decreasing in ϕ and hence can be further simplified by letting $\phi \rightarrow 0$ which yields (36). ■

Proposition 3 reveals that, as expected, the number of eavesdropper antennas that can be tolerated increases with the channel estimation accuracy (i.e., $\bar{\lambda}_k$) and the number of spatial dimensions available for AN (i.e., L). Furthermore, similar to α_{AN} , α_{sec} is a decreasing function of the HWI parameters κ_r^{BS} , κ_t^{MT} , κ_r^{MT} , ξ^{DL} , σ_ψ^2 , and σ_ψ^2 , and an increasing function of κ_t^{BS} . However, unlike α_{AN} , α_{sec} is independent of β .

E. Number of LOs

The number of LOs, N_o , affects the ergodic secrecy rate via the terms a_k , c_k , and μ_k in the achievable ergodic rate in (30). For $N \rightarrow \infty$, a_k and c_k are decreasing functions of N_o , i.e., less multiple access interference is caused if more LOs are employed, whereas the AN leakage term μ_k is an increasing function in N_o . Therefore, considering the specific form of the denominator

of the fraction inside the logarithm in (30), the optimal value of N_o , which maximizes the ergodic secrecy rate, depends on ϕ . In particular, for a given M_o , for $\phi = 1$ no AN is injected and μ_k cancels in the expression for the achievable ergodic rate in (30). Hence, in this case, the ergodic secrecy rate is a monotonically increasing function of N_o , i.e., increasing the number of LOs is beneficial. On the other hand, for a given M_o , for $\phi < 1$, the optimal N_o maximizing the ergodic secrecy rate can be found by performing a numerical search based on (30).

We note that by employing G-NS AN generation and enforcing $M_o = N_o$, we can avoid the harmful effect of the multiple LOs on the AN leakage term μ_k . In this case, the achievable ergodic rate of the MT becomes an increasing function of $M_o = N_o$. However, at the same time, the number of dimensions available for AN injection, $L = N/M_o - K$, is a decreasing function of $M_o = N_o$. Therefore, the optimal $M_o = N_o$ maximizing the ergodic secrecy rate has to be found again by a numerical search.

F. Are HWIs Beneficial for Security?

Since the HWI parameters κ_r^{BS} , κ_t^{MT} , κ_r^{MT} , ξ^{DL} , σ_ψ^2 , and σ_ψ^2 only affect the legitimate user but not the eavesdropper, the corresponding HWIs are always detrimental to the ergodic secrecy rate. However, the additive distortion noise at the BS affects both the achievable ergodic rate of the MT and the capacity of the eavesdropper. Hence, it is not a priori clear if this HWI is beneficial or detrimental to the ergodic secrecy rate. The following Proposition provides a criterion for judging the benefits of the additive BS distortion noise.

Proposition 4: For time interval t , non-zero additive BS distortion noise with small $\kappa_t^{\text{BS}} > 0$, $\kappa_t^{\text{BS}} \rightarrow 0$, is beneficial for the achievable ergodic secrecy rate of the k^{th} MT if and only if

$$(1 - \phi)[1 - N_E/L - (1 - N_E/L - N_E/K)\phi] \times \frac{1 - N_E/L}{1 - (1 - 2\phi)N_E/L} < \frac{\alpha\gamma(N\bar{\lambda}_k\phi + \gamma)}{\beta^2\bar{\lambda}_kN}, \quad (37)$$

where $\gamma = (a_k + c_k)\phi + \beta\mu_k(1 - \phi) + \beta(\kappa_r^{\text{MT}} + \xi^{\text{DL}}/(\beta_k P_T))$.

Proof: For additive BS distortion noise to be beneficial for a given time interval t and small $\kappa_t^{\text{BS}} > 0$, the derivative $\partial \underline{R}_k^{\text{sec}}(t)/\partial \kappa_t^{\text{BS}}$ at $\kappa_t^{\text{BS}} = 0$ has to be positive. Assuming $\underline{R}_k^{\text{sec}}(t) > 0$, this condition leads to $\partial \underline{R}_k(t)/\partial \kappa_t^{\text{BS}}|_{\kappa_t^{\text{BS}}=0} > \partial \bar{C}_E/\partial \kappa_t^{\text{BS}}|_{\kappa_t^{\text{BS}}=0}$, which can be further simplified to (37). ■

Remark 2 : We note that the criterion in Proposition 4 only guarantees that additive BS distortion noise with small positive κ_t^{BS} is beneficial. The ergodic secrecy rate, $\underline{R}_k^{\text{sec}}(t)$, is in

general not monotonic in κ_t^{BS} and larger κ_t^{BS} may be harmful even if small κ_t^{BS} are beneficial, see Section V-E. Furthermore, since the right hand side of (37) is always positive, we conclude that additive BS distortion noise with small κ_t^{BS} is always beneficial when $\phi = 1$, i.e., when AN is not injected.

V. NUMERICAL EXAMPLES

In this section, we provide numerical and simulation results to verify the analysis presented in Sections III and IV and to illustrate the impact of HWIs on the ergodic secrecy rate. For the numerical results, we numerically evaluate the analytical expression for the lower bound on the ergodic secrecy rate obtained by combining (18), (30), and (31). For the simulation results, we employ Monte Carlo simulation and evaluate (18) using $\underline{R}_k^{\text{sec}}(t) = \log_2(1 + \gamma_k(t))$ and $C_E(t) = \log_2(1 + \gamma_E(t))$ with $\gamma_k(t)$ and $\gamma_E(t)$ given by (17) and (15), respectively, for 5,000 independent channel realizations. For simplicity, in this section, we assume that the path-loss for all MTs is identical ², i.e., $\beta_k = 1$, $1 \leq k \leq K$, and the coherence block length is equal to $T = 500$ time slots. Typical values for the phase noise increment standard deviations, σ_ψ , σ_ϕ , used include 0.06° , which was adopted in the long-term evolution (LTE) specifications [33], and 6° , which corresponds to strong phase noise according to [31], [32]. Furthermore, typical values for the additive distortion noise $\kappa_t^{\text{MT}} = \kappa_r^{\text{BS}} = \kappa_t^{\text{BS}} = \kappa_r^{\text{MT}}$ include $\{0, 0.05^2, 0.15^2\}$ [21], whereas the amplified receiver noise was set to $\xi^{\text{UL}} = \xi^{\text{DL}} = 1.58\sigma_n^2$ [26], with $\sigma_n^2 = 1$. The specific values of the adopted system and HWI parameters are provided in the captions of the figures.

A. Capacity of Eavesdropper for G-NS AN Precoding

Fig. 2 depicts the eavesdropper's ergodic capacity, C_E , as a function of β for G-NS AN precoding with $M_o = \{1, 2, 4\}$. Besides results for the analytical upper bound, \overline{C}_E , from (31), we also show simulation results for C_E by averaging $\log_2(1 + \gamma_E)$ over 5,000 independent channel realizations, where γ_E is given by (15). From Fig. 2 we observe that the proposed upper

²Although the analytical results presented in this paper are valid for unequal path-losses, for the presented numerical results, we employ equal path-losses in order to be able to focus on the impact of HWIs on the achievable ergodic secrecy rate. The investigation of this impact is the main objective of this paper, and unequal path-losses do not provide any additional insights in this regard.

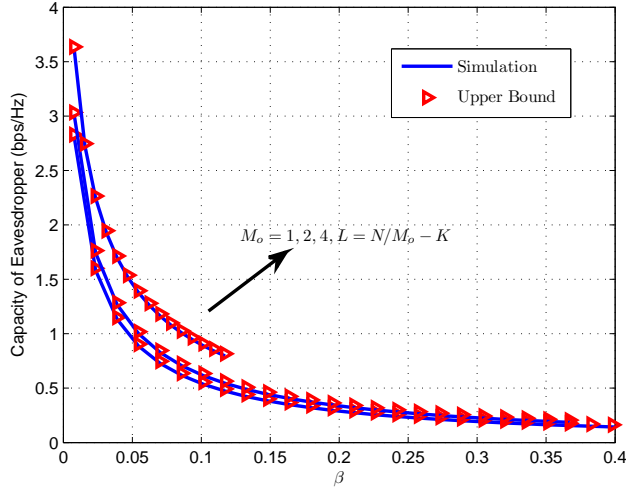


Fig. 2. Capacity of the eavesdropper vs. the normalized number of MTs β for a system with $N = 128$, $N_o = 4$, $N_E = 16$, $P_T = 10$ dB, $\phi = 0.25$, $\kappa_t^{\text{BS}} = 0.15^2$, and G-NS AN precoding with $M_o = \{1, 2, 4\}$.

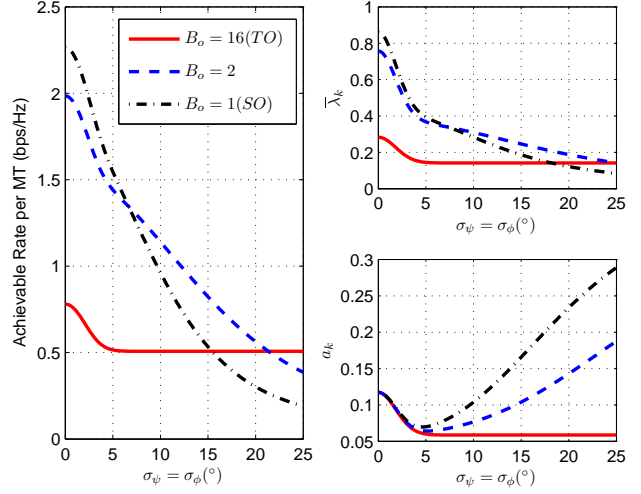


Fig. 3. Achievable ergodic rate, $\bar{\lambda}_k$, and a_k vs. phase noise standard deviation $\sigma_\psi = \sigma_\phi$ for different pilot designs for a system with $N = 128$, $N_o = 2$, $N_E = 16$, $K = B = 16$, $p_\tau = P_T/K$, $P_T = 10$ dB, $\phi = 0.5$, and $\kappa_t^{\text{BS}} = \kappa_r^{\text{BS}} = \kappa_t^{\text{MT}} = \kappa_r^{\text{MT}} = 0.05^2$.

bound on the capacity of the eavesdropper is very tight. Furthermore, as expected, the ergodic capacity of the eavesdropper is an increasing function of M_o since the number of dimensions available for AN generation, $L = N/M_o - K$, is a decreasing function of M_o . In fact, since $L = N/M_o - K > N_E$ is needed for successfully jamming the eavesdropper, for $M_o = 4$, we depict the ergodic capacity of the eavesdropper only for $\beta < 0.125$. Nevertheless, as will be shown below, choosing $M_o > 1$ may still be beneficial as far as the ergodic secrecy rate is concerned as the achievable ergodic rate of the MT is an increasing function of M_o .

B. Achievable Ergodic Rate of MT for Different Pilot Designs

Next, we investigate the impact of the general pilot designs introduced in Section II-A on the lower bound of the achievable ergodic rate of the considered MT given in (30)³. Note that the capacity of the eavesdropper is not affected by the pilot design. For simplicity, we assume equal duration for all training sub-phases, $B_b = B/B_o$, $b \in \{1, \dots, B_o\}$, and $B = K$. The same

³We note that all results obtained by numerically evaluating the analytical expressions derived in this paper were verified by simulations. However, the simulation results are not included in all figures for clarity of presentation.

number of users is assigned to each training sub-phase. In Fig. 3, we show the achievable ergodic rate of a MT in training set \mathcal{S}_{B_o} as well as the corresponding $\bar{\lambda}_k$, which reflects the power of the received useful signal, and a_k , which reflects the multiuser interference. Results for $B_o = 1$ (SO pilots), $B_o = 2$, and $B_o = 16$ (TO pilots) are shown. As predicted in Section IV-A, the multiuser interference, a_k , is monotonically decreasing in B_o as larger B_o improve the robustness against phase noise during the channel estimation phase, which allows better suppression of multiuser interference via MF precoding. Somewhat surprisingly, for $\sigma_\psi = \sigma_\phi \leq 5^\circ$, a_k is a decreasing function of the phase noise variance. This may be attributed to the fact that phase noise prevents the coherent superposition of the multiuser interference generated by different MTs such that large interference values are avoided. On the other hand, for $\sigma_\psi = \sigma_\phi > 5^\circ$, the detrimental effects of the pilot interference caused by the loss of orthogonality for $B_o < 16$ outweigh this positive effect and a_k increases with the phase noise variance. For $\bar{\lambda}_k$, i.e., the received signal power, we observe from Fig. 3 that the optimal B_o depends on the phase noise variance. In particular, for small phase noise variances, small B_o are preferable since the increased pilot power outweighs the loss of orthogonality during training. On the other hand, for large phase noise variances, eventually TO pilots become optimal as the preserved orthogonality during training becomes crucial. The behaviour of $\bar{\lambda}_k$ and a_k is also reflected in the behaviour of the achievable rate of the considered MT. In particular, for the considered system parameters, $B_o = 1$, $B_o = 2$, and $B_o = 16$ are optimal for $\sigma_\psi = \sigma_\phi \leq 6^\circ$, $6^\circ < \sigma_\psi = \sigma_\phi \leq 21^\circ$, and $\sigma_\psi = \sigma_\phi > 21^\circ$ (which is not a practical range), respectively. Hence, in practice, the optimal B_o can be found by evaluating (30).

C. Optimal Power Allocation to Data and AN

Fig. 4 shows the achievable ergodic secrecy rate as a function of the power allocation parameter ϕ for SO and TO pilots and different phase noise variances. G-NS AN precoding with $M_o = N_o = 2$ is adopted. The curve for ideal hardware components, i.e., $\kappa_t^{\text{BS}} = \kappa_r^{\text{BS}} = \kappa_t^{\text{MT}} = \kappa_r^{\text{MT}} = \sigma_\psi = \sigma_\phi = 0$, is also provided for reference. We investigate the optimal power allocation between data transmission and AN emission for the maximization of the ergodic secrecy rate achieved for different phase noise levels. When the phase noise variance is small, i.e., $\sigma_\psi = \sigma_\phi = 0.6^\circ$, SO pilots outperforms TO pilots for all values of ϕ . However, this is not true for stronger phase noise. We also observe that the optimal value for ϕ maximizing the ergodic secrecy rate is only

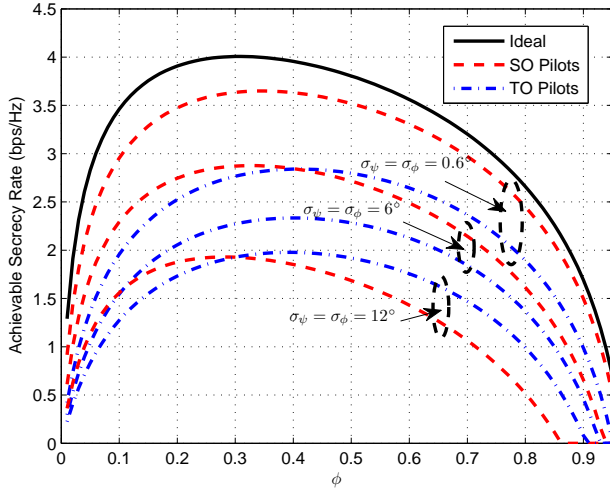


Fig. 4. Achievable ergodic secrecy rate vs. ϕ for SO and TO pilots and a system with $K = B = 4$, $N = 128$, $N_o = M_o = 2$, $N_E = 4$, $p_r = P_T/K$, $P_T = 10$ dB, and $\kappa_t^{\text{BS}} = \kappa_r^{\text{BS}} = \kappa_t^{\text{MT}} = \kappa_r^{\text{MT}} = 0.15^2$.

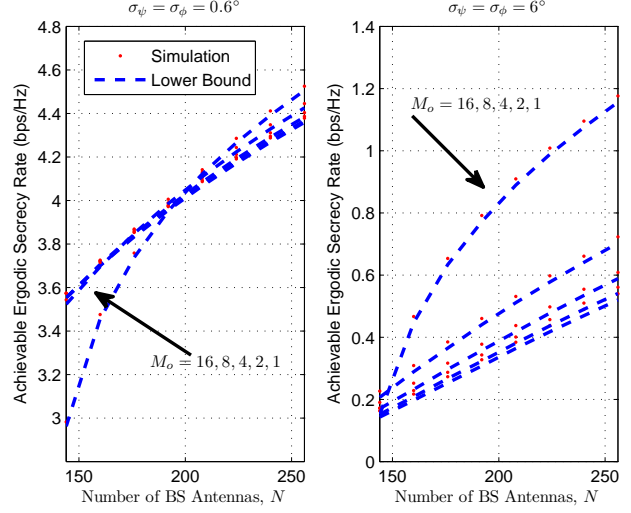


Fig. 5. Achievable ergodic secrecy rate vs. number of BS antennas for G-NS AN precoding and a system with $K = B = 4$, $N_E = 4$, $N_o = 16$, $B_o = 1$, $p_r = P_T/K$, $P_T = 10$ dB, and $\kappa_t^{\text{BS}} = \kappa_r^{\text{BS}} = \kappa_t^{\text{MT}} = \kappa_r^{\text{MT}} = 0.15^2$. The optimal ϕ is adopted.

weakly dependent on the phase noise variance.

D. Achievable Ergodic Secrecy Rate with G-NS AN Precoding

In Fig. 5, we show the ergodic secrecy rate achieved with G-NS AN precoding for different values of M_o as a function of the number of BS antennas. The cases of weak ($\sigma_\psi = \sigma_\phi = 0.6^\circ$) and strong ($\sigma_\psi = \sigma_\phi = 6^\circ$) phase noise are considered. For weak phase noise, using large values of M_o becomes beneficial only for large numbers of antennas, i.e., $N > 200$, as for smaller numbers of antennas the positive effect of larger values of M_o on the AN leakage is outweighed by their negative effect on the number of spatial dimensions available for AN precoding. On the other hand, for strong phase noise, the AN leakage is larger and its mitigation by choosing $M_o = N_o = 16$ is beneficial already for $N > 150$. These observations are in line with our theoretical considerations in Section IV-B. Fig. 5 also confirms the accuracy of the derived analytical expressions for the ergodic secrecy rate.

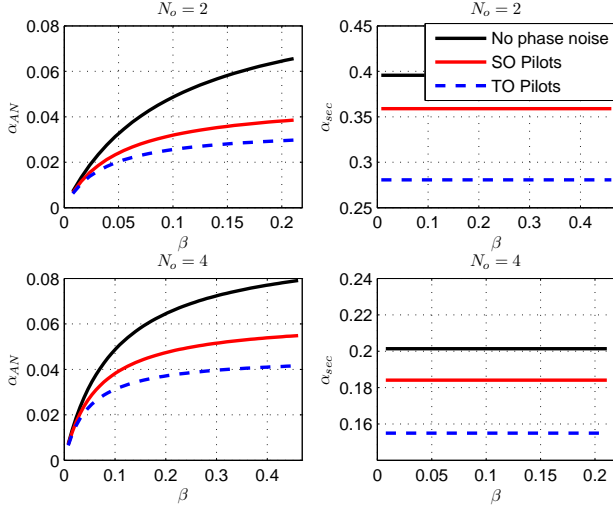


Fig. 6. α_{AN} and α_{sec} vs. the normalized number of MTs β for SO and TO pilots and a system with $N = 128$, $M_o = 2$, $p_r = P_T/K$, $P_T = 10$ dB, $\sigma_\psi = \sigma_\phi = 6^\circ$, and $\kappa_t^{BS} = \kappa_r^{BS} = \kappa_t^{MT} = \kappa_r^{MT} = 0.15^2$.

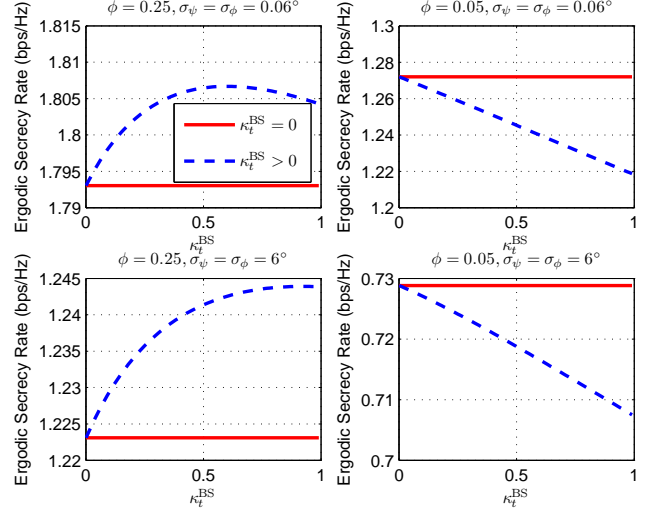


Fig. 7. Achievable ergodic secrecy rate vs. BS distortion noise parameter κ_t^{BS} for a system with $N = 128$, $K = B = 32$, $N_E = 4$, $N_o = M_o = 2$, $p_r = P_T/K$, $P_T = 10$ dB, and $\kappa_r^{BS} = \kappa_t^{MT} = \kappa_r^{MT} = 0.15^2$.

E. Maximum Tolerable Number of Eavesdropper Antennas

Fig. 6 depicts the (normalized) maximum tolerable number of eavesdropper antennas for achieving a positive ergodic secrecy rate for the case without AN generation, α_{AN} , and the case with AN generation, α_{sec} , as a function of the (normalized) number of users, β . Results for channel estimation based on SO and TO pilots as well as the case of no phase noise ($\sigma_\psi = \sigma_\phi = 0^\circ$) are shown for $N_o = 2$ and $N_o = 4$ LOs. First, we note that, as expected from our considerations in Section IV-E, for the case without AN ($\phi = 1$), increasing N_o from 2 to 4 is beneficial, i.e., the number of tolerable eavesdropper antennas increases. In contrast, if AN is injected, $N_o = 2$ is preferable. Second, AN generation is beneficial and improves the robustness against eavesdropping, i.e., $\alpha_{sec} > \alpha_{AN}$. Third, as expected from Sections IV-C and IV-D, α_{AN} is a monotonically increasing function of β whereas α_{sec} is independent of β . Fourth, for the considered example of weak phase noise, SO pilots outperform the TO pilots for all considered cases.

F. Is Additive Distortion Noise at the BS Beneficial for Security?

In Fig. 7, we show the achievable ergodic secrecy rate as a function of the BS distortion noise parameter, κ_t^{BS} , for different phase noise variances and different power allocation factors ϕ . For comparison, the achievable ergodic secrecy rates without BS distortion noise (i.e., $\kappa_t^{\text{BS}} = 0$) are also shown. Fig. 7 shows that if the power allocated to AN is substantial (e.g., $\phi = 0.05$), the additional distortion noise has a negative effect on the ergodic secrecy rate. On the other hand, if the power assigned for AN is not sufficient (e.g., $\phi = 0.25$), non-zero additive distortion noise at the BS is beneficial as the distortion noise acts like additional AN. In particular, for $\phi = 0.25$, $\sigma_\psi = 0.06^\circ$, we obtain for the left hand side and right hand side of (37) 0.52 and 1.66, respectively, which we represent as (0.52, 1.66). Correspondingly, we obtain for $\phi = 0.25$, $\sigma_\psi = 6^\circ$ and $\phi = 0.05$, $\sigma_\psi = 0.06^\circ$ and $\phi = 0.05$, $\sigma_\psi = 6^\circ$ the tuples (0.52, 2.53) and (0.80, 0.16) and (0.80, 0.35), respectively. These values and the results in Fig. 7 suggest that (37) can indeed be used to predict whether or not BS distortion noise is beneficial.

VI. CONCLUSIONS

In this paper, we have investigated the impact of HWIs such as multiplicative phase noise, additive distortion noise, and amplified receiver noise on the secrecy performance of massive MIMO systems employing MF precoding for downlink data transmission. To mitigate the loss of pilot orthogonality during uplink training if multiple MTs emit pilots concurrently, a generalized pilot design was proposed. Furthermore, to avoid the AN leakage caused by the loss of orthogonality between the user channels and the NS AN precoder if multiple noisy LOs are employed at the BS, a novel G-NS AN precoding scheme was introduced. For the considered system, a lower bound on the achievable ergodic secrecy rate of the users was derived. This bound was used to obtain important insights for system design, including the impact of the pilot sequence design, the AN precoder design, the number of LOs, and the various HWI parameters. The following general conclusions can be drawn: 1) Additive distortion noise at the BS may be beneficial for the secrecy performance especially if little or no AN is injected; 2) all other HWIs have a negative impact on the ergodic secrecy rate; 3) despite their susceptibility to pilot interference in the presence of phase noise, SO pilots are preferable except for the case when the phase noise is very strong; 4) if the number of BS antennas is sufficiently large, the proposed

G-NS AN precoder outperforms the conventional NS AN precoder in the presence of phase noise.

Interesting extensions of this paper which could be studied in future research include the impact of HWIs on the physical layer security of multi-cell massive MIMO systems, pilot sequence design under an average power constraint, and optimal AN precoder design for secrecy rate maximization under HWIs.

APPENDIX

A. A Useful Theorem from Free Probability Theory

Theorem 1 [34]: If $(\mathbf{U}, \mathbf{V}) \in \mathbb{C}^N$ are free from $(\mathbf{Y}, \mathbf{Z}) \in \mathbb{C}^N$, then $\text{Tr}(\mathbf{U}\mathbf{Y}\mathbf{V}\mathbf{Z}) =$

$$\text{Tr}(\mathbf{U}) \text{Tr}(\mathbf{V}) \text{Tr}(\mathbf{Y}\mathbf{Z}) + \text{Tr}(\mathbf{Y}) \text{Tr}(\mathbf{Z}) \text{Tr}(\mathbf{U}\mathbf{V}) - \text{Tr}(\mathbf{U}) \text{Tr}(\mathbf{V}) \text{Tr}(\mathbf{Y}) \text{Tr}(\mathbf{Z}), \quad (38)$$

where $\text{Tr}(\cdot) = \lim_{N \rightarrow \infty} \text{tr}(\cdot)/N$.

B. Proof of Lemma 1

The ergodic secrecy rate achieved by the k^{th} MT in symbol interval $t \in \{B+1, \dots, T\}$ is given by [7, Lemma 1]

$$R_k^{\text{sec}}(t) = \mathbb{E} [[R_k(t) - \log_2(1 + \gamma_E(t))]^+] \geq [\mathbb{E}[R_k(t)] - C_E(t)]^+ \stackrel{(a)}{\geq} [\underline{R}_k(t) - C_E(t)]^+ = \underline{R}_k^{\text{sec}}(t), \quad (39)$$

where $\underline{R}_k^{\text{sec}}(t)$ is an achievable lower bound for $R_k^{\text{sec}}(t)$, and (a) uses (16). By averaging $R_k^{\text{sec}}(t)$ over all symbol intervals $t \in \{B+1, \dots, T\}$ we obtain Lemma 1. This completes the proof.

C. Proof of Lemma 2

The expectation given in (19) for $k \in \mathcal{S}_b$ is calculated as $\mathbb{E} \left[\mathbf{g}_k^H \mathbf{\Theta}_k^H(t) \mathbf{f}_k \right]$

$$\begin{aligned} &\stackrel{(a)}{=} \mathbb{E} \left[\frac{\hat{\mathbf{g}}_k^H \mathbf{\Psi}_{t_0}^H(t) \hat{\mathbf{g}}_k}{\|\hat{\mathbf{g}}_k\|} e^{j(\phi_k(t) - \phi_k(t_0))} \right] \stackrel{(b)}{=} \text{tr} \left(\mathbb{E} \left[\frac{\hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H}{\|\hat{\mathbf{g}}_k\|} \right] \mathbb{E} \left[\mathbf{\Psi}_{t_0}^H(t) \right] \right) \mathbb{E} \left[e^{j(\phi_k(t) - \phi_k(t_0))} \right] \\ &= \sqrt{\beta_k N \lambda_k} \cdot e^{-\frac{\sigma_\psi^2 + \sigma_\phi^2}{2} |t - t_0|}, \end{aligned} \quad (40)$$

where $\mathbf{\Psi}_{t_0}(t) = \text{diag} \left(e^{j(\psi_1(t) - \psi_1(t_0))} \mathbf{1}_{1 \times N/N_o}^T, \dots, e^{j(\psi_{N_o}(t) - \psi_{N_o}(t_0))} \mathbf{1}_{1 \times N/N_o}^T \right)$ and λ_k is defined in Lemma 2. In (40), (a) exploits that the channel estimate and the estimation error are uncorrelated [21], and (b) exploits the mutually independence of $\hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H$, $\mathbf{\Psi}_{t_0}^H(t)$, and $e^{j(\phi_k(t) - \phi_k(t_0))}$. This completes the proof.

D. Proof of Lemma 3

In (17), the term reflecting the interference caused by the signal intended for MT $l \in \mathcal{S}_b$ to MT $k \in \mathcal{S}_b$ can be expanded as $\mathbb{E} \left[\left| \mathbf{g}_k^H \boldsymbol{\Theta}_k^H(t) \mathbf{f}_l \right|^2 \right] =$

$$\begin{aligned} & \mathbb{E} \left[\left| \mathbf{g}_k^H(t_0) \boldsymbol{\Psi}_{t_0}^H(t) \frac{\hat{\mathbf{g}}_l}{\|\hat{\mathbf{g}}_l\|} e^{j(\phi_k(t) - \phi_k(t_0))} \right|^2 \right] = \mathbb{E} \left[\text{tr} \left(\mathbf{g}_k(t_0) \mathbf{g}_k^H(t_0) \boldsymbol{\Psi}_{t_0}^H(t) \frac{\hat{\mathbf{g}}_l \hat{\mathbf{g}}_l^H}{\|\hat{\mathbf{g}}_l\|^2} \boldsymbol{\Psi}_{t_0}(t) \right) \right] \\ & \stackrel{(a)}{=} \beta_k + \left(\frac{\mathbb{E} \left[\text{tr} (\mathbf{X}_l^H \mathbf{g}_k(t_0) \mathbf{g}_k^H(t_0) \mathbf{X}_l \boldsymbol{\psi}_b \boldsymbol{\psi}_b^H) \right]}{\beta_l^2 \boldsymbol{\omega}_l^H \boldsymbol{\Theta}_{\sigma(t_0)}^b \boldsymbol{\Sigma}_b^{-1} \boldsymbol{\Theta}_{\sigma(t_0)}^H \boldsymbol{\omega}_l N} - \beta_k \right) \mathbb{E}_\psi \left[\left(\frac{1}{N} \text{tr} (\boldsymbol{\Psi}_{t_0}^H(t)) \right)^2 \right], \end{aligned} \quad (41)$$

where $\mathbf{X}_l = \beta_l \boldsymbol{\omega}_l^H \boldsymbol{\Theta}_{\sigma(t_0)}^b \boldsymbol{\Sigma}_b^{-1} \otimes \mathbf{I}_N$, and (a) exploits Theorem 1 from free probability theory, since the phase drift matrices $\boldsymbol{\Psi}_{t_0}(t)$ and $\boldsymbol{\Psi}_{t_0}^H(t)$ are free from $\mathbf{g}_k(t_0) \mathbf{g}_k^H(t_0)$ and $\frac{\hat{\mathbf{g}}_l \hat{\mathbf{g}}_l^H}{\|\hat{\mathbf{g}}_l\|^2}$. For notational simplicity, we define $I = \mathbb{E} \left[\text{tr} (\mathbf{X}_l^H \mathbf{g}_k(t_0) \mathbf{g}_k^H(t_0) \mathbf{X}_l \boldsymbol{\psi}_b \boldsymbol{\psi}_b^H) \right]$, which can be further expanded as

$$\begin{aligned} I &= \mathbb{E} \left[\text{tr} (\mathbf{Y}_{lk}^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{Y}_{lk} \mathbf{g}_k \mathbf{g}_k^H) \right] + \text{tr} (\beta_k \mathbf{X}_l^H \mathbf{X}_l (\boldsymbol{\Sigma}_b - \beta_k (\mathbf{W}_k^b + \mathbf{U}_k^b)) \otimes \mathbf{I}_N) + \\ & \mathbb{E} \left[\text{tr} \left(\mathbf{X}_l^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{X}_l \left(\mathbf{U}_k^b \otimes \text{diag} (g_k^{(1)}, \dots, g_k^{(N)}) \right) \right) \right], \end{aligned} \quad (42)$$

where

$$\mathbf{Y}_{lk} = \boldsymbol{\Theta}_k^H(t_0) \mathbf{X}_l \left[\boldsymbol{\Theta}_k^H(\bar{B}_{b-1} + 1) \omega_k(\bar{B}_{b-1} + 1), \dots, \boldsymbol{\Theta}_k^H(t_0) \omega_k(t_0) \right]^T. \quad (43)$$

Denoting the t^{th} column of \mathbf{I}_N by $\mathbf{e}_t^N \in \mathbb{C}^{N \times 1}$, the first term on the right hand side of (42), denoted by I_1 , can be expanded as

$$\begin{aligned} I_1 &= \sum_{n_1, n_2, b_1, b_2} [\beta_k \mathbf{X}_l \mathbf{e}_{b_1}^{B_b} \otimes \mathbf{I}_N]_{n_1 n_1} [\beta_k \mathbf{X}_l \mathbf{e}_{b_2}^{B_b} \otimes \mathbf{I}_N]_{n_2 n_2}^H \times \omega_k(b_1) \omega_k^*(b_2) \Theta(n_1, n_2, b_1, b_2, t_0) \\ &= \left| \text{tr} (\beta_k \mathbf{X}_l (\boldsymbol{\Theta}_{\sigma(t_0)}^b \boldsymbol{\omega}_k \otimes \mathbf{I}_N)) \right|^2 + \text{tr} (\beta_k^2 \mathbf{X}_l^H \mathbf{X}_l (\mathbf{W}_k^b \otimes \mathbf{I}_N)) \\ &+ \sum_{|n_1 - n_2| \leq \frac{N}{N_0}} \beta_k^2 (\mathbf{e}_{n_1}^N)^H \mathbf{X}_l ((\mathbf{W}_k^b - \boldsymbol{\Theta}_{\sigma(t_0)}^b \boldsymbol{\omega}_k \boldsymbol{\omega}_k^H \boldsymbol{\Theta}_{\sigma(t_0)}^b) \otimes \mathbf{e}_{n_1}^N (\mathbf{e}_{n_2}^N)^H) \mathbf{X}_l^H \mathbf{e}_{n_2}^N, \end{aligned} \quad (44)$$

where the expectation with respect to the phase drift, $\Theta(n_1, n_2, b_1, b_2, t_0)$, depends on the number of LOs, N_o , and is given by $\Theta(n_1, n_2, b_1, b_2, t_0) =$

$$\mathbb{E} \left[e^{j(\theta_k^{n_1}(b_1) - \theta_k^{n_1}(t_0) - \theta_k^{n_2}(b_2) + \theta_k^{n_2}(t_0))} \right] = \begin{cases} e^{-\frac{\sigma_\psi^2 + \sigma_\phi^2}{2} |b_1 - b_2|} & |n_1 - n_2| \leq \frac{N}{N_o}, \\ e^{-\frac{\sigma_\psi^2 + \sigma_\phi^2}{2} |t_0 - b_1|} e^{-\frac{\sigma_\psi^2 + \sigma_\phi^2}{2} |t_0 - b_2|} & |n_1 - n_2| > \frac{N}{N_o}. \end{cases} \quad (45)$$

Furthermore, we rewrite $\mathbf{U}_k^b = (\kappa_t^{\text{MT}} + \kappa_r^{\text{BS}})p_\tau \sum_{t=1}^{B_b} \mathbf{e}_t^{B_b} (\mathbf{e}_t^{B_b})^H$ and $\text{diag}(g_k^{(1)}, \dots, g_k^{(N)}) = \sum_{n=1}^N |(\mathbf{e}_n^N)^H \mathbf{g}_k|^2 \mathbf{e}_n^N (\mathbf{e}_n^N)^H$. Using these results in the third term on the right hand side of (42), denoted by I_2 , we obtain

$$I_2 = \beta_k^2 \text{tr}(\mathbf{X}_l^H \mathbf{X}_l (\mathbf{U}_k^b \otimes \mathbf{I}_N)) + \sum_{n=1}^N \beta_k^2 (\mathbf{e}_n^N)^H \mathbf{X}_l (\mathbf{U}_k^b \otimes \mathbf{e}_n^N (\mathbf{e}_n^N)^H) \mathbf{X}_l \mathbf{e}_n^N. \quad (46)$$

Applying (44) and (46) in (41) and exploiting $\mathbb{E} \left[\left(\frac{1}{N} \text{tr}(\Psi_{t_0}^H(t)) \right)^2 \right] = \frac{1-\epsilon}{N_o} + \epsilon$, we obtain the result in Lemma 3 for $k, l \in \mathcal{S}_b$.

For the case of $l \notin \mathcal{S}_b$, the multiuser interference term simplifies to

$$\mathbb{E} \left[|\mathbf{g}_k^H \Theta_k^H(t) \mathbf{f}_l|^2 \right] = \mathbb{E} \left[\left| \mathbf{g}_k^H(t_0) \Psi_{t_0}^H(t) \frac{\hat{\mathbf{g}}_l}{\|\hat{\mathbf{g}}_l\|} e^{j(\phi_k(t) - \phi_k(t_0))} \right|^2 \right] = \beta_k, \quad (47)$$

where the last equality follows from the independence of \mathbf{g}_k , $\hat{\mathbf{g}}_l$, $l \notin \mathcal{S}_b$, and $\Psi_{t_0}^H(t)$. This completes the proof.

E. Proof of Lemma 4

The AN leakage power received at the k^{th} MT in time slot t can be expanded as

$$L_{\text{AN}}^k(t) = \mathbb{E} \left[\text{tr}(\hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H \Psi_{t_0}^H(t) \mathbf{A} \mathbf{A}^H \Psi_{t_0}(t)) \right] + \mathbb{E} \left[\mathbf{e}_k^H(t_0) \Psi_{t_0}^H(t) \mathbf{A} \mathbf{A}^H \Psi_{t_0}(t) \mathbf{e}_k(t_0) \right]. \quad (48)$$

By using Theorem 1, the first term in (48) can be further expanded as

$$\beta_k L + (\mathbb{E} [\text{tr}(\hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H \mathbf{A} \mathbf{A}^H)] - \beta_k L) \mathbb{E}_\psi \left[\left(\frac{1}{N} \text{tr}(\Psi_{t_0}(t)) \right)^2 \right] = \beta_k L \left(1 - \frac{1}{N_o} \right) (1 - \epsilon), \quad (49)$$

since phase drift matrices $\Psi_{t_0}(t)$ and $\Psi_{t_0}^H(t)$ are free from $\hat{\mathbf{g}}_k \hat{\mathbf{g}}_k^H$ and $\mathbf{A} \mathbf{A}^H$. Furthermore, we exploited $\hat{\mathbf{g}}_k^H \mathbf{A} = \mathbf{0}$, $1 \leq k \leq K$, which holds for the NS AN precoder.

The second term in (48) is equal to $\beta_k L(1 - \lambda_k)$, with λ_k as defined in Lemma 2, due to the mutual independence of the estimation error vector $\mathbf{e}_k(t_0)$, the phase drift matrix $\Psi_{t_0}(t)$, and the AN precoder \mathbf{A} . Combining these two terms completes the proof.

F. Proof of Lemma 5

For the G-NS AN precoder, we rewrite the leakage power received at the k^{th} MT in time slot t as

$$L_{\text{AN}}^k = \sum_{m=1}^{M_o} \mathbb{E} \left[\left(\mathbf{g}_k^{(m)} \right)^H \left(\Theta_k^{(m)}(t) \right)^H \mathbf{A}_{(m)} \mathbf{A}_{(m)}^H \Theta_k^{(m)}(t) \mathbf{g}_k^{(m)} \right], \quad (50)$$

where $\mathbf{g}_k^{(m)} \in \mathbb{C}^{N/M_o \times 1}$ contains the $((m-1)N/M_o + 1)^{\text{th}}$ to the $(mN/M_o)^{\text{th}}$ elements of vector \mathbf{g}_k , $1 \leq m \leq M_o$, and $\mathbf{\Theta}_k^{(m)}(t) \in \mathbb{C}^{N/M_o \times N/M_o}$ is a diagonal matrix with the $((m-1)N/M_o + 1)^{\text{th}}$ to the $(mN/M_o)^{\text{th}}$ elements of matrix $\mathbf{\Theta}_k(t)$ on its main diagonal. Using similar steps as in Appendix E but with N_o/M_o substituted by N_o for calculation of the expectation terms in (50), we obtain (29). This completes the proof.

G. Proof of Proposition 1

We first adopt Jensen's inequality to upper bound the eavesdropper's capacity at time interval t , as $C_E(t) \leq \log_2(1 + \mathbb{E}[\gamma_E(t)])$, with $\gamma_E(t)$ given in (15). One further step to simplify $\mathbb{E}[\gamma_E(t)]$ requires the statistical independence between $\mathbf{g}_E^k(t)$ and the matrix $\mathbf{X} = \mathbf{G}_E^H \mathbf{\Psi}^H(t)(q\mathbf{A}\mathbf{A}^H + \mathbf{\Upsilon}_t^{\text{BS}})\mathbf{\Psi}(t)\mathbf{G}_E$. We note that the entries of \mathbf{G}_E are independent complex Gaussian random variables. On the other hand, for the G-NS AN precoder, the columns of $\mathbf{\Psi}^H(t)\mathbf{A} \in \mathbb{C}^{N \times L}$ form an orthonormal basis. Hence, $\mathbf{G}_E^H \mathbf{\Psi}^H(t)\mathbf{A}$ also has independent complex Gaussian entries, which are independent from the entries of $\mathbf{g}_E^k(t)$ [9]. Besides, the term $\mathbf{G}_E^H \mathbf{\Psi}^H(t)\mathbf{\Upsilon}_t^{\text{BS}}\mathbf{\Psi}(t)\mathbf{G}_E$ converges to a deterministic diagonal matrix for $N \rightarrow \infty$, which is obviously independent of $\mathbf{g}_E^k(t)$. Therefore, $\mathbb{E}\left[\mathbf{g}_E^k(t)\mathbf{X}^{-1}(\mathbf{g}_E^k(t))^H\right]$ can be rewritten as $\mathbb{E}\left[\mathbf{g}_E^k(t)\mathbb{E}[\mathbf{X}^{-1}](\mathbf{g}_E^k(t))^H\right]$, with $\mathbb{E}[\mathbf{X}^{-1}]$ approximated as a scaled identity matrix as in [7, Appendix C]. This leads to the upper bound given in (31).

REFERENCES

- [1] J. Zhu, R. Schober, and V. K. Bhargava, "Physical layer security for massive MIMO systems impaired by phase noise," *Proc. IEEE Intern. Workshop on Sig. Proc. Advances in Wireless Commun. 2016 (SPAWC 2016)*, pp. 1-5, Edinburgh, UK, Jul. 2016.
- [2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Sig. Proc. Mag.*, vol. 30, no. 1, pp. 40-46, Jan. 2013.
- [3] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of BS antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.
- [4] 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.
- [5] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640-2651, Aug. 2011.
- [6] 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.

- [7] J. Zhu, R. Schober, and V. K. Bhargava, "Secure transmission in multicell massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 476-4781, Sept. 2014.
- [8] J. Zhu and W. Xu, "Securing massive MIMO via power scaling," *IEEE Commun. Letters*, vol. 20, no. 5, pp. 1014-1017, May 2016.
- [9] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas-I: The MISOME wiretap channel," *IEEE Trans. Inform. Theory*, vol. 56, no. 7, pp. 3088-3104, July 2010.
- [10] P. Gopala, L. Lai, and H. El Gamal, "On the secrecy capacity of fading channels," *IEEE Trans. Inform. Theory*, vol. 54, no. 10, pp. 4687-4698, Oct. 2008.
- [11] T. Liu and S. Shamai (Shitz), "A note on the secrecy capacity of the multi-antenna wiretap channel," *IEEE Trans. Inform. Theory*, vol. 55, no. 6, pp. 2547-2553, Jun. 2009.
- [12] G. Geraci, M. Egan, J. Yuan, A. Razi, and I.B. Collings, "Secrecy sum-rates for multi-user MIMO regularized channel inversion precoding," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3472-3482, Nov. 2012.
- [13] G. Geraci, J. Yuan, and I. B. Collings, "Large system analysis of linear precoding in MISO broadcast channels with confidential messages," *IEEE Journal on Sel. Areas in Commun.*, vol. 31, no. 9, pp. 1660-1671, Sept. 2013.
- [14] S. Goel and R. Negi, "Guaranteeing secrecy using artificial noise," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2180-2189, Jun. 2008.
- [15] J. Zhu, R. Schober, and V. K. Bhargava, "Linear precoding of data and artificial noise in secure massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2245-2261, Mar. 2016.
- [16] J. Wang, J. Lee, F. Wang, and T. Q. S. Quek, "Jamming-aided secure communication in massive MIMO Rician channels," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6854-6868, Dec. 2015.
- [17] X. Chen, L. Lei, H. Zhang, and C. Yuen, "Large-scale MIMO relaying techniques for physical layer security: AF or DF?" *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 5135-5146, Sept. 2015.
- [18] S. Im, H. Jeon, J. Choi, and J. Ha, "Secret key agreement with large antenna arrays under the pilot contamination attack," *IEEE Trans. Wireless Commun.*, vol. 14, no. 12, pp. 6579-6594, Dec. 2015.
- [19] Y. Basciftci, C. Koksall, and A. Ashikhmin, "Securing massive MIMO at the physical layer," in *Proc. of IEEE Conference on Communications and Network Security (CNS) 2015*, Florence, Italy, Sept. 2015.
- [20] Y. Wu, R. Schober, D. W. K. Ng, C. Xiao, and G. Caire, "Secure massive MIMO transmission with an active eavesdropper," *IEEE Trans. Inform. Theory*, vol. 62, no. 7, pp. 3880-3900, Jul. 2016.
- [21] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inform. Theory*, vol. 60, no. 11, Nov. 2014.
- [22] T. Schenk, "RF imperfections in high-rate wireless systems: Impact and digital compensation," New York, NY, USA: Springer-Verlag, 2008.
- [23] R. Krishnan, M. R. Khanzadi, N. Krishnan, Y. Wu, A. Graell i Amat, T. Eriksson, and R. Schober, "Linear massive MIMO precoders in the presence of phase noise- A large-scale analysis," *IEEE Trans. Vehicular Tech.*, vol. PP, no. 99, pp. 1-1, Jun. 2015.
- [24] S. K. Mohammed and E. G. Larsson, "Per-antenna constant envelope precoding for large multi-user MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 1059-1071, Mar. 2013.
- [25] J. Zhu, N. Wang, and V. K. Bhargava, "Per-antenna constant envelope precoding for secure transmission in large-scale MISO systems," *Proc. IEEE/CIC Intern. Conf. Commun. in China 2015 (ICCC 2015)*, pp. 1-6, Shenzhen, China, Nov. 2015.

- [26] E. Björnson, M. Matthaiou, and M. Debbah, “Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4353–4368, Aug. 2015.
- [27] X. Xia, D. Zhang, K. Xu, W. Ma, and Y. Xu, “Hardware impairments aware transceiver for full-duplex massive MIMO relaying,” *IEEE Trans. Sig. Proc.*, vol. 63, no. 24, pp. 6565–6580, Nov. 2015.
- [28] M. Wenk, “MIMO-OFDM testbed: Challenges, implementations, and measurement results (Microelectronics),” Konstanz, Germany: Hartung-Gorre, 2010.
- [29] X. Zhou and M. R. McKay, “Secure transmission with artificial noise over fading channels: achievable rate and optimal power allocation,” *IEEE Trans. Veh. Technol.*, vol. 59, pp. 3831–3842, Jul. 2010.
- [30] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, “Uplink performance of time-reversal MRC in massive MIMO systems subject to phase noise,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 711–723, Feb. 2015.
- [31] G. Colavolpe, A. Barbieri, and G. Caire, “Algorithms for iterative decoding in the presence of strong phase noise,” *IEEE J. Sel. Area Commun.*, vol. 23, no. 9, pp. 1748–1757, Sept. 2005.
- [32] R. Krishnan, M. R. Khanzadi, T. Eriksson, and T. Svensson, “Soft metrics and their performance analysis for optimal data detection in the presence of strong oscillator phase noise,” *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2385–2395, Jun. 2013.
- [33] Analog Devices, “FR agile transceiver,” AD9364 datasheet, Feb. 2014. [Revised July 2014]
- [34] F. Hiai and D. Petz, “The semicircle law, free random variables and entropy,” American Mathematical Society, 2006.
- [35] P.-H. Lin, S.-H. Lai, S.-C. Lin, and H.-J. Su, “On optimal artificial-noise assisted secure beamforming for the fading eavesdropper channel,” *IEEE J. Select. Areas Commun.*, vol. 31, no. 9, pp. 1728–1740, Sept. 2013.