Performance Analysis of Self-Interference Cancellation in Full-Duplex Massive MIMO Systems: Subtraction Versus Spatial Suppression

Soo-Min Kim, Student Member, IEEE, Yeon-Geun Lim, Member, IEEE, Linglong Dai, Fellow, IEEE, and Chan-Byoung Chae, Fellow, IEEE

Abstract—Massive multiple-input multiple-output (MIMO) and full-duplex (FD) are promising candidates for achieving the spectral efficiency to meet the needs of 5G communications. One essential key to realizing practical FD massive MIMO systems is how to effectively mitigate the self-interference (SI). Conventionally, however, the performance comparison of different SI methods by reflecting the actual channel characteristics was insufficient in the literature. Accordingly, this paper presents a performance analysis of SI cancellation (SIC) methods in FD massive MIMO systems. Analytical and numerical results confirm that, in an imperfect channel-estimation case, the ergodic rates performance of the spatial suppression in the uplink outperforms those of the SI subtraction, due to the correlation between the precoder and the estimation error of the SI channel. In addition, we discuss which method performs better under different given system constraints such as uplink and downlink sum rates, the total transmit power, and the power scaling law.

Index Terms—Massive MIMO, full-duplex, self-interference cancellation, and spatial suppression.

I. INTRODUCTION

The fifth-generation (5G) communication systems can support various types of services, including enhanced mobile broadband (eMBB), massive machine-type communication, and ultra-reliable low-latency communication [2], [3]. As mobile data traffic is expected to increase, eMBB has become a crucial part of 5G communications. To support the tremendous data traffic of 5G, researchers should jointly consider key issues such as spectral efficiency (bps/Hz) and spectrum extension (Hz/cell). As researchers developed key issues over the past few decades, massive multiple-input multiple-output (MIMO) technology has been suggested as a candidate to achieve high spectral efficiency. At the same time, researchers [4], [5] intending to double the spectrum extension without shifting the system operating frequency band have studied full-duplex (FD) transmission.

To maximize spectral efficiency up to tens of bps/Hz and conserve energy, the researchers in [6], [7], [8], [9], and [10] have proposed massive MIMO with tens of antennas as a core technology of 5G. They considered using simple linear precoders and filters to mitigate the interferences in massive MIMO systems. One key challenge in commercializing massive MIMO systems is the estimation of the high-dimensional MIMO channel matrix. This paper focuses on the channel estimation based on practical channel characteristics [7], [8], [9]. Specifically, the authors in [7] and [8] studied the downlink performance of the maximum ratio transmission (MRT) and zero-forcing (ZF) precoders while assuming either perfect or imperfect channel estimation at the base station (BS). The authors in [9] investigated the uplink performance of the maximum ratio combining (MRC), ZF, and minimum mean square error (MMSE) filters with perfect or imperfect channel estimation.

Another relevant technology is FD transmission, which, compared to half-duplex transmission, can double the spectrum efficiency [5], [11], [12], [13] and reduce transmission latency [14]. In contrast to half-duplex transmission, self-interference (SI), which is a transmitted downlink signal but is directly received at the BS when receiving an uplink signal, should be mitigated. The level of SI is usually higher than that of the receiver noise floor, and thus, any FD system requires SI cancellation (SIC) in both the analog and digital domains [15]. There are typically two types of classical SIC methods in a digital-domain, that is, SI subtraction and spatial suppression. Specifically, SI subtraction subtracts replicated interference signal from the input signal, while spatial suppression applies multi-antenna techniques that utilize extra
degrees of freedom offered by spatial dimensions for linear receive and transmit filtering (it is now considered in the form of null-space projection (NSP) [16]). From the digital-domain SIC perspective, the authors in [17] compared SI subtraction and spatial suppression in a MIMO relay system with a small number of antennas (up to four antennas) at the BS and user. They showed that the performance of the spatial suppression was depended on the rank of the SI channel. The authors paid little attention to the performance analysis of practical massive MIMO systems, such as user rate and uplink/downlink performance, because they focused on introducing advantages and disadvantages for each scheme according to the number of antennas and the rank of the channel. In a practical small MIMO system, SI subtraction is a widely considered SIC technique owing to the limitation of the requirement of additional antennas. The researchers in [18], [19], [20], and [21] prototyped SI subtraction on a software-defined radio platform in real time. They combined a dual-polarization antenna-based analog cancellation with SI subtraction and realized a spectral efficiency 1.9 times higher than that of the half-duplex transmission. Nevertheless, spatial suppression has received limited attention in small MIMO systems.

As aforementioned, in 5G and beyond 5G systems, integrated techniques are required to efficiently fulfill the requirements of eMBB. In particular, we should integrate spectrum extension technologies with massive MIMO to enhance both the data rate of the network and each user. Conventional massive MIMO systems assume a small number of antennas at the users (for example, a single antenna) with a per-user rate constraint, which is limited by the used modulation technologies and allocated bandwidth. In addition, massive MIMO systems are typically launched in the sub-6-GHz range, whereas hybrid beamforming technologies are considered in the millimeter-wave range [22]. Therefore, FD is a potential solution to the spectrum extension of massive MIMO systems. In both [23] and [24], the researchers investigated asymptotic ergodic sum rates and channel estimation methods in FD-assisted massive MIMO systems. They considered the condition of specific system constraints wherein the performance of an FD system outperforms that of a half-duplex system. The authors in [25] proposed spatial suppression-based algorithms such as extended ZF and extended MMSE precoders. As the conventional massive MIMO system serves a relatively smaller number of users than the number of antennas at the BS, the extended precoders can effectively mitigate SI by utilizing the dimensional feasibility in the spatial domain. As an interference-free channel was assumed in [25], the analysis is not general. The researchers in [26] analyzed the ratio of transmit antennas and receive antennas in terms of the numbers of both the downlink and uplink users. The authors in [27] utilized a spatial suppression precoder based on singular-value decomposition and proposed an uplink power control mechanism for cellular systems using a Poisson point process model. Furthermore, the authors in [28] showed the performance trade-off between the degree of freedom and the SI mitigation level, while the authors in [29] presented a combination of spatial suppression and an analog domain cancellation solution that can relax the requirement of the number of antennas and reduce the complexity of SI cancellation. They demonstrated the performance analysis and feasibility of the spatial suppression method under imperfect channel estimation. However, there is insufficient explanation of whether SI suppression is optimal in the considered environments.

In this paper, motivated by the potential of utilizing spatial suppression in massive MIMO systems but the lack of prior studies on its advantages compared to SI subtraction, we compared these two SIC methods in the digital domain. Prior works typically model SI channel coefficients as simple independent identically distributed (i.i.d.) complex Gaussian random variables. In this prior channel model, the researchers assumed that the uplink performances of both SIC methods were the same [25], [26]. Thus, SI subtraction was always a better choice, because spatial suppression requires additional spatial resources at the downlink, which causes a difference in the number of antennas required to obtain a diversity gain for the downlink. In a practical channel, however, an SI channel would be highly correlated because of the higher power in the direct path. Therefore, in the imperfect channel, spatial suppression can be better for the uplink, depending on the magnitude of the channel estimation error. From this point of view, we analyzed the performance according to the diversity gain, number of antennas, and channel estimation error by performing the perfect channel analysis together. Motivated by [17], in small FD MIMO systems, SI subtraction is better because the diversity gain for the downlink is limited because there are fewer antennas. However, in FD massive MIMO systems, because this limit is weaker, spatial suppression may be better. We mathematically prove this content accurately. In addition, the considered small FD MIMO system is not sufficient to meet the requirements of massive MIMO systems in beyond 5G. This paper presents a comparison of the two SIC methods with the following considerations in a system: i) channel estimation error, ii) spatial correlation, and iii) asymmetric data traffic with a non-full buffer, and power scaling law at the downlink. Based on the analysis and numerical results, we determined which method is better in a certain scenario. In our prior work [1], two methods were compared in perfect and imperfect SI channel conditions, particularly in uncorrelated conditions. Furthermore, the two methods were compared through mathematical analysis and proofs in correlated conditions. The performance was also evaluated by considering the practical system constraints mentioned above. Through these analyses, results were obtained using a system-level simulator based on a 3D ray-tracing tool to obtain the real SI channel. The additive quantization noise model (AQNM) was also considered as a channel estimation error model to design the system more practically. Specifically, the main contributions of this study are summarized as follows.

- **Performance analysis with imperfect channel estimation:** First, we provide the performance analysis to show that the uplink performance of spatial suppression is better than that of SI subtraction with imperfect channel estimation. The channel estimation error would be high, particularly in practical full-digital MIMO systems that usually use analog-to-digital converters (ADCs) with few bits (for example, 4-5 bits) for a cost-effective design [30], [31]. The SIC algorithms
having improved performance with imperfect channel estimation are yet to be determined. We also derive the achievable ergodic sum rates of the two SIC methods, then we show that the uplink rate of spatial suppression is better than that of SI subtraction with an imperfect channel estimation error.

• **Performance analysis with spatial correlation:** As noted above, SI channels would be highly correlated. We prove this baseline channel condition by measuring the channel parameters with a 3D ray-tracing tool. Next, we investigate the impact of spatial correlation on the expected SI power. Based on this analysis, we conclude that the uplink rate of spatial suppression improves beyond that of SI subtraction as the spatial correlation increases. Notably, in the conventional analysis, these correlation terms of the SI channel are assumed as independent conditions [25], [26].

• **Performance comparison with system constraints:** Based on our analysis, we compare the two SIC methods when the system constraints are considered. These constraints include the ratio of downlink and uplink traffic, ratio of transmit and receive RF chains, transmit power at the BS, and power scaling law at the BS. Because of the performance trade-off between these two SIC schemes in a certain link direction, our analysis shows that each method exhibits different performances for different system constraints. These comparisons provide insights into the SIC selection criterion for certain system requirements.

The remainder of this paper is organized as follows. In Section II, we introduce the system model and problem statement with respect to precoding/receive combining and SIC methods. In Section III, we present certain mathematical motivations and preliminaries that are useful for analysis. Then we compare the SIC algorithms under different channel conditions. In Section IV, we compare SIC algorithms with respect to the system constraints, such as the transmit power and traffic load. The numerical results are presented in Section V. Finally, Section VI presents our conclusions.

II. SYSTEM MODEL

In this section, we firstly introduce the system model of an asymmetric FD system with massive MIMO and the basic notation used in this paper.¹ Then SINR, precoding/receive combining matrix design, and SIC methods are presented.

### A. FD BS With Massive Arrays

We consider FD multi-user MIMO channels, as illustrated in Fig. 1, with a large number of antennas at the BS that serves \(K_{DL}\) downlink users and \(K_{UL}\) uplink users equipped with one antenna for each. It is assumed that the BS is equipped with \(M_{\text{total}}\) (\(M_{\text{total}} \gg K\)) RF chains that are divided into \(M\) transmit and \(N\) receive RF chains by considering their physical size, whereas each user has a single transmit and single receive RF chain. All users are assumed to operate in a half-duplex mode, so the hardware complexity for the SIC at the user is zero, which is beneficial for their battery uses. This concept has been widely adopted in massive MIMO systems in cellular scenarios [24], [25], [27]. In addition, we assume an asymmetric data traffic load (i.e., the downlink has more data traffic than the uplink). In cellular systems, there hardly exists any user equipment (UE) requiring contents with high data rates for both links within the same coherence time. This stands in contrast to device-to-device scenarios. Thus, supporting the best-performing UE in a half-duplex mode can achieve greater scheduling gain than that in an FD mode. We consider that both the RF chain usage and data traffic load are asymmetric. Therefore, we assume \(M \geq N + K_{DL}\) and \(N \geq K_{UL}\). This asymmetric antenna usage ensures a small and low-cost BS by reducing the receive RF chains that include low-noise amplifiers (LNA) within the fixed size of \(M_{\text{total}}\). Without loss of generality, let us consider that each propagation gain of the SI channels between BS antennas, denoted by \(g_{SI}^{kl}\), is the same as their expectation. In addition, the notations defined in this paper are shown in Table I.

### B. Channel Model

1) **Spatially Correlated Channels:** The downlink, uplink, and SI channels are modeled by the \(K_{DL} \times M\) matrix \(H_{DL} = B_{RX(DL)}^{1/2} H_{DL,\text{id}} R_{TX(DL)}^{1/2} N \times K_{UL}\) matrix \(H_{UL} = B_{RX(UL)}^{1/2}\)

¹Throughout this paper, we use upper- and lower-case boldfaces to denote matrix \(\mathbf{A}\) and vector \(\mathbf{a}\), respectively. We denote the inverse, transpose, and Hermitian of matrix \(\mathbf{A}\) as \(\mathbf{A}^{-1}\), \(\mathbf{A}^T\), and \(\mathbf{A}^\dagger\), respectively. In addition, we denote the conjugate of a scalar \(a\) as \(a^\ast\). \(\text{Var}[X]\) and \(\mathbb{E}[X]\) respectively represent the variance and average of the random variable \(X\), while \(\text{Cov}(X, Y)\) represents the covariance between two random variables \(X\) and \(Y\). \(\text{Sort}(X_k)\) represents the covariance between two random variables \(X\) and \(Y\). \(\text{Sort}(X_k)\) is referred to as a sort function of sequence \(X_k\) in the descending order.
TABLE I
THE BASIC NOTATIONS USED IN THIS PAPER

| Symbol | Description |
|--------|-------------|
| $M_{\text{tot}} = \{M + N\}$ | The number of RF chains (transmit RF chains + receive RF chains) |
| $K_{UL}/K_{UL}$ | The number of downlinks/ uplink users |
| $\rho_{ij} = \rho_{ij}/\rho_{ij}$ | The total transmit signal-to-noise ratio (SNR) at the BS |
| $\rho_{ij} = \rho_{ij}/\rho_{ij}$ | The transmit SNR at the user |
| $\rho_{ij}^{\text{DL}} = \rho_{ij}/\rho_{ij}$ | The received downlink SNR at the users |
| $\rho_{ij}^{\text{UL}} = \rho_{ij}/\rho_{ij}$ | The received SI SNR at the users |
| $\rho_{ij}^{\text{UL}} = \rho_{ij}/\rho_{ij}$ | The received uplink SNR at the BS |
| $\rho_{ij}^{\text{UL}} = \rho_{ij}/\rho_{ij}$ | The received SI SNR at the BS |
| $\rho_{ij}^{\text{UL}} = \rho_{ij}/\rho_{ij}$ | The total transmit power, the transmit power of the user, the noise power of the additive white Gaussian noise (AWGN) of the channel |
| $\vartheta_{ij}^{\text{DL}}/\vartheta_{ij}^{\text{UL}}$ | The propagation gain between the BS and the $k_{ij}$-th DL user / the $k_{ij}$-th UL user |
| $\vartheta_{ij}^{\text{DL}}/\vartheta_{ij}^{\text{UL}}$ | The propagation gain from the $k_{ij}$-th user to the $k_{ij}$-th user |
| $H/R/E$ | The channel/ correlation/ estimated error matrix |
| $\tau$ | The number of pilot symbols for the channel |
| $\epsilon$ | The variance of the estimation error for the channel |
| $\alpha_{\text{nc}}$ | The level of passive analog SIC |
| $\alpha_{\text{nc}}, n_{\text{tx}}$ | The transmitter noise variance, the transmitter noise vector |
| $n_{\text{tx}}, n_{\text{tx}}$ | The AWGN vector at SI |
| $n_{\text{tx}}/n_{\text{tx}}$ | The AWGN vector at downlink/ uplink |
| $s/\pi$ | The transmit symbol at downlink/ uplink |
| $G, W$ | The transmit precoding matrix, the receive combining matrix |
| $\gamma$ | The SNR |
| $\Omega$ | The SI power term |

$H_{\text{UL, id}} \mathcal{R}_{\text{TX(UL)}}^{1/2}$, and $N \times M$ matrix $H_{\text{SI}} = \mathcal{R}_{\text{RX(SI)}}^{1/2}H_{\text{SL, id}} \mathcal{R}_{\text{TX(SI)}}^{1/2}$, which are based on the Kronecker model. The element of $\mathcal{R}_{\text{TX}}$ and $\mathcal{R}_{\text{RX}}$ is $\mathcal{R}_{\text{TX/RX}}^{ij} = J_0(2\pi d_{\text{TX/RX}}^{ij}/\lambda)$, where $d_{\text{TX/RX}}^{ij}$ is the distance between the $i$-th array and the $j$-th array of the transmit/receive antenna at the BS, $J_0$ and $\lambda$ are a zero-order Bessel function of the first kind and a carrier wavelength, respectively. The specific parameters used in the simulation are presented in Section V. The elements of $H_{\text{UL, id}}$, $H_{\text{UL, id}}$, and $H_{\text{SL, id}}$ are i.i.d. complex Gaussian random variables with zero mean and unit variance. The transmit correlation matrices in the downlink, uplink and SI channels are denoted as $R_{\text{TX(DL)}}$, $R_{\text{TX(UL)}}$, and $R_{\text{TX(SI)}}$, respectively, while the receive correlation matrices in these channels are denoted as $R_{\text{RX(DL)}}$, $R_{\text{RX(UL)}}$, and $R_{\text{RX(SI)}}$, respectively. It should be noted that, in contrast to the uplink and downlink channels, the SI channel is highly correlated in the spatial domain. Despite the high level of analog SIC, the power from direct paths (which are line-of-sight (LOS) paths, reflected paths around the antenna, and circuit leakage) is still greater than the power from non-line-of-sight (NLOS) paths. In addition, because the angular spread of the LOS path is very small, the spatial correlation of the SI channel is high. Thus, it is assumed that the SI channel is more highly correlated than the downlink and uplink channels. Furthermore, because the SI channel has a strong LOS component (a Rician channel model), a highly correlated channel and a real SI channel (which are equivalent channel models with a strong LOS component) are considered to supplement the SI channel modeling.

2) MMSE-Based Channel Estimation: Let $\hat{H}_{\text{DL}}, \hat{H}_{\text{UL}}$, and $\hat{H}_{\text{SI}}$ be the estimated channel matrices of $H_{\text{DL}}, H_{\text{UL}}$, and $H_{\text{SI}}$, respectively. Because estimation errors are independent of the original channels, the estimated channels can be modeled using estimated error matrices $(E)$ that have i.i.d. elements with the distribution $\mathcal{CN}(0, e_2)$. The specific parameters used in the simulation are presented in Section V. The elements of $\hat{H}_{\text{UL, id}}$, $\hat{H}_{\text{UL, id}}$, and $\hat{H}_{\text{SL, id}}$ are i.i.d. complex Gaussian random variables with zero mean and unit variance. The transmit correlation matrices in the downlink, uplink and SI channels are denoted as $R_{\text{TX(DL)}}$, $R_{\text{TX(UL)}}$, and $R_{\text{TX(SI)}}$, respectively, while the receive correlation matrices in these channels are denoted as $R_{\text{RX(DL)}}$, $R_{\text{RX(UL)}}$, and $R_{\text{RX(SI)}}$, respectively. It should be noted that, in contrast to the uplink and downlink channels, the SI channel is highly correlated in the spatial domain. Despite the high level of analog SIC, the power from direct paths (which are line-of-sight (LOS) paths, reflected paths around the antenna, and circuit leakage) is still greater than the power from non-line-of-sight (NLOS) paths. In addition, because the angular spread of the LOS path is very small, the spatial correlation of the SI channel is high. Thus, it is assumed that the SI channel is more highly correlated than the downlink and uplink channels. Furthermore, because the SI channel has a strong LOS component (a Rician channel model), a highly correlated channel and a real SI channel (which are equivalent channel models with a strong LOS component) are considered to supplement the SI channel modeling.

2) MMSE-Based Channel Estimation: Let $\hat{H}_{\text{DL}}, \hat{H}_{\text{UL}}$, and $\hat{H}_{\text{SI}}$ be the estimated channel matrices of $H_{\text{DL}}, H_{\text{UL}}$, and $H_{\text{SI}}$, respectively. Because estimation errors are independent of the original channels, the estimated channels can be modeled using estimated error matrices $(E)$ that have i.i.d. elements with the distribution $\mathcal{CN}(0, e_2)$. It is assumed that an MMSE-based channel estimation is used at the BS, which is also used in the 3rd Generation Partnership Project (3GPP) [32]. Other estimation schemes, such as least square-based estimation, can also be adapted to this scheme [9]. In practice, the estimation error of the SI channel is affected by transmitter noise caused by hardware impairments, such as the nonlinearity of the power amplifier, a quantization error, and phase noise [33]. Thus, it is difficult to model the estimated SI channel while taking into consideration hardware impairments. The variance of the estimation error is simply modeled as the normalized-MSE (NMSE) value in (II-B.2) between the estimated channel and the original channel. This error depends on the performance of the estimation algorithm and the specifications of the hardware. Because we consider a scenario of ADCs with a few bits (e.g., 4-5 bits) for realizing a cost-effective design [30], [31], the variance of the estimation error of the SI channel is given by $e_2^2 = \text{NMSE}$, which is derived from the AQNM as shown in II-B.3.

Remark 1: It should be noted that the variance of the estimation error of the SI channel can also be modeled by the number of pilot symbols. Let the pilot symbols be transmitted orthogonally under TDD-based transmission. The numbers of pilot symbols for the SI channels are given by $\tau_{\text{si}} (\geq N)$ ($N$ transmit RF chains are connected to $N$ receive antennas through the switches). From [24], the variance of the estimation error of the SI channel is given by

$$e_2^2 = \frac{\alpha_{\text{tx}}N e_{\text{si}}^2}{\tau_{\text{si}} e_{\text{si}}^2} = \frac{\alpha_{\text{tx}}N e_{\text{si}}^2}{\tau_{\text{si}} e_{\text{si}}^2} + \frac{\alpha_{\text{tx}}N e_{\text{si}}^2}{\tau_{\text{si}} e_{\text{si}}^2} + \frac{\alpha_{\text{tx}}N e_{\text{si}}^2}{\tau_{\text{si}} e_{\text{si}}^2},$$

(1)

where $\alpha_{\text{nc}}$ is the level of the passive analog SIC, and the transmitter noise is a complex Gaussian random variable with zero mean and smaller variance than the receiver noise (denoted by $\alpha_{\text{tx}}$ ($\alpha_{\text{tx}} \ll 1$)).

In the case of the downlink/uplink channel estimation, we assume there are no error terms arising from the hardware impairments. Thus, the variance of the estimation error for the downlink and uplink channels are given by [9]

$$e_{\text{si}}^2 = \frac{1}{K_{\text{DL}} \beta_{\text{dl}} e_{\text{si}}^2} + \frac{1}{K_{\text{UL}} \beta_{\text{ul}} e_{\text{si}}^2},$$

(2)

where $K_{\text{DL}}, K_{\text{UL}}, e_{\text{si}}^2, \beta_{\text{dl}}, \beta_{\text{ul}}$ are system parameters.
Here, the received pilot SNRs of the downlink and uplink users are $K_{DL}\rho_u^{DL}$ and $K_{UL}\rho_u$, respectively. In addition, the variances of the estimated channel for the $k_d$-th and $k_u$-th users are given by

$$\text{Var}[\hat{h}_{k_d}] = \frac{K_{DL}\rho_{NL,k_d}^{DL}}{K_{DL}\rho_{NL,k_d}^{DL} + 1}$$

$$\text{Var}[\hat{h}_{k_u}] = \frac{K_{UL}\rho_{UL,k_u}^{UL}}{K_{UL}\rho_{NL,k_u}^{UL} + 1}.$$  

(3)

If the received pilot SNRs are high (i.e., both SNRs approach infinity), one may consider perfect downlink and uplink channel estimations, which means $\text{Var}[\hat{h}_{k_d}] \to 1$ and $\text{Var}[\hat{h}_{k_u}] \to 1$, and these are the same as $\text{Var}[h_{k_d}]$ and $\text{Var}[h_{k_u}]$, respectively. It should be noted that any estimation algorithm dependent on the transmit pilot power can be modeled similarly.

3) Channel Estimation Error Model: We assume that each RF chain is connected to a $p$ bits ADC at the BS. To represent a quantization error in a channel estimation error, we use the AQNM [30], [31] as follows: $U_{SI} = \alpha U_{SI} + N_q$, where $U_{SI}$ is the $M \times N$ output matrix of the ADC corresponding to the $M \times N$ input matrix $U_{SI}$, which is the received SI signal at the BS. Accordingly, $\alpha$ can be computed as $1 - e^{-\frac{2\pi p}{2}}$. Here, $U_{SI}$ can be modeled as [9]

$$U_{SI} = \sqrt{\frac{\rho_{SI}}{\rho_{anc}}} H_{SI}^T (P + N_{\alpha}) + N_p,$$  

(4)

where $N_{\alpha}$, $P$, and $N_p$ are the transmitter noise, the $M \times N$ matrix that satisfies $PP^H = I$, and the $M \times N$ complex Gaussian random variables with i.i.d. elements following the distribution $CN(0, \sigma^2)$, respectively. $N_q$ is the $M \times N$ additive quantization noise matrix such that $U_{SI}$ and $N_q$ are uncorrelated. It should be noted that the variance of $N_q$ is given by

$$\text{Var}[N_q] = \alpha(1 - \alpha) \text{Var}[U_{SI}],$$

$$\text{Var}[N_q] = \alpha(1 - \alpha)(\frac{\rho_{SI}}{\rho_{anc}}) (1 + \alpha_{tx}) + 1.$$  

(5)

Thus, the MMSE estimate of $\hat{H}_{SI}$ given $U$ is $\hat{H}_{SI} = (\sqrt{\frac{\rho_{SI}}{\rho_{anc}}}U_{SI}^T P \ast D)^T$, where $D = (\frac{\rho_{anc}}{\rho_{SI}} \beta_{SI} + I)^{-1}$.

C. SINR at Uplink and Downlink

Let $g_{k_d}$ and $s_{k_d}$ denote the $k_d$-th column vector of the transmit precoding matrix $G$ and the transmit symbol for the $k_d$-th user at the downlink, respectively. Similarly, let $w_{k_u}$ denote the $k_u$-th column vector of the receive combining matrix $W$ and $x_{k_u}$ represent the transmit symbol for the $k_u$-th user at the uplink. In addition, let $n_{\alpha}$ and $n_{UL}$ be the additive white Gaussian receiver noise vector at the downlink and uplink, respectively. The received signal at the $k_d$-th user is then expressed as

$$y_{k_d} = \sqrt{\rho_{DL}} \hat{h}_{DL,k_d}^T g_{k_d} s_{k_d} + \sqrt{\rho_{UL}} \sum_{\ell=1, \ell \neq k_d}^K \hat{h}_{DL,k_d}^T g_{k_d} e_{\ell}$$

$$\text{desired signal} + \sqrt{\rho_{DL}} \sum_{\ell=1, \ell \neq k_d}^K \hat{h}_{DL,k_d}^T g_{k_d} e_{\ell}$$

$$\text{user-interference} + \sqrt{\rho_{UL}} \sum_{\ell=1, \ell \neq k_u}^K \hat{h}_{UL,k_u}^T u_{\ell}$$

$$\text{desired signal} + \sqrt{\rho_{UL}} \sum_{\ell=1, \ell \neq k_u}^K \hat{h}_{UL,k_u}^T u_{\ell}$$

$$\text{user-interference} + \sqrt{\rho_{UL}} \sum_{\ell=1, \ell \neq k_u}^K \hat{h}_{UL,k_u}^T u_{\ell}$$

$$\text{self-interference}.$$

(6)

where $\hat{h}_{DL,k_d}$ represents the $k_d$-th column vector of $\hat{H}_{DL}$. In addition, $\mathbf{s}$ represents $s + n_{UL}$. To satisfy a power constraint, the normalized transmit beamforming vectors (the columns of the precoding matrix) with vector normalization are given as $g_{k_d} = f_{k_d}/(\sqrt{K_{DL}} ||f_{k_d}||) $, where $1 \leq k_d \leq K_{DL}$ [10].

From (6) and (7), we can derive the SINRs of the $k_d$-th user at the downlink and the $k_u$-th user at the uplink, respectively, as follows:

$$\gamma_{DL,k_d} = \frac{\rho_{DL} ||h_{DL,k_d}^T \sqrt{\rho_{NL,k_d}^{DL}} f_{k_d}||^2}{\tau_{DL} + \tau_{user} + \sum_{k_u = 1}^{K_{UL}} \rho_{UL} ||h_{UL,k_u}^T h_{UL,k_u}||^2 + 1},$$

$$\gamma_{UL,k_u} = \frac{\rho_{UL} ||w_{k_u}^T \hat{h}_{UL,k_u}^T s_{k_u}||^2}{\tau_{UL} + \tau_{user} + \rho_{UL} \Omega (1 + \alpha_{tx}) + ||w_{k_u}||^2},$$

(8)

where

$$\tau_{DL} = \rho_{DL} \sum_{\ell=1}^K ||h_{DL,k_d}^T \sqrt{K_{DL}} ||f_{\ell}||^2, \tau_{UL}$$

$$\tau_{DL} = \rho_{DL} \sum_{\ell=1, \ell \neq k_d}^K ||h_{DL,k_d}^T \sqrt{K_{DL}} ||f_{\ell}||^2, \tau_{UL}$$

$$\tau_{DL} = \sum_{\ell=1, \ell \neq k_u}^K ||h_{UL,k_u}^T ||e_{\ell}||^2, \tau_{user}$$

$$\tau_{UL} = \sum_{\ell=1, \ell \neq k_u}^K ||h_{UL,k_u}^T ||e_{\ell}||^2, \tau_{user}$$

$$\tau_{UL} = \sum_{\ell=1, \ell \neq k_u}^K ||h_{UL,k_u}^T ||e_{\ell}||^2, \tau_{user}$$

$$\tau_{UL} = \sum_{\ell=1, \ell \neq k_u}^K ||h_{UL,k_u}^T ||e_{\ell}||^2, \tau_{user}$$

D. Precoding/Receive Combining Matrix Design and SIC Methods

Because the FD BS always perfectly knows its precoding/receive matrix and transmitted downlink symbols, it can
mitigate SI by directly subtracting \( \sqrt{\rho/\sigma_w^2} w_{\text{ext}}^T \hat{H}_S G \), of which the SIC is referred to as the SI subtraction. The received signal after the SI subtraction is

\[
u_{\text{st},k_0} = u_{k_0} - \sqrt{\rho/\sigma_w^2} w_{\text{ext}}^T \hat{H}_S G \hat{s}.
\]

Here, the subscript \( \text{st} \) indicates the SI subtraction. The ZF precoder for the downlink is assumed at the BS in the cases of the FD BS with the SI subtraction and without SIC:

\[
G_{\text{ZF}} = \hat{H}_{\text{DL}}^* (\hat{H}_{\text{DL}} \hat{H}_{\text{SI}}^*)^{-1}.
\]

In addition, before the precoding/combining matrices are computed only once per coherence block, the complexity difference between MF and ZF/MMSE is relatively small (the bulk of the complexity comes from FFTs and matrix-vector multiplications performed on a per-symbol basis [34]). In addition, because the FD massive MIMO considered has many antennas and a high SNR environment, both ZF and MMSE have similar performance. Furthermore, ZF can provide a clear performance comparison with SI subtraction to satisfy our main contribution. For these reasons, the ZF is considered for downlink transmission. Future work will consider various precoding/combining matrices, including MRT/MRC, generalized MMSE and so on. From the antenna configuration of \( M \geq N + K \), additional antennas can be utilized for spatial suppression [17], [25], [27]. Spatial suppression-based algorithms for massive MIMO systems were proposed in [25] and [27]. We adopt the extended ZF in [25] for spatial suppression, which requires minimum computational complexity for nulling.

\[
G_{\text{sps}} = \hat{H}_{\text{ext}}^* (\hat{H}_{\text{ext}} \hat{H}_{\text{ext}}^*)^{-1}, \quad \hat{H}_{\text{ext}} = \begin{bmatrix} \hat{H}_{\text{DL}} & \hat{H}_{\text{SI}} \end{bmatrix}.
\]

Here, the subscript \( \text{sps} \) indicates the spatial suppression. It should be noted that the vector set, from the beginning of the column vector of \( G_{\text{sps}} \) to the \( K_{\text{DL}} \)-th column vector, is the same as the null-space projection precoder in [17], where the rank of \( H_{\text{SI}} \) is \( N \), and the nullity of that is \( K_{\text{DL}} \).

To eliminate user-interference signals at the uplink, we use the following ZF filter in both the transceiver with the SI subtraction/spatial suppression and without the SIC:

\[
W = (\hat{H}_{\text{UL}}^* \hat{H}_{\text{UL}})^{-1} \hat{H}_{\text{UL}}^*.
\]

III. Performance Comparison in Different Channel Conditions

In this section, we compare the performance of two SIC methods with channel conditions which are correlated/uncorrelated and perfect/imperfect. Specifically, we provide mathematical motivations and preliminaries at first. Then, based on these preliminaries, we compare the two methods by SINR or the expected SI power term in uncorrelated channels. Finally, in subsection III-C, we compare them by the expected SI power term in correlated channels.

A. Mathematical Motivations and Preliminaries

In this subsection, we provide mathematical motivations and preliminaries to investigate the ergodic sum rate and SI power for FD massive MIMO systems. We also present an asymptotic analysis of FD massive MIMO systems under the perfect channel estimation conditions.

1) SI Power Term in Massive MIMO Systems: In the FD system, the SI level degrades the uplink performance as a bottleneck, which is represented as the SI power term \( \Omega \) in (8). Let \( A \) be an \( N \times M \) random matrix, the elements of which are complex random variables with zero mean and \( \sigma_A^2 \). In addition, \( w_{k_0}, g_{nm}, \) and \( g_{mk} \) are the \( n \)-th element of \( w_{k_0} \), the \((m, n)\)-th element of \( A \), and the \((m, \ell)\)-th element of the \( M \times L \) matrix \( G \) (the precoder matrix), respectively. Then, the quadratic form of \( ||w_{k_0}^T A G||^2 \) can be expressed as

\[
||w_{k_0}^T A G||^2 = \sum_{m=1}^{L} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{M} \sum_{\ell=1}^{M} w_{k_0, i} w_{k_0, j} a_{ik} a_{j\ell}^* g_{km} g_{\ell m}. \quad (13)
\]

It should be noted that the SI power term can be derived by replacing \( A \) in (13) with \( H_{\text{SI}} \) or \( E_{\text{SI}} \) according to the SIC method.

Lemma 1: (decomposition of the expected SI power term)

The expectation of (13) can be decomposed into

\[
\mathbb{E} \{ ||w_{k_0}^T A G||^2 \} = \sum_{i=1}^{N} \mathbb{E} \{ ||w_{k_0, i}||^2 \} \times \sum_{m=1}^{L} \sum_{k=1}^{M} \mathbb{E} \{ ||a_{ik}||^2 \} \mathbb{E} \{ ||g_{km}||^2 \} + \Omega_A(i\neq j),
\]

\[
\mathbb{E} \{ ||w_{k_0, i}||^2 \} 
\]

expectation term related to i.i.d. \( H_{UL} \)

\[
\sum_{i=1}^{N} \mathbb{E} \{ ||w_{k_0, i}||^2 \} \times \sum_{m=1}^{L} \sum_{k=1}^{M} \mathbb{E} \{ ||a_{ik}||^2 \} \mathbb{E} \{ ||g_{km}||^2 \} + \zeta
\]

expectation term where \( A \) and \( G \) are independent

\[
\Omega_A(i\neq j),
\]

where

\[
\Omega_A(i\neq j) = \sum_{m=1}^{L} \sum_{i=1}^{N} \sum_{j\neq i}^{N} \sum_{k=1}^{M} \sum_{\ell=1}^{M} \mathbb{E} \{ w_{k_0, i} w_{k_0, j}^* a_{ik} a_{j\ell}^* g_{km} g_{\ell m}^* \},
\]

\[
\zeta = \text{Cov} \{ ||a_{ik}||^2, ||g_{km}||^2 \} + \sum_{\ell=1, \ell \neq k}^{M} \mathbb{E} \{ a_{ik} a_{i\ell}^* g_{km} g_{\ell m}^* \}.
\]
Equation (a) is derived based on the fact that $w_{k,n}$ is independent of both $\theta_{mn}$ and $g_{mn}$.

This lemma implies that the expected SI power varies according to the correlation between $A$ and $G$ as well as the correlation among their own elements. Using this lemma, we compare the two SIC methods in correlated channel models. It should be noted that, the conventional analysis does not focus on these correlation terms by assuming independent conditions [25], [26].

**Lemma 2:** (expected SI power term with independent conditions) The following equation holds when $w_{k,n}$, $A$, and $G$ are independent of each other and, consist of i.i.d. random variables [25], [26]:

$$E\left\{||w_{k,n}^T AG||^2\right\} = \sigma^2 E\left\{||w_{k,n}||^2\right\}.$$

**Proof:** See Appendix A.

2) Approximation of Ergodic Sum Rate of Massive MIMO Systems:

**Remark 2:** Let $S$, $I$, and $N$ be the norms of the random vectors of the desired signal power, interference power, and noise power terms, respectively. From [10, Lemma 4], the ergodic achievable sum rate of massive MIMO systems in the low/high SNR regime can be approximated as follows:

$$E\left\{\log_2 \left(1 + \frac{S}{I + N}\right)\right\} \approx R = \log_2 \left(1 + \frac{E\{S\}}{E\{I + N\}}\right),$$

where $E\{S\}E\left\{\frac{1}{I + N}\right\} = \frac{E\{S\}}{E\{I + N\}}$ or $E\left\{\frac{S}{I + N}\right\} \approx \frac{E\{S\}}{E\{I + N\}}$.

From (15), we let $\gamma = \frac{E\{S\}}{E\{I + N\}}$ be an approximation of $\gamma$.

3) Performance Analysis With Perfect Channel Estimation: First, we begin with a brief overview of the ergodic achievable sum rate of the two SIC schemes with perfect channel state information ($\tau_{DL,UL} = 0$) in spatially uncorrelated channels. In summary, from Remark 2, the approximations of the ergodic rate of the $k_d$-th user with the SI subtraction and the spatial suppression at downlink are expressed, respectively, as follows:

$$R_{k_d, stt} = \log_2 \frac{1 + \rho_{k_d}^{DL}(M - K_{DL} + 1)}{K_{DL}(\sum_{k_s=1}^{K_{UL}} \rho_{k_s,k_d}^{UL} + 1)} + \log_2 \frac{1 + \rho_{k_d}^{UL}(M - K_{UL} + 1)}{K_{UL}(\sum_{k_s=1}^{K_{UL}} \rho_{k_s,k_d}^{UL} + 1)}.$$

$$R_{k_d, sps} = \log_2 \frac{1 + \rho_{k_d}^{DL}(N - K_{DL} + 1)}{K_{DL}(\sum_{k_s=1}^{K_{UL}} \rho_{k_s,k_d}^{UL} + 1)} + \log_2 \frac{1 + \rho_{k_d}^{UL}(N - K_{UL} + 1)}{K_{UL}(\sum_{k_s=1}^{K_{UL}} \rho_{k_s,k_d}^{UL} + 1)}.$$

It should be noted that the approximations of the ergodic sum rate of the half-duplex with ZF precoding and ZF receive combining is the same as $\frac{1}{\delta} R_{k_d, stt}$ with $\rho_{k_d,k_s}^{UL} = 0 \forall k_d, k_s$ [10].

From (19) and (20), the sum rate of the SI subtraction is, in the genie-aided system, always better than that of the spatial suppression owing to the same performance at the uplink and the performance degradation at the downlink. This is because the SI subtraction and spatial suppression transceivers can perfectly cancel out the SI at the uplink, and the spatial suppression requires an additional dimension for the nulling of the ZF precoder at the downlink.

B. Asymptotic Rate in Uncorrelated Channels

In this subsection, we present the asymptotic analysis of FD massive MIMO systems according to the two SIC methods: SI subtraction and spatial suppression. First, we assume uncorrelated channels for simplifying mathematical expressions to obtain insights into how the imperfect channel estimation affects the performance of the two SIC schemes.

In the FD system, the SI level degrades the uplink performance as a bottleneck, which is represented as the SI power term $\Omega$ in (8). Let $\Omega = E\{\Omega\}$ be the expected SI power. If there is no SIC at the BS ($\alpha_{\text{anc}} = 0$ dB), from Lemma 2, the expected SI power is given by

$$\Omega_{\text{sosIC}} = E\{||w_{k_d}^T H_{\text{SI}} G||^2\} = E\{||w_{k_d}||^2\} E\{||H_{\text{SI}}^S||^2\} = E\{||w_{k_d}||^2\}.$$

Similarly, because $E_{\text{SI}}$ and $G_{\text{ZF}}$ are independent, from (14), the quadratic form of the expected SI power of the SI subtraction is given by

$$\Omega_{\text{stt}} = E\{||w_{k_d}^T (H_{\text{SI}} - \hat{H}_{\text{SI}}) G_{\text{ZF}}||^2\} = E\{||w_{k_d}^T E_{\text{SI}} G_{\text{ZF}}||^2\} = \epsilon_{\text{stt}}^2 E\{||w_{k_d}||^2\}.$$

5The right side of $\zeta$ includes the terms that occurred due to the correlated $G$. We discuss this in Lemma 5.
In contrast to the SI subtraction, $E_{\text{St}}$ and $G_{\text{sp}}$ are dependent when spatial suppression is used, because $G_{\text{sp}}$ is a function of $\hat{H}_\text{St} = H_\text{St} + E_{\text{St}}$. Thus, from Lemma 1, we can derive the quadratic form of the expected SI power of the spatial suppression as follows:

$$\Omega_{\text{sp}} = \mathbb{E}\left\{||w_k^T H_\text{St} G_{\text{sp}}||^2\right\} = \mathbb{E}\left\{||w_k^T (\hat{H}_\text{St} - E_{\text{St}}) G_{\text{sp}}||^2\right\}$$

$$= \sum_{i=1}^{N} \mathbb{E}\left\{||w_{ki}||^2\right\} \sum_{m=1}^{K_{UL}} \mathbb{E}\left\{||E_{ik}||^2\right\} \mathbb{E}\left\{||g_{km,sp}||^2\right\}$$

$$+ \sum_{i=1}^{N} \mathbb{E}\left\{||w_{ki}||^2\right\} \sum_{m=1}^{K_{UL}} \sum_{k=1}^{M} \zeta_{\text{sp}}$$

$$= \varepsilon_k^2 \mathbb{E}\left\{||w_{ki}||^2\right\} + \sum_{i=1}^{N} \mathbb{E}\left\{||w_{ki}||^2\right\} \sum_{m=1}^{K_{UL}} \sum_{k=1}^{M} \zeta_{\text{sp}}, \quad (23)$$

where $\Omega_{A(i\neq j)} = 0$, $\zeta_{\text{sp}} = \text{Cov}(||E_{ik}||^2, ||g_{km,sp}||^2) + \sum_{\ell=1, \ell \neq k}^{M} \mathbb{E}\left\{E_{ik} E_{i\ell}^* g_{km,sp} g_{\ell m,sp}^*\right\}$ and $E_{km}$ is the $(n, m)$-th element of $E_{St}$. In addition, $G_{\text{sp}}$ is the subset from the first to the $K_{UL}$-th columns of $G_{\text{sp}}$ and $\hat{H}_\text{St} G_{\text{sp}} = 0$.

**Lemma 3:** The residual expectation term in $\Omega_{\text{sp}}$ owing to the correlation between $E_{\text{St}}$ and $\hat{G}_{\text{sp}}$ is negative:

$$\zeta_{\text{sp}} = \text{Cov}(||E_{ik}||^2, ||g_{km,sp}||^2) + \sum_{\ell=1, \ell \neq k}^{M} \mathbb{E}\left\{E_{ik} E_{i\ell}^* g_{km,sp} g_{\ell m,sp}^*\right\} \leq 0.$$

**Proof:** See Appendix B.

By applying Lemma 3 to the quadratic forms of $\Omega_{\text{st}}$ and $\Omega_{\text{sp}}$ ((22), (23)), we can confirm the inequality

$$\Omega_{\text{st}} \geq \Omega_{\text{sp}}. \quad (24)$$

To illustrate the implicit meaning of Lemma 3, we present Fig. 2, which illustrates an example of the effect of spatial suppression on the residual SI by representing a manifold as a Euclidean space. This shows that, with imperfect channel estimation, the difference vector between the SI channel vector and their estimated vector (i.e., the estimation error vector) is additionally projected onto the null vector of the estimated SI channel vector. Thus, the length of the residual SI vector of the spatial suppression can be shorter than that of SI subtraction. Therefore, the expected SI power of the spatial suppression can be mitigated to a greater extent than that of SI subtraction.

It is trivial from (19) and (20) that the SI subtraction outperforms the spatial suppression in the perfect channel estimation condition, i.e., $\mathcal{R}_{\text{st}} > R_{\text{sp}}$. However, as the ergodic uplink performance of the spatial suppression is the same as that of the SI subtraction in the perfect channel estimation condition, it is better than that of the SI subtraction as determined from (24), i.e., $\Omega_{\text{st}} \geq \Omega_{\text{sp}}$.

Next, we present some approximations of the uplink SINR for each SIC method to support the obtained results. Using the approximation methods of [9] and Remark 2, the approximation of the uplink SINR without the SIC can be derived from (2), (8), and $\Omega_{\text{wosIC}} = \mathbb{E}\left\{||w_{ki}||^2\right\}$ as follows:

$$\bar{\gamma}_{\text{UL}}^{\text{wosIC}, k_n} = \frac{\left(\mathbb{E}\left[||w_{ki}||^2\right] + A A + \mathbb{E}\left[||w_{ki}||^2\right]\right)}{K_{UL}(\sum_{\ell=1}^{K_{UL}} \rho_{\ell \ell}^2 + 1)(N - K_{UL} + 1) + \rho_{\ell \ell}^2 K_{UL}(1 + \alpha_{\ell})}, \quad (25)$$

where $AA = \frac{\rho_{\ell \ell}^2}{\rho_{\ell \ell}^2 + 1} \mathbb{E}\left[||w_{ki}||^2\right] + (1 + \alpha_{\ell})$, $\mathbb{E}\left[||w_{ki}||^2\right] = 1$, $\mathbb{E}\left[||w_{ki}||^2\right] = 0$, and $\mathbb{E}\left[||w_{ki}||^2\right] = \frac{1}{K_{UL} \rho_{\ell \ell}^2 + 1}(N - K_{UL} + 1)$, which is calculated from (3).

Similarly, from (8) and (22), we can derive the approximation of the uplink SINR with the SI subtraction as follows:

$$\bar{\gamma}_{\text{UL}}^{\text{UL}, k_n} = \frac{(K_{UL} \rho_{\ell \ell}^2 + 1)(N - K_{UL} + 1)}{K_{UL}(\sum_{\ell=1}^{K_{UL}} \rho_{\ell \ell}^2 + 1) + \rho_{\ell \ell}^2 K_{UL}(1 + \alpha_{\ell})}. \quad (26)$$

Finally, we present the uplink SINR for spatial suppression. To obtain the independent condition, we should consider, given that $H_{\text{St}}$ and $G_{\text{sp}}$ are not independent of each other, the best and worst estimation errors to obtain the independent condition. In the case of the best channel estimation error ($\epsilon_{\text{st}} \rightarrow 0$), the estimated SI channel tends to the original SI channel ($\hat{H}_{\text{St}} \rightarrow H_{\text{St}}$), and thus $E_{\text{St}}$ and $G_{\text{sp}}$ tend to be independent. Accordingly, from Lemma 2, the expected SI power of the spatial suppression (the best case) is given by

$$\bar{\Omega}_{\text{sp}}(\epsilon_{\text{st}} \rightarrow 0) = \mathbb{E}\left\{||w_k^T H_{\text{St}} G_{\text{sp}}||^2\right\}$$

$$= \mathbb{E}\left\{||w_k^T (\hat{H}_{\text{St}} - E_{\text{St}}) G_{\text{sp}}||^2\right\}$$

$$= \mathbb{E}\left\{||w_k^T E_{\text{St}} G_{\text{sp}}||^2\right\}$$

$$= \varepsilon_k^2 \mathbb{E}\left\{||w_{ki}||^2\right\}. \quad (27)$$

Similarly, in the case of the worst channel estimation error ($\epsilon_{\text{st}} \rightarrow \infty$), the estimated SI channel tends to the channel estimation error matrix ($\hat{H}_{\text{St}} \rightarrow E_{\text{St}}$) and $\hat{H}_{\text{ext}}$ tends to $\frac{\hat{H}_{\text{DL}}}{E_{\text{St}}}$. Therefore, $H_{\text{St}}$ and $G_{\text{sp}}$ tend to be independent. Thus, the
expected SI power of the spatial suppression (the worst case) is given by
\[
\bar{\Omega}_{\text{sps}}(r_{\text{UL}}) = \mathbb{E}\{||w_i^T H_{\text{SI}} G_{\text{sps}}||^2\} = \mathbb{E}\{||w_k^s||^2\}. \quad (28)
\]
From (27) and (28), the approximation of the expected SI power of spatial suppression is expressed as
\[
\tilde{\bar{\Omega}}_{\text{sps}} \approx \frac{\mathbb{E}\{||w_k^s||^2\}}{1/2 + 1}.
\]
This approximation approach is a simple application of harmonic mean. The approximation of the uplink SINR with spatial suppression is then given by
\[
\zeta_{\text{UL},k}^s = \frac{(K_{UL}\rho_{UL}^{UL} + 1)(N - K_{UL} + 1)}{K_{UL}\left(\sum_{\ell=1}^{K_{UL}} \frac{\rho_{UL}^{UL}^{\ell}}{\rho_{UL}^{UL} + 1} + 1\right)} \left(\frac{1}{1/2 + 1}\right) \frac{1}{1 + \alpha_u}.
\]
\[
(29)
\]
C. Expected SI Power in Correlated Channels
In this subsection, we compare the expected SI power of the two SIC methods by considering the channel correlation. The estimated channel \(\hat{H}\) is correlated when the channel is correlated while the estimation errors, elements of \(E\), are uncorrelated [36]. We first investigate how the correlated \(H_{\text{SI}}\) affects the level of SI power when \(G_{\text{sps}}\) is used.

Lemma 4: When considering the correlated \(H_{\text{SI}}, \zeta_{\text{sps}}\) is given by
\[
\zeta_{\text{sps}} = \text{Cov}(\{E_{\ell k}\}^2, |g_{km,\text{sps}}|^2) + \frac{\sum_{j=1}^{M} r_{jm,\text{sps}}^2 \sqrt{q^2_{\ell k,\text{SI}} + 1}}{M.K_{DL}},
\]
where \(r_{jm,\text{sps}} = \text{Corr}(E_{\ell k}^*, g_{km,\text{sps}})\) and \(r_{\ell k,\text{SI}} = \text{Corr}(g_{km,\text{sps}}, g_{\ell m,\text{sps}})\). In addition, \(q\) is a constant that depends on the distribution of \(|g_{km,\text{sps}}|^2\) (if the distribution is approximated to an exponential distribution, then \(q = 2\)).

Proof: See Appendix C.

Because \(r_{\ell k,\text{SI}}^2\) is proportional to the spatial correlation of the highly correlated SI channel, from Lemma 3 and Lemma 4, we can conclude that as the correlation of the SI channel increases, the expected SI power of the spatial suppression decreases. Meanwhile, the correlation of the SI channel does not affect the expected SI power of the SI subtraction because \(r_{jm,\text{sps}} = 0\).

Next, we will show that the uplink performance of spatial suppression is better than that of the SI subtraction in the correlated channels with the practical angle distribution that is based on a Bessel function (e.g., Laplace distribution and wrapped Gaussian distribution [37]).

Lemma 5: In the correlated channels of the \(H_{\text{SI}}\) and \(H_{\text{DL/UL}}, \bar{\Omega}_{\text{stt}} \geq \Omega_{\text{sps}}\) holds, which can be expressed as
\[
\Omega_{\text{sps}} - \bar{\Omega}_{\text{stt}} = \sum_{i=1}^{N} \mathbb{E}\{||w_{k_i}||^2\} \sum_{m=1}^{L} \sum_{k=1}^{M} \zeta_{\text{sps}}.
\]

This approximation approach is a simple application of harmonic mean. The approximation of the uplink SINR with spatial suppression is then given by
\[
\zeta_{\text{UL},k}^s = \frac{(K_{UL}\rho_{UL}^{UL} + 1)(N - K_{UL} + 1)}{K_{UL}\left(\sum_{\ell=1}^{K_{UL}} \frac{\rho_{UL}^{UL}^{\ell}}{\rho_{UL}^{UL} + 1} + 1\right)} \left(\frac{1}{1/2 + 1}\right) \frac{1}{1 + \alpha_u}.
\]

IV. PERFORMANCE COMPARISON WITH SYSTEM CONSTRAINTS
In this section, we compare the performance of the two SIC methods while considering the following system constraints: the downlink-to-uplink data traffic ratio under a non-full buffer scenario, the number of transmit and receive RF chains ratio with transmit power at the BS, and power scaling law at the BS. We hope that this discussion elucidates FD massive MIMO performance under various system constraints.

A. Optimal Antenna Ratio in a Non-Full Buffer Scenario
We first show that an asymmetric RF configuration is advantageous in a non-full-buffer asymmetric data traffic scenario. The traffic load density at the downlink and uplink are denoted by \(\mu_{DL}\) and \(\mu_{UL}\), and \(T_{DL} = \mu_{DL}/R_{DL}\); the transmission time for the given traffic loads is \(T_{UL} = \mu_{UL}/R_{UL}\), represented in [38]. We define the required downlink and uplink rates as the data rates for achieving \(T_{DL} = T_{UL} = T\) in the FD mode within a certain time \((T)\) to avoid half-duplex transmission, and their ratio is represented as \(\eta\), which is the ratio of downlink traffic load to the uplink traffic load, \(\eta = \mu_{DL}/\mu_{UL}\). Simple performance analysis for the optimal RF chain ratio is given as follows.

Remark 3: Assuming that the per-user data traffic is identical under asymmetric data traffic scenarios, the required downlink rate to uplink rate ratio can be expressed as \(\eta = K_{DL}/K_{UL}\). In this scenario, we can obtain insight into the optimal ratio of the numbers of transmit RF chains and receive RF chains, and the optimal ratio of the SI subtraction and the spatial suppression given by \(M/N = \eta\) and \((M-N)/N = \eta\), respectively, derived from the result in [26].

Remark 3 implies that an asymmetric RF configuration is better than a symmetric RF configuration. In addition, this implication holds even if different per-user data traffic is assumed. It should be noted that the performance with the optimal RF chain ratio is not investigated in the following subsections, because the traffic ratio changes while the RF chain ratio at the BS is fixed. Instead, we present insights into the trade-off between the two SIC methods by considering arbitrary data traffic.

B. Performance Gap Between Downlink and Uplink Rate With the Same Transmit Power
We assume that the numbers of downlink and uplink users are equal (i.e., \(K = K_{DL} = K_{UL}\)), while the downlink data traffic and uplink data traffic differ. In addition, the propagation gains have the same values for all \(k\) and \(k\). We analyze the performance from a per-user asymmetric data traffic perspective and consider, for both links, high SINR regimes. SINRs can be expressed in terms of the received
downlink SNR ($\varphi$, proportional to $\rho_d$), where the desired signal power is relatively higher than others at the downlink, and the SI power is relatively smaller than the desired power at the uplink. For example, from (16)-(18), $\varphi$ can be denoted by a linear function of $\rho_d$ ($\varphi = \kappa \rho_d$) in the perfect channel estimation condition, where $\kappa = \frac{R_{\text{DL}}}{\left(\sum_{k=1}^{N} I_{\text{UL}}[k+1]\right)}$. In these regimes, we can approximate the SINRs as follows: $\tilde{\gamma}_{\text{UL}}^{\text{UL}}(\varphi) = (M - K)\varphi$, $\tilde{\gamma}_{\text{UL}}^{\text{UL}}(\varphi) = (M - N - K)\varphi$, $\tilde{\gamma}_{\text{UL}}^{\text{UL}}(\varphi) = \frac{S}{2 + \omega_s - \omega_m}\varphi$, $\tilde{\gamma}_{\text{UL}}^{\text{UL}}(\varphi) = \frac{S}{2 + \omega_s - \omega_m}\varphi$.

Thus, the gaps in the approximated rates between the two SIC methods at each link are expressed as

$$\Delta_{\text{DL}}(\varphi) = -\log_2 \left(1 - \frac{\tilde{\gamma}_{\text{UL}}^{\text{UL}}(\varphi)}{1 + \tilde{\gamma}_{\text{UL}}^{\text{UL}}(\varphi)}\right) \geq 0,$$

$$\Delta_{\text{UL}}(\varphi) = -\log_2 \left(1 - \frac{\tilde{\gamma}_{\text{UL}}^{\text{UL}}(\varphi)}{1 + \tilde{\gamma}_{\text{UL}}^{\text{UL}}(\varphi)}\right) \geq 0.$$

Thus, $\bar{R}_{\text{DL}}(\varphi) - R_{\text{UL}}(\varphi)$ is the cross point when the gaps between the downlink and uplink rates of each SIC method are the same, we obtain

$$R_{\text{DL}}(\varphi) = R_{\text{UL}}(\varphi) = R_{\text{DL}}(\varphi) + \Delta_{\text{UL}}(\varphi),$$

$$R_{\text{DL}}(\varphi) - R_{\text{UL}}(\varphi) = R_{\text{DL}}(\varphi) - \Delta_{\text{DL}}(\varphi),$$

and we can obtain the inequality as $\frac{R_{\text{DL}}}{R_{\text{UL}}} \geq \frac{R_{\text{DL}}^\text{SI}}{R_{\text{UL}}^\text{SI}}$.

Therefore, the gap between the downlink and uplink rates of the SI subtraction is higher than that of the spatial suppression for the given transmit power, the SI subtraction is more desirable when the gap between the downlink and uplink data traffic is higher (i.e., $\eta_{\text{DL}} \geq \eta_{\text{UL}}$), and vice versa.

C. Power Scaling Law in Downlink

Next, we consider a more practical constraint for asymmetric data traffic scenarios. In practical FD systems, achieving the maximum downlink rate can result in a bottleneck for supporting more downlink traffic and reducing the total transmission time of data traffic in the buffer at both links. Here, the maximum downlink rate is defined as the achievable rate by applying the best modulation order and coding rate. We denote it as $\tilde{R}_{\text{max}}^{\text{DL}}$. In addition, we define the minimum required SNRs to achieve the maximum downlink rates as follows:

$$\varphi_{\text{DL}} = \arg \min_{\varphi} \left\{ \sqrt{\log_2 (1 + \tilde{R}_{\text{DL}}^{\text{DL}}(\varphi))} \right\},$$

$$\varphi_{\text{UL}} = \arg \min_{\varphi} \left\{ \sqrt{\log_2 (1 + \tilde{R}_{\text{UL}}^{\text{UL}}(\varphi))} \right\},$$

where $\varphi_{\text{DL}} \leq \varphi_{\text{UL}}$. From the power scaling law in the downlink given by $\tilde{R}_{\text{max}}^{\text{DL}} = R_{\text{DL}}^{\text{DL}}(\varphi_{\text{DL}}) = R_{\text{UL}}^{\text{UL}}(\varphi_{\text{UL}}),$ we have $\varphi_{\text{DL}} = \frac{M - K - N}{M - K - N} \varphi_{\text{UL}}$. We compare the uplink SINR of the SI subtraction at $\varphi_{\text{DL}}$ with that of the spatial suppression at $\varphi_{\text{UL}}$ as follows:

$$\tilde{R}_{\text{UL}}^{\text{UL}}(\varphi_{\text{UL}}) = \frac{I + (\omega_s - \Delta) \varphi_{\text{UL}}}{I + \omega_s \varphi_{\text{UL}}},$$

$$\tilde{R}_{\text{UL}}^{\text{UL}}(\varphi_{\text{UL}}) = \frac{I + (\omega_s - \Delta) \varphi_{\text{UL}}}{I + \omega_s \varphi_{\text{UL}}}.$$
Fig. 3. The digital map for the 3D ray-tracing simulation and the absolute values of the measured $R_{\text{TX(SI)}}$ through the 3D ray-tracing simulation.

Fig. 4. Downlink/uplink sum rate with perfect channel estimation.

B. Channel Model for SI Channel

1) Scenario I: SI Channel With Simple Correlation Matrices: The correlation matrices are simply modeled as

$$
R_{\text{TX(SI)}} = 
\begin{bmatrix}
1 & r_{\text{TX,SI}} & \ldots & r_{\text{TX,SI}}^{M-2} & r_{\text{TX,SI}}^{M-1} \\
\vdots & \ddots & \ddots & \vdots & \vdots \\
M-1 & M-2 & \ldots & 1
\end{bmatrix},
$$

and

$$
R_{\text{RX(SI)}} = 
\begin{bmatrix}
1 & r_{\text{RX,SI}} & \ldots & r_{\text{RX,SI}}^{N-2} & r_{\text{RX,SI}}^{N-1} \\
\vdots & \ddots & \ddots & \vdots & \vdots \\
N-1 & N-2 & \ldots & 1
\end{bmatrix}.
$$

According to [39], we set $r_{\text{TX,SI}}$ as 0.2 at a low correlation and 0.8 at a high correlation.

2) Scenario II: Practical SI Channel Measured From a 3D Ray-Tracing Tool: We develop a system-level simulator based on a 3D ray-tracing tool (WiSE, developed by Bell Labs [40]). We produce a customized digital map of a real building for an indoor scenario. Fig. 3 presents the test site from a third-floor perspective. The digital map includes concrete walls and floors, metallic doors, glass windows, and sheetrock ceilings. The BS is located immediately under a 3 m-high ceiling in a room facing the door to the south. We use

Fig. 5. Ergodic sum rate of the SI subtraction and the spatial suppression at high correlation and low correlation in the SI channel. The legends $H$ and $L$ denote high correlation and low correlation, respectively.
both the co-polarization and cross-polarization patterns of the fabricated dual-polarization antenna, which are measured from an anechoic chamber [19].

To realize an SI channel, we measure the power delay profiles of the $\ell$-th channel tap ($P_\ell$) and angle parameters of the $\ell$-th channel tap, such as the azimuth angle of arrival ($\theta_{AOA}$), azimuth angle of departure ($\phi_{ADO}$), zenith angle of arrival ($\theta_{ZOA}$), and zenith angle of departure ($\phi_{ZOD}$). The BS and users are equipped with the manufactured antenna and the isotropic antenna, respectively.

Next, we realized the SI channel in a realistic indoor scenario. We denote co-polarization and cross-polarization patterns of Port 1 by $G_{tx}^H$ and $G_{tx}^V$, respectively. Using a 3D ray-tracing simulation, we measured the $L_{tap}$ of the channel taps. Based on a clustered channel model [39], the $(n,m)$-th element of the measured SI channel can be obtained as follows:

$$\begin{align*}
[\mathbf{H}_{\text{SI}}]_{nm} = \sum_{\ell=1}^{L_{\text{tap}}} P_\ell & \begin{bmatrix} G_{tx}^H & G_{tx}^V \\ 0 & 0 \end{bmatrix} \\
& \times \begin{bmatrix} e^{j\Phi_{HH}} & e^{j\Phi_{HV}} \\ e^{j\Phi_{VH}} & e^{j\Phi_{VV}} \end{bmatrix} \\
& \times \begin{bmatrix} e^{j\theta_{ZOA}} & e^{j\phi_{AOA}} \\ e^{j\phi_{ZOD}} & e^{j\phi_{ADO}} \end{bmatrix} \\
& \times \begin{bmatrix} e^{2\pi j\hat{\mathbf{r}}_{tx,n},m} & e^{2\pi j\hat{\mathbf{r}}_{tx,n},m} \end{bmatrix} ,
\end{align*}$$

where $\hat{\mathbf{r}}_{tx,n}$ and $\hat{\mathbf{r}}_{tx,m}$ are the location vectors of the transmit antenna $m$ and receive antenna $n$, respectively. In addition, $\hat{\mathbf{r}}_{tx,\ell}$ and $\hat{\mathbf{r}}_{tx,\ell}$ are spherical unit vectors with $\phi_{AOA}$ and $\theta_{ZOA}$ and with $\phi_{ADO}$ and $\theta_{ZOD}$, respectively. From (37), the measured correlation matrices can be calculated from $\mathbf{R}_{\text{TX(SI)}} = \mathbb{E}[\mathbf{H}_{\text{SI}}^* \mathbf{H}_{\text{SI}}]$ and $\mathbf{R}_{\text{RX(SI)}} = \mathbb{E}[\mathbf{H}_{\text{TX(SI)}}^* \mathbf{H}_{\text{SI}}^*]$. The absolute values of the measured $\mathbf{R}_{\text{TX(SI)}}$ are presented in Fig. 3. We can confirm that the practical SI channel is highly correlated owing to the high direct path power and small angular spread.

For modeling a more realistic dual-polarized channel, the XPR values of each ray are applied to co- and cross-polarization antenna patterns instead of using static XPR values in a polarization transfer matrix.

### C. Numerical Result

We first compare the proposed analysis and simulation results where perfect channel estimation is assumed at the BS. Fig. 4 shows that the results from III-A.3 are approximately the same as the ergodic downlink/uplink sum rate of the FD-BS without SIC and with SI subtraction/spatial suppression. Because the SI is perfectly canceled out owing to perfect knowledge of the SI channel, both the uplink performance of the spatial suppression and the SI subtraction are identical, and they use the same ZF receiver. That is to say, there is a performance gap only in the downlink case. An additional spatial domain ($N$) is needed to cancel out the SI channel at the FD-BS with spatial suppression. Thus, we confirm that the SI subtraction outperforms the spatial suppression in the perfect channel estimation condition. For all the results below, imperfect channel estimation is assumed and simulated by using MATLAB.

1) **Scenario I**: Fig. 5 shows the ergodic uplink sum rate of the SI subtraction and spatial suppression at high and low correlations. The SI subtractions always exhibit the same performance irrespective of the correlation of the SI channel. Meanwhile, as the correlation of the SI channel increases, the performance of the spatial suppression increases owing to the projection onto the estimated SI channel manifold. This result verifies Lemma 4.

Fig. 6 presents the ergodic uplink sum rate of the SI subtraction and spatial suppression at high correlation with different RF configurations when the power scaling law at the downlink is assumed. A performance trade-off, as shown in (36), occurs in the RF configuration. In the uplink, spatial suppression outperforms the SI subtraction when $M/N$ is high, and vice versa. Under these constraints, both methods have the same downlink sum rate under these constraints. Furthermore, the opposite result is observed as compared to Remark 3.

2) **Scenario II**: Fig. 7 illustrates the ergodic uplink sum rate of the SI subtraction and spatial suppression when the measured SI channel is considered. To investigate the correlation impact, we additionally evaluated the sum rate with

---

For modeling a more realistic dual-polarized channel, the XPR values of each ray are applied to co- and cross-polarization antenna patterns instead of using static XPR values in a polarization transfer matrix.
From Section IV, \( \Delta \) downlink sum rate and the ergodic uplink sum rate of the channel because of the high correlation. In addition, there is a significant performance gain with spatial suppression in the practical SI channel obtained by the SI subtraction. In addition, there is a significant performance gain with spatial suppression in the practical SI channel.

Fig. 7 presents the performance gap between the ergodic downlink sum rate and the ergodic uplink sum rate of the SI subtraction and the spatial suppression in the practical SI channel.

\[ \text{VI. CONCLUSION} \]

In this paper, two SIC methods in FD massive MIMO systems, i.e., SI subtraction and spatial suppression, have been investigated by considering the SI channel’s estimation error and spatial correlation. In addition, we conducted an SI channel measurement campaign using a 3D ray-tracing simulation. The analysis and numerical results reveal that, in the uplink with imperfect channel estimation, the ergodic performance of the spatial suppression is better than that of the SI subtraction. This performance gain is caused by the correlation between the spatial suppression precoder and the estimation error of the SI channel. Moreover, we provided insight into which SIC algorithm is better under given system constraints, such as the downlink/uplink traffic ratio, number of RF chains, total transmit power, and power scaling law at the BS.

\[ \text{APPENDIX} \]

A. Proof of Lemma 2

From Lemma 1, all residual expectation terms in (14) are zero because of the independent conditions that result in \( \mathbb{E} \{ w_{k,i} w_{k,j}^* \} = \mathbb{E} \{ w_{k,i} \} \mathbb{E} \{ w_{k,j}^* \} = 0 \), \( \text{Cov}(|a_{ik}|^2, |g_{km}|^2) = 0 \), and \( \mathbb{E} \{ a_{ik} g_{km}^* g_{km}^* \} = 0 \), where \( \ell \neq k \). Then, (14) is given by

\[
\mathbb{E} \{ \| \mathbf{w}_{k,i}^{H} \mathbf{A} \mathbf{G} \|^2 \} = \sum_{i=1}^{N} \sum_{k=1}^{M} \mathbb{E} \{ |w_{k,i}|^2 \} \sum_{m=1}^{L} \sigma_{k}^2 \mathbb{E} \{ |g_{km}|^2 \} \geq (b) \sigma_{k}^2 \mathbb{E} \{ |w_{k,i}|^2 \}. \tag{38}
\]

Equation (b) is derived based on \( \sum_{i=1}^{N} \mathbb{E} \{ |w_{k,i}|^2 \} = \mathbb{E} \{ \| \mathbf{w}_{k,i} \|^2 \} \) and \( \sum_{m=1}^{L} \mathbb{E} \{ |g_{km}|^2 \} = 1 \), owing to the definitions of \( L^2 \) norm and vector normalization.

B. Proof of Lemma 3

From \( \mathbb{E} \{ E_{ik} E_{ik}^* \} = 0 \) and \( \mathbb{E} \{ g_{km,sp} g_{km,sp}^* \} = 0 \), \( \zeta_{\text{sp}} \) is given by

\[
\zeta_{\text{sp}} = \text{Cov}(|E_{ik}|^2, |g_{km,sp}|^2) + \sum_{\ell=1, \ell \neq k}^{M} \text{Cov}(E_{ik} E_{i\ell}^*, g_{km,sp} g_{km,sp}^*) + \mathbb{E} \{ E_{ik} E_{ik}^* \}
\]

It is worth noting that power scaling law in the downlink is assumed under the constraints marked with (*). Also (**) works when the gap between the downlink and uplink traffic ratio is greater (i.e., \( \eta_{\text{dl}} \geq \eta_{\text{ul}} \)).
\( \tilde{G}_{\text{sps}} \) is the Pseudo-inverse of \( \tilde{H}_{\text{DL}} \), the random variables, \( E_{ik} \) and \( g_{\text{kmpsps}} \), tend to show opposite behaviors. Thus, each covariance term is negative as follows:

\[
\begin{align*}
\text{Cov}(|E_{ik}|^2, |g_{\text{kmpsps}}|^2) & \leq 0, \\
\text{Cov}(E_{ik}E^*_i, g_{\text{kmpsps}}g^*_{\text{lemsps}}) & \leq 0.
\end{align*}
\]

From (39)-(40), we obtain \( \zeta_{\text{sps}} \leq 0 \).

C. Proof of Lemma 4

Let \( z_k \) and \( z_{\ell} \) be uncorrelated random variables with the same distribution of \( g_{\text{kmpsps}} \). From the Cholesky decomposition, two correlated \( g_{\text{kmpsps}} \) and \( g_{\text{lemsps}} \) can be expressed as

\[
g_{\text{kmpsps}} = z_k, \quad g^*_{\text{lemsps}} = r_kz_k + \sqrt{1 - r^2_k}z_{\ell}.
\]

Then, we obtain the variance of their multiplication as follows:

\[
\begin{align*}
\text{Var}(g_{\text{kmpsps}}g^*_{\text{lemsps}}) &= \mathbb{E}(|g_{\text{kmpsps}}|^2|g_{\text{lemsps}}|^2) + \mathbb{E}(g_{\text{kmpsps}}g^*_{\text{lemsps}}) \\
&= r^2_k\mathbb{E}\{|z_k|^2\} + 2r_k\sqrt{1 - r^2_k}\mathbb{E}\{|z_k|^2z_kz_{\ell}\} + r_k\mathbb{E}\{|z_k|^2\} \\
&+ (1 - r^2_k)\mathbb{E}\{|z_{\ell}|^2\} + \sqrt{1 - r^2_k}\mathbb{E}\{|z_k|z_{\ell}\} \\
&= r^2_k\mathbb{E}\{|g_{\text{kmpsps}}|^4\} + \mathbb{E}(|g_{\text{kmpsps}}|^2),
\end{align*}
\]

where \( \mathbb{E}\{|z_k|^2z_kz_{\ell}\} = 0 \), and \( \mathbb{E}\{|z_kz_{\ell}\} = 0 \). Thus, when \( g_{\text{kmpsps}} \) and \( g_{\text{lemsps}} \) are correlated, the variance of their multiplication is given by

\[
\text{Var}(g_{\text{kmpsps}}g^*_{\text{lemsps}}) = r^2_k\mathbb{E}\{|g_{\text{kmpsps}}|^4\} + \mathbb{E}(|g_{\text{kmpsps}}|^2)^2.
\]

From (39) and (43), we obtain

\[
\zeta_{\text{sps}} = \text{Cov}(|E_{ik}|^2, |g_{\text{kmpsps}}|^2) \\
+ \sum_{\ell=1, \ell \neq k}^M \text{Cov}(E_{ik}E^*_i, g_{\text{kmpsps}}g^*_{\text{lemsps}}) \\
= \text{Cov}(|E_{ik}|^2, |g_{\text{kmpsps}}|^2) \\
+ \sum_{\ell=1, \ell \neq k}^M r_{ik}r_{i\ell}\sqrt{\text{Var}(E_{ik}E^*_i)}\sqrt{\text{Var}(g_{\text{kmpsps}}g_{\text{lemsps}})} \\
= \text{Cov}(|E_{ik}|^2, |g_{\text{kmpsps}}|^2) \\
+ \sum_{\ell=1, \ell \neq k}^M r^2_{ik}\varepsilon^2_{i\ell}r_{ik}\varepsilon^2_{i\ell} + \text{Var}(g_{\text{kmpsps}}g^*_{\text{lemsps}}) \\
= \text{Cov}(|E_{ik}|^2, |g_{\text{kmpsps}}|^2) \\
+ \sum_{\ell=1, \ell \neq k}^M \varepsilon^2_{i\ell}r_{ik}r_{i\ell} \\
\times \sqrt{q^2_{ik}r_{ik} + 1} \\
= \frac{\text{Cov}(|E_{ik}|^2, |g_{\text{kmpsps}}|^2) + \sum_{\ell=1, \ell \neq k}^M \varepsilon^2_{i\ell}r_{ik}r_{i\ell}}{MK_{\text{DL}}},
\]

where \( \mathbb{E}(|g_{\text{kmpsps}}|^2) = \frac{1}{MK_{\text{DL}}} \) and \( \mathbb{E}(|g_{\text{kmpsps}}|^4) = \frac{q^2}{MK_{\text{DL}}} \).

D. Proof of Lemma 5

We assume \( b_{ik}^{+} = \text{Sort}(|\text{Cov}(w_{ik}, w_{ik}^*)|) \) for \( \text{Cov}(w_{ik}, w_{ik}^*) \geq 0 \) and \( b_{ik}^{-} = \text{Sort}(|\text{Cov}(w_{ik}, w_{ik}^*)|) \) for \( \text{Cov}(w_{ik}, w_{ik}^*) \leq 0 \). For practical angle distributions, the correlation function in terms of antenna spacing can be expressed by a zero-order Bessel function [41]. The main lobe of a Bessel function is larger than the side lobes, so \( b_{ik}^{+} > b_{ik}^{-} \) holds. Thus, in the correlated \( H_{\text{DL}} \), we can obtain

\[
\sum_{i=1}^N \sum_{j=1, j \neq i}^N \text{Cov}(w_{ki}, w_{k\ell}^*) \geq 0.
\]

From Lemma 1, we check the quadratic forms of \( \Omega_{\text{st}} \) and \( \Omega_{\text{sps}} \) as follows:

\[
\Omega_{\text{st}} = \sum_{i=1}^N \mathbb{E}\{|w_{ki}|^2\} \sum_{m=1}^L \sum_{k=1}^M \mathbb{E}\{|E_{ik}|^2\} \mathbb{E}\{|g_{\text{kmpsps}}|^2\},
\]

where \( \Omega_{\text{st}}(i \neq j) = 0 \), then \( \mathbb{E}\{E_{ik}E^*_jg_{\text{kmpsps}}g^*_{\text{lemsps}}\} = \mathbb{E}\{E_{ik}E^*_jg_{\text{kmpsps}}g_{\text{lemsps}}\} \). From (46), (47), the expected SI power term of the SI subtraction is not affected by the correlated \( H_{\text{DL}} \). Similar to (39), we have

\[
\mathbb{E}\{E_{ik}E^*_jg_{\text{kmpsps}}g_{\text{lemsps}}\} = \text{Cov}(E_{ik}E^*_j, g_{\text{kmpsps}}g_{\text{lemsps}}).
\]

Then, we can obtain

\[
c = \sum_{i=1}^L \sum_{k=1}^M \sum_{\ell=1}^M \mathbb{E}\{E_{ik}E^*_jg_{\text{kmpsps}}g^*_{\text{lemsps}}\} \leq 0,
\]

where \( i \neq j \). From (46) and (47), the gap between \( \Omega_{\text{st}} \) and \( \Omega_{\text{sps}} \) in the correlated channels can be derived as

\[
\Omega_{\text{sps}} - \Omega_{\text{st}} = \sum_{i=1}^N \mathbb{E}\{|w_{ki}|^2\} \sum_{m=1}^L \sum_{k=1}^M \zeta_{\text{sps}} \\
+ \sum_{i=1}^N \sum_{j=1, j \neq i}^N \mathbb{E}\{w_{ki}w_{k\ell}^*\} \\
\times \sum_{m=1}^L \sum_{k=1}^M \mathbb{E}\{E_{ik}E^*_jg_{\text{kmpsps}}g^*_{\text{lemsps}}\} \\
= \sum_{i=1}^N \mathbb{E}\{|w_{ki}|^2\} \sum_{m=1}^L \sum_{k=1}^M \zeta_{\text{sps}} \\
+ c \sum_{i=1}^N \sum_{j=1, j \neq i}^N \text{Cov}(w_{ki}, w_{k\ell}^*).
\]
Yeon-Geun Lim (Member, IEEE) received the B.S. degree in information and communications engineering from Sungkyunkwan University, South Korea, in 2012, and the Ph.D. degree in engineering from Yonsei University, South Korea, in 2019, under the supervision of Prof. Chan-Byoung Chae. He is currently a Staff Engineer with Samsung Electronics and has worked on 5G NR standardization in 3GPP-RAN1. He was involved in several industrial and national projects sponsored by Samsung Electronics, ITP, and KCA. He has published in various research areas, 17 papers in international journals and conferences. His research interest includes AI/ML algorithms for physical layer wireless communications, massive MIMO, mmwave, full-duplex, new waveform, ray-tracing channel and link, and system level simulation. He was a recipient/co-recipient of Samsung Humantech Paper Award in 2018 and 2019.

Linglong Dai (Fellow, IEEE) received the B.S. degree from Zhejiang University, Hangzhou, China, in 2003, the M.S. degree (Hons.) from the China Academy of Telecommunications Technology, Beijing, China, in 2006, and the Ph.D. degree (Hons.) from Tsinghua University, Beijing, in 2011. From 2011 to 2013, he was a Post-Doctoral Researcher with the Department of Electronic Engineering, Tsinghua University, where he is currently an Associate Professor. He has coauthored the book MmWave Massive MIMO: A Paradigm for 5G (Academic Press, 2016). He has authored or coauthored over 70 IEEE journal papers and over 50 IEEE conference papers. He also holds 19 granted patents. His current research interests include reconfigurable intelligent surface (RIS), massive MIMO, mm-wave and terahertz communications, electromagnetic information theory, and machine learning for wireless communications. He has received five IEEE Best Paper Awards at the IEEE ICC 2013, the IEEE ICC 2014, the IEEE ICC 2017, the IEEE VTC 2017–Fall, and the IEEE ICC 2018. He has also received the Tsinghua University Outstanding Ph.D. Graduate Award in 2011, the Beijing Excellent Doctoral Dissertation Award in 2012, the China National Excellent Doctoral Dissertation Nomination Award in 2018, the IEEE DySPAN Best Demo Award in 2018, the IEEE/KICS JOURNAL OF COMMUNICATIONS AND NETWORKS Best Paper Award in 2018, the Award of Excellence in Leadership of 100 Leading Core Technologies for Korea 2025 from the NAEK in 2017, the Yonam Research Award from the LG Yonam Foundation in 2016, the IEEE INFOCOM Best Demo Award in 2015, the KICS Haedong Young Scholar Award in 2013, IEEE Signal Processing Magazine Best Paper Award in 2013, the IEEE ComSoc AP Outstanding Young Researcher Award in 2012, the IEEE VTS Dan. E. Noble Fellowship Award in 2008. He is currently an Editor-in-Chief of the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL, AND MULTI-SCALE COMMUNICATIONS and a Senior Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS. He has served/serves as the Editor for the IEEE Communications Magazine (Since 2016), the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL, AND MULTI-SCALE COMMUNICATIONS and a Senior Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS. He has served/serves as the Editor for the IEEE Communications Magazine (Since 2016), the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL, AND MULTI-SCALE COMMUNICATIONS and a Senior Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS. He has served/serves as the Editor for the IEEE Communications Magazine (Since 2016), the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL, AND MULTI-SCALE COMMUNICATIONS (2015–2018), and the IEEE WIRELESS COMMUNICATIONS LETTERS (Since 2016). He is an IEEE ComSoc Distinguished Lecturer for the term 2020–2021 and 2022–2023.

Chan-Byoung Chae (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from The University of Texas at Austin in 2008. Prior to joining UT, he was a Research Engineer at the Telecommunications Research and Development Center, Samsung Electronics, Suwon, South Korea, from 2001 to 2005. He is an Underwood Distinguished Professor at the School of Integrated Technology, Yonsei University, South Korea. Before joining Yonsei University, he was with the Bell Laboratories, Alcatel-Lucent, Murray Hill, NJ, USA, from 2009 to 2011, as a Member of Technical Staff; and Harvard University, Cambridge, MA, USA, from 2008 to 2009, as a Post-Doctoral Research Fellow. Dr. Chae was a recipient/co-recipient of the IEEE ICC Best Demo Award in 2022, the IEEE WCNC Best Demo Award in 2020, the Best Young Engineer Award from the National Academy of Engineering of Korea (NAEK) in 2019, the IEEE DySPAN Best Demo Award in 2018, the IEEE/KICS JOURNAL OF COMMUNICATIONS AND NETWORKS Best Paper Award in 2018, the Award of Excellence in Leadership of 100 Leading Core Technologies for Korea 2025 from the NAEK in 2017, the Yonam Research Award from the LG Yonam Foundation in 2016, the IEEE INFOCOM Best Demo Award in 2015, the IEIE/IEEE Joint Award for Young IT Engineer of the Year in 2014, the KICS Haedong Young Scholar Award in 2013, IEEE Signal Processing Magazine Best Paper Award in 2013, the IEEE ComSoc AP Outstanding Young Researcher Award in 2012, the IEEE VTS Dan. E. Noble Fellowship Award in 2008. He is currently an Editor-in-Chief of the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL, AND MULTI-SCALE COMMUNICATIONS and a Senior Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS. He has served/serves as the Editor for the IEEE Communications Magazine (Since 2016), the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL, AND MULTI-SCALE COMMUNICATIONS and a Senior Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS. He has served/serves as the Editor for the IEEE Communications Magazine (Since 2016), the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL, AND MULTI-SCALE COMMUNICATIONS (2015–2018), and the IEEE WIRELESS COMMUNICATIONS LETTERS (Since 2016). He is an IEEE ComSoc Distinguished Lecturer for the term 2020–2021 and 2022–2023.