

Article

Self-Interference Channel Training for Full-Duplex Massive MIMO Systems

Taehyoung Kim ¹, Kyungsik Min ² and Sangjoon Park ^{3,*}

¹ Department of Electrical and Electronic Engineering, Yonsei University, Seoul 03722, Korea; khotdog86@gmail.com

² Samsung Electronics Company Ltd., Suwon 16677, Korea; minkyungsik@gmail.com

³ Department of Electronic Engineering, Kyonggi University, Suwon 16227, Korea

* Correspondence: sj.park@kgu.ac.kr

Abstract: Full-duplex (FD) is a promising technology for increasing the spectral efficiency of next-generation wireless communication systems. A major technical challenge in enabling FD in a real network is to remove the self-interference (SI) caused by simultaneous transmission and reception at the transceiver, and the SI cancellation performance depends significantly on the estimation accuracy of the SI channel. In this study, we proposed a novel partial SI channel training method for minimizing the residual SI power for FD massive multiple-input multiple-output (MIMO) systems. Based on an SI channel training framework under a limited training overhead, using the proposed scheme, the BS estimates only a part of the SI channel vectors, while skipping the channel training for the other remaining SI channel vectors by using their last estimates. With this partial training framework, the proposed scheme finds the optimal partial SI channel training strategy for pilot allocation to minimize the expected residual SI power, considering the time-varying Rician fading channel model for the SI channel. Therefore, the proposed scheme can improve the sum-rate performance compared with other simple partial training schemes for FD massive MIMO systems under a limited training overhead. Numerical results confirm the effectiveness of the proposed scheme for FD massive MIMO systems compared with the full training scheme, as well as other partial training schemes.

Keywords: full-duplex; massive MIMO; self-interference; channel estimation; partial training



Citation: Kim, T.; Min, K.; Park, S. Self-Interference Channel Training for Full-Duplex Massive MIMO Systems. *Sensors* **2021**, *21*, 3250. <https://doi.org/10.3390/s21093250>

Academic Editor: Gianmarco Romano

Received: 26 March 2021

Accepted: 5 May 2021

Published: 7 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Fifth-generation (5G) wireless communication systems, called new radio (NR), have been successfully commercialized at a global level [1,2]. The 5G NR provides a wide range of services with various requirements such as enhanced mobile broadband (eMBB), ultra-reliable low-latency communications (URLLC), and massive machine-type communications (mMTC) [3]. Despite the remarkable improvement in 5G NR in terms of performance and functionalities, there are extensive ongoing studies and standardization efforts to shape next-generation wireless communication systems, namely the sixth-generation (6G) [4]. It is expected that 6G will provide an ultimate experience beyond even that of 5G NR by enabling new services and applications such as multisensory extended reality, mobile holograms, connected robotics, autonomous systems, and wireless brain-computer interactions [5–7]. To meet the explosive data traffic expected in the 6G era, massive multiple-input multiple-output (MIMO) [8–13] and full-duplex (FD) [14–16] are indispensable technologies for improving the network capacity.

In massive MIMO systems, the base station (BS) is equipped with a large number of antennas, and the spectral efficiency can be significantly increased by simultaneously serving frequency resources and a large amount of user equipment (UE) concurrently [8–13]. Meanwhile, in FD systems, downlink (DL) and uplink (UL) transmissions occur simultaneously at the transceiver at the same time and frequency resource [14,15]. Thus, FD

systems can theoretically double the spectral efficiency compared with conventional half-duplex (HD) systems such as time division duplex (TDD) and frequency division duplex (FDD) [16]. Therefore, the joint utilization of FD and massive MIMO can significantly improve the system capacity.

The main challenge for FD systems is the self-interference (SI) phenomenon, in which the signal transmitted from the BS or UE becomes an unwanted interference to the transmitter [17]. In the case of FD-BS, unattenuated DL signals incur SI for UL signals at the BS, where the power of the SI is significantly larger than that of the UL signals attenuated by the path loss. Therefore, to make the FD technique feasible, extensive studies have been conducted on SI cancellation (SIC) schemes. In general, both analog-domain SIC [18–23] and digital-domain SIC [24–28] are required. Regarding analog SIC schemes, radio frequency (RF) and baseband (BB) tapping approaches were studied in [18,19], respectively. Furthermore, a two-stage cancellation architecture that combines RF and BB tapping approaches was proposed in [20,21]. In [22], a time-varying least mean square (LMS) adaptive filtering scheme with step-size parameters decreasing with time was developed. In [23], a practical structure for an analog LMS (ALMS)-based analog SIC scheme was investigated. Regarding digital SIC schemes, the removal of both the fundamental and harmonic components of the SI based on the least square (LS) estimation of an SI channel was investigated in [24]. In [25], a digital SIC scheme that eliminates all transmitter impairments to mitigate the receiver phase noise and nonlinearity effects was proposed. Furthermore, in [26], a hybrid beamforming-based SIC scheme was proposed for FD MIMO systems in millimeter-wave communications. In [27], a digital SIC scheme based on an independent component analysis was proposed. In [28], an iterative nonlinear method was studied for FD systems under mixer imbalance and amplifier nonlinearity.

The level of SI cancellation and suppression depends highly on the estimation accuracy of the SI channel [29–31]. Therefore, improving the SI channel estimation accuracy is a key technical issue for FD systems, and several related studies have been conducted [32–36]. In [32,33], a joint estimation of the SI and data channels based on the maximum-likelihood (ML) approach was studied. In [34], a frequency-domain LS channel estimator was proposed, and an optimal pilot pattern was derived to minimize the sum of the mean squared error (MSE). In [35,36], channel estimation methods for FD systems with large-scale antennas were investigated.

Despite such studies on SI channel estimation, little effort has been made to investigate the SI channel training strategy when considering the training overhead problem for FD massive MIMO systems. Unfortunately, prior works such as [35–38] focused on the full training strategy for the SI channel estimation where all elements of the entire SI channel matrix are estimated. For the FD-BS, the size of the SI channel matrix at the BS increases with the number of BS antennas. In other words, assuming orthogonal training for SI channel estimation, the full training strategy requires the training overhead to linearly increase with the number of BS antennas [39,40]. As a result, a large training overhead is required for full SI channel training in FD massive MIMO systems. This not only decreases the duration of the data transmission phase in each transmission block for a potential spectral efficiency reduction, but also delays the beginning of the data transmission phase in each transmission block, which can be a crucial problem in latency-sensitive services such as URLLC. Therefore, the amount of training overhead for the SI channel can be strictly limited according to the system environment and services, which makes it impossible to apply the full SI channel training strategy. In this case, the BS can simply choose only a part of the SI channel vectors for training in either a random or round-robin manner under a limited training overhead. For random training, the BS arbitrarily chooses a part of the antennas for SI channel training at a given time instance, whereas for the round-robin training, the BS antenna is sequentially selected in a round-robin manner according to previous selections. However, because no optimization is considered for allocation, both partial training strategies can yield poor throughput performance. Thus, an optimized

partial SI channel training strategy needs to be developed instead of an infeasible full training strategy and non-optimized partial training strategies.

Therefore, in this study, we propose an efficient SI channel estimation framework based on a partial training strategy for FD massive MIMO systems. In the proposed scheme, the pilot signals for SI channel training are allocated to a number of massive BS antennas to satisfy the limited training overhead constraint, and only a part of the SI channel vectors among the entire SI channel matrix corresponding to the allocated BS antennas is estimated for each transmission block. Considering that the SI channel can be modeled as a slowly varying Rician fading channel [17,30,38,41,42], we formulated an optimization problem to find an optimal pilot allocation strategy that minimizes the residual SI power after the SIC operation for a given Rician fading channel model. For this purpose, a closed-form expression for the residual SI power after the SIC operation was derived in terms of pilot resource allocation, and a comprehensive algorithm to solve the optimization problem was developed.

The contributions of this study are summarized as follows:

- We developed an efficient partial SI channel training framework for FD massive MIMO under limited training overhead. Considering that the SI channel can be modeled as a quasi-static channel, it is possible to reduce the training overhead by estimating only some of the SI channel vectors corresponding to a set of selected BS antennas, whereas the previous estimates are utilized for the SI channel vectors corresponding to the unselected BS antennas. This can improve the effective throughput in FD massive MIMO systems for a given training overhead.
- To find an optimal training strategy under the proposed SI channel training framework, we formulated an optimization problem to minimize the expected residual SI power after the SIC operation. To this end, we analyzed and derived a closed-form expression of the expected residual SI power in terms of pilot allocation. Based on the reformulated optimization problem, a simple algorithm to find the optimal allocation was applied, where the optimal allocation can be simply conducted through the BS antenna selection and pilot resource allocations.
- The effectiveness of the proposed scheme was verified based on extensive numerical results. It is shown that the proposed scheme can improve the UL sum-rate by approximately 116.7% and 57.8% at a maximum compared with the cases in which the pilot signals for the SI channel training are allocated randomly or in a round-robin manner, respectively. In addition, because the required training overhead for the SI channel training to achieve a given target UL sum-rate is decreased, the duration of the simultaneous DL and UL data transmission phase in each transmission block can be increased using the proposed scheme. As a result, for a given target UL sum-rate, the proposed scheme can also improve the DL sum-rate by approximately 13.2% at a maximum compared with the round-robin training.

The remainder of this paper is organized as follows. Section 2 describes the system model, including the channel model and basic communication procedure for FD massive MIMO systems. Section 3 presents the optimization problem for minimizing the residual self-interference power and provides a simple algorithm for training optimization based on joint antenna selection and pilot allocation. Section 4 provides numerical results to verify the effectiveness of the proposed method compared with other partial training methods. Finally, Section 5 concludes the paper.

Notations: In this paper, matrices and vectors are denoted in bold upper and lower cases, respectively. In addition, $(\cdot)^*$ and $(\cdot)^H$ denote the conjugate and conjugate-transpose operations, respectively; \mathbf{x}_i and x_{ij} denote the i th column vector and the (i, j) th element of \mathbf{X} , and $\mathbb{E}[\cdot]$ denotes the expectation operator. Moreover, $\text{tr}\{\cdot\}$ is the trace operator. \mathbf{I}_N is an $N \times N$ identity matrix. $[\mathbf{I}_N]_n$ is the n th column vector of \mathbf{I}_N . $|\mathcal{S}|$ indicates the cardinality of the set \mathcal{S} , and $\text{diag}\{\cdot\}$ and $\text{sort}\{\cdot\}$ represent the diagonalization and ordering operations, respectively. $\mathcal{CN}(0, 1)$ represents a complex Gaussian random variable with zero mean and unit variance.

2. System Model

We considered a unidirectional FD system consisting of an FD-BS and HD-UE, as shown in Figure 1 [17]. It was assumed that the FD-BS is equipped with M shared antennas, such that the SI is composed of self-talk and cross-talk [14]. We considered $K(= K_u + K_d)$ HD-UEs with a single antenna in which there are K_u UL and K_d DL UEs in the system.

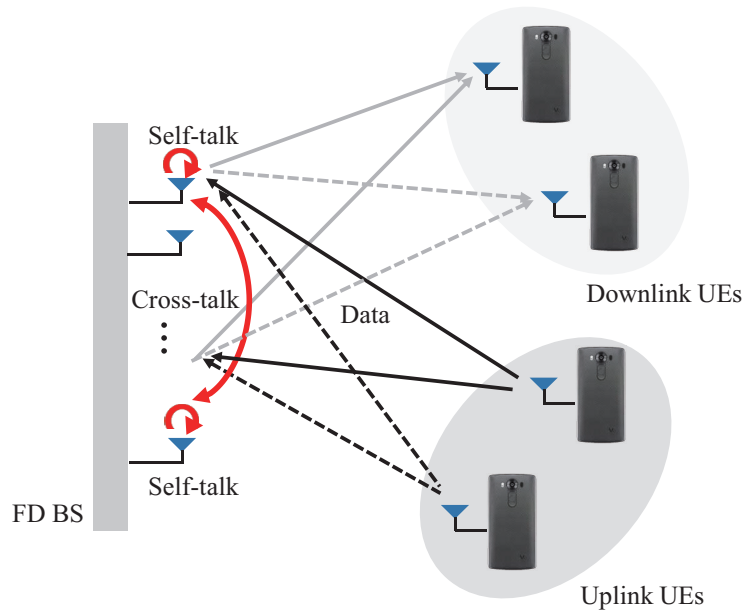


Figure 1. FD massive MIMO system model.

A block-fading channel model was considered, where the channel was assumed to be static during a coherence block of length T channel uses [40]. We denote $\mathbf{H}^u = [\mathbf{h}_1^u, \dots, \mathbf{h}_{K_u}^u]$, $\mathbf{H}^d = [\mathbf{h}_1^d, \dots, \mathbf{h}_{K_d}^d]$, and $\mathbf{H}^s = [\mathbf{h}_1^s, \dots, \mathbf{h}_M^s]$ as an $M \times K_u$ UL channel matrix from the UEs to the BS, an $M \times K_d$ DL channel matrix from the BS to the UEs, and an $M \times M$ SI channel matrix at the BS, respectively. In addition, \mathbf{h}_k^u , \mathbf{h}_k^d , and \mathbf{h}_m^s are the channel vectors for the k th UL UE, the channel vector for the k th DL UE, and the SI channel vector from the m th BS antenna element to all BS antennas, respectively. The desired data signal channels, \mathbf{H}^u and \mathbf{H}^d , are modeled as Rayleigh fading, that is $h_{ij}^u \sim \mathcal{CN}(0, 1)$ and $h_{ij}^d \sim \mathcal{CN}(0, 1)$, where h_{ij}^u and h_{ij}^d are the (i, j) elements of \mathbf{H}^u and \mathbf{H}^d , respectively. Moreover, \mathbf{H}^u and \mathbf{H}^d vary independently between two consecutive blocks. The SI channel matrix \mathbf{H}^s is typically modeled as Rician fading [17,30,38,41,42]. The SI channel is composed of two parts: (i) a strong near-field SI channel representing line-of-sight (LOS) paths and (ii) a weak far-field SI channel representing the reflected non-line-of-sight (NLOS) paths [38,41]. In addition, the NLOS paths are typically slowly changed because the BS is deployed at a high position and does not move. With the Rician fading model, the SI channel element h_{ij}^s is given by the following:

$$h_{ij}^s = \sqrt{\frac{\kappa_{ij}}{\kappa_{ij} + 1}} \bar{h}_{ij}^s + \sqrt{\frac{1}{\kappa_{ij} + 1}} \tilde{h}_{ij}^s \quad (1)$$

where κ_{ij} is the Rician K factor, \bar{h}_{ij}^s is the deterministic part with $|\bar{h}_{ij}^s|^2 = 1$, and \tilde{h}_{ij}^s is a random variable that follows the Rayleigh distribution as $\tilde{h}_{ij}^s \sim \mathcal{CN}(0, 1)$. It is assumed that \tilde{h}_{ij}^s varies between two consecutive blocks, with the following correlation:

$$\tilde{h}_{ij}^s(n) = \sqrt{(c_{ij})^{n'}} \tilde{h}_{ij}^s(n - n') + \sqrt{1 - (c_{ij})^{n'}} q_{ij}, \quad (2)$$

where $\tilde{h}_{ij}^s(n)$ is the random part of the n th block, c_{ij} is the correlation coefficient between two consecutive blocks, and q_{ij} is a new random component, where $q_{ij} \sim \mathcal{CN}(0, 1)$.

We considered a communication procedure that follows the frame structure illustrated in Figure 2. According to the channel model, the channels were assumed to be invariant during a coherent block of T channel uses. During a coherent block, τ channel uses are consumed for SI channel training, and the remaining $(T - \tau)$ channel uses are consumed for the FD data transmission of both the UL and DL. Without loss of generality, to study the effect of knowledge on the SI channel, it was assumed that the BS has perfect knowledge of \mathbf{H}^u and \mathbf{H}^d [17].

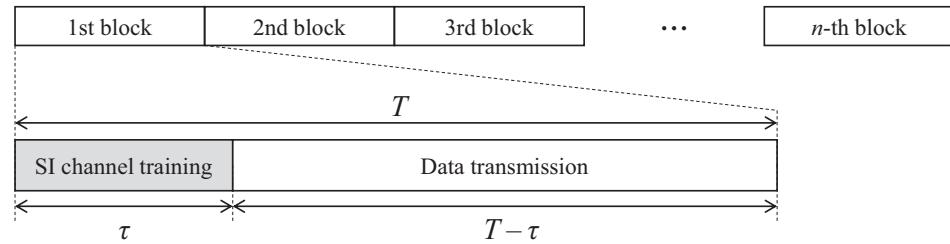


Figure 2. Frame structure for FD massive MIMO systems.

During the SI channel training phase, the BS transmits $\tau \times 1$ pilot sequences $\boldsymbol{\psi}_m$ to obtain $\hat{\mathbf{h}}_m^s$, which is the estimate of the m th SI channel vector \mathbf{h}_m^s . The pilot sequences are pairwise orthogonal, that is $\boldsymbol{\psi}_m^H \boldsymbol{\psi}_m = 1$ and $\boldsymbol{\psi}_m^H \boldsymbol{\psi}_{m'} = 0$, where $m \neq m'$.

After the SI channel training phase, the data transmission phase begins for FD transmission. The BS receives UL data from the K_u UL UEs and transmits DL data to the K_d DL UEs simultaneously. For the UL signal, the FD-BS conducts a series of signal processing procedures to eliminate the SI. In general, the FD receiver is composed of an analog SIC, analog-to-digital conversion (ADC), and digital SIC [14,15]. By passing through the FD receiver, the received UL signal is gradually restored. The received UL data signal vector for the K_u UEs before SIC is given by:

$$\mathbf{y}_u = \underbrace{\sqrt{\rho_u} \mathbf{H}^u \mathbf{x}^u}_{\text{desired signal term}} + \underbrace{\sqrt{\rho_s} \mathbf{H}^s \mathbf{W} \mathbf{x}^d}_{\text{SI term}} + \mathbf{z}^u, \quad (3)$$

where $\mathbf{x}^u = [x_1^u, \dots, x_{K_u}^u]^T$ is a UL data signal vector, ρ_u is the UL received power, $\mathbf{x}^d = [x_1^d, \dots, x_{K_d}^d]^T$ is a DL data signal vector, ρ_s is the DL power received at the BS (i.e., SI power), $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_{K_d}]$ with $\|\mathbf{w}_k\|^2 = 1$ a DL precoding matrix, and $\mathbf{z}^u = [z_1^u, \dots, z_M^u]^T$ is a Gaussian noise vector with zero mean and unit variance. After SIC, the received UL signal can be expressed as:

$$\begin{aligned} \mathbf{y}_{u,\text{SIC}} &= \mathbf{y}_u - \alpha \sqrt{\rho_s} \hat{\mathbf{H}}^s \mathbf{W} \mathbf{x}^d \\ &= \sqrt{\rho_u} \mathbf{H}^u \mathbf{x}^u + \underbrace{\alpha \sqrt{\rho_s} (\mathbf{H}^s - \hat{\mathbf{H}}^s) \mathbf{W} \mathbf{x}^d}_{\text{residual SI term}} + \tilde{\mathbf{z}}^u, \end{aligned} \quad (4)$$

where $\hat{\mathbf{H}}^s = [\hat{\mathbf{h}}_1^s, \dots, \hat{\mathbf{h}}_M^s]$ is the estimated SI channel matrix, α is the analog SIC gain, and $\tilde{\mathbf{z}}^u = [\tilde{z}_1, \dots, \tilde{z}_M]^T$ is the effective Gaussian noise, including the quantization error with zero mean and $\sigma_{\tilde{z}}^2$ variance. By applying the MIMO linear receiver to $\mathbf{y}_{u,\text{SIC}}$, we obtain the following:

$$\begin{aligned} \mathbf{y}_{u,\text{rec}} &= \mathbf{G}^H \mathbf{y}_{u,\text{SIC}} \\ &= \sqrt{\rho_u} \mathbf{G}^H \mathbf{H}^u \mathbf{x}^u + \alpha \sqrt{\rho_s} \mathbf{G}^H (\mathbf{H}^s - \hat{\mathbf{H}}^s) \mathbf{W} \mathbf{x}^d + \mathbf{G}^H \tilde{\mathbf{z}}^u, \end{aligned} \quad (5)$$

where $\mathbf{G} = [\mathbf{g}_1, \dots, \mathbf{g}_{K_u}]$ with $\|\mathbf{g}_k\|^2 = 1$ is the MIMO receiver matrix. Accordingly, the received UL signal for the k th UE is given by:

$$r_{u,k} = \sqrt{\rho_u} \mathbf{g}_k^H \mathbf{h}_k^u x_k^u + \sqrt{\rho_u} \mathbf{g}_k^H \sum_{j \neq k} \mathbf{h}_j^u x_j^u + \alpha \sqrt{\rho_s} \mathbf{g}_k^H (\mathbf{H}^s - \hat{\mathbf{H}}^s) \mathbf{W} \mathbf{x}^d + \mathbf{g}_k^H \mathbf{z}^u. \quad (6)$$

The UL SINR for the k th UE is obtained by the following:

$$\gamma_{u,k} = \frac{\rho_u |\mathbf{g}_k^H \mathbf{h}_k^u|^2}{\rho_u \sum_{j \neq k} |\mathbf{g}_k^H \mathbf{h}_j^u|^2 + \alpha^2 \rho_s \|\mathbf{g}_k^H (\mathbf{H}^s - \hat{\mathbf{H}}^s) \mathbf{W}\|^2 + \sigma_z^2}, \quad (7)$$

and the UL achievable rate for the k th UE is given by:

$$R_{u,k} = \log_2(1 + \gamma_{u,k}). \quad (8)$$

Meanwhile, the received DL data signal for the k th UE is given by:

$$r_{d,k} = \sqrt{\rho_d} (\mathbf{h}_k^d)^H \mathbf{w}_k x_k + \sqrt{\rho_d} \sum_{j \neq k} (\mathbf{h}_j^d)^H \mathbf{w}_j x_j + z_{d,k}, \quad (9)$$

where ρ_d and $z_{d,k}$ are the DL power received at the k th UE and Gaussian noise with zero mean and unit variance, respectively. Thus, the DL SINR for the k th UE is obtained as follows:

$$\gamma_{d,k} = \frac{\rho_d |(\mathbf{h}_k^d)^H \mathbf{w}_k|^2}{\rho_d \sum_{j \neq k} |(\mathbf{h}_j^d)^H \mathbf{w}_j|^2 + 1}, \quad (10)$$

and the corresponding DL achievable rate is given by:

$$R_{d,k} = \log_2(1 + \gamma_{d,k}). \quad (11)$$

Finally, the total sum-rate reflecting the SI channel training overhead can be written as follows:

$$\begin{aligned} \eta_{\text{sum}} &= \eta_u + \eta_d \\ &= \left(1 - \frac{\tau}{T}\right) \left(\sum_{k=1}^{K_u} R_{u,k} + \sum_{k=1}^{K_d} R_{d,k} \right). \end{aligned} \quad (12)$$

3. Proposed Self-Interference Channel Training

SIC performance at the FD receiver relies heavily on the accuracy of the SI channel estimation. To achieve a better SI channel estimation performance, a larger training overhead is required. However, increasing the training overhead for the SI channel reduces the duration of the data training phase, which can cause a large throughput loss for a large training overhead. The SI channel estimation issue is more problematic for massive MIMO systems because the training length for SI channel estimation based on orthogonal training linearly increases with the number of antennas [8]. For a massive MIMO, it is infeasible to employ full training overhead for SI channel estimation, where the training length is larger than or equal to the number of BS antennas. Therefore, we considered a more practical scenario for SI channel estimation at the FD-BS, where the training length is smaller than the number of BS antennas, that is $\tau < M$. Considering the limited training overhead, we developed a partial SI channel training framework.

3.1. Partial Self-Interference Channel Training Framework

Figure 3 illustrates an example of the proposed partial SI channel training method for FD massive MIMO systems under a limited training overhead. Because $\tau < M$ at a given transmission block, it is possible to estimate at most τ SI channel vectors among the M vectors in the entire SI channel matrix. Let $\mathcal{S}_{tr}(n) = \{i_1, i_2, \dots, i_L\}$ be the set of antenna

indices for SI channel vectors estimated at the n th block and $\mathcal{S}_{no}(n) = \{i'_1, i'_2, \dots, i'_{M-L}\}$ be the set of antenna indices for SI channel vectors not estimated at the n th block, where the number of estimated SI channel vectors $L = |\mathcal{S}_{tr}(n)|$ is no greater than τ because multiple training overhead can be used for a given SI channel vector. We can then write $\mathbf{H}_{tr}^s(n) = [\mathbf{h}_{i_1}(n), \mathbf{h}_{i_2}(n), \dots, \mathbf{h}_{i_L}(n)]$ as the $M \times L$ aggregated SI channel matrix to be estimated and $\mathbf{H}_{no}^s(n) = [\mathbf{h}_{i'_1}(n), \mathbf{h}_{i'_2}(n), \dots, \mathbf{h}_{i'_{M-L}}(n)]$ as the remaining $M \times (M - L)$ aggregated SI channel matrix not to be estimated at the n th block. As shown in Figure 3, the BS consumes τ channel uses to estimate a set of SI channel vectors in $\mathbf{H}_{tr}^s(n)$ by properly transmitting pilot signals. In contrast, the BS skips the channel estimation operation for the other remaining SI channel vectors included in $\mathbf{H}_{no}^s(n)$. Instead, it is possible to reuse the last estimations of SI channel vectors obtained during the SI training phase in the previous transmission blocks.

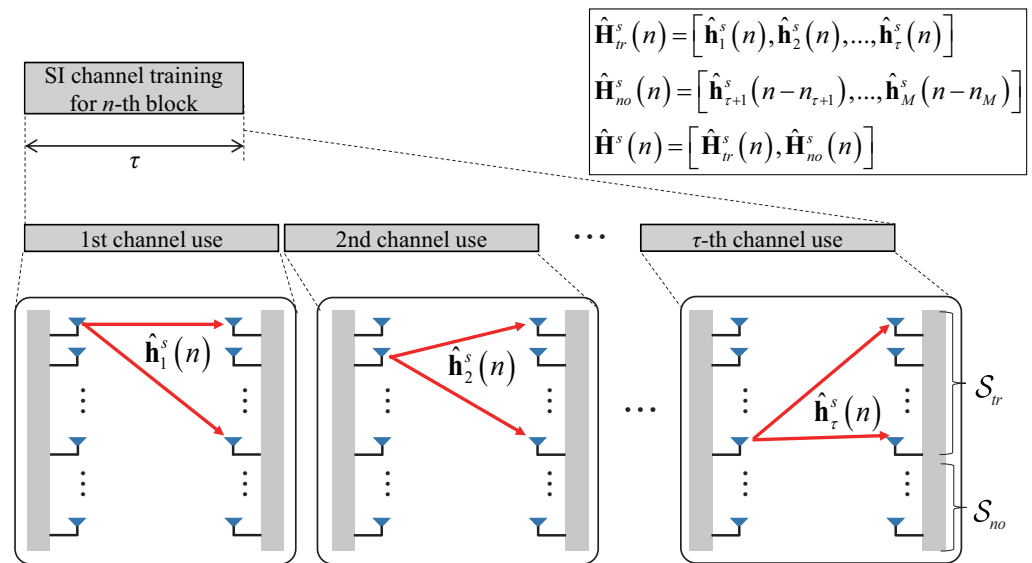


Figure 3. An example of partial SI channel training with limited training length, assuming $\tau = L$ for simplicity.

To estimate $\mathbf{H}_{tr}^s(n)$, the BS can transmit $\tau \times L$ pilot sequences $\mathbf{\Psi} = [\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_L]$, where $\mathbf{\Psi}^H \mathbf{\Psi} = \mathbf{I}_L$. The $M \times \tau$ signal matrix received during the training phase is then given by the following:

$$\mathbf{Y}_{tr}^s(n) = \sqrt{\rho_p} \mathbf{H}_{tr}^s(n) \mathbf{P} \mathbf{\Psi}^H + \mathbf{Z}^s, \quad (13)$$

where ρ_p is the pilot power, $\mathbf{P} = \text{diag}(p_1, \dots, p_L)$ is an $L \times L$ matrix representing the pilot allocation with $\sum_{i=1}^L p_i = \tau$, and \mathbf{Z}^s is an $M \times \tau$ noise matrix with Gaussian elements of $\mathcal{CN}(0, 1)$. According to \mathbf{P} , the BS can consume p_i channel uses for transmitting $\boldsymbol{\psi}_i$. From (13), the BS can obtain $\hat{\mathbf{H}}_{tr}^s(n)$, which is the estimate of $\mathbf{H}_{tr}^s(n)$, by applying the existing conventional linear estimator in the manner of the LS or minimum MSE (MMSE) [40,43]. For the other remaining SI channel vectors in $\mathbf{H}_{no}^s(n)$, by using the nature of slowly-varying SI channels [29–35], the BS can reuse the previous estimated SI channel vectors, i.e.,

$$\begin{aligned} \hat{\mathbf{H}}_{no}^s(n) &= [\hat{\mathbf{h}}_{i'_1}^s(n), \hat{\mathbf{h}}_{i'_2}^s(n), \dots, \hat{\mathbf{h}}_{i'_{M-L}}^s(n)] \\ &= [\hat{\mathbf{h}}_{i'_1}^s(n - n_{i'_1}), \hat{\mathbf{h}}_{i'_2}^s(n - n_{i'_2}), \dots, \hat{\mathbf{h}}_{i'_{M-L}}^s(n - n_{i'_{M-L}})], \end{aligned} \quad (14)$$

where n_i is the number of blocks passed after the last estimation of the i th SI channel vector. Consequently, the total estimated SI channel matrix at the n th block for the proposed partial training method is given as follows:

$$\hat{\mathbf{H}}^s(n) = \text{sort}\{[\hat{\mathbf{H}}_{tr}^s(n), \hat{\mathbf{H}}_{no}^s(n)]\}. \quad (15)$$

3.2. Problem Formulation

To further maximize the throughput, the set of SI channel vectors to be estimated at the given block should be appropriately selected. To formulate an optimization problem with explicit optimization parameters, we define an $M \times L$ selection matrix \mathbf{M} , where the elements $m_{ij} = 1$ if i is the j th element of $\mathcal{S}_{tr}(n)$, and $m_{ij} = 0$, otherwise. Then, (13) can be reformulated as:

$$\begin{aligned} \mathbf{Y}_{tr}^s(n) &= \sqrt{\rho_p} \mathbf{H}^s(n) \mathbf{M} \mathbf{P} \mathbf{\Psi}^H + \mathbf{Z}^s \\ &= \sqrt{\rho_p} \mathbf{H}^s(n) \tilde{\mathbf{M}} \mathbf{\Psi}^H + \mathbf{Z}^s, \end{aligned} \quad (16)$$

where $\tilde{\mathbf{M}} \triangleq \mathbf{M} \mathbf{P}$ is an $M \times L$ matrix representing the antenna selection with a pilot allocation.

By considering the residual SI term in (4) as the metric, after removing the constant parameters α and $\sqrt{\rho_s}$, an optimization problem to minimize the expected residual SI power can be formulated as follows:

$$\arg \min_{\tilde{\mathbf{M}}} \mathbb{E} \left[\left\| \mathbf{G}^H(n) (\mathbf{H}^s(n) - \hat{\mathbf{H}}^s(n)) \mathbf{W}(n) \mathbf{x}^d(n) \right\|^2 \right], \quad (17)$$

subject to $\sum_i \sum_j \tilde{m}_{ij} = \tau$.

To solve the optimization problem in (17), we analyzed the expected residual SI power and derived a closed-form expression of the expected residual SI power in terms of the SI channel estimation error as the following lemma.

Lemma 1. Let v_{ij} be the (i, j) element of $\mathbf{V} = \mathbf{W} \mathbf{W}^H$, u_{ij} be the (i, j) element of $\mathbf{U} = \mathbf{G} \mathbf{G}^H$, and ε_{ij}^2 be the channel estimation error variance, defined as $\varepsilon_{ij}^2 = \mathbb{E}[(h_{ij}^s - \hat{h}_{ij}^s)(h_{ij}^s - \hat{h}_{ij}^s)^*]$. The expected residual SI power is then given as follows:

$$\mathbb{E} \left[\left\| \mathbf{G}^H (\mathbf{H}^s - \hat{\mathbf{H}}^s) \mathbf{W} \mathbf{x}^d \right\|^2 \right] = \sum_{i=1}^M v_{ii} \cdot \left(\sum_{j=1}^M \varepsilon_{ij}^2 u_{jj} \right). \quad (18)$$

Proof. See Appendix A. \square

As shown in Lemma 1, the residual SI power is determined based on the coefficients of the DL precoder v_{ii} , the coefficients of the UL receiver u_{jj} , and the channel estimation error variance ε_{ij}^2 . The BS already has knowledge of v_{ii} and u_{jj} for a given transmission block, and thus, v_{ii} and u_{jj} are deterministic. In contrast, ε_{ij}^2 varies depending on the SI channel training methodology. Given that a partial training strategy is employed, the channel estimation error variance for a given SI channel vector becomes different depending on whether the SI channel vector corresponds to the antenna in \mathcal{S}_{tr} or \mathcal{S}_{no} . To quantify the estimation error variance, we derived the following lemma:

Lemma 2. Assuming the LS estimator, the SI channel estimation error variance at the n th transmission block is given by:

$$\varepsilon_{ij}^2(n) = \begin{cases} \frac{1}{\sum_j \tilde{m}_{ij}(n) \cdot \rho_p} & \text{for } i \in \mathcal{S}_{tr}(n) \\ \frac{2^{-2\sqrt{(c_{ij})^{n_i}}}}{\kappa_{ij} + 1} + \frac{1}{\sum_j \tilde{m}_{ij}(n-n_i) \cdot \rho_p} & \text{for } i \in \mathcal{S}_{no}(n) \end{cases}. \quad (19)$$

Proof. See Appendix B. \square

Figure 4 compares the simulation results for the error variance ε_{ij}^2 with Lemma 2 when $M = 64$. It is shown that the analysis in Lemma 2 is correctly matched with the simulations. According to the results in Figure 4, it is shown that the estimation error variance decreases as the Rician factor increases because the effect of the deterministic part

increases as the Rician factor increases. In addition, the estimation error variance decreases as the correlation factor increases. Therefore, we can conclude that our proposed SI channel estimation framework is a feasible solution for FD massive systems because the SI channel typically has a semi-static characteristic with large Rician and correlation factors.

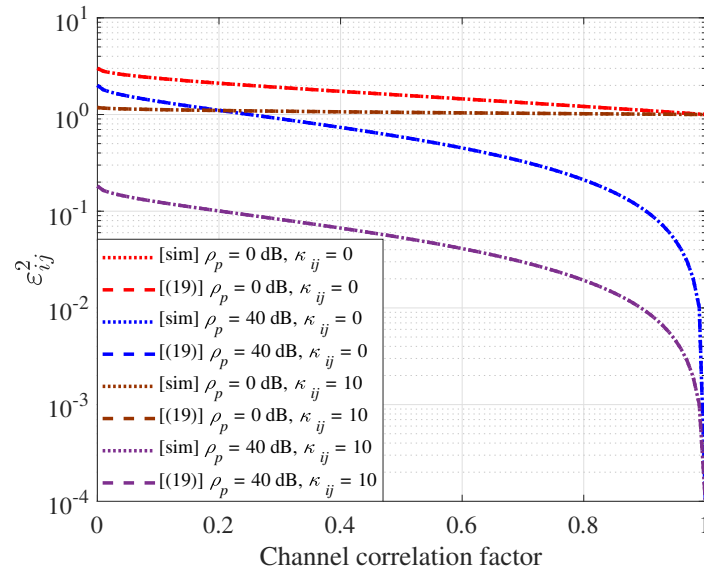


Figure 4. Comparison of the analyzed error variance in (19) and simulated values with $M = 64$. The channel correlation factor along the x-axis corresponds to $(c_{ij})^{n_i}$.

Next, we reformulated the optimization problem to find an optimal pilot allocation $\tilde{\mathbf{M}}^*$ by substituting (18) and (19) into (17) as:

$$\arg \min_{\tilde{\mathbf{M}}(n)} \sum_{i=1}^M v_{ii}(n) f(a_i(n), i), \quad (20)$$

where:

$$f(a_i(n), i) = \begin{cases} \sum_{j=1}^M \left(\frac{1}{a_i(n) \cdot \rho_p} \right) \cdot u_{jj}(n), & a_i(n) > 0 \\ \sum_{j=1}^M \left(\frac{2 - 2\sqrt{(c_{ij})^{n_i}}}{\kappa_{ij} + 1} + \frac{1}{a_i(n - n_i) \cdot \rho_p} \right) \cdot u_{jj}(n), & a_i(n) = 0 \end{cases} \quad (21)$$

subject to $\sum_i a_i(n) = \tau$, where $a_i(n) = \sum_j \tilde{m}_{ij}(n)$ represents the number of allocated channel uses to estimate the i th SI channel vector.

3.3. Proposed Optimal SI Channel Training Strategy

The optimization problem in (20) is a mixed-integer nonlinear programming (MINLP) problem, which is generally difficult to solve [44]. To find an optimal solution of (20), an exhaustive search with a computational complexity of $\mathcal{O}(2^{ML})$ is required, which is infeasible considering a large M of massive MIMO. Therefore, as an alternative, we proposed a simple step-by-step algorithm to solve the optimization problem in (20), which is summarized as follows:

- Step 0 (preparation): Obtain $\mathbf{V}(n)$ and $\mathbf{U}(n)$ for the current transmission block. It is assumed that the BS already has knowledge of the Rician factors (κ_{ij}) , channel correlation coefficients (c_{ij}) , and the previous optimal pilot allocation strategy, that is $\tilde{\mathbf{M}}^*(1), \dots, \tilde{\mathbf{M}}^*(n-1)$.
- Step 1 (initialization): Initialize the parameter $a_i(n) = 0$ for $1 \leq i \leq M$.

- Step 2 (calculation): Calculate the decrease in the residual SI power based on the objective function in (20) when the number of allocated channel uses increases by one, i.e.,

$$\Delta_i(n) = v_{ii}(n)f(a_i(n), i) - v_{ii}(n)f(a_i(n) + 1, i)$$

for all antennas, $1 \leq i \leq M$. Find the antenna index that maximizes the decrement $\Delta_i(n)$, that is,

$$i^* = \arg \max_i \Delta_i(n). \quad (22)$$

Increase the allocated channel uses for the i^* th antenna by:

$$a_{i^*}(n) = a_{i^*}(n) + 1.$$

- Step 3 (next allocation) : Repeat Step 2 until $\sum_i a_i(n) = \tau$.

The proposed algorithm jointly finds the set of antenna indices to be estimated at the n th block and the number of channel uses for the corresponding pilot allocation. Owing to the closed-form expression of the objective function, it is possible to conjecture that the SI channel vector mostly contributes to the minimization of the residual SI power according to the pilot allocation without any trial on the real SI channel estimation. To obtain the initial sets of $\tilde{\mathbf{M}}^*(1), \dots, \tilde{\mathbf{M}}^*(n-1)$, the BS can simply estimate the SI channel vectors in a round-robin manner to ensure that the SI channel matrix has been estimated at least once before the n th block.

It is worth mentioning that the proposed algorithm was designed to find an optimal solution in a greedy manner for low complexity. In each iteration of the algorithm, it is ensured that there always exists only one solution i^* to maximize Δ_i . In addition, whenever an optimal i^* is determined at a given iteration, the metric of the optimization problem in (20) decreases because i^* is selected to reduce the metric based on (22). Consequently, the proposed algorithm always converges. Furthermore, the computational complexity of the proposed algorithm is dominated by the computational complexity of the matrix multiplication to obtain $\mathbf{V}(n)$ and $\mathbf{U}(n)$, that is $\max(\mathcal{O}(M^2K_d), \mathcal{O}(M^2K_u))$. Therefore, although the proposed algorithm has a higher computational complexity compared with the simple random and round-robin training strategies, the proposed algorithm entails a significantly smaller computational complexity from an exhaustive search with a computational complexity of $\mathcal{O}(2^{ML})$.

4. Simulation Results

In this section, the results of the numerical evaluation are presented to verify the performance benefits of the proposed scheme. Random and round-robin training schemes were considered in addition to the proposed scheme. The BS antenna and pilot resource allocations were arbitrarily applied for every transmission block during random training, whereas the BS antenna was sequentially selected in a round-robin manner according to the selection of the last transmission block during the round-robin training. Furthermore, it was assumed that the BS has the outdated SI channel information before the initial transmission block. The LS channel estimator was considered for the SI channel estimation. Further, the zero-forcing (ZF) beamformer was employed for DL transmission, whereas the ZF receiver was employed for UL reception. In addition, unless specified otherwise, the following parameters were considered: $M = 64$, $T = 128$, $\rho_p = 40$ dB, $\rho_d = 30$ dB, $\rho_u = 15$ dB, $c_{ij} = 0.9$, and 10 bit quantization for ADC [17,41,45]. Finally, considering the limited resources for the training overhead, the maximum τ was set to 48 among the $T (= 128)$ channel uses for each transmission block [46].

Figure 5 shows the UL sum-rate ($\sum_k R_{u,k}$) of the training strategies per channel use according to the number of UEs (K_u) and the SI training overhead (τ), where $\alpha = 70$ dB [47,48] and $\kappa_{ij} = 3$ [38], and an 80 dB path loss (obtained by assuming a 100 m distance between the BS and UE with a path loss exponent of four [17]) between the BS and UE was considered. The UL sum-rates per channel use in Figure 5 are not normalized by the amount of training overhead τ , and the SI channel estimation accuracy improves regardless of the training strategy as τ increases. Therefore, the UL sum-rates per channel use in Figure 5 improve regardless of the training strategy as τ increases. Nevertheless, by considering the pilot resource allocation to minimize the residual SI power, the proposed scheme outperforms the random and round-robin schemes. Furthermore, because the minimization of the residual SI power can be more effectively performed, the performance gap of the proposed scheme over the other schemes increases with τ . Specifically, when $\tau = 48$, the UL sum-rate increments of the proposed scheme over the random training are approximately 3.69, 5.29, and 8.15 bits/s/Hz for $K_u = 2, 4$, and 8, respectively, whereas those of the proposed scheme over the round-robin training are approximately 2.51, 2.98, and 3.61 bits/s/Hz for $K_u = 2, 4$, and 8, respectively.

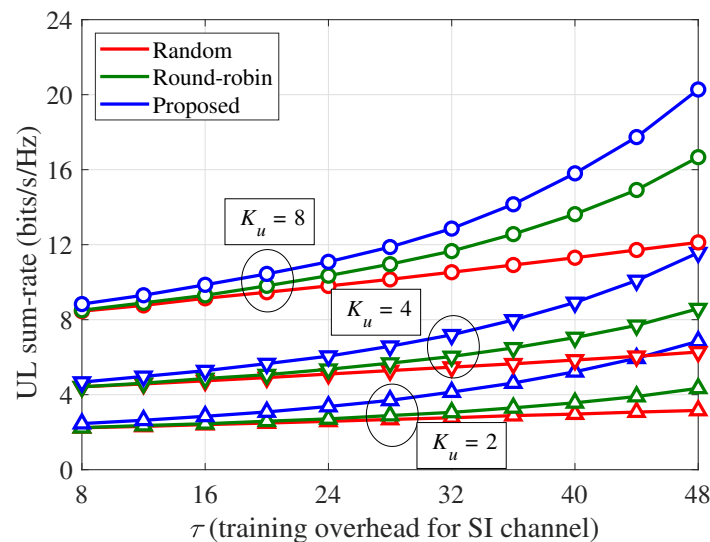


Figure 5. UL sum-rate ($\sum_k R_{u,k}$) per channel use according to the number of UEs (K_u) and SI training overhead (τ): $\alpha = 70$ dB, $\kappa_{ij} = 3$, and 80 dB path loss between the BS and UE.

Figure 6 shows the normalized UL sum-rate (η_u) of the training strategies per channel use according to the number of UEs (K_u) and the SI training overhead (τ), where the system parameters are identical to those in Figure 5. Unlike the results in Figure 5, the UL sum-rate of the random training decreases as τ increases because of the arbitrary pilot allocation in each transmission block. Meanwhile, as the estimation of each SI channel vector is sequentially applied in each block, the round-robin training can prioritize the update of more outdated estimated SI channel vectors, which results in a gradual performance improvement for a larger τ . In contrast, for the given τ and K_u , the proposed scheme outperforms the round-robin training, as well as the random training, which implies that the pilot resource allocations for the minimization of the residual SI power are more effective than those for an update of more outdated SI channel information. When $\tau = 48$, the normalized UL sum-rate improvements of the proposed scheme over the random training are approximately 116.7%, 84.7%, and 67.3% for $K_u = 2, 4$, and 8, respectively, and those of the proposed scheme over the round-robin training are approximately 57.8%, 34.7%, and 21.7% for $K_u = 2, 4$, and 8, respectively.

In Figure 7, the effects of the Rician factor (κ_{ij}) on the normalized UL sum-rate of the partial training strategy are shown, where $K_u = 8$, $\alpha = 70$ dB, and 80 dB path loss between the BS and UE are considered. Because the SI channel becomes more deterministic as κ_{ij} increases, the utilization of the previous estimates yields fewer errors, which leads to a performance improvement of the partial training strategy for a larger κ_{ij} . Meanwhile, as κ_{ij} decreases, the residual SI power can be significantly affected by the pilot resource allocations, and the performance improvement ratio of the proposed scheme over the other schemes can be increased for a smaller κ_{ij} . Specifically, when $\tau = 48$, the normalized UL sum-rate improvements of the proposed scheme over the random training are approximately 92.1%, 64.0%, and 54.2% for $\kappa_{ij} = 0, 4$, and 8, respectively, and those of the proposed scheme over the round-robin training are approximately 29.4%, 20.7%, and 18.1% for $\kappa_{ij} = 0, 4$, and 8, respectively.

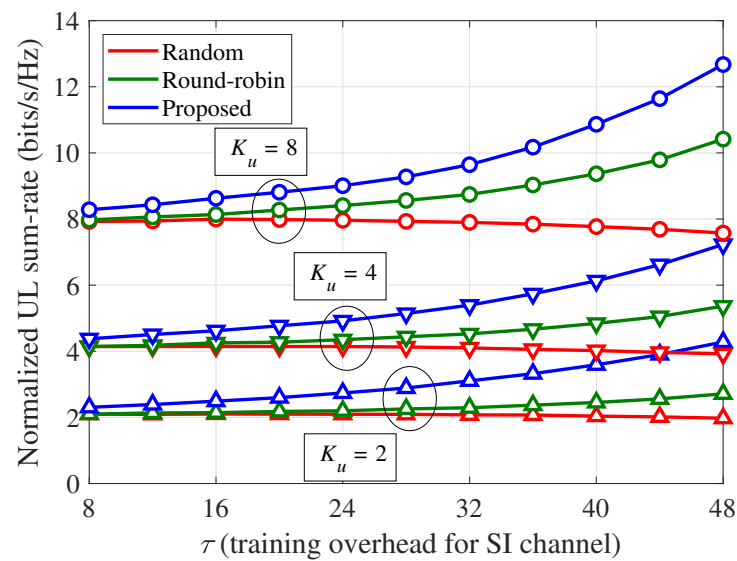


Figure 6. Normalized UL sum-rate (η_u) per channel use according to the number of UEs (K_u) and SI training overhead (τ): $\alpha = 70$ dB, $\kappa_{ij} = 3$, and 80 dB path loss between the BS and UE.

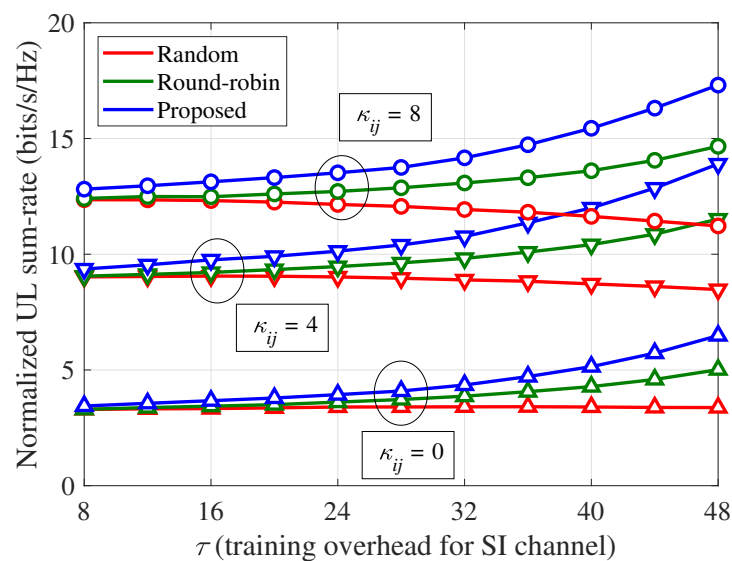


Figure 7. Normalized UL sum-rate (η_u) per channel use according to the Rician factor (κ_{ij}) and SI training overhead (τ): $K_u = 8$, $\alpha = 70$ dB, and 80 dB path loss between the BS and UE.

In Figure 8, the effects of the path loss between the BS and UE on the normalized UL sum-rate of the partial training strategy are illustrated, where $K_u = 8$, $\alpha = 70$ dB, and

$\kappa_{ij} = 3$. As the path loss increases, more DL power is required to achieve the same ρ_d , which results in severe SI problems owing to the increased SI power. Therefore, as the path loss increases, the UL sum-rates of FD massive MIMO systems are significantly degraded, although the minimization of the residual SI power becomes more important. Therefore, the normalized UL sum-rate improvement ratios of the proposed scheme over the other schemes are more significant as the path loss increases.

Figure 9 shows the normalized UL sum-rates per channel use according to the analog SIC gain (α), where $K_u = 8$, $\kappa_{ij} = 3$, $\tau = 48$, and 80 dB path loss between the BS and UE is considered. It is shown that the proposed scheme obtains better UL sum-rates than the other schemes regardless of α . The performance improvement ratios of the proposed scheme increase for a smaller α , because a smaller α indicates more severe SI problems at the FD-BS, similar to the case of the larger path loss shown in Figure 8.

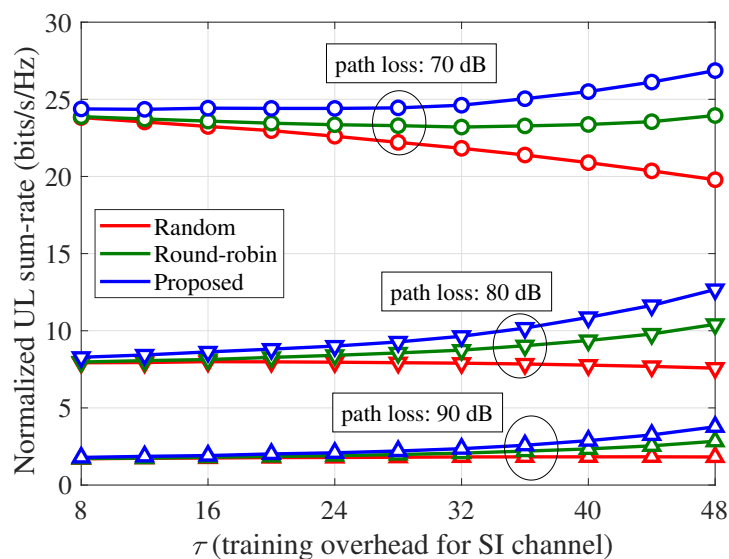


Figure 8. Normalized UL sum-rate (η_u) per channel use according to the path loss between the BS and UE and SI training overhead (τ): $K_u = 8$, $\alpha = 70$ dB, and $\kappa_{ij} = 3$.

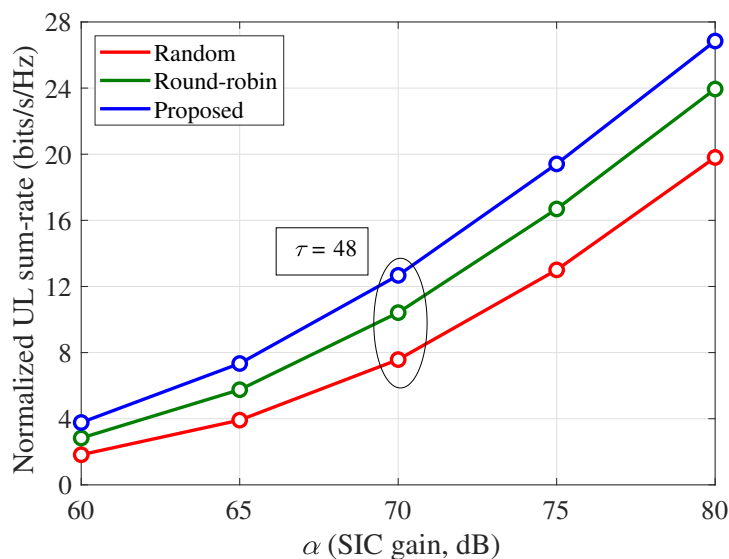


Figure 9. Normalized UL sum-rate (η_u) per channel use according to the analog SIC gain (α): $K_u = 8$, $\kappa_{ij} = 3$, $\tau = 48$, and 80 dB path loss between the BS and UE.

The results in Figures 5–9 show that the proposed scheme can achieve a larger UL sum-rate for a given SI channel training length compared to the other training schemes. This

implies that, to achieve a target UL sum-rate, a smaller training overhead for the SI channel is required for the proposed scheme compared to the other training schemes. Because the proposed scheme requires a smaller training overhead to estimate the SI channel for a given target UL sum-rate, by having a larger data transmission phase, the proposed scheme can also achieve a larger normalized DL sum-rate than the other training schemes.

Therefore, in Figure 10, the required training overhead and corresponding normalized DL sum-rate according to the target normalized UL sum-rate are shown for the proposed scheme and round-robin training. The results for the random training were omitted because they cannot meet the target normalized UL sum-rate of interest. As shown in Figure 10, to achieve a given target normalized UL sum-rate, the proposed scheme requires approximately 26.1% to 76.2% smaller training overhead (τ) compared with the round-robin training. That is, the portion of the data transmission phase ($1 - \tau/T$) in each symbol block can also be increased by the proposed scheme. As a result, for a given target normalized UL sum-rate, the proposed scheme can achieve a maximum improvement of approximately 13.2% on the normalized DL sum-rate from the round-robin training.

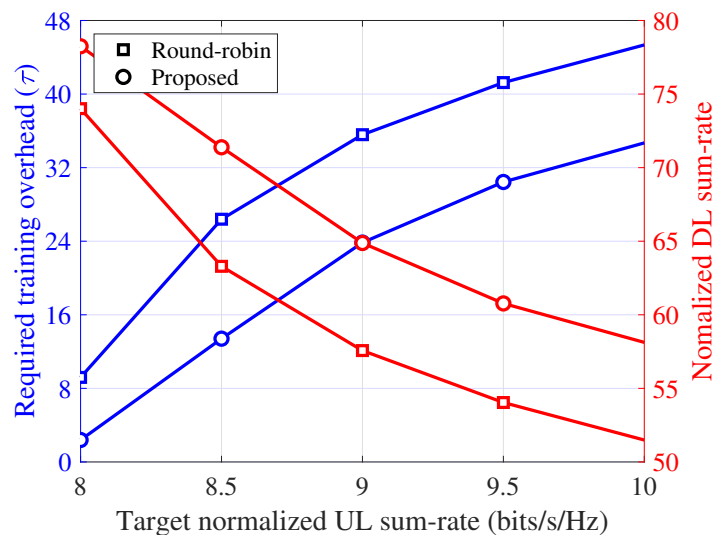


Figure 10. Required training overhead (τ) and normalized DL sum-rate (η_d) per channel use according to the target normalized UL sum-rate: $K_u = 8$, $\kappa_{ij} = 3$, and 80 dB path loss between the BS and UE.

Finally, in Figure 11, the normalized DL and UL sum-rate (η_{sum}) of the partial training strategies per channel use according to the number of BS antennas (M) is compared with that of the full training strategy. Because the full training requires $\tau = M$ to estimate all SI channel vectors, the portion of the data transmission phase, which is used for the full-duplex data transmission of both DL and UL, in each symbol block decreases as M increases. As a result, the normalized sum-rate of full training rapidly decreases for a larger M and approaches zero when $M = T$. In contrast, in cases of partial training strategies, the portion of the data transmission phase decreases as τ increases instead of M in full training. Therefore, although the normalized UL sum-rate of the partial training strategies increases with τ for a given M and T as shown in Figures 5–9, the normalized DL and UL sum-rate of the partial training strategies can be decreased for a larger τ . This implies that the selection of τ for the proposed scheme should be decided based on the target service application with a given requirement for the UL and DL rates.

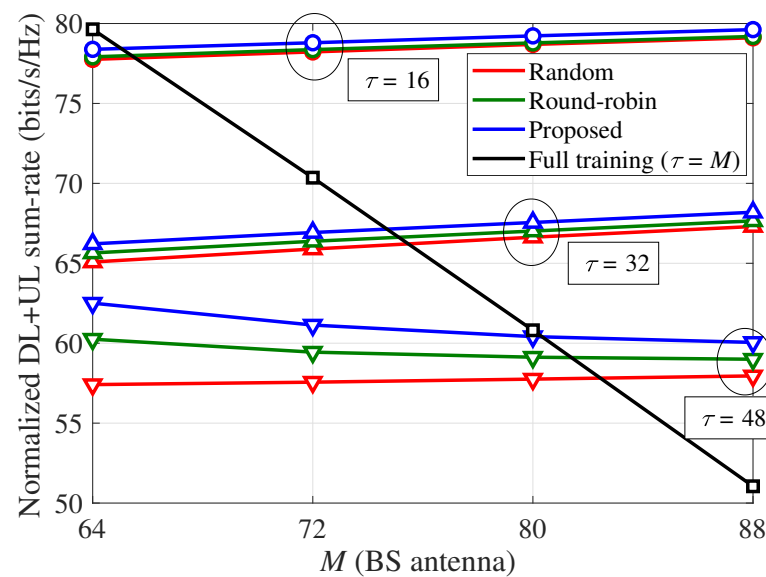


Figure 11. Normalized DL and UL sum-rate (η_{sum}) per channel use according to the BS antenna (M): $K_u = 8$, $\alpha = 70$ dB, $\kappa_{ij} = 3$, and 80 dB path loss between the BS and UE.

5. Conclusions

In this study, we investigated a novel SI channel training framework for FD massive MIMO systems with limited training overhead. The proposed scheme enables a partial training strategy to minimize the expected residual SI power, and it is shown that the objective function of the optimization problem can be reformulated in terms of the parameters of antenna selection and pilot resource allocations. From numerical evaluations, it was verified that the proposed scheme is an effective SI channel training strategy for FD massive MIMO systems, particularly when the system suffers from severe SI signals. In this study, it was assumed that the training overhead and pilot power are already provided by the system. Thus, the proposed scheme can be extended to jointly optimize the training overhead and pilot power with antenna selection and pilot resource allocation. This remains within the scope of future studies.

Author Contributions: Conceptualization, K.M. and S.P.; methodology, T.K. and S.P.; software, S.P.; validation, T.K., K.M. and S.P.; formal analysis, T.K., K.M. and S.P.; investigation, K.M. and S.P.; writing—original draft preparation, T.K. and S.P.; writing—review and editing, T.K. and S.P.; visualization, T.K. and S.P.; supervision, S.P.; funding acquisition, S.P. All authors read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2016R1D1A1B03934546) and the Ministry of Science and ICT (NRF-2019R1C1C1003202).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Lemma 1

By defining $\mathbf{E} \triangleq \mathbf{H}^s - \hat{\mathbf{H}}^s$, the expected residual SI power is calculated as follows:

$$\begin{aligned} \mathbb{E} \left[\left\| \mathbf{G}^H \mathbf{E} \mathbf{W} \mathbf{x}^d \right\|^2 \right] &= \mathbb{E} \left[\text{tr} \left((\mathbf{x}^d)^H \mathbf{W}^H \mathbf{E}^H \mathbf{G} \mathbf{G}^H \mathbf{E} \mathbf{W} \mathbf{x}^d \right) \right] \\ &= \mathbb{E} \left[\text{tr} \left(\mathbf{W} \mathbf{x}^d (\mathbf{x}^d)^H \mathbf{W}^H \mathbf{E}^H \mathbf{G} \mathbf{G}^H \mathbf{E} \right) \right] \\ &= \text{tr} \left(\mathbb{E}[\mathbf{W}] \mathbb{E} \left[\mathbf{x}^d (\mathbf{x}^d)^H \right] \mathbb{E}[\mathbf{W}^H] \mathbb{E} \left[\mathbf{E}^H \mathbf{G} \mathbf{G}^H \mathbf{E} \right] \right) \\ &\stackrel{(a)}{=} \text{tr} \left(\mathbf{W} \mathbf{W}^H \mathbb{E} \left[\mathbf{E}^H \mathbf{G} \mathbf{G}^H \mathbf{E} \right] \right), \end{aligned} \quad (\text{A1})$$

where (a) is from $\mathbb{E} \left[\mathbf{x}^d (\mathbf{x}^d)^H \right] = \mathbf{I}_{K_d}$. Let $\mathbf{U} \triangleq \mathbf{G} \mathbf{G}^H$, and the term $\mathbb{E} \left[\mathbf{E}^H \mathbf{U} \mathbf{E} \right]$ in (A1) can then be formulated as follows:

$$\begin{aligned} \mathbb{E} \left[\mathbf{E}^H \mathbf{U} \mathbf{E} \right] &= \mathbb{E} \left[\begin{bmatrix} \boldsymbol{\varepsilon}_1^H \\ \vdots \\ \boldsymbol{\varepsilon}_M^H \end{bmatrix} \mathbf{U} \begin{bmatrix} \boldsymbol{\varepsilon}_1 & \cdots & \boldsymbol{\varepsilon}_M \end{bmatrix} \right] \\ &= \mathbb{E} \begin{bmatrix} \boldsymbol{\varepsilon}_1^H \mathbf{U} \boldsymbol{\varepsilon}_1 & \cdots & \boldsymbol{\varepsilon}_1^H \mathbf{U} \boldsymbol{\varepsilon}_M \\ \vdots & \ddots & \vdots \\ \boldsymbol{\varepsilon}_M^H \mathbf{U} \boldsymbol{\varepsilon}_1 & \cdots & \boldsymbol{\varepsilon}_M^H \mathbf{U} \boldsymbol{\varepsilon}_M \end{bmatrix} \\ &\stackrel{(b)}{=} \mathbb{E} \begin{bmatrix} \boldsymbol{\varepsilon}_1^H \mathbf{U} \boldsymbol{\varepsilon}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\varepsilon}_M^H \mathbf{U} \boldsymbol{\varepsilon}_M \end{bmatrix}, \end{aligned} \quad (\text{A2})$$

where (b) is derived from the fact that the estimation errors for the different channel vectors are uncorrelated. Here, $\mathbb{E} \left[\boldsymbol{\varepsilon}_i^H \mathbf{U} \boldsymbol{\varepsilon}_i \right]$ for $1 \leq i \leq M$ can be calculated as follows:

$$\begin{aligned} \mathbb{E} \left[\boldsymbol{\varepsilon}_i^H \mathbf{U} \boldsymbol{\varepsilon}_i \right] &= \mathbb{E} \left[\text{tr} \left(\boldsymbol{\varepsilon}_i^H \mathbf{U} \boldsymbol{\varepsilon}_i \right) \right] \\ &= \mathbb{E} \left[\text{tr} \left(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^H \mathbf{U} \right) \right] \\ &= \text{tr} \left(\mathbb{E} \left[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^H \right] \mathbf{U} \right) \\ &= \text{tr} \left(\mathbb{E} \begin{bmatrix} \boldsymbol{\varepsilon}_{i1} \boldsymbol{\varepsilon}_{i1}^* & \cdots & \boldsymbol{\varepsilon}_{i1} \boldsymbol{\varepsilon}_{iM}^* \\ \vdots & \ddots & \vdots \\ \boldsymbol{\varepsilon}_{iM} \boldsymbol{\varepsilon}_{i1}^* & \cdots & \boldsymbol{\varepsilon}_{iM} \boldsymbol{\varepsilon}_{iM}^* \end{bmatrix} \mathbf{U} \right) \\ &\stackrel{(c)}{=} \text{tr} \left(\mathbb{E} \begin{bmatrix} \boldsymbol{\varepsilon}_{i1} \boldsymbol{\varepsilon}_{i1}^* & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{\varepsilon}_{iM} \boldsymbol{\varepsilon}_{iM}^* \end{bmatrix} \mathbf{U} \right) \\ &= \sum_{j=1}^M \boldsymbol{\varepsilon}_{ij}^2 u_{jj}, \end{aligned} \quad (\text{A3})$$

where (c) is derived from the fact that the estimation errors for different channel elements are uncorrelated. By substituting (A2) and (A3) into (A1), we obtain Lemma 1.

Appendix B. Proof of Lemma 2

For $i \in \mathcal{S}_{tr}(n)$, the i th SI channel vector for the n th block is estimated based on the LS estimation. In this case, the channel estimation error variance $\boldsymbol{\varepsilon}_{ij}^2$ is given by the following [43]:

$$\boldsymbol{\varepsilon}_{ij}^2 = \frac{1}{a_i(n) \rho_p}, \quad (\text{A4})$$

where $a_i(n) = \sum_j \tilde{m}_{ij}(n)$ is the number of channel uses consumed for transmitting the pilot signal to estimate the i th SI channel vector.

For $i \in \mathcal{S}_{tr}(n)$, the BS skips the estimation of the i th SI channel vector for the n th block. Instead, the previously estimated i th SI channel vector, that is $\hat{\mathbf{h}}_i^s(n - n_i)$, replaces $\hat{\mathbf{h}}_i^s(n)$. Therefore, the error variance is defined as follows:

$$\varepsilon_{ij}^2(n) = \mathbb{E} \left[\left(h_{ij}^s(n) - \hat{h}_{ij}^s(n - n_i) \right)^* \left(h_{ij}^s(n) - \hat{h}_{ij}^s(n - n_i) \right) \right]. \quad (\text{A5})$$

By substituting the channel characteristics defined in (1) and (2) into (A5), we obtain:

$$\begin{aligned} & \mathbb{E} \left[\left(h_{ij}^s(n) - \hat{h}_{ij}^s(n - n_i) \right)^* \left(h_{ij}^s(n) - \hat{h}_{ij}^s(n - n_i) \right) \right] \\ &= \frac{(1 - (c_{ij})^{n_i})}{\kappa_{ij} + 1} \mathbb{E} \left[q_{ij}^* q_{ij} \right] \\ &+ \mathbb{E} \left[\begin{aligned} & \left(h_{ij}^s(n - n_i) - \hat{h}_{ij}^s(n - n_i) + \sqrt{\frac{1}{\kappa_{ij} + 1}} (\sqrt{(c_{ij})^{n_i}} - 1) \tilde{h}_{ij}^s(n - n_i) \right)^* \\ & \times \left(h_{ij}^s(n - n_i) - \hat{h}_{ij}^s(n - n_i) + \sqrt{\frac{1}{\kappa_{ij} + 1}} (\sqrt{(c_{ij})^{n_i}} - 1) \tilde{h}_{ij}^s(n - n_i) \right) \end{aligned} \right] \\ &= \frac{(1 - (c_{ij})^{n_i})}{\kappa_{ij} + 1} + \mathbb{E} \left[\left(h_{ij}^s(n - n_i) - \hat{h}_{ij}^s(n - n_i) \right)^* \left(h_{ij}^s(n - n_i) - \hat{h}_{ij}^s(n - n_i) \right) \right] \\ &+ \frac{(\sqrt{(c_{ij})^{n_i}} - 1)^2}{\kappa_{ij} + 1} \mathbb{E} \left[\left(\tilde{h}_{ij}^s \right)^* (n - n_i) \tilde{h}_{ij}^s(n - n_i) \right] \\ &+ \sqrt{\frac{1}{\kappa_{ij} + 1}} (\sqrt{(c_{ij})^{n_i}} - 1) \left(\begin{aligned} & \mathbb{E} \left[\left(h_{ij}^s(n - n_i) - \hat{h}_{ij}^s(n - n_i) \right)^* \tilde{h}_{ij}^s(n - n_i) \right] \\ & + \mathbb{E} \left[\left(\tilde{h}_{ij}^s \right)^* (n - n_i) \left(h_{ij}^s(n - n_i) - \hat{h}_{ij}^s(n - n_i) \right) \right] \end{aligned} \right) \\ &\stackrel{(a)}{=} \frac{(1 - (c_{ij})^{n_i})}{\kappa_{ij} + 1} + \frac{1}{a_i(n - n_i) \cdot \rho_p} + \frac{(\sqrt{(c_{ij})^{n_i}} - 1)^2}{\kappa_{ij} + 1} \\ &+ \sqrt{\frac{1}{\kappa_{ij} + 1}} (\sqrt{(c_{ij})^{n_i}} - 1) \left(\mathbb{E} \left[(\delta_{ij}(n - n_i))^* \tilde{h}_{ij}^s(n - n_i) \right] + \mathbb{E} \left[\left(\tilde{h}_{ij}^s \right)^* (n - n_i) \delta_{ij}(n - n_i) \right] \right) \\ &\stackrel{(b)}{=} \frac{(1 - (c_{ij})^{n_i})}{\kappa_{ij} + 1} + \frac{1}{a_i(n - n_i) \cdot \rho_p} + \frac{(\sqrt{(c_{ij})^{n_i}} - 1)^2}{\kappa_{ij} + 1} \\ &= \frac{2 - 2\sqrt{(c_{ij})^{n_i}}}{\kappa_{ij} + 1} + \frac{1}{a_i(n - n_i) \cdot \rho_p}. \quad (\text{A6}) \end{aligned}$$

In (A6), (a) is derived from $\mathbb{E} \left[\left(h_{ij}^s(n - n_i) - \hat{h}_{ij}^s(n - n_i) \right)^* \left(h_{ij}^s(n - n_i) - \hat{h}_{ij}^s(n - n_i) \right) \right] = \frac{1}{a_i(n - n_i) \cdot \rho_p}$ and $\mathbb{E} \left[\left(\tilde{h}_{ij}^s \right)^* (n - n_i) \tilde{h}_{ij}^s(n - n_i) \right] = 1$, and (b) is derived from $\mathbb{E} \left[\delta_{ij}(n - n_i) \right] = 0$ with $\delta_{ij}(n - n_i) \triangleq h_{ij}^s(n - n_i) - \hat{h}_{ij}^s(n - n_i)$. Consequently, we obtain Lemma 2.

References

- Kim, Y.; Kim, Y.; Oh, J.; Ji, H.; Yeo, J.; Choi, S.; Ryu, H.; Noh, H.; Kim, T.; Sun, F.; et al. New Radio (NR) and its Evolution toward 5G-Advanced. *IEEE Wirel. Commun.* **2019**, *26*, 2–7. [CrossRef]
- Guevara, L.; Cheein, F.A. The Role of 5G Technologies: Challenges in Smart Cities and Intelligent Transportation Systems. *Sustainability* **2020**, *12*, 6469. [CrossRef]
- Yeo, J.; Kim, T.; Oh, J.; Park, S.; Kim, Y.; Lee, J. Advanced Data Transmission Framework for 5G Wireless Communications in the 3GPP New Radio Standard. *IEEE Commun. Stand. Mag.* **2019**, *3*, 38–43. [CrossRef]
- Qamar, F.; Siddiqui, M.U.A.; Hindia, M.N.; Hassan, R.; Nguyen, Q.N. Issues, Challenges, and Research Trends in Spectrum Management: A Comprehensive Overview and New Vision for Designing 6G Networks. *Electronics* **2020**, *9*, 1416. [CrossRef]
- Zhang, L.; Liang, Y.; Niyato, D. 6G Visions: Mobile Ultra-Broadband, Super Internet-of-Things, and Artificial Intelligence. *China Commun.* **2019**, *16*, 1–14.
- Saad, W.; Bennis, M.; Chen, M. A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems. *IEEE Netw.* **2020**, *34*, 134–142. [CrossRef]
- Jiang, W.; Han, B.; Habibi, M.A.; Schotten, H.D. The Road Towards 6G: A Comprehensive Survey. *IEEE Open J. Commun. Soc.* **2021**, *2*, 334–366. [CrossRef]

8. Marzetta, T.L. Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas. *IEEE Trans. Wirel. Commun.* **2010**, *9*, 3590–3600. [[CrossRef](#)]
9. Ngo, H.Q.; Larsson, E.G.; Marzetta, T.L. Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems. *IEEE Trans. Commun.* **2013**, *61*, 1436–1449.
10. Lu, L.; Li, G.Y.; Swindlehurst, A.L.; Ashikhmin, A.; Zhang, R. An Overview of Massive MIMO: Benefits and Challenges. *IEEE J. Sel. Top. Signal Process.* **2014**, *8*, 742–758. [[CrossRef](#)]
11. Onggosanusi, E.; Rahman, M.S.; Guo, L.; Kwak, Y.; Noh, H.; Kim, Y.; Faxer, S.; Harrison, M.; Frenne, M.; Grant, S. Modular and High-Resolution Channel State Information and Beam Management for 5G New Radio. *IEEE Commun. Mag.* **2018**, *56*, 48–55. [[CrossRef](#)]
12. Kim, T.; Park, S. Scaling Laws of Scheduling Gain for Uplink Massive MIMO Systems: Is User Scheduling Still Beneficial for Massive MIMO? *Electronics* **2020**, *9*, 1650. [[CrossRef](#)]
13. Kim, T.; Park, S. Statistical Beamforming for Massive MIMO Systems with Distinct Spatial Correlations. *Sensors* **2020**, *20*, 6255. [[CrossRef](#)] [[PubMed](#)]
14. Sabharwal, A.; Schniter, P.; Guo, D.; Bliss, D.W.; Rangarajan, S.; Wichman, R. In-Band Full-Duplex Wireless: Challenges and Opportunities. *IEEE J. Sel. Areas Commun.* **2014**, *32*, 1637–1652. [[CrossRef](#)]
15. Kim, D.; Lee, H.; Hong, D. A Survey of In-Band Full-Duplex Transmission: From the Perspective of PHY and MAC Layers. *IEEE Commun. Surv. Tutor.* **2015**, *17*, 2017–2046. [[CrossRef](#)]
16. Zhang, Z.; Long, K.; Vasilakos, A.V.; Hanzo, L. Full-Duplex Wireless Communications: Challenges, Solutions, and Future Research Directions. *Proc. IEEE* **2016**, *104*, 1369–1409. [[CrossRef](#)]
17. Park, E.; Bae, J.; Ju, H.; Han, Y. Resource Allocation for Full-Duplex Systems with Imperfect Co-Channel Interference Estimation. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 2388–2400. [[CrossRef](#)]
18. Duarte, M.; Dick, C.; Sabharwal, A. Experiment-Driven Characterization of Full-Duplex Wireless Systems. *IEEE Trans. Wirel. Commun.* **2012**, *11*, 4296–4307. [[CrossRef](#)]
19. Kolodziej, K.E.; McMichael, J.G.; Perry, B.T. Multitap RF Canceller for In-Band Full-Duplex Wireless Communications. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 4321–4334. [[CrossRef](#)]
20. Quan, X.; Liu, Y.; Pan, W.; Tang, Y.; Kang, K. A Two-Stage Analog Cancellation Architecture for Self-Interference Suppression in Full-Duplex Communications. In Proceedings of the 2017 IEEE MTT-S International Microwave Symposium (IMS), Honolulu, HI, USA, 4–9 June 2017; pp. 1–4.
21. Liu, Y.; Roblin, P.; Quan, X.; Pan, W.; Shao, S.; Tang, Y. A Full-Duplex Transceiver with Two-Stage Analog Cancellations for Multipath Self-Interference. *IEEE Trans. Microw. Theory Tech.* **2017**, *65*, 5263–5273. [[CrossRef](#)]
22. Sun, Z.-Y.; Zhao, Y.-J. A Novel Self-Interference Cancellation Method Using an Optimized LMS Algorithm in CCFD Systems for a 5G Communication Network. *Appl. Sci.* **2019**, *9*, 3308. [[CrossRef](#)]
23. Le, A.T.; Tran, L.C.; Huang, X.; Guo, Y.J. Analog Least Mean Square Loop for Self-Interference Cancellation: A Practical Perspective. *Sensors* **2020**, *20*, 270. [[CrossRef](#)] [[PubMed](#)]
24. Bharadia, D.; McMilin, E.; Katti, S. Full Duplex Radios. In Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM, Hong Kong, China, 12–16 August 2013; pp. 375–386.
25. Ahmed, E.; Eltawil, A.M. All-Digital Self-Interference Cancellation Technique for Full-Duplex Systems. *IEEE Trans. Wirel. Commun.* **2015**, *14*, 3519–3532. [[CrossRef](#)]
26. Liu, X.; Xiao, Z.; Bai, L.; Choi, J.; Xia, P.; Xia, X.-G. Beamforming Based Full-Duplex for Millimeter-Wave Communication. *Sensors* **2016**, *16*, 1130. [[CrossRef](#)]
27. Li, J.; Zhang, H.; Fan, M. Digital Self-Interference Cancellation Based on Independent Component Analysis for Co-Time Co-frequency Full-Duplex Communication Systems. *IEEE Access* **2017**, *5*, 10222–10231. [[CrossRef](#)]
28. Komatsu, K.; Miyaji, Y.; Uehara, H. Iterative Nonlinear Self-Interference Cancellation for In-Band Full-Duplex Wireless Communications Under Mixer Imbalance and Amplifier Nonlinearity. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 4424–4438. [[CrossRef](#)]
29. Kim, D.; Ju, H.; Park, S.; Hong, D. Effects of Channel Estimation Error on Full-Duplex Two-Way Networks. *IEEE Trans. Veh. Technol.* **2013**, *62*, 4666–4672.
30. Jang, Y.; Min, K.; Park, S.; Jung, M.; Ko, K.; Choi, S. Dispersed Signal Transmission and Reception Scheme for Full-Duplex Systems. *IEEE Access* **2019**, *7*, 138771–138778. [[CrossRef](#)]
31. Nguyen, B.C.; Thang, N.N.; Tran, X.N.; Dung, L.T. Impacts of Imperfect Channel State Information, Transceiver Hardware, and Self-Interference Cancellation on the Performance of Full-Duplex MIMO Relay System. *Sensors* **2020**, *20*, 1671. [[CrossRef](#)] [[PubMed](#)]
32. Masmoudi, A.; Le-Ngoc, T. A Maximum-Likelihood Channel Estimator for Self-Interference Cancellation in Full-Duplex Systems. *IEEE Trans. Veh. Technol.* **2016**, *65*, 5122–5132. [[CrossRef](#)]
33. Qiao, G.; Gan, S.; Liu, S.; Song, Q. Self-Interference Channel Estimation Algorithm Based on Maximum-Likelihood Estimator in In-Band Full-Duplex Underwater Acoustic Communication System. *IEEE Access* **2018**, *6*, 62324–62334. [[CrossRef](#)]
34. Shu, F.; Wang, J.; Li, J.; Chen, R.; Chen, W. Pilot Optimization, Channel Estimation, and Optimal Detection for Full-Duplex OFDM Systems with IQ Imbalances. *IEEE Trans. Veh. Technol.* **2017**, *66*, 6993–7009. [[CrossRef](#)]
35. Xiong, X.; Wang, X.; Riihonen, T.; You, X. Channel Estimation for Full-Duplex Relay Systems with Large-Scale Antenna Arrays. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 6925–6938. [[CrossRef](#)]

36. Chen, D.; Liu, F.; Feng, C.; Zhao, W. Coupling Channel Estimation Based Self-Interference Cancellation in Massive MIMO Full-Duplex System. In Proceedings of the 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, Canada, 8–13 October 2017; pp. 1–5.
37. Koh, J.; Lim, Y.; Chae, C.; Kang, J. On the Feasibility of Full-Duplex Large-Scale MIMO Cellular Systems. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 6231–6250. [[CrossRef](#)]
38. Wang, S.; Liu, Y.; Zhang, W.; Zhang, H. Achievable Rates of Full-Duplex Massive MIMO Relay Systems Over Rician Fading Channels. *IEEE Trans. Veh. Technol.* **2017**, *66*, 9825–9837. [[CrossRef](#)]
39. Ngo, H.Q.; Larsson, E.G.; Marzetta, T.L. Massive MU-MIMO Downlink TDD Systems with Linear Precoding and Downlink Pilots. In Proceedings of the 2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 2–4 October 2013; pp. 293–298.
40. Kim, T.; Min, K.; Jung, M.; Choi, S. Scaling Laws of Optimal Training Lengths for TDD Massive MIMO Systems. *IEEE Trans. Veh. Technol.* **2018**, *67*, 7128–7142. [[CrossRef](#)]
41. Koc, A.; Le-Ngoc, T. Full-Duplex mmWave Massive MIMO Systems: A Joint Hybrid Precoding/Combining and Self-Interference Cancellation Design. *IEEE Open J. Commun. Soc.* **2021**, *2*, 754–774. [[CrossRef](#)]
42. Satyanarayana, K.; El-Hajjar, M.; Kuo, P.; Mourad, A.; Hanzo, L. Hybrid Beamforming Design for Full-Duplex Millimeter Wave Communication. *IEEE Trans. Veh. Technol.* **2019**, *68*, 1394–1404. [[CrossRef](#)]
43. Wu, J.; Xiao, C. Optimal Diversity Combining Based on Linear Estimation of Rician Fading Channels. *IEEE Trans. Commun.* **2008**, *56*, 1612–1615. [[CrossRef](#)]
44. Yu, G.; Xu, L.; Feng, D.; Yin, R.; Li, G.Y.; Jiang, Y. Joint Mode Selection and Resource Allocation for Device-to-Device Communications. *IEEE Trans. Commun.* **2014**, *62*, 3814–3824. [[CrossRef](#)]
45. Kurose, D.; Ito, T.; Ueno, T.; Yamaji, T.; Itakura, T. 55-mW 200-MSPS 10-Bit Pipeline ADCs for Wireless Receiver. *IEEE J. Solid-State Circ.* **2006**, *41*, 1589–1595. [[CrossRef](#)]
46. 3GPP. 3GPP Technical Specification TS 36.211. *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 16)*, V16.5.0.; 3GPP: Sophia, France, 2021.
47. Everett, E.; Sahai, A.; Sabharwal, A. Passive Self-Interference Suppression for Full-Duplex Infrastructure Nodes. *IEEE Trans. Wirel. Commun.* **2014**, *13*, 680–694. [[CrossRef](#)]
48. Ji, H.; Kim, Y.; Muhammad, K.; Tarver, C.; Tonnemacher, M.; Kim, T.; Oh, J.; Yu, B.; Xu, G.; Lee, J. Extending 5G TDD Coverage With XDD: Cross Division Duplex. *IEEE Access* **2021**, *9*, 51380–51392. [[CrossRef](#)]