Packet Structure and Receiver Design for Low Latency Wireless Communications with Ultra-Short Packets

Byungju Lee, *Member, IEEE*, Sunho Park, *Member, IEEE*, David J. Love, *Fellow, IEEE*, Hyoungju Ji, *Member, IEEE*, and Byonghyo Shim, *Senior Member, IEEE*

Abstract

Fifth generation (5G) wireless standards require much lower latency than what current wireless systems can guarantee. The main challenge in fulfilling these requirements is the development of short packet transmission, in contrast to most of the current standards which use a long data packet structure. Since the available training resources are limited by the packet size, reliable channel and interference covariance estimation with reduced training overhead is crucial to any system using short data packets. In this paper, we propose an efficient receiver that exploits useful information available in the data transmission period to enhance the reliability of the short packet transmission. In the proposed method, the receive filter (i.e., the sample covariance matrix) is estimated using the received samples from the data transmission without using an interference training period. A channel estimation algorithm to use the most reliable data symbols as virtual pilots is employed to improve quality of the channel estimate. Simulation results verify that the proposed receiver algorithms enhance the reception quality of the short packet transmission.

B. Lee and D. J. Love are with School of Electrical and Computer Engineering, Purdue Univ., West Lafayette, IN, USA (email:{byungjulee, djlove}@purdue.edu). S. Park, H. Ji, and B. Shim (corresponding author) are with Institute of New Media and Communications and School of Electrical and Computer Engineering, Seoul National University, Seoul, Korea (e-mail:{shpark,hyoungjuji}@islab.snu.ac.kr,bshim@snu.ac.kr).

This work was supported in part by the National Science Foundation (NSF) grant CNS1642982 and the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIP) (2016R1A2B3015576).

Parts of this paper was presented at the GLOBECOM, Washington, USA, December 4-8, 2016 [1].

Packet Structure and Receiver Design for Low Latency Wireless Communications with Ultra-Short Packets

I. Introduction

Fifth generation (5G) communication networks will be a key enabler in realizing the Internet of Things (IoT) era and hyper-connected society [2]–[4]. To support real-time applications with stringent delay requirements, communication systems supporting ultra low latency are needed [5]–[8]. For this reason, International telecommunication union (ITU) defined ultra reliable and low latency communication (uRLLC) as one of key use cases for 5G wireless communications.[1] One direct way to meet the stringent low latency requirements is to use a short-sized packet. This is in contrast to most current wireless systems whose sole purpose is to transmit long data packets efficiently. Indeed, most current physical layer design relies heavily on long codes to approach Shannon capacity. Sensors and devices in an IoT network transmit a very small amount of information such as environmental data (e.g., temperature, humidity, and pollution density), locations, and emergency alarm, and thus asymptotic capacity-achieving principles are not relevant to the transmission of short packets. In view of this, 5G physical layer design (e.g., pilot transmission strategy, coding scheme, hybrid automatic repeat request) needs to be reevaluated and potentially redesigned as a whole. Many papers have been dedicated to analyzing performance metrics that are relevant to short packet communication, including the maximal achievable rate at finite packet length and finite packet error probability instead of using two classic information-theoretic metrics, the ergodic capacity and the outage capacity [9]–[12]. Finite block-length analysis has been extended to spectrum sharing networks using rate adaptation [13] and wireless energy and information transmission using feedback [14].

In this paper, we consider practical constraints that are encountered when implementing a short packet transmission framework. First, massive and simultaneous communications among

[1]Three key use cases include enhance mobile broadband (eMBB), massive machine-type communications (mMTC), and ultra reliable and low latency communication (uRLLC) [6].

autonomous devices will induce inter-device interference at the receiver in an IoT network. In order to control the inter-device interference, a technique to use multiple receive antennas has been proposed by utilizing the spatial degrees of freedom (DoF) provided by multiple receive antennas to balance interference suppression and desired signal power improvement [15]. To implement this approach, the receiver needs to acquire the desired CSI and the interference plus noise covariance. While the desired CSI can be estimated using pilot signals, direct estimation of the interfering CSI is difficult due to the large number of interfering devices in the IoT network. One way to tackle this problem is to use an *interference training period* in which the interference (plus noise) covariance matrix can be estimated by listening to interference-only transmissions [16]. The main drawback of this approach is that it incurs a severe transmission rate loss since the target transmit device should remain silent during the interference training period. Also, the interference training period would not be large enough to reliably estimate the interference covariance for this short-sized packets, resulting in severe degradation in performance.

Moreover, since current systems are designed to carry long data packets, the pilot transmission period can be made relatively small even though the actual period of pilot signals is large. However, the portion of time used for pilot signals is unduly large in a short-sized packet framework if the training period is not reduced. On the other hand, if the pilot length is shortened, there would be a significant degradation in performance. Since there is a trade-off between the duration of the pilot training period and the data transmission period, the main challenge is to perform reliable channel estimation while affecting the minimal impact on the duration of data transmission period.

An aim of this paper is to propose an efficient receiver that exploits information obtained during the *data transmission period* to improve the reception quality of the system in the short packet transmission. Intuitively, when time resources are limited, we need to use *all* received data aggressively in order to optimize the receiver performance. With this goal in mind, we propose a receive filtering algorithm that estimates the interference covariance matrix using the received signal from the normal data transmission period in the IoT network. In doing so, the interference training period becomes unnecessary and the transmission rate loss caused by the interference training period can be avoided. We show from analysis and numerical experiments that the proposed method achieves a linear scaling of SINR in the number of receive antennas.

Next, we propose a frame structure for stringent latency services. In the proposed framework, pilot and data are transmitted simultaneously, which is referred to as one-shot random access.

We propose a strategy to exploit the data symbols received from the target transmit device to improve the channel estimation quality at the receiver. Some works using data signals for the channel estimation purpose have been investigated [17]–[20]. We note that conventional schemes are optimized for traditional systems such as the non-OFDM system with frequency-selective channel [17], single and multiuser SIMO system [18], wireless LAN (IEEE 802.11n) [19], and LTE systems with long packet [20], do not perfectly match to ultra-short packet scenario. In this paper, we choose the most reliable received data symbols, referred to as *virtual pilots*, among all possible soft decision data symbols and then use them to re-estimate the channel. Our proposed method is distinct from previous efforts that we select a virtual pilot group making a dominant contribution to the channel estimation quality. Towards this end, we design a mean square error (MSE) based virtual pilot selection strategy. We show from numerical simulations that the performance gain of the proposed method over the conventional receiver technique is 2 dB gain in the packet error rate for short packet based systems (e.g., packet size is set to tens of bytes).

The remainder of the paper is organized as follows. In Section II, we describe the structure of a packet and review conventional receiver techniques. In Section III, we present the proposed interference covariance matrix estimation technique suited to the short packet transmission. In Section IV, we describe the proposed virtual pilot based channel estimation technique. In Section V, we present simulation results to verify the performance of the proposed scheme. We conclude the paper in Section VI.

We briefly summarize notation used in this paper. We employ uppercase boldface letters for matrices and lowercase boldface letters for vectors. The superscripts $(\cdot)^H$ and $(\cdot)^T$ denote the conjugate transpose and transpose, respectively. $\mathbb{C}$ denotes the field of complex numbers. $\|\cdot\|_p$ indicates the p-norm. $\mathbf{I}_N$ is the $N \times N$ identity matrix. $E[\cdot]$ is the expectation operator. $\otimes$ is the Kronecker product operator. $\mathcal{CN}(m, \sigma^2)$ denotes a complex Gaussian random variable with mean m and variance σ^2.

II. System Description

A. System Model and Packet Structure

We consider uplink communication in an IoT network, with an example setup shown in Fig. 1. We assume that the target transmit device has a single antenna and the paired access point (AP)

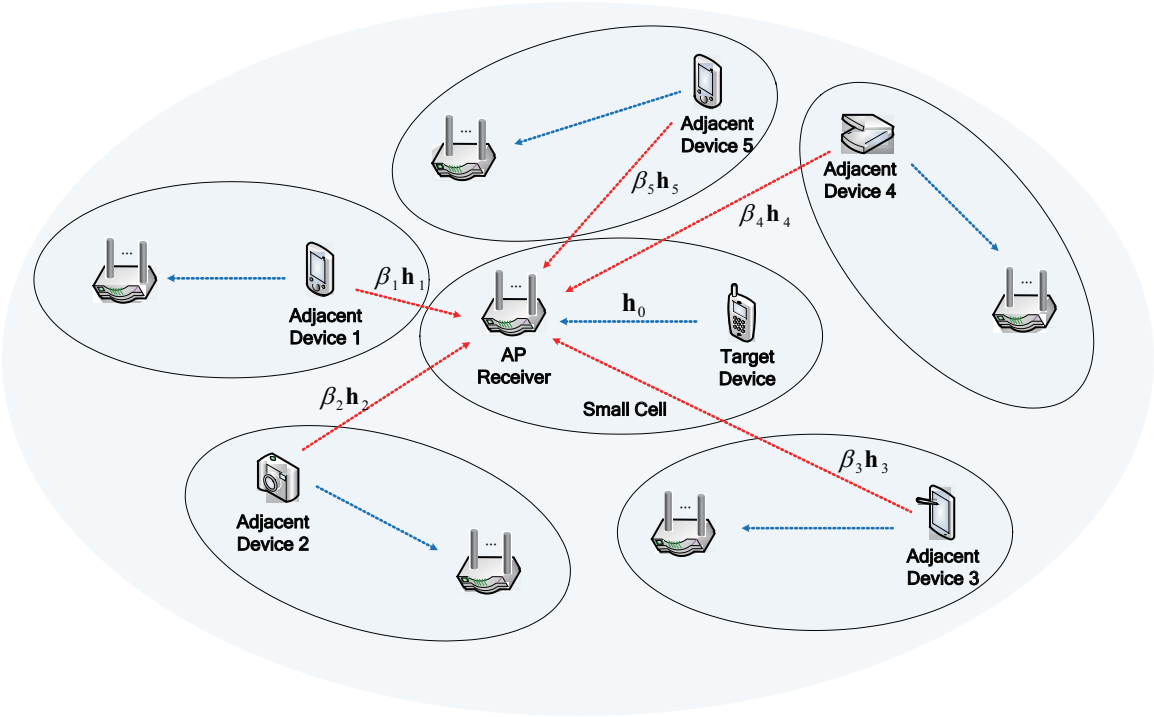


Fig. 1. A paired AP receives the desired signal from a target device and is subject to interference (dashed lines) from neighboring devices.

receiver has N antennas. Since there are a large number of devices in the radio transmission range, the received signal contains interference from adjacent devices as well as the desired information.[2] We assume a block-fading channel model, and each block consists of N_b channel uses, among which N_p uses are for the pilot training and N_d uses are for the data transmission (see Fig. 2(a)).

In this setup, the received signal for the ℓ-th channel use in the i-th fading block is given by

$$\mathbf{y}_i[\ell] = \beta_0 \mathbf{h}_{0,i} s_{0,i}[\ell] + \sum_{j=1}^{J} \beta_j \mathbf{h}_{j,i} s_{j,i}[\ell] + \mathbf{n}_i[\ell] \tag{1}$$

where $\beta_0 = d_0^{-\alpha/2}$ is the attenuation factor for the target transmit device ($\beta_0 > 0$), α is the path-loss exponent ($\alpha > 2$), $\mathbf{h}_{0,i} \in \mathbb{C}^{N\times 1}$ is the channel vector between the target transmit device and

[2]In this work, we assume that the AP receiver experiences interference from devices in neighboring cells for both channel estimation phase and data transmission phase. In short packet transmission, the packet length would be far shorter than the channel coherent time and thus the channel can be fairly static and at least slowly varying within a packet. Therefore, in the short packet regime, impact of mis-aligned coherence block would not be significant. For the sake of simplicity, we assume that the coherence times of the desired channel and interfering channels are equal.

AP receiver, $s_{0,i}[\ell]$ is the symbol transmitted by the target transmit device ($E[|s_{0,i}[\ell]|^2] = \rho$), J is the number of interfering devices from neighboring cells, $\beta_j = d_j^{-\alpha/2}$ is the attenuation factor for the j-th interferer ($\beta_j > 0$), $\mathbf{h}_{j,i} \in \mathbb{C}^{N\times 1}$ is the channel vector between the j-th interferer and AP receiver, $s_{j,i}[\ell]$ is the symbol transmitted by the j-th interferer ($E[|s_{j,i}[\ell]|^2] = \rho$), and $\mathbf{n}_i[\ell] \in \mathbb{C}^{N\times 1}$ is the complex Gaussian noise vector ($\mathbf{n}_i[\ell] \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$). Also, without loss of generality, we assume that the distances $d_0, d_1, \ldots, d_J$ are sorted in ascending order. In this work, we assume that the channel remains unchanged within a block and changes independently from block-to-block. In the sequel, we skip the fading block index i for notational convenience.

B. Conventional Receiver Technique

If a unit norm receive filter $\mathbf{v}$ is applied to the received signal vector $\mathbf{y}[\ell]$, an estimate of the desired symbol is $\hat{s}_0[\ell] = \mathbf{v}^H\mathbf{y}[\ell]$ and the resulting SINR is

$$\mathrm{SINR}(\mathbf{v}) = \frac{\rho\mathbf{v}^H\mathbf{h}_0\mathbf{h}_0^H\mathbf{v}}{\mathbf{v}^H(\sigma^2\mathbf{I} + \rho\sum_{j=1}^{J}\beta_j\mathbf{h}_j\mathbf{h}_j^H)\mathbf{v}}. \tag{2}$$

Under the condition that the target rate of the transmit device is $R = \log_2(1+\tau)$, we say a communication is successful if the received SINR is larger than τ. Hence, the outage probability at SINR threshold τ is expressed as $P_{\text{out}} = P[\mathrm{SINR}(\mathbf{v}) \le \tau]$. Note that the receive filter $\mathbf{v}$ can be designed to remove the interference or to boost the power of the desired signal. In order to minimize the outage probability, a receive filter weight should be designed to maximize the SINR of the receiver, which is commonly referred to as an *MMSE receiver* [21]–[28][3]. Denoting the covariance of the interference plus noise as $\Sigma = \sigma^2\mathbf{I} + \rho\sum_j \beta_j\mathbf{h}_j\mathbf{h}_j^H$, the receive filter of the conventional MMSE receiver is [29]

$$\mathbf{v}^* = \underset{\mathbf{v}}{\mathrm{argmax}}\,\frac{\mathbf{v}^H\mathbf{h}_0\mathbf{h}_0^H\mathbf{v}}{\mathbf{v}^H\Sigma\mathbf{v}} = \frac{\Sigma^{-1}\mathbf{h}_0}{\|\Sigma^{-1}\mathbf{h}_0\|_2}. \tag{3}$$

Plugging (3) into (2), we obtain the best achievable SINR as

$$\begin{aligned}\mathrm{SINR}^* &= \frac{\rho(\mathbf{h}_0^H\Sigma^{-1}\mathbf{h}_0)^2}{\mathbf{h}_0^H\Sigma^{-1}(\sigma^2\mathbf{I} + \rho\sum_{j=1}^{J}\beta_j\mathbf{h}_j\mathbf{h}_j^H)\Sigma^{-1}\mathbf{h}_0} \\ &= \rho\mathbf{h}_0^H\Sigma^{-1}\mathbf{h}_0. \end{aligned}\tag{4}$$

[3]The conventional performance measures such as ergodic capacity may not be suitable for short-packet communication systems because these metrics pertain to the asymptotic regime of long data packets [11], [12]. Nonetheless, this quantity is still simple and useful tool to analyze the behavior of the proposed scheme.

Note that the conventional MMSE receiver usually requires an interference training period to estimate Σ. Recall that the covariance matrix Σ is a statistic of the noise and interference. Since orthogonality among the large number of pilot sequences cannot be guaranteed, it is not easy to estimate the interfering channels $\mathbf{h}_j$ $(j = 1, \cdots, J)$ individually. To handle this issue, an approach to estimate the sample covariance $\hat{\Sigma}$ using a specially designed interference training period was proposed [16]. In order to observe the covariance associated with the interference and noise, the target transmit device should remain silent in this period. The sample covariance $\hat{\Sigma}$ obtained in the interference training period is

$$\hat{\Sigma} = \frac{1}{K} \sum_{i=1}^{K} \mathbf{r}[i]\mathbf{r}[i]^H \tag{5}$$

where K is the duration of the training period and $\mathbf{r}[i]$ is the i-th received sample ($\mathbf{r}[i] = \sum_{j=1}^{J} \beta_j \mathbf{h}_j s_j[i] + \mathbf{n}[i]$). By replacing Σ with $\hat{\Sigma}$ in (3), i.e., using $\hat{\mathbf{v}} = \frac{\hat{\Sigma}^{-1}\mathbf{h}_0}{\|\hat{\Sigma}^{-1}\mathbf{h}_0\|_2}$ instead of $\mathbf{v}^*$, we obtain the SINR estimate

$$\mathrm{SINR} = \frac{\rho(\mathbf{h}_0^H \hat{\Sigma}^{-1}\mathbf{h}_0)^2}{\mathbf{h}_0^H \hat{\Sigma}^{-H} \Sigma \hat{\Sigma}^{-1} \mathbf{h}_0}. \tag{6}$$

Using a Gaussian approximation for the interference[4], the expected SINR[5] of (6) becomes [30]

$$\begin{aligned} E_{\hat{\Sigma}}[\mathrm{SINR}] &= E_{\hat{\Sigma}}\left[\frac{\rho(\mathbf{h}_0^H \hat{\Sigma}^{-1}\mathbf{h}_0)^2}{\mathbf{h}_0^H \hat{\Sigma}^{-H} \Sigma \hat{\Sigma}^{-1} \mathbf{h}_0}\right] \\ &= \left(1 - \frac{N-1}{K+1}\right) \mathrm{SINR}^*. \end{aligned} \tag{7}$$

Note that the expected SINR in (7) contains an additional scaling factor $1 - \frac{N-1}{K+1}$. Since the interference training period K for the short packet framework would be very small, $E[\mathrm{SINR}]$ would be much smaller than SINR^*, the best achievable SINR in (4). Further, when the number of antennas N increases, there would also be a loss in the SINR and achievable rate, which makes this approach especially unsuitable for the small packet transmission.

[4]Gaussian approximation of interferences becomes more accurate when the density of machine-type devices becomes higher.

[5]Since the average analysis is relevant only if there are sufficiently many packets, and thus the expected SINR might not be a good fit for the short packet based communication systems. Nevertheless, unless the packet length is extremely short, hundreds of samples might be used in the computation of the expected SINR and thus average SINR is meaningful to some extent.

III. Receiver Design

Recall that two main ingredients of the receive filter in (3) are the channel vector $\mathbf{h}_0$ and the covariance matrix Σ of the interference and noise. In this section, we present a strategy to estimate the covariance matrix Σ using measurements in the data transmission period and show that the achievable SINR of this strategy equals SINR^*. By estimating the covariance matrix in the normal data transmission period instead of the interference training period, we can avoid the transmission rate loss caused by the interference training period.

Note that the received signal $\mathbf{y}[\ell] = \mathbf{h}_0 s_0[\ell] + \sum_{j=1}^{J} \beta_j \mathbf{h}_j s_j[\ell] + \mathbf{n}[\ell]$ contains interference from adjacent devices and noise as well as the desired signal. Using the receive filter $\mathbf{v} = \frac{\Sigma_0^{-1}\mathbf{h}_0}{\|\Sigma_0^{-1}\mathbf{h}_0\|_2}$, the estimate of the desired symbol becomes

$$\begin{aligned}\hat{s}_0[\ell] &= \mathbf{v}^H \mathbf{y}[\ell] \\ &= \frac{\rho \mathbf{h}_0^H (\rho \mathbf{h}_0 \mathbf{h}_0^H + \rho \sum_{j=1}^{J} \beta_j \mathbf{h}_j \mathbf{h}_j^H + \sigma^2 \mathbf{I})^{-1}}{\|\rho \mathbf{h}_0^H (\rho \mathbf{h}_0 \mathbf{h}_0^H + \rho \sum_{j=1}^{J} \beta_j \mathbf{h}_j \mathbf{h}_j^H + \sigma^2 \mathbf{I})^{-1}\|_2} \mathbf{y}[\ell] \\ &= \frac{\mathbf{h}_0^H \Sigma_0^{-1}}{\|\mathbf{h}_0^H \Sigma_0^{-1}\|_2} \mathbf{y}[\ell] \end{aligned} \tag{8}$$

where $\Sigma_0 = \Sigma + \rho \mathbf{h}_0 \mathbf{h}_0^H$. In the following theorem, we show that the best possible SINR in (4) can be achieved even with the inclusion of the desired signal in the covariance matrix.[6]

Theorem 1. *When the covariance matrix Σ_0 is employed in the receive filter $\mathbf{v} = \frac{\Sigma_0^{-1}\mathbf{h}_0}{\|\Sigma_0^{-1}\mathbf{h}_0\|_2}$, the SINR of the proposed strategy is $\mathrm{SINR}^* = \rho \mathbf{h}_0^H \Sigma^{-1} \mathbf{h}_0$.*

Proof. Let $\mathrm{SINR}_{\mathrm{prop}}$ be the SINR of the proposed strategy. Then, using (2), we have

$$\mathrm{SINR}_{\mathrm{prop}} = \frac{\rho \left(\mathbf{h}_0^H \left(\Sigma + \rho \mathbf{h}_0 \mathbf{h}_0^H \right)^{-1} \mathbf{h}_0 \right)^2}{\mathbf{h}_0^H \left(\Sigma + \rho \mathbf{h}_0 \mathbf{h}_0^H \right)^{-1} \Sigma \left(\Sigma + \rho \mathbf{h}_0 \mathbf{h}_0^H \right)^{-1} \mathbf{h}_0}. \tag{9}$$

We first consider the numerator. Using the matrix inversion lemma[7], we have

$$\left(\Sigma + \rho \mathbf{h}_0 \mathbf{h}_0^H \right)^{-1} = \Sigma^{-1} - \frac{\Sigma^{-1} \mathbf{h}_0 \mathbf{h}_0^H \Sigma^{-1}}{\rho + \mathbf{h}_0^H \Sigma^{-1} \mathbf{h}_0}. \tag{10}$$

From (9) and (10), we have

[6]In this section, we focus on the effect of covariance matrix Σ_0 and the sample covariance matrix $\hat{\Sigma}_0$ on SINR under the assumption that $\mathbf{h}_0$ is perfectly known at the receiver. We discuss the effect of estimated channel $\hat{\mathbf{h}}_0$ in the next section.

[7]Note that (10) can be obtained by letting $A = \Sigma$, $B = \mathbf{h}_0$, $C = \mathbf{h}_0^H$ and $D = \rho \mathbf{I}$ in the matrix inversion lemma $\left(A - BD^{-1}C\right)^{-1} = A^{-1} + A^{-1}B\left(D - CA^{-1}B\right)^{-1}CA^{-1}$.

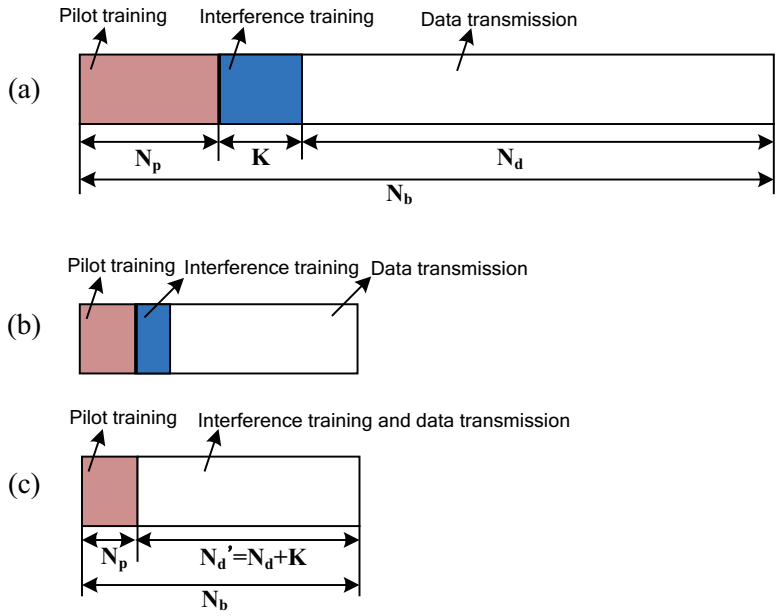


Fig. 2. Packet structure of (a) the conventional MMSE receiver for long data packets, (b) the conventional MMSE receiver for short data packets, and (c) the proposed linear receiver for short data packets.

$$\begin{aligned}\left(\mathbf{h}_0^H\left(\Sigma+\rho\mathbf{h}_0\mathbf{h}_0^H\right)^{-1}\mathbf{h}_0\right)^2 &= \left(\mathbf{h}_0^H\Sigma^{-1}\mathbf{h}_0-\frac{(\mathbf{h}_0^H\Sigma^{-1}\mathbf{h}_0)^2}{\rho+\mathbf{h}_0^H\Sigma^{-1}\mathbf{h}_0}\right)^2\\ &= \left(\frac{\rho\mathbf{h}_0^H\Sigma^{-1}\mathbf{h}_0}{\rho+\mathbf{h}_0^H\Sigma^{-1}\mathbf{h}_0}\right)^2\\ &= \left(\frac{\rho\mathrm{SINR}^*}{\rho^2+\mathrm{SINR}^*}\right)^2 \end{aligned} \tag{11}$$

where $\mathrm{SINR}^*=\rho\mathbf{h}_0^H\Sigma^{-1}\mathbf{h}_0$ (see (4)). Now, by plugging (10) into the denominator of (9), we have

$$\begin{aligned}&\mathbf{h}_0^H\left(\Sigma+\rho\mathbf{h}_0\mathbf{h}_0^H\right)^{-1}\Sigma\left(\Sigma+\rho\mathbf{h}_0\mathbf{h}_0^H\right)^{-1}\mathbf{h}_0\\ &=\mathbf{h}_0^H\left(\Sigma^{-1}-\frac{\Sigma^{-1}\mathbf{h}_0\mathbf{h}_0^H\Sigma^{-1}}{\rho+\mathbf{h}_0^H\Sigma^{-1}\mathbf{h}_0}\right)\Sigma\left(\Sigma^{-1}-\frac{\Sigma^{-1}\mathbf{h}_0\mathbf{h}_0^H\Sigma^{-1}}{\rho+\mathbf{h}_0^H\Sigma^{-1}\mathbf{h}_0}\right)\mathbf{h}_0\\ &=\left(\mathbf{h}_0^H-\frac{\mathbf{h}_0^H\Sigma^{-1}\mathbf{h}_0\mathbf{h}_0^H}{\rho+\mathbf{h}_0^H\Sigma^{-1}\mathbf{h}_0}\right)\left(\Sigma^{-1}\mathbf{h}_0-\frac{\Sigma^{-1}\mathbf{h}_0\mathbf{h}_0^H\Sigma^{-1}\mathbf{h}_0}{\rho+\mathbf{h}_0^H\Sigma^{-1}\mathbf{h}_0}\right)\\ &=\frac{\rho^2\mathbf{h}_0^H\Sigma^{-1}\mathbf{h}_0}{\left(\rho+\mathbf{h}_0^H\Sigma^{-1}\mathbf{h}_0\right)^2}\\ &=\frac{\rho^3\mathrm{SINR}^*}{(\rho^2+\mathrm{SINR}^*)^2}. \end{aligned} \tag{12}$$

Plugging (11) and (12) into (9), we have

$$\mathrm{SINR}_{\mathrm{prop}}=\frac{\rho\left(\frac{\rho\mathrm{SINR}^*}{\rho^2+\mathrm{SINR}^*}\right)^2}{\frac{\rho^3\mathrm{SINR}^*}{(\rho^2+\mathrm{SINR}^*)^2}}=\mathrm{SINR}^*, \tag{13}$$

which is the desired result. □

It is worth noting that the main goal of Theorem 1 is to validate the method of estimating the covariance matrix in using only normal data transmission. By achieving the maximum SINR and avoiding the interference training period, the proposed method can exploit an additional K channel uses for the data transmission period compared to the conventional MMSE receiver.

By slightly modifying this result and using a Gaussian approximation for the interference, we can obtain the SINR expression for a realistic scenario where the sample covariance matrix is employed. In this scenario, we use $\hat{\Sigma}_0 = \hat{\Sigma} + \rho\hat{\mathbf{h}}_0\hat{\mathbf{h}}_0^H$ instead of Σ_0, and thus the modified receive filter is[8]

$$\hat{\mathbf{v}} = \frac{\hat{\Sigma}_0^{-1}\mathbf{h}_0}{\|\hat{\Sigma}_0^{-1}\mathbf{h}_0\|_2}. \tag{14}$$

Theorem 2. *When the sample covariance in (14) is employed, the expected SINR of the proposed linear MMSE receiver is*

$$\begin{aligned} E_{\hat{\Sigma}_0}\left[\mathrm{SINR}_{\mathrm{prop}}\right] &= E_{\hat{\Sigma}_0}\left[\frac{\rho(\mathbf{h}_0^H\hat{\Sigma}_0^{-1}\mathbf{h}_0)^2}{\mathbf{h}_0^H\hat{\Sigma}_0^{-1}\Sigma\hat{\Sigma}_0^{-1}\mathbf{h}_0}\right] \\ &= \left(1-\frac{N-1}{N_d+1}\right)\mathrm{SINR}^*. \end{aligned} \tag{15}$$

Proof. By the direct substitution of $\hat{\Sigma}_0 = \hat{\Sigma} + \rho\hat{\mathbf{h}}_0\hat{\mathbf{h}}_0^H$ into (6), we have

$$\mathrm{SINR} = \frac{\left(\mathbf{h}_0^H\hat{\Sigma}_0^{-1}\mathbf{h}_0\right)^2}{\left(\mathbf{h}_0^H\hat{\Sigma}_0^{-1}\Sigma\hat{\Sigma}_0^{-1}\mathbf{h}_0\right)}. \tag{16}$$

Let $\widetilde{\Sigma} = \hat{\Sigma} + \rho(\hat{\mathbf{h}}_0\hat{\mathbf{h}}_0^H - \mathbf{h}_0\mathbf{h}_0^H)$, then $\hat{\Sigma}_0 = \widetilde{\Sigma} + \rho\mathbf{h}_0\mathbf{h}_0^H$ and $\hat{\Sigma}_0^{-1} = \widetilde{\Sigma}^{-1} - \frac{\widetilde{\Sigma}^{-1}\mathbf{h}_0\mathbf{h}_0^H\widetilde{\Sigma}^{-1}}{\rho+\mathbf{h}_0^H\widetilde{\Sigma}^{-1}\mathbf{h}_0}$ by the matrix inversion lemma. Therefore,

$$\begin{aligned} \mathrm{SINR} &= \frac{\rho\left(\mathbf{h}_0^H\left(\widetilde{\Sigma}^{-1} - \frac{\widetilde{\Sigma}^{-1}\mathbf{h}_0\mathbf{h}_0^H\widetilde{\Sigma}^{-1}}{\rho+\mathbf{h}_0^H\widetilde{\Sigma}^{-1}\mathbf{h}_0}\right)\mathbf{h}_0\right)^2}{\left(\mathbf{h}_0^H\left(\widetilde{\Sigma}^{-1} - \frac{\widetilde{\Sigma}^{-1}\mathbf{h}_0\mathbf{h}_0^H\widetilde{\Sigma}^{-1}}{\rho+\mathbf{h}_0^H\widetilde{\Sigma}^{-1}\mathbf{h}_0}\right)\Sigma\left(\widetilde{\Sigma}^{-1} - \frac{\widetilde{\Sigma}^{-1}\mathbf{h}_0\mathbf{h}_0^H\widetilde{\Sigma}^{-1}}{\rho+\mathbf{h}_0^H\widetilde{\Sigma}^{-1}\mathbf{h}_0}\right)\mathbf{h}_0\right)} \\ &= \frac{\rho\left(\frac{\rho\mathbf{h}_0^H\widetilde{\Sigma}^{-1}\mathbf{h}_0}{\rho+\mathbf{h}_0^H\widetilde{\Sigma}^{-1}\mathbf{h}_0}\right)^2}{\left(\frac{\rho^2\mathbf{h}_0^H\widetilde{\Sigma}^{-1}\Sigma\widetilde{\Sigma}^{-1}\mathbf{h}_0}{(\rho+\mathbf{h}_0^H\widetilde{\Sigma}^{-1}\mathbf{h}_0)^2}\right)} \\ &= \frac{\rho\left(\mathbf{h}_0^H\widetilde{\Sigma}^{-1}\mathbf{h}_0\right)^2}{\mathbf{h}_0^H\widetilde{\Sigma}^{-1}\Sigma\widetilde{\Sigma}^{-1}\mathbf{h}_0}. \end{aligned} \tag{17}$$

[8]Here we use $\hat{\mathbf{h}}_0$ instead of $\mathbf{h}_0$in $\hat{\Sigma}_0$ in order to observe the effect of $\hat{\Sigma}_0$ on SINR.

Further, by denoting $\kappa = \frac{\text{SINR}}{\text{SINR}^*}$, $\mathbf{\Xi} = (\frac{\text{SINR}^*}{\rho})^{-\frac{1}{2}}\Sigma^{-\frac{1}{2}}\mathbf{h}_0$, and $\hat{\Sigma}_1 = \Sigma^{-\frac{1}{2}}\widetilde{\Sigma}\Sigma^{-\frac{1}{2}}$, we have

$$\begin{aligned}\kappa &= \frac{1}{\text{SINR}^*}\frac{\rho\left(\mathbf{h}_0^H\tilde{\Sigma}^{-1}\mathbf{h}_0\right)^2}{\mathbf{h}_0^H\tilde{\Sigma}^{-1}\Sigma\tilde{\Sigma}^{-1}\mathbf{h}_0} \\ &= \frac{\left(\mathbf{\Xi}^H\hat{\Sigma}_1^{-1}\mathbf{\Xi}\right)^2}{\mathbf{\Xi}^H\hat{\Sigma}_1^{-2}\mathbf{\Xi}}. \end{aligned} \tag{18}$$

Note that the joint distribution of the elements of $\hat{\Sigma}$ follows the central complex Wishart distribution $CW(M, N; \Sigma)$ [31]. Hence, $E[\hat{\Sigma}_1] = \Sigma^{-\frac{1}{2}}E[\widetilde{\Sigma}]\Sigma^{-\frac{1}{2}} = \Sigma^{-\frac{1}{2}}\Sigma\Sigma^{-\frac{1}{2}} = \mathbf{I}$, and $\hat{\Sigma}_1$ is a complex Wishart distribution $CW(N_d, N; \mathbf{I})$. The probability density function of κ is given by [30]

$$P(\kappa) = \frac{\Gamma(N_d+1)}{\Gamma(N-1)\Gamma(N_d+2-N)}\kappa^{(N_d+2-N)-1}(1-\kappa)^{(N-1)-1} \tag{19}$$

where κ follows the regularized incomplete beta distribution. Since $\Gamma(i) = (i-1)!$ for an integer i, the expectation of κ is

$$\begin{aligned}E[\kappa] &= \int_0^1 \kappa P(\kappa)d\kappa \\ &= \frac{N_d!}{(N_d-2)!(N_d+1-N)!}\int_0^1 \kappa^{N_d+2-N}(1-\kappa)^{N-2}d\kappa \\ &= 1 - \frac{N-1}{N_d+1}. \end{aligned} \tag{20}$$

Recalling that $\kappa = \frac{1}{\text{SINR}^*}\frac{\rho\left(\mathbf{h}_0^H\hat{\Sigma}_0^{-1}\mathbf{h}_0\right)^2}{\mathbf{h}_0^H\hat{\Sigma}_0^{-1}\Sigma\hat{\Sigma}_0^{-1}\mathbf{h}_0}$, we further have

$$\begin{aligned}E_{\hat{\Sigma}_0}\left[\text{SINR}_{\text{prop}}\right] &= E_{\hat{\Sigma}_0}\left[\frac{\rho(\mathbf{h}_0^H\hat{\Sigma}_0^{-1}\mathbf{h}_0)^2}{\mathbf{h}_0^H\hat{\Sigma}_0^{-1}\Sigma\hat{\Sigma}_0^{-1}\mathbf{h}_0}\right] \\ &= \left(1 - \frac{N-1}{N_d+1}\right)\text{SINR}^*. \end{aligned}$$

□

Since the interference training period is unnecessary, the data transmission period is improved from N_d to $N_d' = N_d + K$ and at the same time the covariance matrix is estimated using samples in the data transmission period (see Fig. 2). Therefore, we obtain a more accurate estimation of the interference covariance in the short packet framework. The effect of the sample covariance matrix on the received SINR shows the efficacy of the proposed receiver by obtaining a larger scaling factor $\left(1 - \frac{N-1}{N_d'+1}\right)$. In Fig. 4, we plot the received SINR as a function of the number of receive antennas N. When the packet size N_b is large (e.g., $N_b = 1000$), which corresponds to the

packet length of current wireless systems, $K = 0.1N_b$ is also sufficiently larger than the number of antennas N so that scaling factor of the proposed scheme $\left(1 - \frac{N-1}{N_d'+1}\right)$ and the conventional MMSE receiver $\left(1 - \frac{N-1}{K+1}\right)$ are not much different. This behavior, however, does not hold true for short-sized packet regime. In fact, as shown in Fig. 4(a), we see that the performance of the proposed scheme improves linearly with N. Since the interference training is needed for the conventional MMSE receiver, we observe that when N_b is small (e.g., $N_b = 100$) the performance of the conventional MMSE SINR receiver with sample covariance matrix does not scale with N.

IV. Joint Pilot and Data Symbol based Channel Estimation

In the previous sections, we assumed that the desired channel $\mathbf{h}_0$ is perfectly known. As mentioned, when we consider the short packet transmission, the training period should also be reduced to avoid an excessive amount of training overhead. In this section, we propose a channel estimation technique that jointly uses the pilot signals and data symbols to improve channel estimation quality. In a nutshell, the proposed channel estimation technique picks a small number of reliable data symbols among all available data symbols. Using the chosen data symbols (which we call *virtual pilots*) together with the pilot signals in the re-estimation of the channel vector, we achieve great improvement of the channel estimation performance in the short packet regime.

A. The Joint Pilot and Data Symbol based Channel Estimation

Before we proceed, we briefly review conventional MMSE-based channel estimation. The received pilot signals in the training period are expressed as

$$\begin{aligned} \mathbf{y}^{(1)} &= \mathbf{h}_0 p_0^{(1)} + \sum_{j=1}^{J} \beta_j \mathbf{h}_j s_j^{(1)} + \mathbf{n}^{(1)} \\ &\vdots \\ \mathbf{y}^{(N_p)} &= \mathbf{h}_0 p_0^{(N_p)} + \sum_{j=1}^{J} \beta_j \mathbf{h}_j s_j^{(N_p)} + \mathbf{n}^{(N_p)} \end{aligned} \tag{21}$$

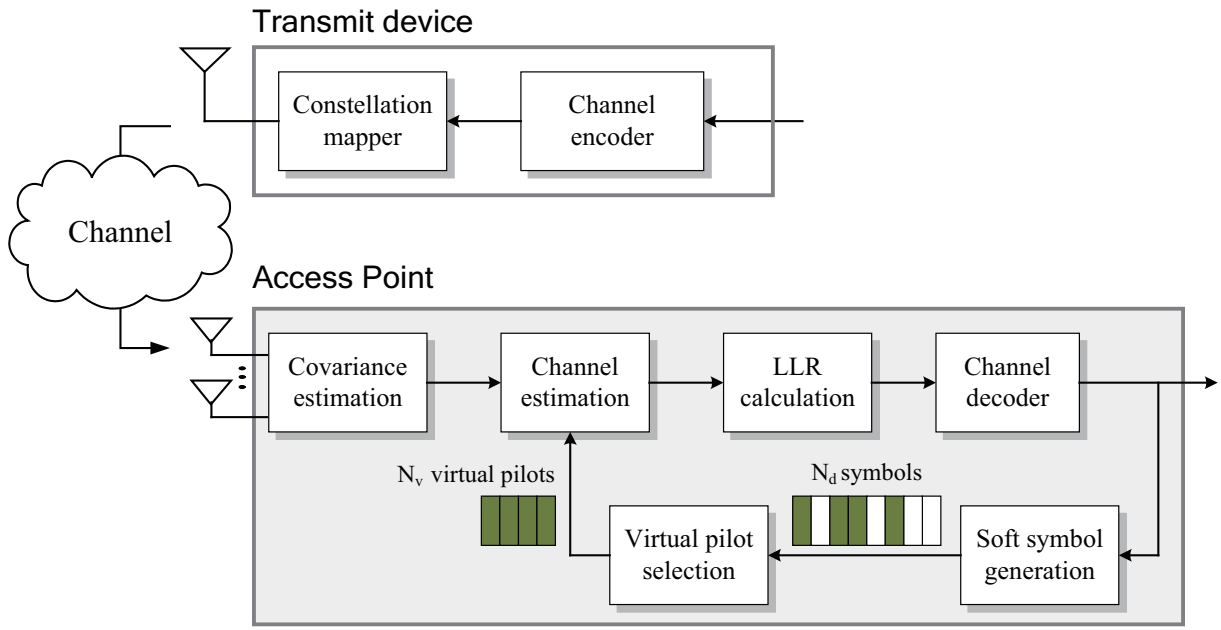


Fig. 3. Block diagram of the receiver using virtual pilot and pilot signal in the re-estimation of the channel. Among all the data symbols, symbols with small MSE are chosen as virtual pilots. Note that symbols with small MSE are green colored.

where $\mathbf{y}^{(i)}$ is the i-th observation, $\mathbf{n}^{(i)}$ is the i-th noise, $p_0^{(i)}$ is the i-th pilot signal in the target transmit device, and N_p is the total number of pilot observations. The $N_pN \times 1$ vectorized received pilot signals are

$$\mathbf{y}_p = \mathbf{P}_0\mathbf{h}_0 + \mathbf{z}_p \tag{22}$$

where $\mathbf{P}_0 = \mathbf{p}_0 \otimes \mathbf{I}_N$ is $N_pN \times N$-dimensional training matrix with $\mathbf{p}_0 = [p_0^{(1)}, \ldots, p_0^{(N_p)}]^T$, and $\mathbf{z}_p = \left[(\sum_{j=1}^{J} \beta_j \mathbf{h}_j s_j^{(1)} + \mathbf{n}^{(1)})^T \cdots (\sum_{j=1}^{J} \beta_j \mathbf{h}_j s_j^{(N_p)} + \mathbf{n}^{(N_p)})^T\right]^T$ is the interference plus noise vector over the training period. We assume that $\mathbf{h}_0$ follows a circular symmetric complex normal distribution, i.e., $\mathbf{h}_0 \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\mathrm{hh}})$ where $\mathbf{C}_{\mathrm{hh}} = \mathrm{Cov}(\mathbf{h}_0, \mathbf{h}_0)$. The MMSE weight matrix minimizing the mean square error between the original channel vector $\mathbf{h}_0$ and the estimate $\hat{\mathbf{h}}_0 = \mathbf{W}\mathbf{y}_p$ is [32]

$$\begin{aligned}
\mathbf{W} &= \mathrm{Cov}(\mathbf{h}_0, \mathbf{y}_p)\,\mathrm{Cov}(\mathbf{y}_p, \mathbf{y}_p)^{-1} \\
&= E[\mathbf{h}_0\mathbf{y}_p^H]E[\mathbf{y}_p\mathbf{y}_p^H] \qquad (23) \\
&= E[\mathbf{h}_0\mathbf{h}_0^H\mathbf{P}_0^H + \mathbf{h}_0\mathbf{z}_p^H] \\
&\quad \times E[\mathbf{P}_0\mathbf{h}_0\mathbf{h}_0^H\mathbf{P}_0^H + \mathbf{P}_0\mathbf{h}_0\mathbf{z}_p^H + \mathbf{z}_p\mathbf{h}_0^H\mathbf{P}_0^H + \mathbf{z}_p\mathbf{z}_p^H] \\
&= \mathbf{C}_{\mathrm{hh}}\mathbf{P}_0^H\left(\mathbf{P}_0\mathbf{C}_{\mathrm{hh}}\mathbf{P}_0^H + \eta_{\mathrm{p}}\mathbf{I}_N\right)^{-1} \qquad (24)
\end{aligned}$$

where (23) is due to $\mathrm{Cov}(\mathbf{a},\mathbf{b}) = E[\mathbf{a}\mathbf{b}^H] - E[\mathbf{a}]E[\mathbf{b}^H]$ and η_{p} is the variance of the interference plus noise.[9] The corresponding MMSE-based channel estimate $\hat{\mathbf{h}}_0$ is

$$\begin{aligned}\hat{\mathbf{h}}_0 &= \mathbf{C}_{\mathrm{hh}}\mathbf{P}_0^H\left(\mathbf{P}_0\mathbf{C}_{\mathrm{hh}}\mathbf{P}_0^H + \eta_{\mathrm{p}}\mathbf{I}_N\right)^{-1}(\mathbf{P}_0\mathbf{h}_0 + \mathbf{z}_p)\\ &= (N_p\mathbf{C}_{\mathrm{hh}} + \eta_{\mathrm{p}}\mathbf{I}_N)^{-1}\mathbf{C}_{\mathrm{hh}}\mathbf{P}_0^H(\mathbf{P}_0\mathbf{h}_0 + \mathbf{z}_p)\\ &= \left(\frac{\eta_{\mathrm{p}}}{N_p}\mathbf{I}_N + \mathbf{C}_{\mathrm{hh}}\right)^{-1}\mathbf{C}_{\mathrm{hh}}\left(\mathbf{h}_0 + \frac{1}{N_p}\mathbf{P}_0^H\mathbf{z}_p\right). \end{aligned} \tag{25}$$

We observe from (25) that the estimated channel vector $\hat{\mathbf{h}}_0$ converges to the original channel vector $\mathbf{h}_0$ as the training period N_p increases. This clearly demonstrates that there would be a degradation in the channel estimation quality for the short packet regime. To enhance the channel estimation quality without increasing the pilot overhead, we exploit the virtual pilots in the re-estimation of channels. Using the deliberately chosen virtual pilots together with the conventional pilots, the channel is re-estimated, and the newly generated channel estimate will be used for the symbol detection. When N_v data symbols are selected for the virtual pilot purpose, the received signals are

$$\begin{aligned}\mathbf{y}^{(1)} &= \mathbf{h}_0 s_0^{(1)} + \sum_{j=1}^{J}\beta_j\mathbf{h}_j s_j^{(1)} + \mathbf{n}^{(1)}\\ &\vdots\\ \mathbf{y}^{(N_v)} &= \mathbf{h}_0 s_0^{(N_v)} + \sum_{j=1}^{J}\beta_j\mathbf{h}_j s_j^{(N_v)} + \mathbf{n}^{(N_v)}\end{aligned} \tag{26}$$

where $\mathbf{y}^{(i)}$ is the i-th observation, $\mathbf{n}^{(i)}$ is the i-th noise, $s_0^{(i)}$ is the data symbol of the target transmit device, and $s_j^{(i)}$ is the data symbol of the j-th interferer. The $N_vN\times 1$ vectorized virtual pilot observations can be expressed as

$$\mathbf{y}_v = \mathbf{S}_0\mathbf{h}_0 + \mathbf{z}_v \tag{27}$$

where $\mathbf{S}_0 = \mathbf{s}_0 \otimes \mathbf{I}_N$ (of size $N_vN\times N$) is the virtual pilot matrix where $\mathbf{s}_0 = [s_0^{(1)},\ldots,s_0^{(N_v)}]^T$ is the data symbol sequence and $\mathbf{z}_v = \left[(\sum_{j=1}^{J}\beta_j\mathbf{h}_j s_j^{(1)} + \mathbf{n}^{(1)})^T \cdots (\sum_{j=1}^{J}\beta_j\mathbf{h}_j s_j^{(N_v)} + \mathbf{n}^{(N_v)})^T\right]^T$

[9]Note that the receiver needs to incorporate η_{p} in the MMSE-based channel estimation phase. Since it is hard to obtain exact variance $\sum_{j=1}^{J}\beta_j^2 N\rho + \sigma^2$, we instead use $\eta_{\mathrm{p}} = J\bar{\beta}^2 N\rho + \sigma^2$ where $\bar{\beta} = \bar{d}^{\alpha/2}$ is obtained by using the cell-radius $\bar{d}$ under the assumption that the interferers are closer to the receiver (i.e., $\bar{\beta} > \beta_j$). Thus, our result in channel estimation might be slightly pessimistic. Also, the reason to treat interference as noise in (24) is because it is difficult to acquire the covariance of the interference. Note that the MSE formula in (37) is obtained under the assumption that $\mathbf{h}_0$ as well as the interfering channels are spatially uncorrelated.

is the interference plus noise signal vector in the virtual pilot transmission period. By stacking the pilot observation vector $\mathbf{y}_p$ and virtual pilot observation vector $\mathbf{y}_v$, we obtain the composite observation vector $\mathbf{y}_c$ for the channel re-estimation as

$$\mathbf{y}_c = \begin{bmatrix} \mathbf{y}_p \\ \mathbf{y}_v \end{bmatrix} = \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{S}_0 \end{bmatrix} \mathbf{h}_0 + \begin{bmatrix} \mathbf{z}_p \\ \mathbf{z}_v \end{bmatrix}. \tag{28}$$

The MMSE estimate of the proposed scheme is expressed as

$$\begin{aligned} \hat{\mathbf{h}}_0 &= \mathrm{Cov}\left(\mathbf{h}_0, \mathbf{y}_c\right) \mathrm{Cov}\left(\mathbf{y}_c, \mathbf{y}_c\right)^{-1} \mathbf{y}_c \\ &= \mathbf{\Omega}\mathbf{\Sigma}^{-1}\mathbf{y}_c \end{aligned} \tag{29}$$

where

$$\begin{aligned} \mathbf{\Omega} &= \mathrm{Cov}\left(\mathbf{h}_0, \mathbf{y}_c\right) \\ &= \begin{bmatrix} E[\mathbf{h}_0\mathbf{y}_p^H] & E[\mathbf{h}_0\mathbf{y}_v^H] \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{\mathrm{hh}}\mathbf{P}_0^H & \mathbf{C}_{\mathrm{hh}}\bar{\mathbf{S}}_0^H \end{bmatrix} \end{aligned} \tag{30}$$

and

$$\begin{aligned} \mathbf{\Sigma} &= \mathrm{Cov}\left(\mathbf{y}_c, \mathbf{y}_c\right) \\ &= \begin{bmatrix} E[\mathbf{y}_p\mathbf{y}_p^H] & E[\mathbf{y}_p\mathbf{y}_v^H] \\ E[\mathbf{y}_v\mathbf{y}_p^H] & E[\mathbf{y}_v\mathbf{y}_v^H] \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{P}_0\mathbf{C}_{\mathrm{hh}}\mathbf{P}_0^H + \eta_{\mathrm{p}}\mathbf{I} & \mathbf{P}_0\mathbf{C}_{\mathrm{hh}}\bar{\mathbf{S}}_0^H \\ \bar{\mathbf{S}}_0\mathbf{C}_{\mathrm{hh}}\mathbf{P}_0^H & \mathbf{\Lambda}_0 \otimes \mathrm{diag}(\mathbf{C}_{\mathrm{hh}}) + \eta_{\mathrm{v}}\mathbf{I} \end{bmatrix}. \end{aligned} \tag{31}$$

Note that $\bar{\mathbf{S}}_0 = E[\mathbf{S}_0] = \bar{\mathbf{s}}_0 \otimes \mathbf{I}_N$ where $\bar{\mathbf{s}}_0 = [\bar{s}_0^{(1)}, \ldots, \bar{s}_0^{(N_v)}]^T$ is obtained from the first order moment of $s_0^{(i)}$ [33]. That is,

$$\bar{s}_0^{(i)} = E[s_0^{(i)}] = \sum_{\theta \in \Theta} \theta \prod_{k=1}^{Q} \frac{1}{2}\left(1 + c_{0,k}^{(i)} \tanh\left(\frac{1}{2}L(c_{0,k}^{(i)})\right)\right) \tag{32}$$

where Θ is a constellation set, $c_{0,k}^{(i)}$ is the k-th coded bit, Q is the number of (coded) bits mapped to a data symbol $s_0^{(i)}$ in 2^Q-ary quadrature amplitude modulation (QAM) constellations, and $L(c_{0,k}^{(i)})$ is the log-likelihood ratio (LLR) of the k-th coded bit mapped from a data symbol $s_0^{(i)}$. $\mathbf{\Lambda}_0 = [\bar{\lambda}_0^{(1)}, \ldots, \bar{\lambda}_0^{(N_v)}]^T$ is the vector of the second order moments of $s_0^{(i)}$ given by

$$\begin{aligned} \bar{\lambda}_0^{(i)} &= E[|s_0^{(i)}|^2] \\ &= \sum_{\theta \in \Theta} |\theta|^2 \prod_{k=1}^{Q} \frac{1}{2}\left(1 + c_{0,k}^{(i)} \tanh\left(\frac{1}{2}L(c_{0,k}^{(i)})\right)\right). \end{aligned} \tag{33}$$

Plugging (29) and (30) into (28), we obtain the channel estimate $\hat{\mathbf{h}}_0$ as

$$\hat{\mathbf{h}}_0 = \begin{bmatrix} \mathbf{C}_{\rm hh}\mathbf{P}_0^H & \mathbf{C}_{\rm hh}\bar{\mathbf{S}}_0^H \end{bmatrix} \times \begin{bmatrix} \mathbf{P}_0\mathbf{C}_{\rm hh}\mathbf{P}_0^H + \eta_{\rm p}\mathbf{I} & \mathbf{P}_0\mathbf{C}_{\rm hh}\bar{\mathbf{S}}_0^H \\ \bar{\mathbf{S}}_0\mathbf{C}_{\rm hh}\mathbf{P}_0^H & \mathbf{\Lambda}_0 \otimes \mathrm{diag}(\mathbf{C}_{\rm hh}) + \eta_{\rm v}\mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y}_p \\ \mathbf{y}_v \end{bmatrix}.$$

While the conventional MMSE estimate in (25) uses only the received pilot sequence for the channel estimation, the proposed channel estimator in (29) utilizes the most reliable data symbols, i.e., the soft symbol estimate $\bar{\mathbf{S}}_0$ and second order moments $\mathbf{\Lambda}_0$, as virtual pilots. Clearly, reliability of the soft symbols directly affects the quality of the proposed channel estimation so that we need to choose virtual pilots with caution.

B. Virtual Pilot Selection

Since the virtual pilots can improve the channel estimation quality of the proposed method, the best way to select virtual pilots would be to compare the performance metric (e.g., MSE of the estimated channel) of all possible $\binom{N_d}{N_v}$ data symbol combinations and then choose the combination achieving the minimum MSE. Since this procedure is computationally demanding and hence not pragmatic, we use a simple suboptimal approach to compute the MSE when single data symbol is used for the virtual pilot. Among all data symbols, we choose the N_v-best data symbols generating the smallest MSE as virtual pilots (see Fig. 3). Although this approach does not consider the correlation among virtual pilot symbols and hence is not optimal, computational complexity is much smaller than the approach using all possible symbol combinations. Also, this approach is effective in improving the quality of channel re-estimation.

Let $\hat{\mathbf{h}}_0^{(n)}$ be the estimated channel vector when the n-th data symbol is used as a virtual pilot, then the MSE metric ε_n for the hypothetical selection of the n-th data symbol is expressed as

$$\begin{aligned} \varepsilon_n = E\left[\|\mathbf{h}_0 - \hat{\mathbf{h}}_0^{(n)}\|_2^2\right] &= \mathrm{tr}\left(\mathrm{Cov}\left(\mathbf{h}_0 - \tilde{\mathbf{\Omega}}\tilde{\mathbf{\Sigma}}^{-1}\begin{bmatrix}\mathbf{z}_p \\ \mathbf{y}_v^{(n)}\end{bmatrix}\right)\right) \\ &= \mathrm{tr}\left(\mathrm{Cov}\left(\mathbf{h}_0\right) - \mathrm{Cov}\left(\tilde{\mathbf{\Omega}}\tilde{\mathbf{\Sigma}}^{-1}\begin{bmatrix}\mathbf{y}_p \\ \mathbf{y}_v^{(n)}\end{bmatrix}\right)\right) \\ &= \mathrm{tr}\left(\mathrm{Cov}\left(\mathbf{h}_0\right) - \tilde{\mathbf{\Omega}}\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{\Omega}}^H\right) \end{aligned} \tag{34}$$

where

$$\begin{aligned}\tilde{\mathbf{\Omega}} &= \mathrm{Cov}\left(\mathbf{h}_0, \begin{bmatrix}\mathbf{y}_p \\ \mathbf{y}_v^{(n)}\end{bmatrix}\right) \\ &= \begin{bmatrix}E[\mathbf{h}_0\mathbf{y}_p^H] & E[\mathbf{h}_0\mathbf{y}_v^{(n)H}]\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{C}_{\mathrm{hh}}\mathbf{P}_0^H & (\bar{s}_0^{(n)})^*\mathbf{C}_{\mathrm{hh}}\end{bmatrix} \end{aligned} \tag{35}$$

and

$$\begin{aligned}\tilde{\mathbf{\Sigma}} &= \mathrm{Cov}\left(\begin{bmatrix}\mathbf{y}_p \\ \mathbf{y}_v^{(n)}\end{bmatrix}\right) \\ &= \begin{bmatrix}E[\mathbf{y}_p\mathbf{y}_p^H] & E[\mathbf{y}_p\mathbf{y}_v^{(n)H}] \\ E[\mathbf{y}_v^{(n)}\mathbf{y}_p^H] & E[\mathbf{y}_v^{(n)}\mathbf{y}_v^{(n)H}]\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{P}_0\mathbf{C}_{\mathrm{hh}}\mathbf{P}_0^H + \eta_{\mathrm{p}}\mathbf{I} & (\bar{s}_0^{(n)})^*\mathbf{P}_0\mathbf{C}_{\mathrm{hh}} \\ (\bar{s}_0^{(n)})^*\mathbf{C}_{\mathrm{hh}}\mathbf{P}_0^H & \bar{\lambda}_0^{(n)}\mathbf{C}_{\mathrm{hh}} + \eta_{\mathrm{v}}\mathbf{I}\end{bmatrix} \end{aligned} \tag{36}$$

using $\mathbf{y}_v^{(n)} = s_0^{(n)}\mathbf{h}_0 + \mathbf{z}_v^{(n)}$ is the virtual pilot observation vector for the n-th data symbol. From (34), (35), and (36), ε_n is expressed as (see Appendix A)

$$\begin{aligned}\varepsilon_n = N(1 &- \gamma_p^2(1+\gamma_p)\frac{\bar{\lambda}_0^{(n)}}{\bar{\lambda}_0^{(n)} + \eta_{\mathrm{v}}}(\bar{\lambda}_0^{(n)})^2 \\ &- \gamma_p^2\left(\frac{2\eta_{\mathrm{v}}\bar{\lambda}_0^{(n)} + \eta_{\mathrm{v}}^2}{\bar{\lambda}_0^{(n)} + \eta_{\mathrm{v}}}\right)\bar{\lambda}_0^{(n)} + \gamma_p\frac{(\bar{\lambda}_0^{(n)})^3}{(\bar{\lambda}_0^{(n)} + \eta_{\mathrm{v}})^3} \\ &+ \frac{(\bar{\lambda}_0^{(n)})^2}{(\bar{\lambda}_0^{(n)} + \eta_{\mathrm{v}})^2}\frac{N_p^2}{\gamma_p} + \frac{\bar{\lambda}_0^{(n)}}{\bar{\lambda}_0^{(n)} + \eta_{\mathrm{v}}}(2\gamma_p + N_p^2 - 1) - \gamma_p)\end{aligned} \tag{37}$$

where $\gamma_p = \frac{N_p}{N_p + \eta_{\mathrm{p}}}$. Note that ε_n depends on the reliability of soft decisions (i.e., the second order statistic of data symbol $\bar{\lambda}_0^{(n)}$ in (33)). Hence, in order to achieve the minimum MSE, the desired data symbol maximizing $\bar{\lambda}_0^{(n)}$ should be chosen as the virtual pilot. Once ε_n for all n are computed, we choose N_v data symbols minimizing ε_n (see Fig. 7). Observations of virtual pilots and normal pilots are used for the re-estimation of the channel vector.

V. Simulation Results and Discussions

In this section, we evaluate the performance of the proposed algorithm. In our setup, we assume that adjacent interfering devices are randomly located on a square (of area $100m^2$). The target receiver is located at center and the target transmit device is located 1 meter away from

the receiver. The packet size is set to 100 bits.[10] As a performance measure, we consider a packet error rate (PER). We assume that elements of channel vector for each device are i.i.d. zero mean complex Gaussian random variables with unit variance.

In our simulations, we study the performance of the following receiver techniques:

1) MMSE receiver with estimated CSI (realistic MMSE receiver) [16]: we set the interference training period as K.
2) Proposed method with estimated CSI: we set the interference training period as $N_d = N_b - N_p$.
3) Proposed method with virtual pilot signals (VPS): we choose N_v-best virtual pilots with the smallest MSE for the channel re-estimation.
4) Proposed method with accurate VPS: we use N_v-accurate data symbols as virtual pilots in the re-estimation of the channel vector.

In Fig. 5, we plot the PER performance of all techniques under consideration. In this simulations, we use the half rate ($r = \frac{1}{2}$) convolutional code with feedback polynomial $1+D+D^2$ and feedforward polynomial $1 + D^2$.[11] In the proposed method, 20 data symbols are used as VPS ($N_v = 20$). We set $J = 6$ as dominant neighboring interfering devices. Since the proposed method employs deliberately chosen reliable symbols to estimate the channel, it achieves substantial performance gain. For example, the proposed method with VPS achieves more than 2 dB gain over the realistic MMSE receiver at 10^{-3} PER point. We also observe that the addition of VPS achieves more than 1 dB gain over the proposed method without VPS.

In Fig. 6, we plot the PER performance with code rate $r = \frac{3}{4}$ convolutional code with feedback polynomial $1 + D^2 + D^3$ and feedforward polynomial $1 + D + D^3$. We observe from the figure that the performance gain of the proposed method with VPS is 2 dB over the realistic MMSE receiver and more than 0.8 dB over the proposed method without VPS at 10^{-3} PER point.

In Fig. 8, we plot the PER performance result of the schemes we considered with polar code and convolutional codes. We set the block length N_b as 64 since the block length of the polar code should be always a power of 2. We observe from the figure that the proposed method with polar code is superior to the convolutional code based schemes. Specifically, we observe that use of polar code will bring one dB over the convolutional code at 10^{-3} PER point.

[10]In order to meet the stringent delay requirements, packet size of the uRLLC use case is set to $100 \sim 200$ bits [36], [37].

[11]In IoT systems, convolutional codes are preferred over turbo or LDPC codes [5].

We next consider the performance of the proposed method over the Rician fading channel. Note that Rician fading includes the line-of-sight (LOS) signal propagation, which can be more general and realistic than Rayleigh fading [34]. In this simulations, we use the Rician K-factor $K = 6$ dB. Due to the LOS components of the interference channels, we observe from Fig. 7 that the performance of all receivers under consideration is worse than the case using the Rayleigh fading channel. From numerical results, the proposed method with VPS is still effective and achieving more than 1.5 dB gain over the realistic MMSE receiver at 10^{-3} PER point. We also observe that the addition of VPS achieves around 0.8 dB gain over the proposed method without VPS.

In Fig. 9, we plot the throughput as a function of the number of receive antennas N for temporally correlated channels. We consider temporally correlated channels that are modeled by a first order Gauss-Markov process [35] as $\mathbf{h}[k] = \xi\mathbf{h}[k-1] + \sqrt{1-\xi^2}\mathbf{g}[k]$ where $\mathbf{g}[k]$ is the innovation process, which is modeled as having i.i.d entries distributed with $\mathcal{CN}(0,1)$ and $0 \leq \xi \leq 1$ is the temporal correlation coefficient. In this simulations, we use $\xi = 0.9881$ for the moderate mobile speed. To evaluate the throughput, we present Monte-Carlo simulation results with 10000 iterations. Since each iteration consists of 10 fading blocks, the maximum throughput is about 5 Mbps. As observed in (7) and (15), when the number of receive antennas N increases, the training period used for the sample covariance matrix computation should also be increased to attain the target SINR. Thus, when the interfering training period K is fixed, the scaling factor $\left(1 - \frac{N-1}{K+1}\right)$ decreases with the number of receiver antennas N. As a result, we observe the throughput degradation on the realistic MMSE receiver when number of receive antennas is large ($N \geq 7$). Whereas, the throughput of the proposed VPS scheme increases with the number of receive antennas N.

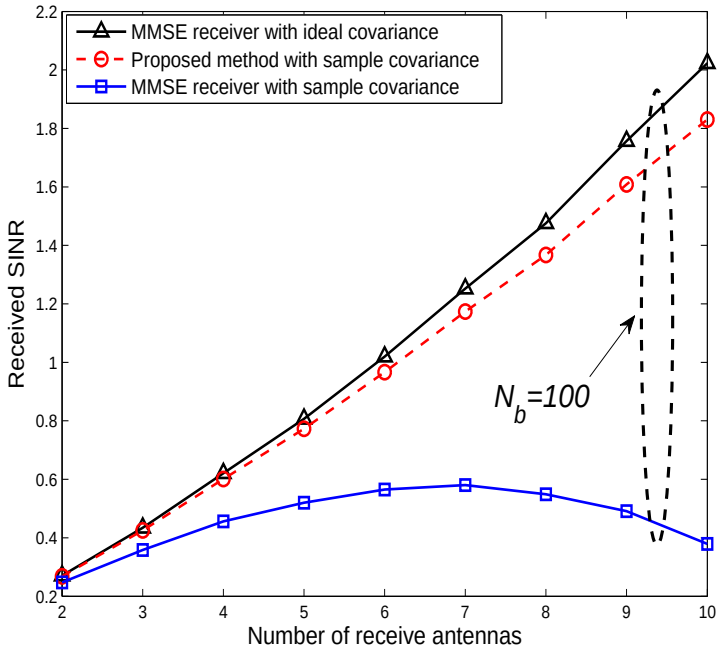
In Fig. 10, we plot the throughput as a function of the number of adjacent interfering devices J. Since the accuracy of the covariance matrix estimation deteriorates as the number of interfering devices J grows large, we see that the throughput decreases with J for all methods simulated. We observe that the proposed method using VPS performs close to the MMSE receiver using the perfect CSI ($5 \sim 7\%$ throughput loss) and also provides 21% gain over the realistic MMSE receiver when $J \leq 30$.

Finally, in Fig. 11, we show the throughput as a function of packet size N_b. We observe that the rate loss of the proposed method over the MMSE receiver with perfect CSI is around $6\% \sim 13\%$, which is far better than the rate loss of the realistic MMSE receiver ($12\% \sim 32\%$).

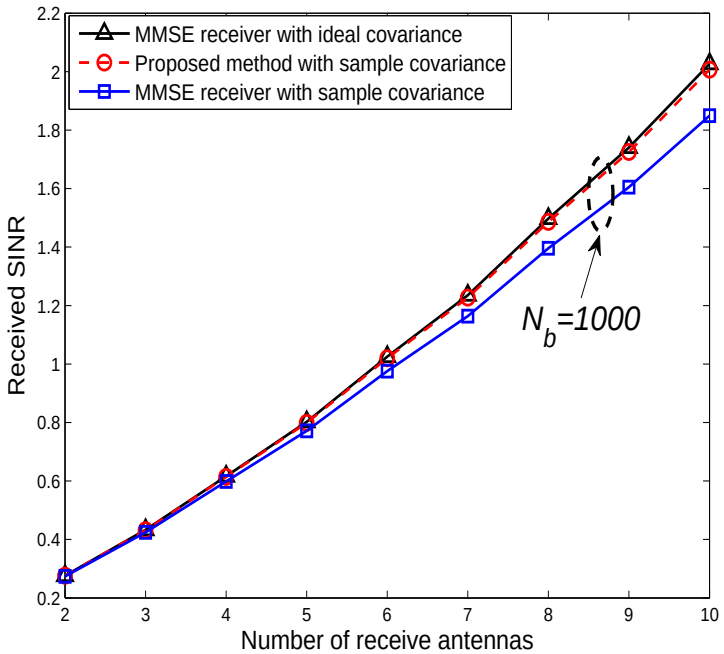
These results demonstrate that the proposed method has clear benefit over the realistic MMSE receiver in the short packet transmission.

VI. Conclusion

Short packet transmission systems are critical to fulfilling the stringent low latency requirements of the fifth generation (5G) wireless standards. We proposed a receiver technique suited to short packet transmission. Our work was motivated by the observation that the insufficient training period resulting from a small packet structure causes severe degradation in the desired channel and interference covariance matrix estimation. By exploiting reliable symbols in the data transmission period, the proposed receiver algorithm achieves improved estimation and eventual throughput gain. Although our study in this work focused on the low latency communication, the main concept can be readily extended to massive machine-type communications scenario (mMTC use case) and high throughput massive MIMO scenario (eMBB use case) of 5G wireless communications. In both scenarios, the channel estimation quality would be crucial to achieve the desired goal. A noncoherent approach that does not rely on the pilot signal might be alternative option for uRLLC communication. Since there are many issues such as signal design, receiver design, performance evaluation, harmonization with other 5G scenario (eMBB and mMTC use cases), these issues need further study. We leave these interesting explorations for our future work.



(a)



(b)

Fig. 4. The received SINR of the proposed method and conventional MMSE receiver for $N_b = 100$ and 1000. We set $J = 30$, SNR $= 0$ dB, $N_p = 0.1N_b$, $K = 0.1N_b$, $N_d = 0.8N_b$, $N_d' = 0.9N_b$.

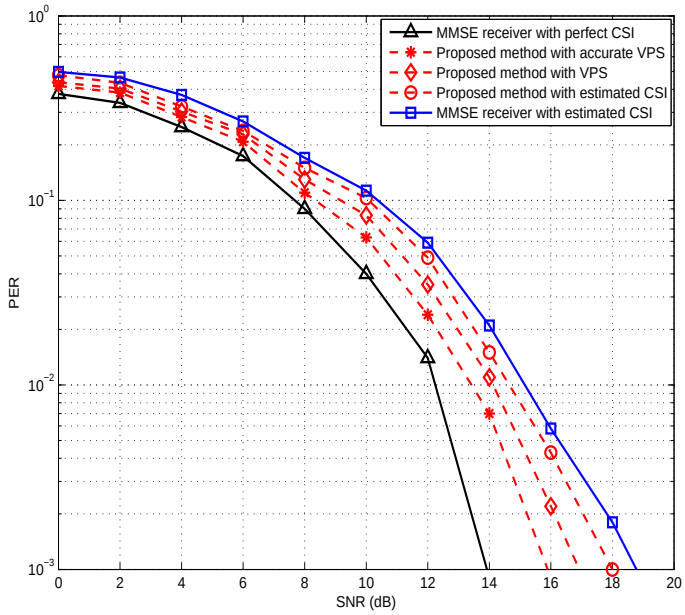


Fig. 5. PER performance of receiver techniques as a function of SNR. We set $r = 1/2$, $N = 6$, $J = 6$, $N_p = 10$, $K = 10$, $N_d = 40$, $N_d' = 50$, $N_b = 60$, $N_v = 20$.

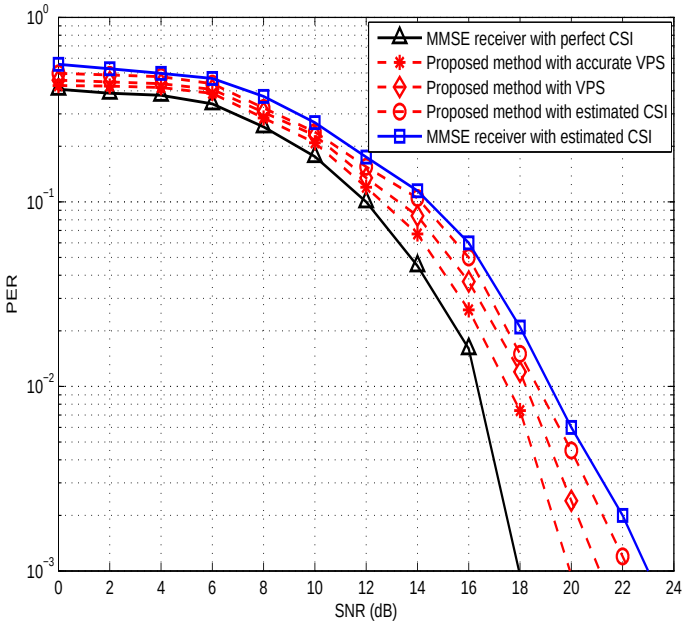


Fig. 6. PER performance of receiver techniques as a function of SNR. We set $r = 3/4$, $N = 6$, $J = 6$, $N_p = 10$, $K = 10$, $N_d = 40$, $N_d' = 50$, $N_b = 60$, $N_v = 20$.

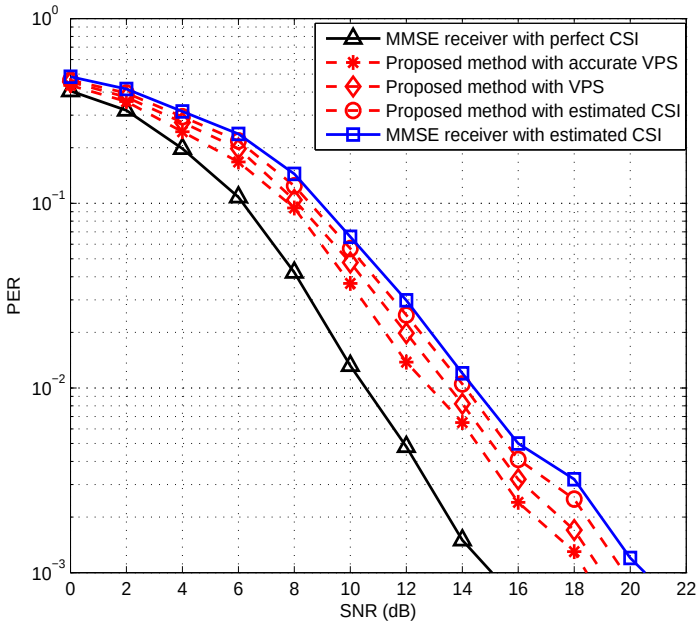


Fig. 7. PER performance of receiver techniques as a function of SNR over the Rician fading channel. We set $r = 1/2$, $N = 6$, $J = 6$, $N_p = 10$, $K = 10$, $N_d = 40$, $N'_d = 50$, $N_b = 60$, $N_v = 20$.

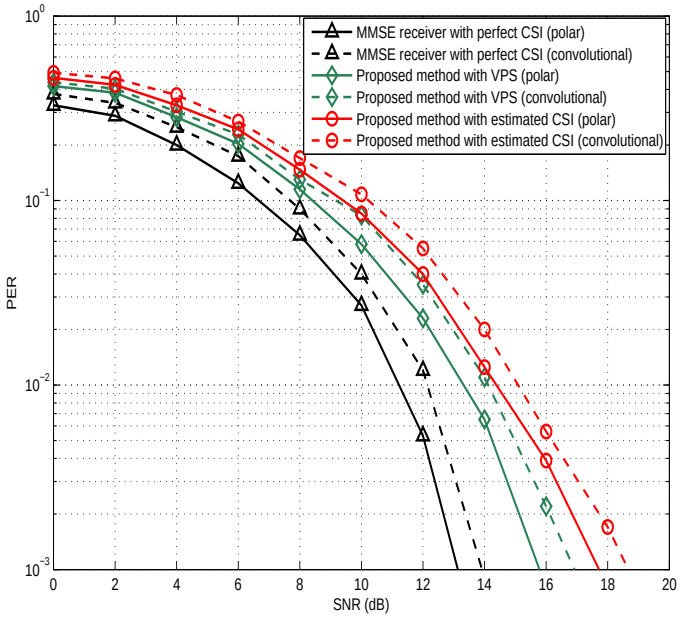


Fig. 8. PER performance of receiver techniques as a function of SNR with polar code and convolutional codes. We set $r = 1/2$, $N = 6$, $J = 6$, $N_p = 10$, $N'_d = 54$, $N_b = 64$, $N_v = 20$.

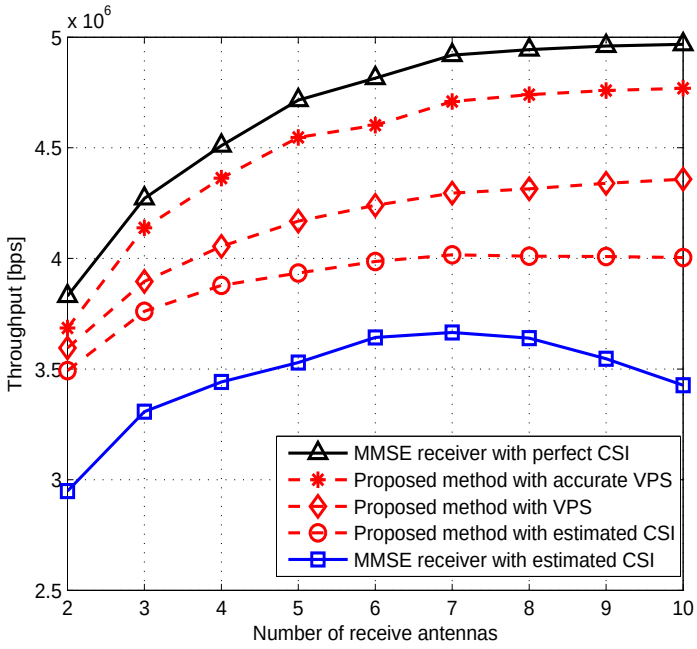


Fig. 9. Throughput of receiver techniques as a function of number of receive antennas N for temporally-correlated block fading channel. We set $SNR = 0$ dB, $r = \frac{1}{2}$, $J = 6$, $N_p = 10$, $K = 10$, $N_d = 40$, $N'_d = 50$, $N_b = 60$, $N_v = 20$.

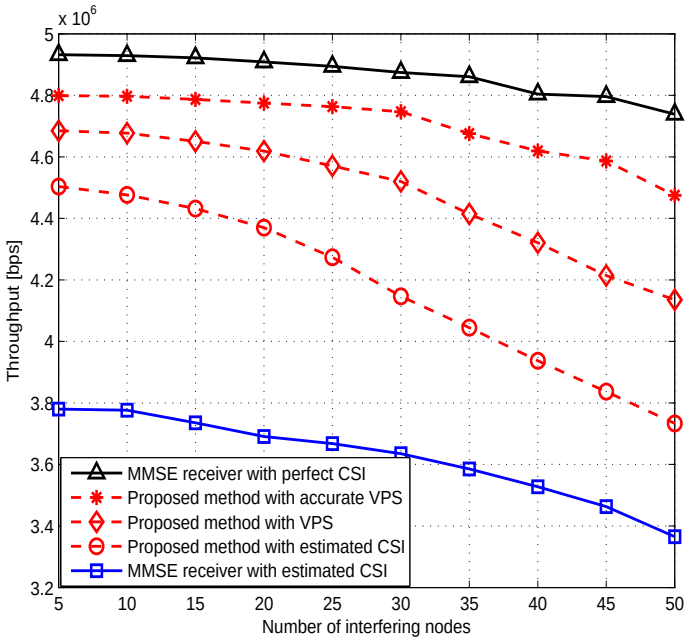


Fig. 10. Throughput of receiver techniques as a function of interfering devices for temporally-correlated block fading channel. We set $SNR = 10$ dB, $r = \frac{1}{2}$, $N = 6$, $N_p = 10$, $K = 10$, $N_d = 40$, $N'_d = 50$, $N_b = 60$, $N_v = 20$.

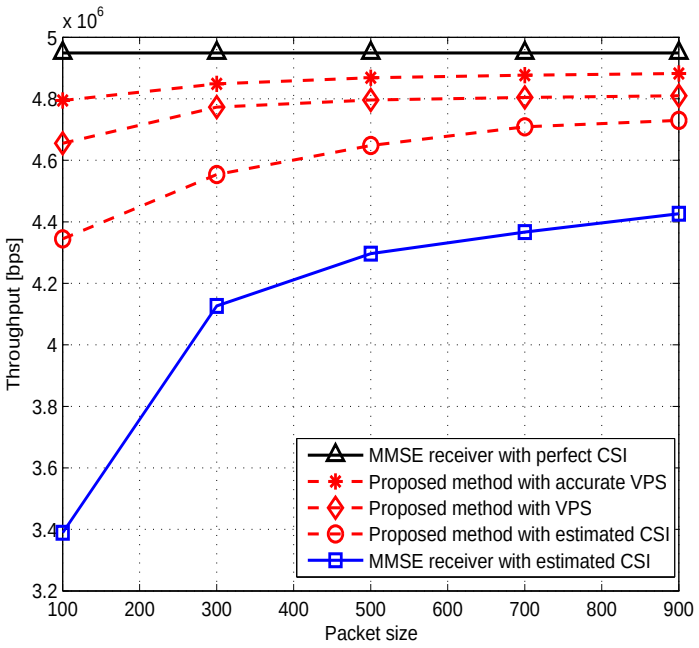


Fig. 11. Throughput of receiver techniques as a function of N_b for temporally-correlated block fading channel. We set $J = 6$, SNR $= 0$ dB, $N = 8$, $N_p = 0.1N_b$, $K = 0.1N_b$, $N_d = 0.8N_b$, $N_d' = 0.9N_b$, $N_v = 0.2N_d$.

APPENDIX A
DERIVATION OF (37)

Recall from (34) that ε_n is given by

$$\varepsilon_n = \text{tr}\left(\text{Cov}\left(\mathbf{h}_0\right) - \tilde{\mathbf{\Omega}}\tilde{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{\Omega}}^H\right) \tag{38}$$

where

$$\tilde{\mathbf{\Omega}} = \text{Cov}\left(\mathbf{h}_0, \begin{bmatrix}\mathbf{y}_p \\ \mathbf{y}_v^{(n)}\end{bmatrix}\right) = \begin{bmatrix}\mathbf{C}_{\text{hh}}\mathbf{P}_0^H & (\bar{s}_0^{(n)})^*\mathbf{C}_{\text{hh}}\end{bmatrix} \tag{39}$$

and

$$\begin{aligned}\tilde{\mathbf{\Sigma}} &= \text{Cov}\left(\begin{bmatrix}\mathbf{y}_p \\ \mathbf{y}_v^{(n)}\end{bmatrix}, \begin{bmatrix}\mathbf{y}_p \\ \mathbf{y}_v^{(n)}\end{bmatrix}\right) \\ &= \begin{bmatrix}\mathbf{P}_0\mathbf{C}_{\text{hh}}\mathbf{P}_0^H + \eta_{\text{p}}\mathbf{I} & \mathbf{P}_0\mathbf{C}_{\text{hh}}(\bar{s}_0^{(n)})^* \\ \bar{s}_0^{(n)}\mathbf{C}_{\text{hh}}\mathbf{P}_0^H & \bar{\lambda}_0^{(n)}\mathbf{C}_{\text{hh}} + \eta_{\text{v}}\mathbf{I}\end{bmatrix}.\end{aligned} \tag{40}$$

For convenience, we let

$$\begin{bmatrix}\mathbf{\Omega}_1 & \mathbf{\Omega}_2\end{bmatrix} = \begin{bmatrix}\mathbf{C}_{\text{hh}}\mathbf{P}_0^H & (\bar{s}_0^{(n)})^*\mathbf{C}_{\text{hh}}\end{bmatrix}, \tag{41}$$

$$\begin{bmatrix}\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22}\end{bmatrix} = \begin{bmatrix}\mathbf{P}_0\mathbf{C}_{\text{hh}}\mathbf{P}_0^H + \eta_{\text{p}}\mathbf{I} & \mathbf{P}_0\mathbf{C}_{\text{hh}}(\bar{s}_0^{(n)})^* \\ \bar{s}_0^{(n)}\mathbf{C}_{\text{hh}}\mathbf{P}_0^H & \bar{\lambda}_0^{(n)}\mathbf{C}_{\text{hh}} + \eta_{\text{v}}\mathbf{I}\end{bmatrix}, \tag{42}$$

and

$$\begin{bmatrix}\mathbf{\Gamma}_{11} & \mathbf{\Gamma}_{12} \\ \mathbf{\Gamma}_{21} & \mathbf{\Gamma}_{22}\end{bmatrix} = \begin{bmatrix}\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22}\end{bmatrix}^{-1}, \tag{43}$$

then

$$\begin{aligned}\varepsilon_n &= \text{tr}\left(\mathbf{C}_{\text{hh}} - \begin{bmatrix}\mathbf{\Omega}_1 & \mathbf{\Omega}_2\end{bmatrix}\begin{bmatrix}\mathbf{\Gamma}_{11} & \mathbf{\Gamma}_{12} \\ \mathbf{\Gamma}_{21} & \mathbf{\Gamma}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{\Omega}_1 & \mathbf{\Omega}_2\end{bmatrix}^H\right) \\ &= \text{tr}(\mathbf{C}_{\text{hh}} - \mathbf{\Omega}_1\mathbf{\Gamma}_{11}\mathbf{\Omega}_1^H - \mathbf{\Omega}_2\mathbf{\Gamma}_{21}\mathbf{\Omega}_1^H - \mathbf{\Omega}_1\mathbf{\Gamma}_{12}\mathbf{\Omega}_2^H - \mathbf{\Omega}_2\mathbf{\Gamma}_{22}\mathbf{\Omega}_2^H).\end{aligned} \tag{44}$$

Using the inversion formula of partitioned matrices [22], we have

$$\begin{aligned}\mathbf{\Gamma}_{11} &= (\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})^{-1} \\ \mathbf{\Gamma}_{12} &= -\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{21})^{-1} \\ \mathbf{\Gamma}_{21} &= -\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}(\mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21})^{-1} \\ \mathbf{\Gamma}_{22} &= (\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})^{-1}.\end{aligned} \tag{45}$$

Using (45), we further have

$$\begin{aligned}
\mathbf{\Omega}_1\mathbf{\Gamma}_{11}\mathbf{\Omega}_1^H &= \mathbf{C}_{\rm hh}\mathbf{P}_0^H\mathbf{\Gamma}_{11}\mathbf{P}_0\mathbf{C}_{\rm hh} \\
\mathbf{\Omega}_2\mathbf{\Gamma}_{21}\mathbf{\Omega}_1^H &= (\bar{s}_0^{(n)})^*\mathbf{C}_{\rm hh}\mathbf{\Gamma}_{21}\mathbf{P}_0\mathbf{C}_{\rm hh} \\
\mathbf{\Omega}_1\mathbf{\Gamma}_{12}\mathbf{\Omega}_2^H &= \bar{s}_0^{(n)}\mathbf{C}_{\rm hh}\mathbf{P}_0^H\mathbf{\Gamma}_{12}\mathbf{C}_{\rm hh} \\
\mathbf{\Omega}_2\mathbf{\Gamma}_{22}\mathbf{\Omega}_2^H &= \bar{\lambda}_0^{(n)}\mathbf{C}_{\rm hh}\mathbf{\Gamma}_{22}\mathbf{C}_{\rm hh}. \qquad (46)
\end{aligned}$$

Using the matrix inversion lemma ($(A-BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D-CA^{-1}B)CA^{-1}$) and noting that

$$\begin{aligned}
\mathbf{\Phi} &= \mathbf{P}_0^H\mathbf{\Sigma}_{11}\mathbf{P}_0 \\
\mathbf{\Phi}' &= \mathbf{P}_0^H\mathbf{\Sigma}_{11}^{-1}\mathbf{P}_0 \\
\mathbf{\Psi} &= \bar{\lambda}_0^{(n)}\mathbf{\Sigma}_{22} \\
\mathbf{\Psi}' &= \bar{\lambda}_0^{(n)}\mathbf{\Sigma}_{22}^{-1}, \qquad (47)
\end{aligned}$$

we further have

$$\begin{aligned}
\mathbf{\Omega}_1\mathbf{\Gamma}_{11}\mathbf{\Omega}_1^H &= \mathbf{C}_{\rm hh}\mathbf{\Phi}'(\mathbf{I} + \mathbf{C}_{\rm hh}\mathbf{\Psi}\mathbf{C}_{\rm hh}\mathbf{\Phi}' - (\bar{\lambda}_0^{(n)})^2\mathbf{C}_{\rm hh} \\
&\quad \times \mathbf{\Phi}'\mathbf{C}_{\rm hh}\mathbf{C}_{\rm hh}\mathbf{\Phi}')\mathbf{C}_{\rm hh} \\
\mathbf{\Omega}_2\mathbf{\Gamma}_{21}\mathbf{\Omega}_1^H &= -\mathbf{C}_{\rm hh}\mathbf{\Psi}'\mathbf{C}_{\rm hh}(\mathbf{I} + \mathbf{\Phi}'\mathbf{C}_{\rm hh}\mathbf{\Psi}\mathbf{C}_{\rm hh} - (\bar{\lambda}_0^{(n)})^2\mathbf{\Phi}'\mathbf{C}_{\rm hh} \\
&\quad \times \mathbf{C}_{\rm hh}\mathbf{\Phi}'\mathbf{C}_{\rm hh}\mathbf{C}_{\rm hh})\mathbf{\Phi}'\mathbf{C}_{\rm hh} \\
\mathbf{\Omega}_1\mathbf{\Gamma}_{12}\mathbf{\Omega}_2^H &= -\mathbf{C}_{\rm hh}\mathbf{\Phi}'\mathbf{C}_{\rm hh}(\mathbf{I} + \mathbf{\Psi}'\mathbf{C}_{\rm hh}\mathbf{\Phi}\mathbf{C}_{\rm hh} - \mathbf{\Psi}' \\
&\quad \times \mathbf{C}_{\rm hh}^2\mathbf{\Psi}'\mathbf{C}_{\rm hh}^2)\mathbf{\Psi}'\mathbf{C}_{\rm hh} \\
\mathbf{\Omega}_2\mathbf{\Gamma}_{22}\mathbf{\Omega}_2^H &= \mathbf{C}_{\rm hh}\mathbf{\Psi}'(\mathbf{I} + \mathbf{C}_{\rm hh}\mathbf{\Phi}\mathbf{C}_{\rm hh}\mathbf{\Psi}' - \mathbf{C}_{\rm hh}^2\mathbf{\Psi}'\mathbf{C}_{\rm hh}^2\mathbf{\Psi}')\mathbf{C}_{\rm hh}. \qquad (48)
\end{aligned}$$

Under the assumption that there is no correlation among receive antennas, $\mathbf{C}_{\rm hh} = \mathbf{I}$, and hence (47) becomes

$$\begin{aligned}
\mathbf{\Phi} &= \mathbf{P}_0^H(\mathbf{P}_0\mathbf{P}_0^H + \eta_{\rm p}\mathbf{I})\mathbf{P}_0 = N_p(N_p+\eta_{\rm p})\mathbf{I} \\
\mathbf{\Phi}' &= \mathbf{P}_0^H(\mathbf{P}_0\mathbf{P}_0^H + \eta_p\mathbf{I})^{-1}\mathbf{P}_0 = \frac{N_p}{N_p+\eta_{\rm p}}\mathbf{I} \\
\mathbf{\Psi} &= (\bar{s}_0^{(n)})^*\mathbf{I}(\bar{\lambda}_0^{(n)}\mathbf{I} + \eta_{\rm v}\mathbf{I})\bar{s}_0^{(n)}\mathbf{I} = \bar{\lambda}_0^{(n)}(\bar{\lambda}_0^{(n)} + \eta_{\rm v})\mathbf{I} \\
\mathbf{\Psi}' &= (\bar{s}_0^{(n)})^*\mathbf{I}(\bar{\lambda}_0^{(n)}\mathbf{I} + \eta_{\rm v}\mathbf{I})^{-1}\bar{s}_0^{(n)}\mathbf{I} = \frac{\bar{\lambda}_0^{(n)}}{\bar{\lambda}_0^{(n)} + \eta_{\rm v}}\mathbf{I}. \qquad (49)
\end{aligned}$$

Plugging (48) and (49) into (44) and after some manipulations, we have

$$\begin{aligned}\varepsilon_n &= \mathrm{tr}(\mathbf{I} - \left((\mathbf{\Phi}')^2\mathbf{\Psi}' + (\mathbf{\Phi}')^2\right)\mathbf{\Psi} - (\bar{\lambda}_0^{(n)})^2((\mathbf{\Phi}')^3\mathbf{\Psi}' + (\mathbf{\Phi}')^2) \\ &\quad + (\mathbf{I} - \mathbf{\Phi}')\,(\mathbf{\Psi}')^3 + \mathbf{\Phi}(\mathbf{\Psi}')^2 - (\mathbf{I} - \mathbf{\Phi}'\mathbf{\Phi} - 2\mathbf{\Phi}')\,\mathbf{\Psi}' - \mathbf{\Phi}') \\ &= N(1 - \gamma_p^2(1+\gamma_p)\frac{\bar{\lambda}_0^{(n)}}{\bar{\lambda}_0^{(n)} + \eta_{\mathrm{v}}}(\bar{\lambda}_0^{(n)})^2 - \gamma_p^2(\frac{2\eta_{\mathrm{v}}\bar{\lambda}_0^{(n)} + \eta_{\mathrm{v}}^2}{\bar{\lambda}_0^{(n)} + \eta_{\mathrm{v}}}) \\ &\quad \times \bar{\lambda}_0^{(n)} + \gamma_p\frac{(\bar{\lambda}_0^{(n)})^3}{(\bar{\lambda}_0^{(n)} + \eta_{\mathrm{v}})^3} + \frac{(\bar{\lambda}_0^{(n)})^2}{(\bar{\lambda}_0^{(n)} + \eta_{\mathrm{v}})^2}\frac{N_p^2}{\gamma_p} \\ &\quad + \frac{\bar{\lambda}_0^{(n)}}{\bar{\lambda}_0^{(n)} + \eta_{\mathrm{v}}}(2\gamma_p + N_p^2 - 1) - \gamma_p) \end{aligned} \tag{50}$$

where $\gamma_p = \frac{N_p}{N_p + \eta_{\mathrm{p}}}$.

REFERENCES

[1] B. Lee, S. Park, D. J. Love, H. Ji, and B. Shim, "Packet structure and receiver design for low-latency communications with ultra-small packets," *Proc. of IEEE Global Communications (GLOBECOM) Conference*, Dec. 2016.

[2] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787-2805, Oct. 2010.

[3] E. L*ä*hetkangas, K. Pajukosi, J. Vihri*ä*l*ä*, G. Berardinelli, M. Lauridsen, E. Tiirola, and P. Mogensen, "Achieving low latency and energy consumption by 5G TDD mode optimization," in *Proc. of IEEE International Conference on Communications (ICC)*, Jun. 2014.

[4] O. Yilmaz, Y. Wang, N. Johansson, N. Brahmi, S. Ashraf, and J. Sachs, "Analysis of ultra-reliable and low-latency 5G commuinication for a factory automaiton use case," in *Proc. of IEEE International Conference on Communications (ICC)*, Jun. 2015.

[5] N. Johansson, Y. Wang, E. Eriksson, and M. Hessler, "Radio access for ultra-reliable and low-latency 5G communications," in *Proc. of IEEE International Conference on Communications (ICC)*, Jun. 2015.

[6] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74-80, Feb. 2014.

[7] A. Osserian, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. Uusitalo, B. Timus, and M. Fallgren, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 26-35, May 2014.

[8] E. Dahlman, G. Mildh, S. Parkvall, J. Peisa, J. Sachs, and Y. Sel*é*n, "5G radio access," *Ericsson Review*, no. 6, Jun. 2014.

[9] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307-2359, May 2010.

[10] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static multiple-antenna fading channels at finite blocklength," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232-4265, July 2014.

[11] G. Durisi, T. Koch, J. *Ö*stman, Y. Polyanskiy, and W. Yang, "Short-packet communications over multiple-antenna rayleigh-fading channels," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 618-629, Feb. 2016.

[12] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711-1726, Sep. 2016.

[13] B. Makki, T. Svensson, and M. Zorzi, "Finite block-length analysis of spectrum sharing networks using rate adaptation," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2823-2835, Aug. 2015.

[14] B. Makki, T. Svensson, and M. Zorzi, "Wireless energy and information transmission using feedback: infinite and finite block-length analysis," *IEEE Trans. Commun.*, vol. 64, no. 12, pp. 5304-5318, Dec. 2016.

[15] J. Andrews, W. Choi, and R. Heath, Jr., "Overcoming Interference in Spatial Multiplexing MIMO Cellular Networks," *IEEE Wireless Commun.*, vol. 14, no. 6, pp. 95-104, Dec. 2007.

[16] N. Jindal, J. G. Andrews, and S. Weber, "Multi-antenna communication in ad hoc networks: achieving MIMO gains with SIMO transmission," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 529-540, Feb. 2011.

[17] J. Gao and H. Liu, "Decision-directed estimation of MIMO time-varying Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1412-1417, July 2005.

[18] H. Vikalo, B. Hassibi, and P. Stoica, "Efficient joint maximum-likelihood channel estimation and signal detection," *IEEE Trans. Wireless Commun.*, vol. 5, no. 7, pp. 1838-1845, July 2006.

[19] D. Yoon and J. Moon, "Soft-decision-driven channel estimation for pipelined turbo receivers," *IEEE Trans. Commun.*, vol. 59, no. 8, pp. 2141-2151, Aug. 2011.

[20] S. Park, B. Shim, and J. Choi, "Iterative channel estimation using virtual pilot signals for MIMO-OFDM systems," *IEEE Trans. Sig. Proc.*, vol. 63, no. 12, pp. 3032-3045, Jun. 2015.

[21] H. V. Poor and S. Verdu, "Probability of error in MMSE multiuser detection," *IEEE Trans. Inf. Theory.*, vol. 43, no. 3, pp. 858-871, May 1997.

[22] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Thoery.*, EngleWood Cliffs, NJ, USA: Prentice-Hall, 1998.

[23] H. Gao, P. J. Smith, and M. V. Clark, "Theoretical reliability of MMSE linear diversity combining in Rayleigh-fading additive interference channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 666-672, May 1998.

[24] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications.*, Cambridge, U.K.: Cambridge Univ. Press, 2005.

[25] P. Li, D. Paul, R. Narasimhan, and J. Cioffi, "On the distribution of SINR for the MMSE MIMO receiver and performance analysis," *IEEE Trans. Inf. Theory.*, vol. 52, no. 1, pp. 271-286, Jan. 2006.

[26] N. Kim, Y. Lee, and H. Park, "Performance analysis of MIMO system with linear MMSE receiver," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11, pp. 4474-4478, Nov. 2008.

[27] S. A. M. Ghanem, "Multiple access gaussian channels with arbitrary inputs: optimal precoding and power allocation," *arXiv preprint arXiv:1411.0446*, 2014.

[28] S. T. Veetil, K. Kuchi, and R. K. Ganti, "Performance of PZF and MMSE receivers in cellular networks with multi-user spatial multiplexing," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 4867-4878, Sept. 2015.

[29] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. 35, no. 10, pp. 1365-1375, 1987.

[30] I. Reed, J. Mallet, and L. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 10, no. 6, pp. 853-863, Nov. 1974.

[31] N. Goodman, "Statistical analysis based on a certain multivariate complex gaussian distribution," *The Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 152-177, March 1963.

[32] G. Fordor, P. D. Marco, M. Telek, "On the impact of antenna correlation on the pilot-data balance in multiple antenna systems," in *Proc. of IEEE International Conference on Communications (ICC)*, Jun. 2015.

[33] B. Hochwald and S. T. Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389-399, Mar. 2003.

[34] J. Zhang, L. Dai, Z. He, S. Jin, and X. Li, "Performance analysis of mixed-ADC massive MIMO systems over Rician fading channels," *arXiv preprint arXiv:1703.03642*, 2017.

[35] W. C. Jakes, "Microwave Mobile Communications," New York, John Wiley & Sons Inc, 1975.

[36] F. Fettweis and S. Alamouti, "5G: personal mobile internet beyond what cellular did to telephony," *IEEE Commun. Mag.*, vol. 52, no. 2 pp. 140-145, Feb. 2014.

[37] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59-65, Sept. 2016.