Short-Packet Transmission over a Bidirectional Massive MIMO link

Citation for the original published paper (version of record):
Östman, J., Lancho Serrano, A., Durisi, G. (2019)
Short-Packet Transmission over a Bidirectional Massive MIMO link
Conference Record - Asilomar Conference on Signals, Systems and Computers

Short-Packet Transmission over a Bidirectional Massive MIMO link

Johan Östman, Alejandro Lancho, and Giuseppe Durisi
Chalmers University of Technology, Gothenburg, Sweden

Abstract—**We consider the transmission of short packets over a bidirectional communication link where multiple devices, e.g., sensors and actuators, exchange small-data payloads with a base station equipped with a large antenna array. Using results from finite-blocklength information theory, we characterize the minimum SNR required to achieve a target error probability for a fixed packet length and a fixed payload size. Our nonasymptotic analysis, which applies to the scenario in which the bidirectional communication is device-initiated, and also to the more challenging case when it is base-station initiated, provides guidelines on the design of massive multiple-input multiple-output links that need to support sporadic ultra-reliable low-latency transmissions. Specifically, it allows us to determine the optimal amount of resources that need to be dedicated to the acquisition of channel state information.**

I. Introduction

Because of its ability to accommodate many parallel high-throughput links in the same time-frequency resources, massive multiple-input multiple-output (MIMO) has been identified as a key technology for next-generation wireless systems [1], [2]. Furthermore, the potentially large spatial diversity provided by massive MIMO makes this technology also relevant for some of the new use cases in next generation's wireless systems, where reliability and latency, rather than throughput, are in focus [3].

One such use case is ultra-reliable low-latency communications (URLLC), where small data payloads need to be transmitted under stringent latency and reliability constraints. For example, in the context of factory automation, one may need to deliver packets of 100 bits, conveying, e.g., readings from sensors or commands to actuators, within hundreds of microseconds and with a reliability no smaller than 99.999%. In this scenario, the stringent delay constraint prohibits the exploitation of diversity in time; furthermore, there may be only limited diversity in frequency. Thus, spatial diversity offered by multiple antennas is critical to achieve the desired reliability [4].

The purpose of this paper is to provide a characterization of the error probability achievable in a bidirectional massive MIMO link as a function of the SNR, the number of active user equipments (UEs), the number of the available antennas at the base station (BS), and the size of the information payload. Previous results reported in the literature [5]–[7] rely on asymptotic performance metrics, such as ergodic and outage capacity, to characterize the performance of latency-constrained communication systems. Our analysis relies instead on tools from finite-blocklength information theory, which are more suited to the blocklengths of interest in URLLC than asymptotic performance metrics [8].

Literature review: Most of the information-theoretic characterizations of massive MIMO communication links deal with bounds on the ergodic capacity [9]. These bounds are typically obtained under specific assumptions on the signaling scheme and on the operations performed at the receiver side, which are motivated by practical considerations. Specifically, it is common to postulate that the system operates in time-division duplex (TDD) mode, that pilot symbols are transmitted in the uplink (UL), and that the BS performs minimum mean-square error (MMSE) channel estimation followed by linear combining in the UL and linear precoding in the downlink (DL). In both the UL and the DL, the channel estimate is treated as perfect. Furthermore, the most commonly used bounds assume implicitly that the receiver (be it the BS or the UE) performs *mismatched scaled-nearest-neighbor decoding* [10] by treating channel estimation errors and residual multiuser interference as noise.

These ergodic bounds are, however, unsuitable for URLLC. Indeed, they rely on the assumption that each codeword spans a large number of diversity branches over time—an assumption that is not valid in low-latency scenarios.

An alternative approach, recently followed in [5], [7] is to use instead outage capacity as performance metric. The outage capacity is an asymptotic performance metric that pertains to the setup in which the channel stays constant (or varies only a finite number of times), as the blocklength grows large—a setting often referred to as *quasi-static* fading. The use of the outage capacity in [5], [7] is motivated by the *zero-dispersion* result obtained in [11] which we shall briefly review next. Let the maximum coding rate be the largest rate at which one can transmit information for a given constraint on the blocklength and the packet error probability. In [11], it is shown that the speed at which the maximum coding rate converges to the outage capacity for quasi-static fading channels, as the blocklength increases, is much faster than the speed at which the maximum coding rate converges to Shannon's capacity for nonfading additive white Gaussian noise (AWGN) channels. Intuitively, the reason is that errors in quasi-static fading channels are caused by deep-fade events, which cannot be alleviated through coding.

However, this result relies on a Taylor expansion of the maximum coding rate, in which high-order terms that depend on the fading distribution are ignored. In particular, it is known that these high-order terms become increasingly large as the fading distribution becomes more concentrated around its mean [12],

This work was supported by the Swedish Research Council under grants 2014-6066 and 2016-03293.

which is exactly what happens in massive MIMO links when channel hardening occurs. This makes the use of outage capacity questionable.

Another unsatisfactory consequence resulting from using outage capacity is that the channel can be estimated perfectly at no rate penalty, both in the UL and in the DL. Indeed, it is sufficient to transmit a number of pilot symbols that grows sub-linearly with the blocklength [13, p. 2632]. This is dissatisfying as the performance of massive MIMO systems in the URLLC regime are expected to depend heavily on the channel estimation accuracy [3]. This issue is partially addressed in [5], [7] by utilizing outage-probability approximations in which the rate is multiplied by a correction factor that accounts for pilot overhead. However, the validity of such approximations is unclear.

Contributions: We provide a finite-blocklength framework to analyze the performance of massive MIMO systems in the URLLC regime. Specifically, we present finite-blocklength bounds on the error probability that capture the main features of massive MIMO links, i.e., UL pilot transmissions, linear combining/precoding, and mismatched nearest-neighbor detection. The bounds are based on random coding, pertain to Gaussian codebooks, and rely on the random-coding union bound with parameter s (RCUs) [14]. Furthermore, by generalizing the analysis presented in [8], [15], we also obtain finite-blocklength bounds for the setup in which an (inner) orthogonal space-time block code (OSTBC) is used at the transmitter side to provide spatial diversity for the case in which channel state information (CSI) at the transmitter is not available.

We then apply these bounds to two scenarios that are relevant for URLLC: a UE-initiated bidirectional communication link, and a BS-initiated bidirectional communication link. While the first scenario is somewhat standard in massive MIMO analyses, the second scenario is less investigated in the literature. In the second scenario, similar to the initial-access problem considered in [7], the BS cannot perform beamforming based on UL-pilot channel estimation. Hence, it needs to resort to space-time block-codes to achieve spatial diversity. Furthermore, as already pointed out in [7], the significant overhead caused by DL pilot transmission prevents the BS from using all available transmit antennas. For both scenarios, our bounds allow one to determine the optimal number of pilot symbols to be transmitted in order to minimize the SNR required to sustain a target error probability.

Notation: Boldface lower-case letters denote vectors and boldface upper-case letters are used for matrices. We denote by $\mathbf{0}_n$ and $\mathbf{I}_n$, the all-zero vector of size n and the identity matrix of size $n \times n$, respectively. The superscripts $(\cdot)^{\mathsf{T}}$, $(\cdot)^{\mathsf{H}}$, and $(\cdot)^*$ are used for transposition, Hermitian transposition, and complex conjugation. The distribution of a standard circularly symmetric Gaussian random variable is denoted by $\mathcal{CN}(0, 1)$. Finally, $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary part, the expectation operator is denoted by $\mathbb{E}[\cdot]$, and the ℓ_2-norm is written as $\|\cdot\|$.

II. Finite-Blocklength Bounds for a Simplified Channel Model

We start by presenting our finite-blocklength framework for a simplified channel model that, as we shall see, captures the main features of the massive MIMO setup we are interested in. Consider the complex-valued additive channel

$$v_k = g t_k + w_k, \quad k = 1, \dots, n. \tag{1}$$

Here, t_k denotes the channel input, g is a deterministic channel gain, and w_k denotes the additive noise. The channel output is represented by v_k and n stands for the blocklength. To transmit the message $m \in \{1, \dots, M\}$, the encoder maps it to one out of M n-dimensional codewords $\{\mathbf{t}(m)\}_{m=1}^{M}$, where $\mathbf{t} = [t_1, \dots, t_n]$.

The key step to obtain finite-blocklength bounds that are relevant for the massive MIMO setup we are interested in, is to model appropriately the operations that the decoder is allowed to perform. In what follows, we will assume that:

- The receiver has an estimate $\hat{g}$ of the channel gain g that is treated as perfect.
- To decode the transmitted message, the receiver seeks the codeword $\mathbf{t}(m)$ that, once scaled by $\hat{g}$ is the closest to the received vector $\mathbf{v} = [v_1, \dots, v_n]$ in Euclidean distance. Mathematically, the estimated message $\hat{m}$ at the receiver is given by

$$\hat{m} = \underset{\tilde{m} \in \{1,\dots,M\}}{\arg\min} \|\mathbf{v} - \hat{g}\mathbf{t}(\tilde{m})\|^2. \tag{2}$$

Some comments are in order. The receiver just described is the maximum likelihood (ML) receiver if and only if $\hat{g} = g$ and w_k is an i.i.d. $\mathcal{CN}(0, \sigma^2)$ sequence. This means, that the receiver just introduced treats the additive noise (which is not necessarily Gaussian) as Gaussian. We refer to this decoder as a mismatched scaled nearest-neighbor (SNN) decoder [10].

We are now interested in determining a bound on the message error probability ϵ achieved by this receiver. To do so, we follow a standard practice in information theory and use a random-coding approach, where we analyze the error probability of an ensemble of random codes, which are generated by drawing the elements of each codeword independently from a given distribution. Specifically, we consider a Gaussian random code ensemble, where the elements of each codeword are drawn independently from a $\mathcal{CN}(0, \rho)$ distribution.[1] Here, ρ can be thought of as the average transmit power. A simple generalization of the random coding union bound in [17, Thm. 16] to the mismatched SNN decoder (2) results in the following bound

$$\epsilon \leq \mathbb{E}[\min\{1, (M-1) f(\mathbf{t}, \mathbf{v})\}] \tag{3}$$

where $f(\mathbf{t}, \mathbf{v}) = \Pr\{\|\mathbf{v} - \hat{g}\bar{\mathbf{t}}\|^2 \leq \|\mathbf{v} - \hat{g}\mathbf{t}\|^2 | \mathbf{t}, \mathbf{v}\}$. The random variables involved in the bound have the following joint distribution: $P_{\mathbf{t},\mathbf{v},\bar{\mathbf{t}}}(\mathbf{a}, \mathbf{b}, \mathbf{c}) = P_{\mathbf{t}}(\mathbf{a}) P_{\mathbf{v}|\mathbf{t}}(\mathbf{b}|\mathbf{a}) P_{\mathbf{t}}(\mathbf{c})$. Coarsely speaking, $\mathbf{t}$ denotes the transmitted codeword, whereas $\bar{\mathbf{t}}$ denotes another codeword. Clearly if $\bar{\mathbf{t}}$ is closer to $\mathbf{v}$ in Euclidean distance after being scaled by $\hat{g}$, the decoded message will be wrong. The bound (3) then follows from a tightened version of the union bound.

Although tight, this bound is difficult to compute numerically. This is because M is typically very large (e.g., $M = 2^{50}$ for a

[1] Note that this ensemble is not optimal at finite blocklength, not even when $\hat{g} = g$ and the additive noise is Gaussian [16]. We chose it because it results in simple expressions. The analysis can be easily extended to other ensembles.

code of rate $1/2$ and blocklength 100). Hence, the probability term inside the expectation needs to be computed with very high precision—something that is not possible using plain vanilla Monte-Carlo methods. The approach proposed in [14] to solve this issue is to upper-bound the probability term using the Chernoff bound. This results in the so-called RCUs bound:

$$\epsilon \leq \inf_{s>0} \mathbb{E}[\exp\left(-\max\{0, \imath_s(\mathbf{t},\mathbf{v}) - \ln(M-1)\}\right)] \quad (4)$$

where

$$\imath_s(\mathbf{t},\mathbf{v}) = s\|\mathbf{v} - \hat{g}\mathbf{t}\|^2 - s\frac{\|\mathbf{v}\|^2}{1+s\rho|\hat{g}|^2} - n\ln(1+s\rho|\hat{g}|^2). \quad (5)$$

Next, we will discuss how to use (4) to assess the finite-blocklength performance of massive MIMO links.

III. UE-Initiated Communication

A. Uplink

We assume that transmissions are scheduled using TDD. Each TDD frame is divided into an UL and a DL phase, with each phase lasting for n channel uses. We assume that U single-antenna UEs are simultaneously active and that the BS has B antennas. The UEs initiate the transmission by sending orthogonal pilot sequences consisting of $U \leq n_\mathrm{p} < n$ symbols, each of power ρ_ul. Once the training phase is over, the UEs transmit coded data on the remaining $n - n_\mathrm{p}$ channel uses.

The received signal corresponding to the kth transmitted data symbols from the U UEs is

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{z}_k, \quad k = 1, \ldots, n - n_\mathrm{p}. \quad (6)$$

Here, $\mathbf{x}_k \sim \mathcal{CN}(\mathbf{0}_U, \rho_\mathrm{ul}\mathbf{I}_U)$ denotes the transmitted symbols from all UEs at time k, $\mathbf{H} \in \mathbb{C}^{B\times U}$ is the fading matrix, which is random but remains constant over the TDD frame, and $\mathbf{z}_k \sim \mathcal{CN}(\mathbf{0}_B, \mathbf{I}_B)$ is the AWGN at the BS. For sake of simplicity, we assume that the entries of $\mathbf{H}$ are drawn independently from a $\mathcal{CN}(0,1)$ distribution. However, our framework is general, and can be readily applied to arbitrary fading distributions.

The BS uses the n_p pilot symbols to estimate the channel matrix. Throughout the paper, we focus on MMSE channel estimation, which results in the estimate

$$\widehat{\mathbf{H}} = \frac{\sqrt{n_\mathrm{p}\rho_\mathrm{ul}}}{1+n_\mathrm{p}\rho_\mathrm{ul}}\left(\sqrt{n_\mathrm{p}\rho_\mathrm{ul}}\,\mathbf{H} + \mathbf{Z}\right). \quad (7)$$

Here, $\mathbf{Z}$ is a $B \times U$ matrix with i.i.d. $\mathcal{CN}(0,1)$ entries, which captures the impact of the additive noise on the channel estimate.

Next, the BS uses $\widehat{\mathbf{H}}$ to construct a $B \times U$ linear combiner (e.g., a maximum-ratio combiner) $\mathbf{W}$ that is used to separate the signals from the U users. Specifically, the output of the combiner corresponding to the signal transmitted by UE u at time k is

$$r_k^{(u)} = \mathbf{w}_u^\mathsf{H}\mathbf{h}_u x_k^{(u)} + \sum_{u'\neq u}\mathbf{w}_u^\mathsf{H}\mathbf{h}_{u'}x_k^{(u')} + \mathbf{w}_u^\mathsf{H}\mathbf{z}_k. \quad (8)$$

Here, $\mathbf{w}_u$ and $\mathbf{h}_u$ denote the uth column of the matrix $\mathbf{W}$ and $\mathbf{H}$, respectively. Furthermore, $x_k^{(u)}$ stands for the uth entry of the vector $\mathbf{x}_k$. Note that the first term in (8) corresponds to the desired signal from UE u, the second term is the residual multiuser interference after linear combining, and the third term is due to additive noise. Furthermore, note that (8) is structurally similar to (1): just set $v_k = r_k^{(u)}$, $t_k = x_k^{(u)}$, $g = \mathbf{w}_u^\mathsf{H}\mathbf{h}_u$, and $w_k = \sum_{u'\neq u}\mathbf{w}_u^\mathsf{H}\mathbf{h}_{u'}x_k^{(u')} + \mathbf{w}_u^\mathsf{H}\mathbf{z}_k$.

We assume that the BS decodes the message from each UE separately (no joint decoding). Furthermore, we assume that the BS treats the acquired channel estimate as perfect, and the residual multiuser interference as additive noise. In the notation introduced in Section II, this corresponds to performing mismatched SNN decoding with $\hat{g} = \mathbf{w}_u^\mathsf{H}\widehat{\mathbf{h}}_u$. It follows that the error probability bound (4) applies to this setup, once the substitutions described above are performed, and after taking an additional expectation over $\mathbf{H}$ and over $\mathbf{Z}$ in (7). Indeed, different from the setup in Section II, the channel is now random.

B. Downlink

In the DL phase, the BS multiplies the U-dimensional symbol vector $\mathbf{x}_k \sim \mathcal{CN}(\mathbf{0}_U, \rho_\mathrm{dl}\mathbf{I}_U)$ at time k by the $B \times U$ linear precoding matrix $\mathbf{P}$, constructed on the basis of the channel estimate $\widehat{\mathbf{H}}$ obtained in the UL phase. We assume that each column of $\mathbf{P}$ is normalized so that the expected value of its ℓ_2 norm is 1. The received signal at UE u corresponding to the kth transmitted data vector from the BS is

$$y_k^{(u)} = \mathbf{h}_u^\mathsf{T}\mathbf{p}_u x_k^{(u)} + \sum_{u'\neq u}\mathbf{h}_u^\mathsf{T}\mathbf{p}_{u'}x_k^{(u')} + z_k^{(u)}. \quad (9)$$

Here, $\mathbf{p}_u$ denotes the uth column of the linear precoding matrix $\mathbf{P}$ and $z_k^{(u)} \sim \mathcal{CN}(0,1)$ denotes the AWGN at UE u. Similar to (8), the first term in (9) corresponds to the desired signal from the BS while the second term contains the residual multiuser interference after linear precoding. Again, we can put (9) in the form given in (1) by setting $v_k = y_k^{(u)}$, $t_k = x_k^{(u)}$, $g = \mathbf{h}_u^\mathsf{T}\mathbf{p}_u$, and $w_k = \sum_{u'\neq u}\mathbf{h}_u^\mathsf{T}\mathbf{p}_{u'}x_k^{(u')} + z_k^{(u)}$.

We assume that each UE performs mismatched SNN decoding where the multiuser interference is treated as noise. Since no pilot symbols are transmitted in the downlink phase, no knowledge of the channel gain $g = \mathbf{h}_u^\mathsf{T}\mathbf{p}_u$ is available at the UEs. We assume, however, that each UE has some statistical knowledge of the channel; specifically, as commonly done in the massive MIMO literature, we assume that each UE knows the mean $\mathbb{E}\left[\mathbf{h}_u^\mathsf{T}\mathbf{p}_u\right]$ of the channel gain and uses this quantity to perform mismatched SNN decoding. Specifically, we set $\hat{g} = \mathbb{E}\left[\mathbf{h}_u^\mathsf{T}\mathbf{p}_u\right]$. Obvioulsy, channel hardening is critical for this choice to result in good performance.

The error probability in the DL can be readily evaluated using (4) after taking an additional expectation over $\mathbf{H}$ and over the matrix $\mathbf{Z}$ in (7).

IV. BS-Initiated Communication

In this section, we consider a BS-initiated bidirectional communication link. In the DL, the BS needs to deliver a common message to all UEs. Each UE then replies individually with a potentially distinct message. We will focus in this section exclusively on the first DL phase, since the UL phase is similar to the one described in Section III-A.

The initial DL phase is challenging, since no CSI is available to the BS. Hence, no beamforming is possible. This means that the spatial diversity required to achieve the target reliability needs to be provided through the use of space-time codes. Furthermore, the UEs cannot rely on channel hardening in the decoding process, and instead need to estimate (implicitly or explicitly) the fading channel. In this section, we consider explicit channel estimation based on downlink pilot symbols.

As noted previously in the massive MIMO literature (see [7] and references therein), it is not feasible to transmit orthogonal pilot sequences from all available antennas. Indeed, in short-packet transmissions, the blocklength n may be of the same order as the number of available antennas B, which makes orthogonal pilot transmission from all antennas unattractive since too few resources would be left for the transmission of the data symbols.

Following the strategy in [7], [18], we assume that the BS relies on an OSTBC that uses only B' of the B available BS antennas. Our aim is to use the error probability bound (4) in order to characterize the trade-off between B' and n_p. On the one hand, increasing B' results in more spatial diversity, which lowers the error probability; on the other hand, increasing B' results in an increased pilot overhead, which yields to a reduction of the number of symbols that can be used for data transmission.

To adapt (4) to the scenario just described, we will use the OSTBC to space-time encode the coded symbols generated by a Gaussian random code. Then, we will apply the error probability bound in (4) to characterize the performance achievable using downlink pilot transmission and mismatched SNN decoding.

We assume that the OSTBC produces matrix-valued symbols $\mathbf{X} \in \mathbb{C}^{B' \times n_\mathrm{c}}$, each one encoding $n_\mathrm{s} \leq n_\mathrm{c}$ complex-valued input symbols $\{q_i\}_{i=1}^{n_\mathrm{s}}$ generated independently from a $\mathcal{CN}(0, \rho_\mathrm{dl}/B')$ distribution. Each OSTBC codeword is transmitted over n_c channel uses and across B' antennas where $B' \ll B$. The rate of the OSTBC is given by $R_\mathrm{ostbc} = n_\mathrm{s}/n_\mathrm{c}$. We follow [19] and express each OSTBC symbol $\mathbf{X}$ as

$$\mathbf{X} = \sum_{i=1}^{n_\mathrm{s}} \Re(q_i)\mathbf{A}_i + j\Im(q_i)\mathbf{B}_i. \tag{10}$$

The orthogonality assumption implies that

$$\mathbb{E}\left[\mathbf{X}\mathbf{X}^\mathsf{H}\right] = \frac{n_\mathrm{c}\rho_\mathrm{dl}}{B'}\mathbf{I}_{B'}. \tag{11}$$

Hence, ρ_dl can be thought of as the total transmit power in each time instant. Similar to [7], for $B' = 4$, we choose $\{\mathbf{A}_i, \mathbf{B}_i\}$ so that the resulting OSTBC is the one given in [19, Example 7.4] and for larger values of B' we construct $\{\mathbf{A}_i, \mathbf{B}_i\}$ following the procedure outlined in [20] (although a higher-rate OSTBC might be available).

A dimension-reducing matrix $\mathbf{U}$ of size $B \times B'$ is used to map each OSTBC symbol $\mathbf{X}$ to the B BS antennas. For simplicity, we assume that this matrix is obtained by eliminating the last $B - B'$ columns of a randomly generated unitary matrix of dimension $B \times B$.[2]

The DL transmission consists of a training phase and a data phase. In the training phase, orthogonal pilot sequences of length $n_\mathrm{p} \geq B'$ are transmitted from each BS antenna. These pilot sequences are used at each UE to estimate the effective channel $\mathbf{h}_u^{(\mathrm{eff})} = \mathbf{U}^\mathsf{T}\mathbf{h}_u$, where $\mathbf{h}_u \in \mathbb{C}^B$ denotes the channel from the B antennas at the BS to UE u, and $\mathbf{h}_u^{(\mathrm{eff})} \in \mathbb{C}^{B'}$. We assume that channel estimation is performed using the MMSE principle (see (7)), and denote by $\widehat{\mathbf{h}}_u^{(\mathrm{eff})} \in \mathbb{C}^{B'}$ the MMSE estimate.

In the data phase, ℓ space-time-coded symbols are transmitted from the BS. We assume that the overall DL phase lasts at most n channel uses. Hence, for a given choice of the OSTBC, the integers n_p and ℓ need to be chosen such that $n_\mathrm{p} + \ell n_\mathrm{c} \leq n$. The received signal at the uth UE corresponding to the kth OSTBC symbol $\mathbf{X}_k$, $k = 1, \ldots, \ell$, is given by

$$\mathbf{y}_k^{(u)} = \mathbf{h}_u^\mathsf{T}\mathbf{U}\mathbf{X}_k + \mathbf{z}_k^{(u)} = (\mathbf{h}_u^\mathrm{eff})^\mathsf{T}\mathbf{X}_k + \mathbf{z}_k^{(u)}. \tag{12}$$

Here, $\mathbf{y}_k^{(u)}$ is an n_c-dimensional vector, and the additive noise is denoted by $\mathbf{z}_k^{(u)} \sim \mathcal{CN}(\mathbf{0}_{n_\mathrm{c}}, \mathbf{I}_{n_\mathrm{c}})$.

We assume that the uth UE obtains an estimate $r_{k,i}^{(u)}$ of the ith coded symbol $q_{k,i}$ transmitted on the kth OSTBC symbol as follows [19]:

$$\begin{aligned} r_{k,i}^{(u)} = \Re\left(\frac{(\widehat{\mathbf{h}}_u^\mathrm{eff})^\mathsf{T}\mathbf{A}_i}{\|\widehat{\mathbf{h}}_u^\mathrm{eff}\|}(\mathbf{y}_k^{(u)})^\mathsf{H}\right) \\ + j\Im\left(\frac{(\widehat{\mathbf{h}}_u^\mathrm{eff})^\mathsf{T}\mathbf{B}_i}{\|\widehat{\mathbf{h}}_u^\mathrm{eff}\|}(\mathbf{y}_k^{(u)})^\mathsf{H}\right). \end{aligned} \tag{13}$$

Then mismatched SNN decoding on the basis of $\widehat{\mathbf{h}}_u^\mathrm{eff}$ is performed with $\hat{g} = \|\widehat{\mathbf{h}}_u^\mathrm{eff}\|$, channel inputs given by the transmitted symbols $\{q_{k,i}\}$, and channel outputs given by the corresponding estimates $\{r_{k,i}^{(u)}\}$. We refer the reader to [7] for a decomposition of (13) into useful-signal part and intersymbol-interference terms that result from channel-estimation errors.

V. Numerical Results

In this section, we consider a scenario where $B = 100$ and $U = 10$. Furthermore, we assume that $n = 288$ and that $\log_2 M = 30$, i.e., each message consists of 30 bits. These two values are in agreement with the so-called compact downlink control information transmission scenario [21]. We assume, throughout this section, an i.i.d. Rayleigh fading channel.

First, we consider the UE-initiated-transmission scenario and assume a target error probability of $\epsilon = 10^{-5}$ on the bi-directional link. To satisfy the error-probability target, we require $\epsilon_\mathrm{ul} = \epsilon_\mathrm{dl} = \epsilon/2$, where ϵ_ul and ϵ_dl denote the error probability on the UL and DL, respectively.

In Fig. 1, we illustrate the minimum SNR (obtained via (4)) required for both the UL and the DL transmission to achieve ϵ_ul and ϵ_dl, respectively. In the figure, we assume maximum-ratio combining and maximum-ratio precoding. We see that the UL SNR decreases as the number of pilot symbols increase up until $n_\mathrm{p} = 100$. For $n_\mathrm{p} > 100$, the required SNR increases because the channel estimation overhead offsets the performance gain resulting from a more accurate channel estimate. Not surprisingly, the picture is different for the DL. Since pilot overhead penalizes only the UL, the downlink SNR decreases as the number of pilot symbols increase. We also see from the figure that the optimum number of pilot symbols that minimizes the total SNR is $n_\mathrm{p} = 150$.

Next, we consider the BS-initiated-transmission scenario. In Fig. 2, we illustrate the minimum SNR (obtained via (4)) required

[2]Better designs may be possible, especially if information about the statistical properties of the propagation channel is available at the BS.

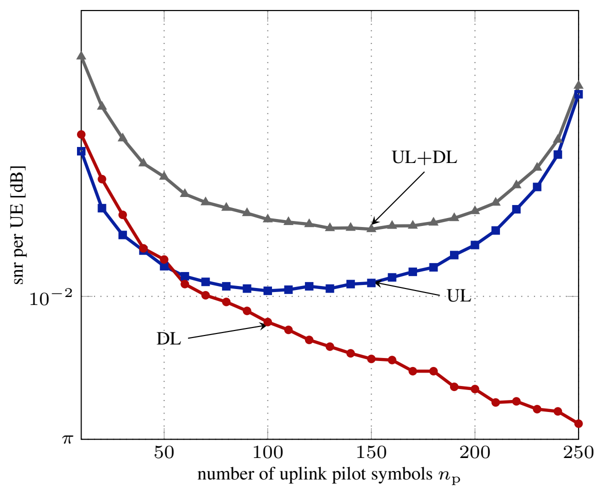

Fig. 1. The minimum SNR required in the UL and in the DL to achieve $\epsilon_{\mathrm{ul}} = \epsilon/2$ and $\epsilon_{\mathrm{dl}} = \epsilon/2$, respectively. Here, $\epsilon = 10^{-5}$, $B = 100$, $U = 10$, $n = 288$, and $\log_2 M = 30$. The propagation channel is modeled as i.i.d. spatially-white Rayleigh fading.

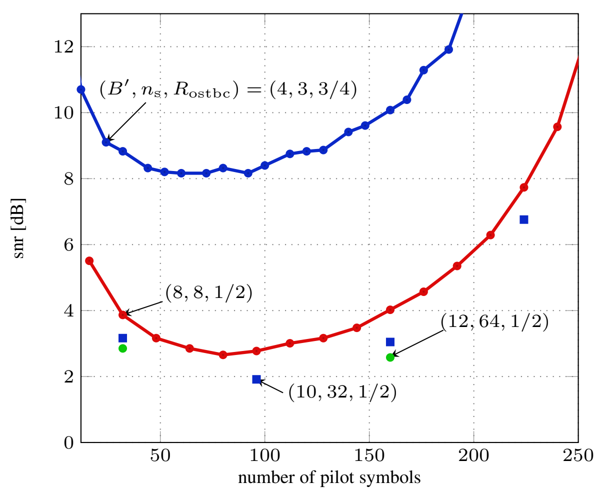

Fig. 2. The minimum SNR required for $B = 100$ and $n = 288$ to achieve $\epsilon_{\mathrm{ostbc}} = 10^{-5}$ as a function of the number of pilot symbols n_{p} per antenna, for four different OSTBCs. The propagation channel is modeled as i.i.d. spatially-white Rayleigh fading.

to achieve $\epsilon = 10^{-5}$, for four different OSTBCs as a function of the number of pilot symbols transmitted from each active BS antenna. It can be seen that setting $B' = 4$ results in a high required SNR, because this space-time code offers very limited spatial diversity. As we increase B' to 10, the required minimum SNR can be reduced by about $6.3\,\mathrm{dB}$. Increasing B' further is not helpful because of the channel-estimation overhead. In particular, we see from the figure that, as the number of active antennas increases, setting n_{p} appropriately is critical. For example, for $B' = 10$, the required SNR to achieve 10^{-5} is about $1.9\,\mathrm{dB}$ when $n_{\mathrm{p}} = 96$ but $6.7\,\mathrm{dB}$ when $n_{\mathrm{p}} = 224$.

VI. Conclusion

We have presented a framework based on finite-blocklength information theory that is suitable for determining the error probability achievable on massive MIMO links in URLLC scenarios. Through numerical simulations involving bi-directional UE-initiated and BS-initiated communication links, we have illustrated how to use this framework to optimize the number of pilot symbols to minimize the transmit power given a reliability and a latency constraint.

References

[1] 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.

[2] A. Zaidi, F. Athley, J. Medbo, U. Gustavsson, G. Durisi, and X. Cheng, *5G Physical Layer: principles, models and technology components*. London, U.K.: Academic Press, 2018.

[3] P. Popovski, Č. Stefanović, J. J. Nielsen, E. de Carvalho, M. Angjelichinoski, K. F. Trillingsgaard, and A.-S. Bana, "Wireless access in ultra-reliable low-latency communication (URLLC)," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5783–5801, Aug. 2019.

[4] N. A. Johansson, Y.-P. E. Wang, E. Eriksson, and M. Hessler, "Radio access for ultra-reliable and low-latency 5G communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 1184–1189.

[5] A. Bana, G. Xu, E. D. Carvalho, and P. Popovski, "Ultra Reliable Low Latency Communications in Massive Multi-Antenna Systems," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Oct. 2018, pp. 188–192.

[6] A.-S. Bana, E. de Carvalho, B. Soret, T. Abrão, J. C. Marinello, E. G. Larsson, and P. Popovski, "Massive MIMO for Internet of Things (IoT) Connectivity," *arXiv:1905.06205 [cs, math]*, May 2019.

[7] M. Karlsson, E. Björnsson, and E. G. Larsson, "Performance of in-band transmission of system information in massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1700–1712, Mar. 2018.

[8] G. Durisi, T. Koch, J. Östman, 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.

[9] E. Björnsson, J. Hoydis, and L. Sanguinetti, ser. Foundations and Trends in Signal Processing. Delft, The Netherlands: now Publishers, 2019, vol. 11, no. 3-4.

[10] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need 'perfect side information' be?" *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[11] 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, Jul. 2014.

[12] ——, "Quasi-static SIMO fading channels at finite blocklength," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 1531–1535.

[13] E. Biglieri, J. G. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[14] A. Martinez and A. Guillén i Fàbregas, "Saddlepoint approximation of random–coding bounds," in *Proc. Inf. Theory Applicat. Workshop (ITA)*, San Diego, CA, USA, Feb. 2011, pp. 1–6.

[15] G. C. Ferrante, J. Östman, G. Durisi, and K. Kittichokechai, "Pilot-assisted short-packet transmission over multiantenna fading channels: A 5G case study," in *Conf. Inf. Sci. Sys. (CISS)*, New Jersey, USA, 2018, pp. 1–6.

[16] J. Scarlett, V. Y. F. Tan, and G. Durisi, "The dispersion of nearest-neighbor decoding for additive non-Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 81–92, Jan. 2017.

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

[18] X. Meng, X. Gao, and X.-G. Xia, "Omnidirectional precoding based transmission in massive MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, Jan. 2016.

[19] E. G. Larsson and P. Stoica, *Space-time block coding for wireless communications*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[20] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.

[21] "R1-1720997: On PDCCH for ultra-reliable transmission," 3GPP RAN1#91, 2017.