Low-Complexity Detection of Uplink NOMA by Exploiting Properties of the Propagation Channel

Bashar Tahir†, Stefan Schwarz†, and Markus Rupp
Institute of Telecommunications, Technische Universität (TU) Wien, Vienna, Austria

***Abstract*—Uplink non-orthogonal multiple access (NOMA) has been proposed as an efficient technique to support massive connectivity and reduce access-latency. However, due to the inherent multiuser interference within such a system, iterative joint detection is required, which is of high-complexity. In this paper, we exploit the propagation properties of wireless channels to reduce the detection complexity. In particular, when neighboring spreading-blocks on the time-frequency grid experience similar channel conditions, then it is possible to reuse the calculated filter weights between them. We propose four detection strategies and compare them across a wide range of time- and frequency-selectively. Then, assuming the base station is equipped with a sufficient number of antennas, we replace the MMSE filter with a lower-complexity approximation using Neumann series expansion. The results show that our strategies incur only a small performance loss, while substantially cutting down complexity.**

I. Introduction

Massive connectivity and low-latency operation are main drivers for the 5th generation (5G) and future wireless systems [1]. Example applications utilizing such capabilities, are massive machine-type communications [2], and cellular-assisted vehicular communications [3], [4]. Supporting such large number of devices comes at the cost of high signaling overhead at the cellular network [5]. For current systems, e.g., long-term evolution (LTE), the devices go through a schedule-grant procedure [6], in which they first request a permission to transmit, and then are only able to do so once the base station (BS) assign them certain time-frequency resources in an orthogonal manner, i.e., they are allocated separate resources. Such handshake procedure can be highly inefficient when the devices data-payload is comparable in size to the required control signaling. This is a possible case if we consider machine-type communications, e.g., internet-of-things (IoT) [5]. Moreover, due to the large number of devices, processing those requests might introduce an extra delay.

Uplink non-orthogonal multiple access (NOMA) is an attractive solution to tackle those issues [7], [8]. It allows the devices to access the network in a grant-free manner, in which the BS does not have to coordinate the transmission of each device. Rather, the devices pick time-frequency resources on their own and transmit immediately. Since the transmissions are not coordinated, multiple devices might contest the same resources, leading to collisions. A class of uplink NOMA schemes are those that operate in the code-domain, such as those using short spreading signatures at the symbol-level [9]. They have been shown to achieve high overloading rate and are robust to collisions, especially when combined with multiple receive antennas at the BS. However, since the devices share the same resources, joint detection via iterative receivers is required in order to achieve a good performance.

In this work, we utilize the time-frequency correlation of wireless channels to reduce the detection complexity of uplink NOMA. We make use of the fact that neighboring spreading-blocks on the time-frequency grid exhibit similar propagation conditions, and thus the filters used in the detection process can be reused between those blocks. Our proposed detection strategies are applicable to all uplink NOMA schemes that perform linear filtering, e.g., minimum mean square error (MMSE) filtering, during the iterative detection. Moreover, given that the BS is equipped with a sufficient number of receive antennas, we replace the MMSE filter with a lower-complexity approximate filter using a first-order Neumann series expansion in a fashion similar to [10], and show that it performs very close to the exact filter.

II. System Model

We consider an uplink in which K devices are active in a synchronous operation using orthogonal frequency-division multiplexing (OFDM). The synchronization could be achieved via a downlink broadcast signal transmitted by the BSs. Each device spreads its transmit symbols by a spreading signature of length L, which is then mapped onto the time-frequency grid. The devices pick their spreading signature on their own, and thus collisions could occur. After applying the inverse fast-Fourier-transform (FFT) at the devices and the FFT at the BS, the received signal at the i^{th} spreading-block is given by

$$\mathbf{y}_i = \sum_{k=1}^{K} \sqrt{P_k L}\, (\mathbf{h}_{k,i} \circ \mathbf{s}_k)\, x_{k,i} + \mathbf{n}_i,\ i = 1, 2, \ldots, M, \quad (1)$$

where

- $P_k \in \mathbb{R}^+$ is the average transmit power of the k^{th} device,
- $\mathbf{h}_{k,i} \in \mathbb{C}^{L\times 1}$ is the fading coefficients vector of the k^{th} device at the i^{th} the spreading-block,
- $\mathbf{s}_k \in \mathbb{C}^{L\times 1}$ is the unit-norm spreading signature of the k^{th} device,
- $\circ$ denotes the element-wise (Hadamard) product,
- $x_{k,i} \in \mathbb{C}$ is the i^{th} transmit symbol of the k^{th} device,
- $\mathbf{n}_i \in \mathbb{C}^{L\times 1}$ is the Gaussian noise vector at the i^{th} block,
- and M is the total number of the transmit symbols.

†Bashar Tahir and Stefan Schwarz are with the Christian Doppler Laboratory for Dependable Wireless Connectivity for the Society in Motion.

The financial support by the Austrian Federal Ministry for Digital and Economic Affairs and the National Foundation for Research, Technology and Development is gratefully acknowledged.

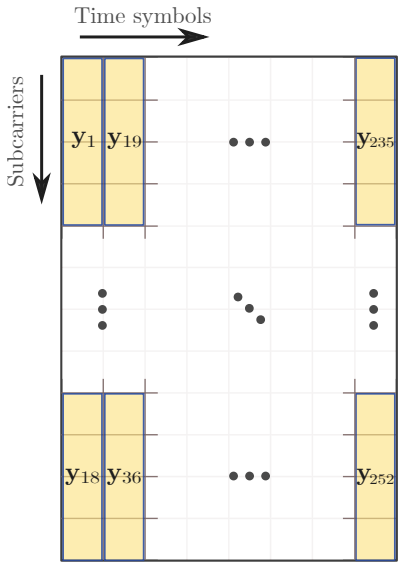

Fig. 1: Mapping of the spreading-blocks for 72 subcarriers and 14 time (OFDM) symbols, with a spreading length of $L = 4$.

The spreading-block mapping is elaborated in Figure 1, where the spreading is performed across frequency, such that the device maps the spreading-blocks into groups of L subcarriers. The block index i spans the OFDM symbols one after another. In other words, if the total number of allocated subcarriers is 72 and the spreading length is 4, then every OFDM symbol will contain $72/4 = 18$ spreading-blocks. Block $\mathbf{y}_1$ would be the first block in the first OFDM symbol, and block $\mathbf{y}_{19}$ would be the first block in the second OFDM symbol, etc, as shown in the figure.

Let the BS be equipped with R receive antennas, and denote by $\mathbf{y}_i^{(r)}$ the i^{th} received block at the r^{th} antenna, then from (1) we have

$$\begin{bmatrix} \mathbf{y}_i^{(1)} \\ \mathbf{y}_i^{(2)} \\ \vdots \\ \mathbf{y}_i^{(R)} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{K} \sqrt{P_k L}\ (\mathbf{h}_{k,i}^{(1)} \circ \mathbf{s}_k)\, x_{k,i} \\ \sum_{k=1}^{K} \sqrt{P_k L}\ (\mathbf{h}_{k,i}^{(2)} \circ \mathbf{s}_k)\, x_{k,i} \\ \vdots \\ \sum_{k=1}^{K} \sqrt{P_k L}\ (\mathbf{h}_{k,i}^{(R)} \circ \mathbf{s}_k)\, x_{k,i} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_i^{(1)} \\ \mathbf{n}_i^{(2)} \\ \vdots \\ \mathbf{n}_i^{(R)} \end{bmatrix}, \tag{2}$$

where $\mathbf{h}_{k,i}^{(r)}$ and $\mathbf{n}_i^{(r)}$ now have a superscript index r indicating the corresponding receive antenna. Let $\mathbf{g}_{k,i}^{(r)} = (\mathbf{h}_{k,i}^{(r)} \circ \mathbf{s}_k)$ be the effective signature at antenna r, we develop (2) as follows

$$\underbrace{\begin{bmatrix} \mathbf{y}_i^{(1)} \\ \mathbf{y}_i^{(2)} \\ \vdots \\ \mathbf{y}_i^{(R)} \end{bmatrix}}_{\widetilde{\mathbf{y}}_i} = \underbrace{\begin{bmatrix} \mathbf{g}_{1,i}^{(1)} & \mathbf{g}_{2,i}^{(1)} & \cdots & \mathbf{g}_{K,i}^{(1)} \\ \mathbf{g}_{1,i}^{(2)} & \mathbf{g}_{2,i}^{(2)} & \cdots & \mathbf{g}_{K,i}^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{g}_{1,i}^{(R)} & \mathbf{g}_{2,i}^{(R)} & \cdots & \mathbf{g}_{K,i}^{(R)} \end{bmatrix}}_{\mathbf{G}_i} \underbrace{\begin{bmatrix} \sqrt{P_1 L} x_{1,i} \\ \sqrt{P_2 L} x_{2,i} \\ \vdots \\ \sqrt{P_K L} x_{K,i} \end{bmatrix}}_{\mathbf{P}^{1/2}\mathbf{x}_i} + \underbrace{\begin{bmatrix} \mathbf{n}_i^{(1)} \\ \mathbf{n}_i^{(2)} \\ \vdots \\ \mathbf{n}_i^{(R)} \end{bmatrix}}_{\widetilde{\mathbf{n}}_i},$$

where $\mathbf{P}$ is a diagonal matrix defined as

$$\mathbf{P} = \text{diag}(P_1 L,\ P_2 L,\ \ldots\ ,\ P_K L)\,,$$

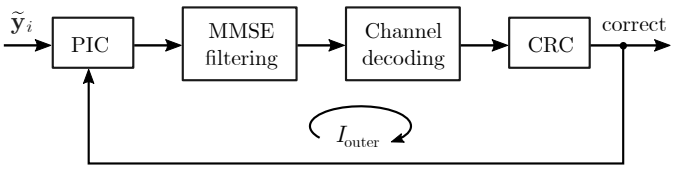

Fig. 2: Iterative CWL-MMSE-PIC receiver.

and $\mathbf{x}_i$ is the transmit symbols vector defined as

$$\mathbf{x}_i = [x_{1,i},\ x_{2,i},\ \ \ldots\ \ , x_{K,i}]^T\,.$$

Finally, we can write (2) as

$$\begin{aligned} \widetilde{\mathbf{y}}_i &= \mathbf{G}_i \mathbf{P}^{1/2} \mathbf{x}_i + \widetilde{\mathbf{n}}_i \\ &= \widetilde{\mathbf{G}}_i \mathbf{x}_i + \widetilde{\mathbf{n}}_i\ . \end{aligned} \tag{3}$$

In order to perform multiuser detection (MUD), the BS employs an iterative codeword level (CWL) MMSE receiver with parallel interference cancellation (PIC). The interference cancellation we consider here is hard, i.e., only the decoded devices that pass the cyclic-redundancy-check (CRC) are canceled in the next iteration. Figure 2 shows the block digram of the MUD receiver, where I_{outer} denotes the maximum number of PIC iterations.

At every iteration, the BS calculates an MMSE weight for each spreading-block i, $i = 1, 2, \ldots, M$, i.e.,

$$\mathbf{x}_i^{\text{MMSE}} = \widetilde{\mathbf{G}}_i^H (\widetilde{\mathbf{G}}_i \widetilde{\mathbf{G}}_i^H + \sigma_{\mathbf{n}}^2 \mathbf{I})^{-1}\ \widetilde{\mathbf{y}}_i\ , \tag{4}$$

where $\sigma_{\mathbf{n}}^2$ is the noise variance, which is assumed to be equal across all spreading-blocks and receive antennas, and $\mathbf{I}$ is the identity matrix of size $LR \times LR$. Once the M symbol estimates of the entire subframe are obtained, channel decoding is performed and the resulting messages are checked for CRC, which is then followed by the cancellation step.

III. Reducing Detection Complexity by Exploiting Time-Frequency Correlation

A. Overview

In the following, we adopt a structure of resources that is similar to LTE. The time-frequency grid is divided into resource-blocks (RBs). Each RB consists of 12 subcarriers and 7 symbols. A device can have an allocation of multiple RBs. The seven time symbols of the RBs form a slot, and two slots together form a subframe. We adopt a spreading length of $L = 4$, and similarly to the previous section, we assume that the spreading is carried out in the frequency direction. Therefore, a RB would consist of $12 \times 7/4 = 21$ spreading-blocks. Assuming that the whole subframe is used for data transmission (no control signaling), the BS needs to calculate an MMSE weight 21 times in every RB, since there are 21 spreading-blocks in each RB. However, by inspecting $\widetilde{\mathbf{G}}_i$, we notice that the only changing quantity from one spreading-block to another are the channel fading coefficients $\mathbf{h}_{k,i}^{(r)}$. The spreading signature $\mathbf{s}_k$ and the transmit power P_k are fixed during the entire subframe transmission. This motivates us to

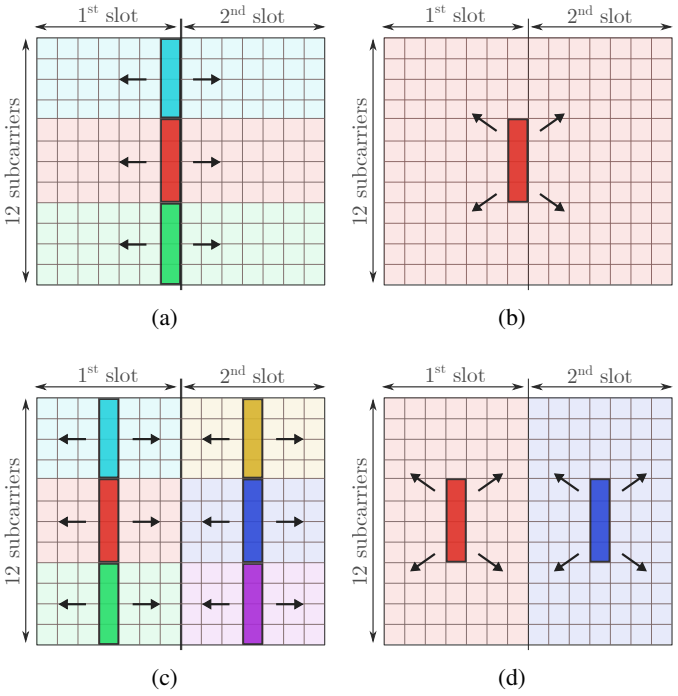

Fig. 3: Receiver strategies under consideration. The solid blocks indicate the place where the MMSE weights are calculated. The shades indicate where they are reused.

look into the dynamic behavior of $\mathbf{h}_{k,i}^{(r)}$, because if the whole subframe has a constant $\mathbf{h}_{k,i}^{(r)}$, then $\widetilde{\mathbf{G}}_i$ would be constant in i, and the BS would only need to calculate a single MMSE weight and reuse it for the entire subframe, reducing the computational complexity.

For wireless transmission, multipath propagation induces delay dispersion between the time samples of the signal, which translates into the channel being selective (varying) in frequency. Motion, on the other hand, induces Doppler spread across the subcarriers, and this translates into the channel being selective in time. Therefore, $\widetilde{\mathbf{G}}_i$, in general, will vary from one i to another. However, unless the delay spread is very severe and the devices are moving at very high velocities, the wireless channel will exhibit correlation in time and frequency.

Let us analyze the effect of reusing weights on the post-filtering signal-to-interference-plus-noise ratio (SINR). Consider the detection of spreading-block i using a weight calculated at another block j. Let $\widetilde{\mathbf{g}}_{k,i}$ be the k^{th} column of $\widetilde{\mathbf{G}}_i$, then from (4), the post-filtering SINR of the k^{th} device at block i is given by

$$\text{SINR}_{k,i}^{\text{MMSE}} = \frac{\left|\widetilde{\mathbf{g}}_{k,j}^H \mathbf{Z}_j^{-1} \widetilde{\mathbf{g}}_{k,i}\right|^2}{\sum_{l\neq k}\left|\widetilde{\mathbf{g}}_{k,j}^H \mathbf{Z}_j^{-1} \widetilde{\mathbf{g}}_{l,i}\right|^2 + \sigma_{\mathbf{n}}^2 \left\|\mathbf{Z}_j^{-1}\widetilde{\mathbf{g}}_{k,j}\right\|^2} , \quad (5)$$

where

$$\mathbf{Z}_j = \widetilde{\mathbf{G}}_j \widetilde{\mathbf{G}}_j^H + \sigma_{\mathbf{n}}^2 \mathbf{I} . \quad (6)$$

Under the assumption that the fading is independent across the different devices, the interference power term in (5) will not be affected by the choice of j with respect to i. The noise power term is also unaffected by the choice of j. Therefore, we only need to consider the desired signal power term. Since $\mathbf{Z}_j^{-1}$ is Hermitian positive semi-definite, we can apply the Cholesky decomposition $\mathbf{Z}_j^{-1} = \mathbf{U}_j^H \mathbf{U}_j$, where $\mathbf{U}_j$ is an upper-triangular matrix. The nominator of (5) becomes

$$\left|\widetilde{\mathbf{g}}_{k,j}^H \mathbf{Z}_j^{-1} \widetilde{\mathbf{g}}_{k,i}\right|^2 = \left|\widetilde{\mathbf{g}}_{k,j}^H \mathbf{U}_j^H \mathbf{U}_j \widetilde{\mathbf{g}}_{k,i}\right|^2 . \quad (7)$$

Using Cauchy-Schwarz inequality, we formulate an upper bound on the desired signal power as follows

$$\left|\widetilde{\mathbf{g}}_{k,j}^H \mathbf{U}_j^H \mathbf{U}_j \widetilde{\mathbf{g}}_{k,i}\right|^2 \leq \left\|\mathbf{U}_j \widetilde{\mathbf{g}}_{k,j}\right\|_2^2 \left\|\mathbf{U}_j \widetilde{\mathbf{g}}_{k,i}\right\|_2^2 , \quad (8)$$

with the maximum achieved in the case of proportionality, i.e., for some constant α, we have

$$\begin{aligned} \mathbf{U}_j \widetilde{\mathbf{g}}_{k,j} &= \alpha \, \mathbf{U}_j \widetilde{\mathbf{g}}_{k,i} \\ \widetilde{\mathbf{g}}_{k,j} &= \alpha \, \widetilde{\mathbf{g}}_{k,i} , \end{aligned} \quad (9)$$

where in the last step, the inverse of $\mathbf{U}_j$ was applied to both sides, as $\mathbf{U}_j$ is always invertible. One maximizer is the choice $j = i$, i.e., we filter block i using a weight calculated at block i, which is the trivial case. However, we can also see that as long as $\widetilde{\mathbf{g}}_{k,j}$ is similar to $\widetilde{\mathbf{g}}_{k,i}$, that is, they are correlated, then the loss in the post-filtering SINR might be acceptable. In general, we would like to choose an optimal j^*, such that the minimum post-filtering SINR across all the blocks and all devices, is maximized, i.e.,

$$j^* = \underset{j}{\text{argmax}} \min_{k,i} \left|\widetilde{\mathbf{g}}_{k,j}^H \mathbf{Z}_j^{-1} \widetilde{\mathbf{g}}_{k,i}\right|^2 , \quad (10)$$

which is difficult to solve, especially when considering correlated fading. Of course, we can also perform this search on a per-RB basis, instead of the entire subframe, which is then expected to perform better in the case of high selectivity. Still, it does not make it easier to solve. We thus turn our attention to low-complexity suboptimal strategies based on the correlation properties of the channels. We consider the following four strategies, which are depicted in Figure 3:

- Receiver (a): assumes the channels are highly selective in frequency, but correlated in time. It only calculates MMSE weights for the spreading-blocks in the middle of the subframe (7^{th} OFDM symbol), and then those weights are reused for the other time symbols.
- Receiver (b): assumes the channels are moderately selective in frequency, but correlated in time. It only calculates MMSE weights for the middle subcarrier group of each RB at the middle time symbol (7^{th} OFDM symbol).
- Receiver (c): assumes the channels are highly selective in frequency, but moderately selective in time. It calculates MMSE weights for all spreading-blocks of the 4^{th} and 11^{th} OFDM symbols. Then, for the first slot, the 4^{th} symbol weights are reused, and for the second slot, the weights from the 11^{th} symbol are reused.
- Receiver (d): assumes the channels are moderately selective in both frequency and time. It calculates a single MMSE weight in the middle of every RB, and then the entire RB is filtered with that weight.

B. Setup

We now put our different receiver strategies to test, and compare them against a full receiver that calculates MMSE weights for all spreading-blocks. The setup is as follows: $K = 12$ devices are simultaneously transmitting in a grant-free manner with a spreading length of $L = 4$. Since there is no coordination by the BS, collisions can occur. We design the spreading signatures such that the maximum cross-correlation between each other is minimized using the algorithm in [11]. More specifically, we construct a 4×16 Grassmannian codebook. The signature design can also be extended to multi-BS deployments as in [12]. The channel realizations were generated according to the tapped-delay-line-C (TDL-C) channel model [13] with Jakes spectrum using the Vienna 5G Link-Level Simulator[1] [14]. The devices transmit powers are chosen such that their average signal-to-noise ratios (SNRs) at the BS are uniformly distributed in the range $[1, 11]$ dB, that is, 6 dB plus a uniform spread of ± 5 dB. This would correspond to the difference in pathloss between the devices. The simulation parameters are summarized in Table I below.

All the results we show in this paper are accompanied by the performance of a single user occupying the time-frequency resources alone, detected without any of the approximations that we propose here. This represents the performance of a perfect orthogonal multiple access (OMA) spreading system that does not suffer from multiuser interference, and thus serves as a baseline to our NOMA results. Note that for $L = 4$, an OMA system can only support up to four devices.

Parameter	Value
# active devices K	12
Spreading length L	4
Collisions allowed	Yes
Center frequency	2 GHz
Subcarrier spacing	15 kHz
Allocated subcarriers	72
Transmission duration	2 slots (14 OFDM symbols)
# receive antennas R	2
Average SNR	6 dB
SNR spread	± 5 dB
Channel model	TDL-C with Jakes Doppler spectrum
Modulation	4-QAM
Channel coding	New-radio LDPC, code-rate $= 1/2$
# PIC iterations	$I_{\text{outer}} = 6$

TABLE I: Simulation parameters.

C. Performance over Delay Spread

First, we investigate the performance of such a system in terms of the average block error ratio (BLER) at various levels of the root-mean-square (RMS) delay spread. All the devices are moving at a fixed velocity of 50 km/h. The results are shown in Figure 4. We observe that receivers (b) and (d) exhibit a bad performance at very high delay spreads. This is expected since the MMSE weights for those receivers are only calculated at the middle subcarrier groups, and therefore when the channels are highly frequency-selective, the calculated weights fail to represent the other subcarrier groups. However, per 3GPP [13], 300 ns is already considered a long delay spread, and they declare the nominal delay spread to be 100 ns. Therefore, above 400 ns would represent extreme cases. Also, in the simulation, we set all the devices to have the same delay spread, i.e., the results represent a worse-case scenario. In practice, some devices will experience short delay spread, others will experience a long one. Based on that, we conclude that for moderate to long delay spreads, any of the receiver strategies is applicable with a small difference in performance compared to a full receiver that calculates an MMSE weight for every spreading-block. Also, we notice that receivers (a) and (b) have a slightly worse performance than (c) and (d) at low delay spreads. The difference is due to the devices moving at 50 km/h, i.e., the channels are time-variant, for which the later receivers are better suited, as we see next.

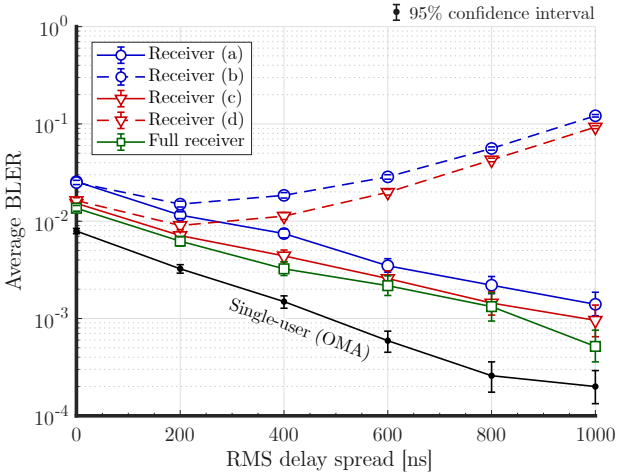

Fig. 4: Performance of the proposed receiver strategies at different levels of delay spread.

D. Performance over Velocity

Now, we fix the delay spread of the devices to be 300 ns, and we investigate the BLER performance at different velocities. The results are shown in Figure 5. We observe that receivers (c) and (d) are more robust to the time-selectively, especially at high velocities. This is due to the MMSE weights being calculated in each time-slot of the subframe, which allow them to better approximate the weights in each slot, compared to the case where a single weight is used to approximate the entire subframe (two slots) in the case of receivers (a) and (b).

E. Discussion

From the previous results, we propose employing receiver strategy (d). It exhibits only a small performance loss in moderate to long delay spreads, and it is very robust to high Doppler spreads. In terms of complexity saving, it calculates

[1] Available: https://www.nt.tuwien.ac.at/research/mobile-communications/vccs/vienna-5g-simulators/.

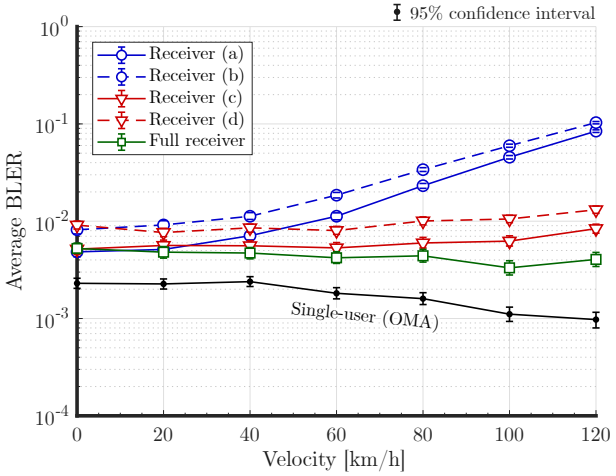

Fig. 5: Performance of the proposed receiver strategies at different velocities.

a single weight in the middle of each RB, which is then applied to the entire RB. Since each RB has 21 spreading-blocks as mentioned before, then receiver (d) would provide a complexity reduction of $21/1 = 21$ times, compared to a full blown receiver that calculates the MMSE weight at every spreading-block. If the environment is expected to be highly frequency- selective, then receiver (c) would be a better choice, but it will have a reduction in complexity of only $21/3 = 7$ times, compared to a full receiver.

It can also be noticed that in both figures, the performance of the full receiver gets better as the channels get more selective. The reason for that is due to the diversity gain harvested by the channel code at the bit-level [15], and also by the short spreading signatures when the channel becomes highly selective in the direction of spreading (frequency in our case). Moreover, in the case of practical channel estimation (e.g., with pilots), we expect the performance to become worse at high delay and Doppler spreads, due to the increase in channel estimation errors [15].

IV. Reducing Detection Complexity by Exploiting the Spatial Domain

Adding more receive antennas at the BS makes the columns of $\mathbf{G}_i$ in (3) longer, i.e., the effective signatures of the devices get longer. Combined with the assumption that the antennas experience sufficiently uncorrelated fading, those resultant signatures will in turn have a low cross-correlation, i.e., less interference between the devices. This behavior is well-known in the context of massive multiple-input multiple-output (MIMO), which results in linear receivers performing close to optimal, and even a matched filter (MF) can provide a good performance [10]. In our case, we do not have a massive MIMO assumption, and we employ iterative receivers. The question then is, as we add more antennas, do we really need to use an exact MMSE weight for the NOMA detection in (4), or is it sufficient to employ an approximation, such as a MF?

To answer that, consider the MMSE weight given by

$$\mathbf{W}_i^{\text{MMSE}} = \widetilde{\mathbf{G}}_i^H (\widetilde{\mathbf{G}}_i \widetilde{\mathbf{G}}_i^H + \sigma_{\mathbf{n}}^2 \mathbf{I})^{-1} \ .$$

Let $\mathbf{Z}_i = \widetilde{\mathbf{G}}_i \widetilde{\mathbf{G}}_i^H + \sigma_{\mathbf{n}}^2 \mathbf{I}$ as in (6). Using a Neumann series expansion, its inverse can be approximated as [10], [16], [17]

$$\mathbf{Z}_i^{-1} \approx \beta_i \sum_{n=0}^{N-1} (\mathbf{I} - \beta_i \mathbf{Z}_i)^n \ , \tag{11}$$

where N is the number of terms used for the approximation, and β_i is a parameter that should be chosen such that the sum converges as $N \to \infty$. For guaranteed convergence, the spectral radius of $\mathbf{I} - \beta_i \mathbf{Z}_i$ must satisfy [16]

$$\rho(\mathbf{I} - \beta_i \mathbf{Z}_i) < 1 \ , \tag{12}$$

where $\rho(.)$ denotes the spectral radius. Since $\mathbf{Z}_i$ is positive semi-definite, then condition (12) is equivalent to

$$|1 - \beta_i \lambda_{\max}(\mathbf{Z}_i)| < 1 \ , \tag{13}$$

where $\lambda_{\max}(\mathbf{Z}_i)$ denotes the maximum eigenvalue of $\mathbf{Z}_i$. Solving (13) yields the following range for β_i

$$0 < \beta_i < 2/\lambda_{\max}(\mathbf{Z}_i) \ . \tag{14}$$

The idea is to choose a β_i that results in a fast convergence, that is, with as few sum terms as possible. However, this can be computationally problematic, since it involves the calculation of eigenvalues. Therefore, we use a low-complexity approximation given by the trace, i.e., we choose β_i as

$$\beta_i = 2/\text{trace}(\mathbf{Z}_i) \ . \tag{15}$$

Now, if we take a zeroth-order approximation, we end up with the MF. Let us instead consider a first-order approximation

$$\mathbf{Z}_i^{-1} \approx \beta_i(\mathbf{I} + (\mathbf{I} - \beta_i \mathbf{Z}_i)) = \beta_i(2\mathbf{I} - \beta_i \mathbf{Z}_i) \ . \tag{16}$$

Our approximate filter then is given by

$$\mathbf{W}_i^{\text{Approx}} = \beta_i \widetilde{\mathbf{G}}_i^H (2\mathbf{I} - \beta_i \mathbf{Z}_i) \ . \tag{17}$$

Note that for modulation orders higher than 4-QAM, both the MF and approximate filter outputs need to be scaled down by the effective filter-channel gains to get correct amplitude scaling for the demapping operation. In our case, we only consider 4-QAM, and therefore we ignore the exact scaling.

We consider a simulation scenario with a similar setup as in Table 1. All the devices' channels have an RMS delay spread of 300 ns, and all the devices are moving at a velocity of 50 km/h. We employ receiver strategy (d), in which only a single weight is calculated in every RB. We compare the case where the BS is equipped with two antennas (as in previous simulations) against the case where it is equipped with four. In each case, we compare exact MMSE filtering against the first-order approximation in (17) and against a MF. The x-axis this time is the average SNR of the devices, and similarly to the previous section, an SNR spread of ± 5 dB is applied.

Figure 6 shows the BLER performance of the different filters. It can be observed that for the two antennas case, a substantial gap exists between the approximate and exact

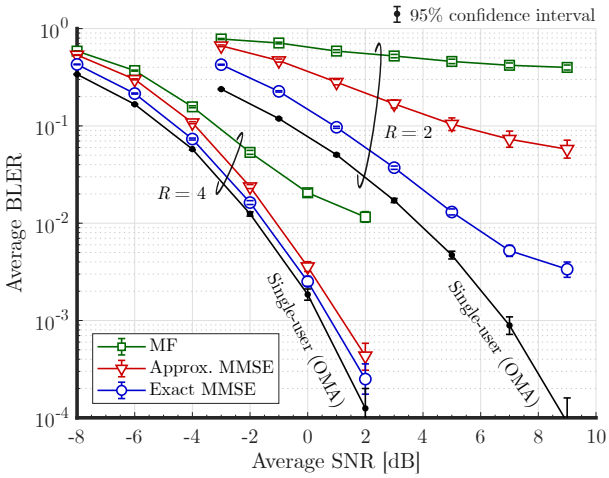

Fig. 6: Performance of the approximate filter and the MF for the cases of $R = 2$ and 4. Strategy (d) is employed here.

filters. However, the gap becomes almost non-existent once we switch to four antennas. Similarly, for the MF, the gap is much smaller compared to the two antennas case. The interesting point here is that by adding only two additional antennas, we almost closed the gap to the exact MMSE filter. We did not need to go for a large number of antennas. This trend suggests that for even a higher number of receive antennas, a MF would be a proper replacement of the exact MMSE filter, with only a small impact on performance. Also, we see that the gap to the OMA user is reduced as well, suggesting that receiver strategy (d) now incurs even less performance loss compared to a full receiver.

Complexity-wise, it might seem that the approximate filter has a cubic complexity order, as it involves a matrix-matrix multiplication. However, since we are only interested in the filter output, it can be implemented as a cascade of two MFs [10], i.e., first apply $(2\mathbf{I}-\beta_i\mathbf{Z}_i)$, and then filter the output with $\beta_i\widetilde{\mathbf{G}}_i^H$. This requires only matrix-vector multiplications which have a complexity order of $\mathcal{O}(L^2R^2 + KLR)$. On the other hand, exact MMSE filtering can also be implemented through cascade filters, however, one of the cascade filters requires a matrix inverse computation, and therefore the exact weight has a complexity order of $\mathcal{O}(L^3R^3 + L^2R^2 + KLR)$. As can be seen, it is much higher compared to the first-order approximate filter, especially in the range where $K \leq LR$.

V. Conclusion

In this paper, we propose low-complexity detection strategies for uplink code-domain NOMA. Namely, we utilize the time-frequency correlation properties of the wireless channel to reduce the amount of filter calculations. The rationally behind it, is that neighboring spreading-blocks on the time-frequency grid experience similar channel conditions, and therefore the calculated filter can be reused across those correlated blocks. Moreover, if the BS is equipped with a sufficient number of receive antennas, then we show that low-complexity filtering based on first-order Neumann series expansion provides a very close performance to exact MMSE filtering, at least for the considered NOMA configuration.

References

[1] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai, "A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions," *IEEE Communications Surveys Tutorials*, vol. 20, no. 4, pp. 3098–3130, Fourthquarter 2018.

[2] 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 Communications Magazine*, vol. 54, no. 9, pp. 59–65, September 2016.

[3] S. Schwarz, T. Philosof, and M. Rupp, "Signal Processing Challenges in Cellular-Assisted Vehicular Communications: Efforts and developments within 3GPP LTE and beyond," *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 47–59, March 2017.

[4] B. Di, L. Song, Y. Li, and Z. Han, "V2X Meets NOMA: Non-Orthogonal Multiple Access for 5G-Enabled Vehicular Networks," *IEEE Wireless Communications*, vol. 24, no. 6, pp. 14–21, Dec 2017.

[5] L. Tian, C. Yan, W. Li, Z. Yuan, W. Cao, and Y. Yuan, "On uplink non-orthogonal multiple access for 5g: opportunities and challenges," *China Communications*, vol. 14, no. 12, pp. 142–152, December 2017.

[6] N. Abu-Ali, A. M. Taha, M. Salah, and H. Hassanein, "Uplink Scheduling in LTE and LTE-Advanced: Tutorial, Survey and Evaluation Framework," *IEEE Communications Surveys Tutorials*, vol. 16, no. 3, pp. 1239–1265, Third 2014.

[7] K. Au, L. Zhang, H. Nikopour, E. Yi, A. Bayesteh, U. Vilaipornsawai, J. Ma, and P. Zhu, "Uplink contention based SCMA for 5G radio access," in *2014 IEEE Globecom Workshops (GC Wkshps)*, Dec 2014, pp. 900–905.

[8] B. Wang, L. Dai, Y. Yuan, and Z. Wang, "Compressive Sensing Based Multi-User Detection for Uplink Grant-Free Non-Orthogonal Multiple Access," in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, Sep. 2015, pp. 1–5.

[9] Z. Wu, K. Lu, C. Jiang, and X. Shao, "Comprehensive Study and Comparison on 5G NOMA Schemes," *IEEE Access*, vol. 6, pp. 18 511–18 519, 2018.

[10] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan 2013.

[11] B. Tahir, S. Schwarz, and M. Rupp, "Constructing Grassmannian Frames by an Iterative Collision-Based Packing," *IEEE Signal Processing Letters*, vol. 26, no. 7, pp. 1056–1060, July 2019.

[12] ——, "Joint Codebook Design for Multi-cell NOMA," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 4814–4818.

[13] 3rd Generation Partnership Project (3GPP), "Technical Specification Group Radio Access Network; Study on channel model for frequency spectrum above 6 GHz," 3rd Generation Partnership Project (3GPP), TR 38.900, Jun. 2018.

[14] S. Pratschner, B. Tahir, L. Marijanovic, M. Mussbah, K. Kirev, R. Nissel, S. Schwarz, and M. Rupp, "Versatile mobile communications simulation: the Vienna 5G Link Level Simulator," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, p. 226, Sep. 2018.

[15] E. Zöchmann, S. Schwarz, S. Pratschner, L. Nagel, M. Lerch, and M. Rupp, "Exploring the physical layer frontiers of cellular uplink," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, pp. 1–18, 2016. [Online]. Available: http://dx.doi.org/10.1186/s13638-016-0609-1

[16] D. Zhu, B. Li, and P. Liang, "On the matrix inversion approximation based on neumann series in massive MIMO systems," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 1763–1769.

[17] J. Minango and C. de Almeida, "Low-complexity MMSE detector based on the first-order Neumann series expansion for massive MIMO systems," in *2017 IEEE 9th Latin-American Conference on Communications (LATINCOM)*, Nov 2017, pp. 1–5.