



Single Carrier Modulation Resource Allocation for Massive MIMO with Virtual Antennas

October 2022

Changing the World's Energy Future

Brent A. Kenney, Behrouz Farhang-Boroujeny, Arslan J Majid, Hussein Moradi



DISCLAIMER

This information was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness, of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. References herein to any specific commercial product, process, or service by trade name, trade mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

Single Carrier Modulation Resource Allocation for Massive MIMO with Virtual Antennas

Brent A. Kenney, Behrouz Farhang-Boroujeny, Arslan J Majid, Hussein Moradi

October 2022

**Idaho National Laboratory
Idaho Falls, Idaho 83415**

<http://www.inl.gov>

**Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517**



Single Carrier Modulation Resource Allocation for Massive MIMO with Virtual Antennas

Journal:	<i>IEEE Transactions on Communications</i>
Manuscript ID	TCOM-TPS-22-1219
Manuscript Type:	Transactions Paper Submissions
Date Submitted by the Author:	29-Sep-2022
Complete List of Authors:	Kenney, Brent Majid, Arslan; Idaho National Laboratory Moradi, Hussein; Idaho National Laboratory Farhang-Boroujeny, Behrouz; University of Utah, ECE
Keyword:	MIMO systems

SCHOLARONE™
Manuscripts

Single Carrier Modulation Resource Allocation for Massive MIMO with Virtual Antennas

Brent A. Kenney*, Arslan J. Majid†, Hussein Moradi†, and Behrouz Farhang-Boroujeny*

*Electrical and Computer Engineering Department, University of Utah, Salt Lake City,
Utah, USA †Idaho National Laboratory, Salt Lake City, Utah, USA

Abstract

Today's Massive MIMO cellular operation is dominated by orthogonal frequency division multiplexing (OFDM) modulation. One of the advantages of OFDM is the flexibility to carve up the available spectrum into resource blocks (RBs) that can operate adjacent to one another. Massive MIMO adds spatial multiplexing layers on top of the RBs, enabling the simultaneous operation of dozens of UEs. Cyclic prefixed single carrier modulation (CP-SCM) is an alternative to OFDM that also benefits from massive MIMO and is useful as an OFDM alternative in specific scenarios. However, until recently, no resource allocation method has been proposed for CP-SCM. In this paper, we present a CP-SCM resource allocation scheme using virtual antennas in a massive MIMO time domain duplexed scenario. By creating several virtual antennas for each physical antenna, the benefits of massive MIMO are achieved with a smaller physical antenna count. Resources are quantized into data streams, and each user can be assigned a variable number of simultaneous streams. This paper presents a detailed development and analysis of multi-stream processing (MSP) for both uplink detection and downlink precoding. We also introduce heterogeneous MSP, where CP-SCM and OFDM signals can be processed in the same MSP framework.

I. INTRODUCTION

The ability to allocate different amounts of radio resources to various users is an essential capability of cellular network management. Resource allocation allows for efficient use of scarce radio frequency (RF) spectrum and supports a diverse network with a wide range of data requirements between users. Orthogonal frequency domain multiplexing (OFDM) divides the spectrum into resource blocks (RBs) that

This manuscript has in part been authored by Battelle Energy Alliance, LLC under Contract No. DE-AC07-05ID14517 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the paper for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. **STI Number: INL/JOU-22-65760.**

1 occupy adjacent portions of the band [1]. In addition to multiplexing users by frequency, OFDM also
2 supports massive multiple-input multiple-output (MIMO) processing, which uses spatial multiplexing to
3 create layers of user signaling within the same frequency resource [2], [3], [4], [5], [6], [7]. By combining
4 RBs and massive MIMO processing, OFDM-based systems can service dozens of users simultaneously.
5
6

7 As an alternative to OFDM, single carrier modulation (SCM) has been reintroduced in a number of
8 scenarios [8], including massive MIMO processing. Cyclic prefixed single carrier modulation (CP-SCM)
9 shares the same basic framing structure as OFDM since both use a cyclic prefix (CP) to preserve circular
10 convolution. The CP allows CP-SCM to completely compensate for inter-symbol interference (ISI) through
11 frequency domain (FD) processing, similar to OFDM. In addition, the CP makes each frame of CP-SCM
12 a single period of a periodic signal. Hence, the frame can be characterized by its Fourier series, and a
13 per-tone equalizer (as in OFDM) is applicable. Although CP-SCM has not been adopted in the 5G NR
14 (new radio) standard, it has been shown to be a useful alternative to OFDM in a massive MIMO scenario
15 with excellent processing gains when processed in the FD [9], [10], [11]. Explanations of detection and
16 precoding for CP-SCM are found in [12] and [13], respectively. One reason for investigating CP-SCM
17 is that it has been shown to possess many desirable advantages over OFDM, including low peak-to-
18 average-power ratio (PAPR) and robustness to carrier frequency offset [8]. However, it was not until [14]
19 and this paper that resource allocation was introduced for CP-SCM. The main point of this paper is to
20 present a CP-SCM solution that facilitates resource allocation, similar to what is done in current standards
21 with RBs. It should be noted that the intent of this paper is not to replace OFDM-based waveforms in
22 future standards. Rather, this paper provides a robust and flexible alternative that provides advantages
23 to specialized scenarios, such as massive machine type communication (mMTC), where there are many
24 low-rate, low-power devices in a small area.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

40 Resource allocation for CP-SCM is based upon the notions of data streams and virtual antennas. In
41 previous CP-SCM descriptions (e.g., [11]), N data symbols were transmitted per frame for each user
42 equipment (UE). In order to enable resource allocation, the N potential samples are divided into L
43 data streams of N/L symbols, where one or more data streams can be assigned to a UE in a given frame.
44 As illustrated in Fig. 1, each data stream is upsampled by L and interleaved with the other data streams
45 from the same UE. The upsampling process replicates the information-bearing signal spectrum L times,
46 which gives rise to the virtual antennas. Since there are L copies of a data stream's spectrum, each of the
47 M base station (BS) antennas receives the spectrum L times. Each spectral copy sees a different channel
48 (i.e., different channel coefficients). Hence, the multi-stream processing (MSP) technique presented here
49 increases the effective number of antennas from M to ML . Given ML virtual antennas, the number of
50
51
52
53
54
55
56
57
58
59
60

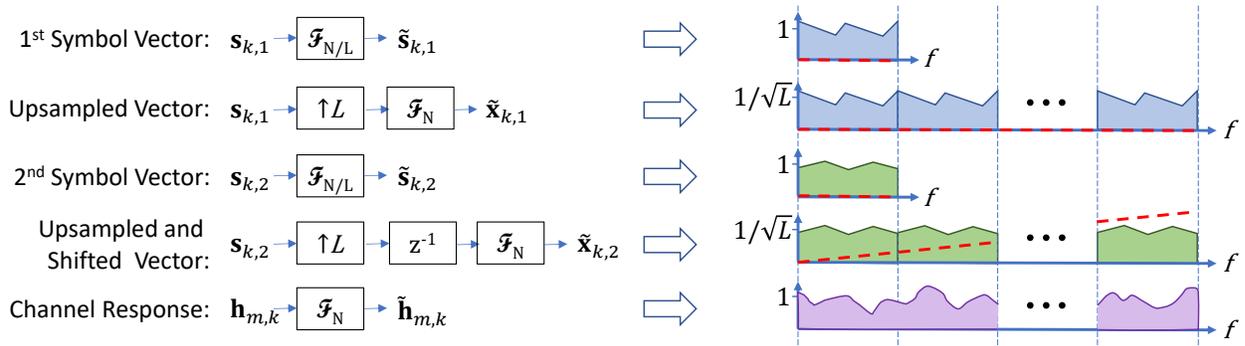


Fig. 1. TD (left) and FD amplitude (right) representations of an example of UL data streams. The N/L -length data streams $\mathbf{s}_{k,1}$ and $\mathbf{s}_{k,2}$ are represented before and after being prepared for UL multi-stream transmission. The N -length vectors $\tilde{\mathbf{x}}_{k,1}$ and $\tilde{\mathbf{x}}_{k,2}$ are the upsampled (and shifted) data streams in the FD. Upsampling in the TD results in spectral replication in the FD, albeit with reduced amplitude. A delay in the TD results in a linear phase in the FD, where the additional phase rotation is illustrated by the dashed red line. An example of the magnitude of the channel between UE k and BS antenna m shows that each of the L spectral replicas sees a different part of the channel. Hence, L virtual antennas are created from each physical BS antenna.

physical antennas, M , can be reduced while still operating in the massive MIMO regime. The key to processing multiple data streams from the same UE is to properly account for the linear phase shift due to the delay in the time domain (TD). For the remainder of this paper, MSP is understood to enable resource allocation in CP-SCM because it allows each UE to convey an independent number of data streams.

In the literature, the use of virtual antennas was first used by [15], [16], and [17] for the downlink (DL) precoding technique called waveforming. Waveforming uses time-reversal (TR) precoding and wide bandwidths to take advantage of individual multipath components. The signal is upsampled, effectively replicating the information bandwidth across the available bandwidth. Precoding is applied in the form of TR (aka conjugate beamforming for massive MIMO). At the receiver, the signal is first downsampled, which collapses the power back to the information bandwidth. This is similar to the single-stream approach of MSP, except that the precoding is not conducted in the FD. We point out that the virtual antennas referenced in the MSP development are due to the copies of the data stream's spectrum with uncorrelated channel coefficients, which differ from the virtual antennas created by multipath reflections in the waveforming case.

One instance of CP-SCM that will serve as an example throughout this paper is the cyclic prefix direct sequence spread spectrum (CP-DSSS) waveform. CP-DSSS was first proposed in [18] as a control channel to facilitate ultra-reliable low-latency communications (URLLC). It is characterized by its CP-SCM construction and Zadoff-Chu (ZC) spreading sequence. These two features create a circulant matrix for spreading and despreading, resulting in simplified FD signal processing. In [19], CP-DSSS was

1 introduced as a data channel, where a single data stream was allocated per UE. As in [14], we use
2 the expanded definition of CP-DSSS, where multiple data streams are supported. Although we use CP-
3 DSSS as the example waveform, we emphasize that the details of this paper apply to the broad class of
4 CP-SCM waveforms.
5
6
7

8 As an example of CP-SCM in an mMTC scenario, CP-DSSS was used in [20] as the waveform for
9 a low-power secondary network that operates using the same frequency resources as the OFDM-based
10 primary network. In this scenario, the secondary network is divided into several femtocells to keep the
11 transmit power low for the secondary network users. Each femtocell is serviced by a femtocell gateway
12 (FGW), which acts as a hub for the femtocell and also provides a wireless backhaul to the primary
13 network. When the FGW is receiving, it must simultaneously detect the OFDM signal from the primary
14 BS and the CP-DSSS signal from the femtocell terminals (FTs). Likewise, when the FGW is transmitting,
15 it must precode the OFDM signal for the primary network and the CP-DSSS signal for the secondary
16 network at the same time. We will show how these heterogeneous signals can be processed using MSP
17 detection and precoding algorithms. We note that if MSP were used in [20], then the realized processing
18 gain of the secondary network would have been higher. In addition, the secondary network resources
19 could have been divided among more users.
20
21
22
23
24
25
26
27
28

29 In addition to the scenario with heterogeneous signal types, we introduce two other use cases. The first
30 example is a low-interference secondary network that operates within the same spectrum as the primary
31 network. This is another aspect of the initial work in [20]. In this scenario, the FGW uses multiple antennas
32 through MSP to spread a small number of data symbols per user across each signal frame to reduce the
33 power in the secondary network. With reduced data rates in the secondary network and close proximity
34 between the FGW and its FTs, the interference power received by the BS of the primary network can be
35 managed to an acceptable level. The second example is a scenario where the spectrum targeted for MSP
36 communication is sparsely occupied by other signals [21]. In this application, the interfered portions of
37 the spectrum can be masked to avoid interfering with the incumbent signals. Likewise, spectral masking
38 benefits the MSP receiver by ignoring the interference contribution from the incumbent signals. Both of
39 these use cases, as well as the heterogeneous signal scenario, are explained in more detail in Section V.
40
41
42
43
44
45
46
47

48 This paper is organized as follows. The system model is presented in Section II. Section III provides
49 the MSP details for uplink (UL) detection. Section IV follows with MSP details for efficient DL precoding
50 using the matrix inverses that are calculated during UL detection. In Section V, we explain some of the
51 benefits of MSP and discuss three use cases where MSP is well-suited. The details of one of these use
52 cases are presented in Section VI, where OFDM signals are processed within the same framework as
53
54
55
56
57
58
59
60

MSP signals. Section VII provides concluding remarks.

Notation: Italic letters represent scalars. Bold lowercase letters represent column vectors. FD vectors are capped with a tilde. Bold uppercase letters represent matrices. \mathbf{I}_j is the $j \times j$ identity matrix. $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ represent the complex conjugate, transpose, and Hermitian operators, respectively. $\mathbb{E}[\cdot]$ is the expected value taken over all channel and noise realizations. Finally, $\text{tr}\{\cdot\}$ is the trace operator.

II. SYSTEM MODEL

We recall that the presence of CP in OFDM is an enabler to performing most of the signal processing steps in the frequency domain, separately for each frequency bin, e.g., [22], [23]. As a result of the CP, each signal frame may be thought of as one period of a periodic signal and hence, can be perfectly expressed by its Fourier series components which are obtained through the application of a Discrete Fourier Transform (DFT). This property is applicable to CP-SCM because of the same reasons. It is on this basis that most of the analytical developments in this paper have been developed.

The scenarios modeled in this paper assume that the UE does not have any channel state information (CSI). Each UE is supposed to know the average channel power to maintain a power target at the BS, averaged across all BS antennas. Likewise, the BS is aware of the large-scale path loss such that it can tailor power levels for each UE during DL operation. The BS has CSI between each of the K UEs and the M BS antennas, which can be obtained from the UE's pilot signals. This paper assumes perfect CSI at the BS, but the addition of channel estimation error should be relatively straightforward by following the conventions presented in [11].

The UL and DL transmissions consist of frames of length N potential symbols per user. The over-the-air transmission includes a CP to preserve circular convolution, but the precoding and detection algorithms are performed after the CP is removed. When the TD samples are converted to the FD, they are referred to as bins.

The following variables are used repeatedly throughout the paper:

- M : Number of physical antennas at the BS
- K : Number of physical UEs
- N : Number of time samples without CP (or frequency bins)
- L : Upsample factor for data streams
- ℓ_k : Number of data streams for user k

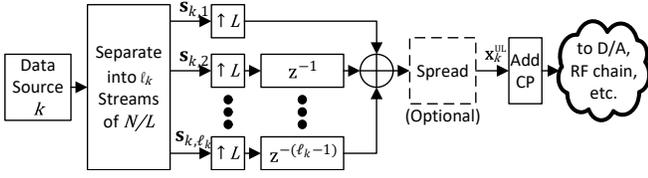


Fig. 2. UE transmitter for UL MSP. The ℓ_k streams allocated to the k^{th} UE are upsampled and interleaved as illustrated. Spreading can be applied as an option.

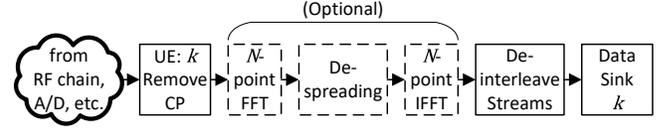


Fig. 3. UE receiver for DL MSP. This receiver structure follows the simpler TD deinterleaving from Section IV-C.

A. UL Model

In the UL model, the k^{th} UE transmits ℓ_k streams of N/L unit variance symbols, represented as $s_{k,1}$ to s_{k,ℓ_k} . The maximum number of streams for each user is L . Each stream of N/L symbols is upsampled by L . This operation is represented mathematically through the expander matrix \mathbf{E}_L , which is formed by inserting $L-1$ rows of zeros after each row of the identity matrix $\mathbf{I}_{N/L}$, resulting in a matrix with dimensions $N \times (N/L)$. After upsampling, the symbols are interleaved together in the TD by shifting and then summing the streams as depicted in Fig. 2. The resulting N -length sequence of each stream is $\mathbf{x}_{k,l} = \mathbf{E}_{L(l-1)} \mathbf{s}_{k,l}$, where the parenthetical subscript represents the number of circular shifts applied to each column of \mathbf{E}_L . Note that l is the stream index used throughout this paper, which starts at 1. For convenience, we assume that each subsequent stream is shifted by one. The transmitted sequence from the k^{th} UE prior to the addition of the CP is given as

$$\mathbf{x}_k^{\text{UL}} = \mathbf{Z} \sum_{l=1}^{\ell_k} \mathbf{E}_{L(l-1)} \mathbf{s}_{k,l} \quad (1)$$

where \mathbf{Z} is the optional spreading matrix. The spreading operation fills in the gaps of the data stream created during the upsample process. In this paper, we assume that the same \mathbf{Z} is used for each UE. Hence, spreading in this context is not used for code division multiple access (CDMA), although it can be used to suppress interference from UEs in adjacent cells that use a different \mathbf{Z} matrix. We assume that \mathbf{Z} is a unitary matrix (i.e., $\mathbf{Z}^H \mathbf{Z} = \mathbf{I}_N$). Consequently, \mathbf{Z} does not alter the transmit power. As in the case of CP-DSSS, we also assume that \mathbf{Z} is circulant. This property will prove useful for efficient despreading in the FD.

The received signal from the K UEs at the m^{th} antenna after CP removal is expressed as

$$\mathbf{y}_m^{\text{UL}} = \sum_{k=1}^K \mathbf{H}_{m,k} \mathbf{x}_k^{\text{UL}} + \mathbf{w}_m, \quad (2)$$

where $\mathbf{H}_{m,k}$ is the $N \times N$ convolutional channel matrix between antenna m and user k , and \mathbf{w}_m is the receiver noise ($\mathbf{w}_m \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I}_N)$).

Let $\mathbf{h}_{m,k}$ represent the channel impulse response vector between antenna m and user k , which is of length ℓ_h . $\mathbf{H}_{m,k}$ is formed by first taking $\mathbf{h}_{m,k}$ and appending $N-\ell_h$ zeros to form the base vector $\mathbf{h}_{m,k(0)}$. We then take downward circular shifts of $\mathbf{h}_{m,k(0)}$ to create

$$\mathbf{H}_{m,k} = \begin{bmatrix} \mathbf{h}_{m,k(0)} & \mathbf{h}_{m,k(1)} & \cdots & \mathbf{h}_{m,k(N-2)} & \mathbf{h}_{m,k(N-1)} \end{bmatrix}, \quad (3)$$

where the parenthetical subscript again represents the number of downward circular shifts applied to the base vector. $\mathbf{H}_{m,k}$ is a circulant matrix based on its construction.

Detection is performed in the FD, and the conversion is made through the DFT. We represent the scaled DFT as \mathcal{F}_N , where N is the dimension of the matrix. The scaling includes a $1/\sqrt{N}$ factor such that $\mathcal{F}_N \mathcal{F}_N^{-1} = \mathbf{I}_N$. It is also worth noting that $\mathcal{F}_N^{-1} = \mathcal{F}_N^H$. We will also refer to the unscaled DFT, which can be represented as $\sqrt{N}\mathcal{F}_N$.

In order to represent (2) in the FD, we first examine the FD representation of $\mathbf{E}_{L(l-1)}$ from (1). Based on the construction of the expander matrix, the product $\mathcal{F}_N \mathbf{E}_{L(l-1)}$ results in N/L columns of \mathcal{F}_N , which are equally spaced such that every L^{th} column is selected based on the indicated starting column. By periodically sampling the DFT, L spectral copies of the N/L -length spectrum are created. That is, by expanding/upsampling a sequence in the TD by a factor of L , the repeated spectrum in the FD is compressed by a factor of L [24]. This phenomenon can be termed the spectral compression property of upsampling. We find that when \mathcal{F}_N is sampled with a spacing of L , starting with the first column, the result is L stacked versions of $\mathcal{F}_{N/L}$ with a scale factor of $1/\sqrt{L}$. This scale factor results from the $\sqrt{L/N}$ factor associated with $\mathcal{F}_{N/L}$. We represent the stacked versions of $\mathcal{F}_{N/L}$ by introducing the $N \times (N/L)$ vertical tiling matrix, $\mathbf{T}_{N,L}$, which is defined as $\mathbf{T}_{N,L} = [\mathbf{I}_{N/L} \ \mathbf{I}_{N/L} \ \cdots \ \mathbf{I}_{N/L}]^T$. Hence, we specify the FD representation of the unshifted expander matrix as $\mathcal{F}_N \mathbf{E}_{L(0)} = \mathbf{T}_{N,L} \mathcal{F}_{N/L} / \sqrt{L}$.

The circular time shift property of the DFT states that a circular shift of any TD sequence will result in a linear phase in the FD [25]. We note that the circular shift of \mathbf{E}_L determines the starting column of the DFT matrix. The first column of the unscaled DFT matrix is all ones. Each successive column of the unscaled DFT matrix is obtained by multiplying the previous column by the second column of the unscaled DFT matrix. We define $\omega = e^{-j2\pi/N}$ and create a vector $\boldsymbol{\psi} = [\omega^0 \ \omega^1 \ \cdots \ \omega^{N-1}]^T$ to represent the second column of the unscaled DFT matrix. We define the diagonal matrix $\boldsymbol{\Psi}$ with elements that equal the elements of $\boldsymbol{\psi}$ (i.e., $\text{diag}(\boldsymbol{\Psi}) = \boldsymbol{\psi}$). By raising $\boldsymbol{\Psi}$ to the power of the circular shift, we can represent the linear phase associated with the circular shift of the expander matrix. We now express the product of the scaled DFT matrix and the circularly shifted expander matrix as

$$\mathcal{F}_N \mathbf{E}_{L(l-1)} = \frac{1}{\sqrt{L}} \boldsymbol{\Psi}^{l-1} \mathbf{T}_{N,L} \mathcal{F}_{N/L}. \quad (4)$$

Note that the $\mathcal{F}_{N/L}$ matrix factored on the right will be applied to the symbol vector resulting in $\mathcal{F}_{N/L}\mathbf{s}_{k,l} = \tilde{\mathbf{s}}_{k,l}$.

The FD representation of the transmitted UL signal from (1) can now be given as

$$\tilde{\mathbf{x}}_k^{\text{UL}} = \mathbf{\Omega} \frac{1}{\sqrt{L}} \sum_{l=1}^{\ell_k} \mathbf{\Psi}^{l-1} \mathbf{T}_{N,L} \tilde{\mathbf{s}}_{k,l}, \quad (5)$$

where $\tilde{\mathbf{x}}_k^{\text{UL}} = \mathcal{F}_N \mathbf{x}_k^{\text{UL}}$ and $\mathbf{\Omega}$ is the FD representation of \mathbf{Z} . The FD representation of a matrix may not be a common term, but it is applicable in cases where a matrix can be decomposed using a DFT matrix and an IDFT matrix. Any $N \times N$ circulant matrix (e.g., $\mathbf{H}_{m,k}$ and \mathbf{Z}) is diagonalized by \mathcal{F}_N . In other words, the FD representation of circulant matrices are diagonal [26]. We find that $\mathbf{H}_{m,k} = \mathcal{F}_N^{-1} \mathbf{\Lambda}_{m,k} \mathcal{F}_N$ and $\mathbf{Z} = \mathcal{F}_N^{-1} \mathbf{\Omega} \mathcal{F}_N$. It follows that taking the N -point DFT of $\mathbf{H}_{m,k}$ results in $\mathcal{F}_N \mathbf{H}_{m,k} = \mathbf{\Lambda}_{m,k} \mathcal{F}_N$, where $\mathbf{\Lambda}_{m,k}$ is a diagonal matrix containing the eigenvalues of $\mathbf{H}_{m,k}$. Let $\lambda_{m,k,i}$ represent the i^{th} value along the diagonal of $\mathbf{\Lambda}_{m,k}$. The eigenvalues can also be obtained by taking the N -point DFT of the channel impulse response $\mathbf{h}_{m,k}$, which is used to form $\mathbf{H}_{m,k}$. For more efficient computation, it is noted that all of the FD conversions can be performed with the Fast Fourier Transform (FFT) instead of the DFT.

We now represent the received UL signal from (2) in the FD as

$$\begin{aligned} \tilde{\mathbf{y}}_m^{\text{UL}} &= \sum_{k=1}^K \mathbf{\Lambda}_{m,k} \tilde{\mathbf{x}}_k^{\text{UL}} + \tilde{\mathbf{w}}_m \\ &= \sum_{k=1}^K \mathbf{\Lambda}_{m,k} \mathbf{\Omega} \frac{1}{\sqrt{L}} \sum_{l=1}^{\ell_k} \mathbf{\Psi}^{l-1} \mathbf{T}_{N,L} \tilde{\mathbf{s}}_{k,l} + \tilde{\mathbf{w}}_m, \end{aligned} \quad (6)$$

where $\tilde{\mathbf{w}}_m$ is the FD representation of the noise vector associated with antenna port m . Note that the DFT does not change the noise statistics (i.e., $\tilde{\mathbf{w}}_m \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I}_N)$). We use the fact that $\mathbf{\Omega}$ is diagonal to show that $\mathbf{\Lambda}_{m,k}$ and $\mathbf{\Omega}$ are commutable. As a consequence, despreading the received signal is performed in the FD by multiplying the received signal by $\mathbf{\Omega}^*$ since $\mathbf{\Omega}^* \mathbf{\Omega} = \mathbf{I}_N$ as a result of \mathbf{Z} being unitary. Thus, despreading by a circulant and unitary matrix does not change the underlying signal properties, so $\mathbf{\Omega}$ will be dropped from future use of (6) in order to focus on the processing after despreading.

Here, we point out that the majority of the processing specified in this paper is performed in the FD on a bin-by-bin basis. We use the variable n to represent the bin index, and we append this index to indicate an element of a given vector. Since bin processing involves inputs from multiple antennas and multiple streams, we use a colon to represent when a vector index spans all possible values. For example, we define the vector of received UL signals for all BS antennas corresponding to the n^{th} bin as $\tilde{\mathbf{y}}_{:,n}^{\text{UL}} = [\tilde{y}_{1,n}^{\text{UL}} \tilde{y}_{2,n}^{\text{UL}} \dots \tilde{y}_{M,n}^{\text{UL}}]^T$. The noise vector is $\tilde{\mathbf{w}}_{:,n} = [\tilde{w}_{1,n} \tilde{w}_{2,n} \dots \tilde{w}_{M,n}]^T$. Likewise, we define the vector of FD symbols for all streams corresponding to the n^{th} bin as $\tilde{\mathbf{s}}_{k,:,n} = [\tilde{s}_{k,1,n} \tilde{s}_{k,2,n} \dots \tilde{s}_{k,\ell_k,n}]^T$. These, and related terms, will be used for detection at the BS.

In order to perform FD processing on a bin-by-bin basis, we begin by representing the FD channel matrix for bin n , which is based on the diagonal elements of the $\mathbf{\Lambda}_{m,k}$ matrices. The $M \times K$ FD channel matrix is defined as

$$\mathbf{A}_n = \begin{bmatrix} \lambda_{1,1,n} & \lambda_{1,2,n} & \dots & \lambda_{1,K,n} \\ \lambda_{2,1,n} & \lambda_{2,2,n} & \dots & \lambda_{2,K,n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{M,1,n} & \lambda_{M,2,n} & \dots & \lambda_{M,K,n} \end{bmatrix}. \quad (7)$$

We now express the length- M vector of FD received signals for bin n , based on the first expression in (6), as

$$\tilde{\mathbf{y}}_{:,n}^{\text{UL}} = \mathbf{A}_n \tilde{\mathbf{x}}_{:,n}^{\text{UL}} + \tilde{\mathbf{w}}_{:,n}. \quad (8)$$

B. DL Model

The DL model follows the same conventions as the UL model, but we assume that this is a time domain duplex (TDD) system, which allows us to use the channel coefficients determined via UL transmissions for the DL. Each BS antenna transmits a CP and the associated N -length vector, denoted as \mathbf{x}_m^{DL} for the m^{th} antenna. The received signal vector at the k^{th} UE after CP removal is given as

$$\mathbf{y}_k^{\text{DL}} = p_k^{-\frac{1}{2}} \sum_{m=1}^M \mathbf{H}_{m,k} \mathbf{x}_m^{\text{DL}} + \mathbf{w}_k, \quad (9)$$

where $\mathbf{H}_{m,k}$ is the circulant channel matrix defined previously, \mathbf{w}_k is the noise vector associated with the k^{th} UE's receiver ($\mathbf{w}_k \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I}_N)$), and $p_k^{-1/2}$ is the amplitude scaling corresponding to the large-scale path loss between the BS and the k^{th} UE (i.e., p_k^{-1}). Note that the large-scale path loss is not included in the UL model because the UEs power control such that the UE signals are received with the same power per symbol. Upon receipt of the DL signal, the UE is responsible for despreading the signal, if needed, and deinterleaving the streams. Most of the complexity of the DL model is encapsulated in the formation of the precoded vector $\tilde{\mathbf{x}}_m^{\text{DL}}$. The receiver at the UE is relatively simple and may be implemented in the FD or the TD. Further details are given in Section IV. For UE detection in the FD, we represent the received signal at the k^{th} UE as

$$\tilde{\mathbf{y}}_k^{\text{DL}} = p_k^{-\frac{1}{2}} \sum_{m=1}^M \mathbf{\Lambda}_{m,k} \tilde{\mathbf{x}}_m^{\text{DL}} + \tilde{\mathbf{w}}_k. \quad (10)$$

Precoding for the DL is calculated in the FD on a per-bin basis. Let $\tilde{\mathbf{x}}_{:,n}^{\text{DL}}$ be the vector of FD precoded values for the n^{th} bin across all M antennas. The n^{th} bin of the received vector at the k^{th} UE is defined as

$$\tilde{y}_{k,n}^{\text{DL}} = p_k^{-\frac{1}{2}} \boldsymbol{\lambda}_{:,k,n}^{\text{T}} \tilde{\mathbf{x}}_{:,n}^{\text{DL}} + \tilde{w}_{k,n}, \quad (11)$$

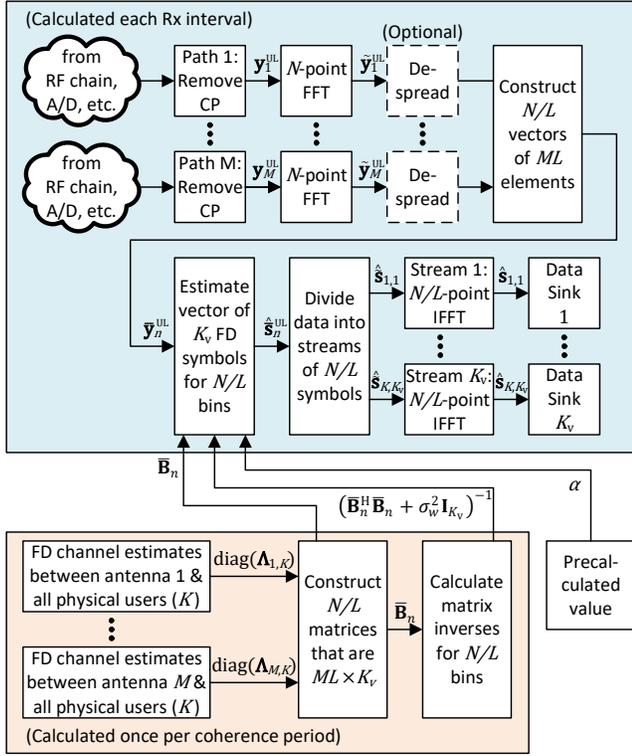


Fig. 4. BS receiver using MSP detection from (15). The FD channel estimates are used to construct the composite channel matrices, $\bar{\mathbf{B}}_n$, for each coherence period. The associated matrix inverses for N/L bins are also computed once per coherence period. MSP produces a vector of symbol estimates for each of the K_v virtual users.

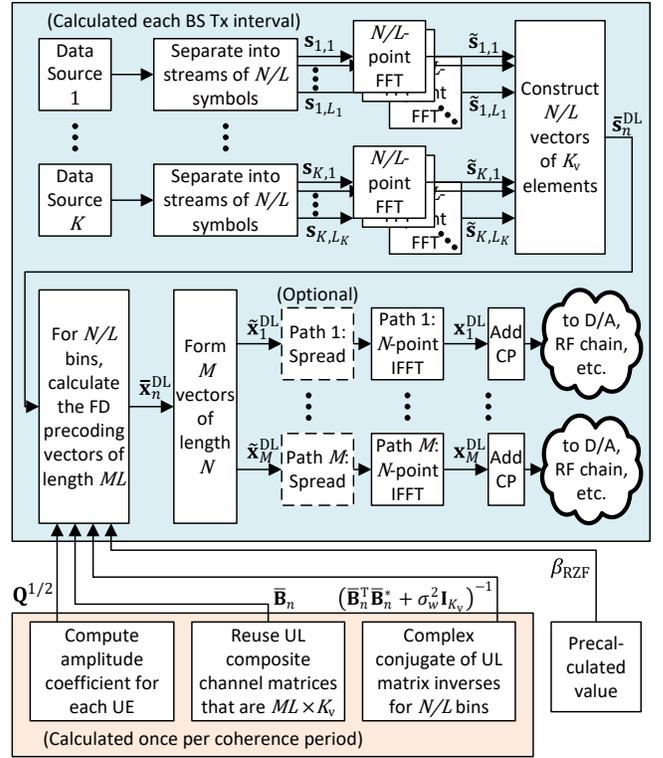


Fig. 5. BS Transmitter using MSP DL precoding from (18). The computational complexity of the precoder is significantly reduced by reusing the matrix inverses that were calculated for the MSP UL detector. Power optimization need only be computed once per coherence period, and it is not a significant driver for complexity.

where $\tilde{y}_{k,n}^{\text{DL}}$ is the n^{th} element of $\tilde{\mathbf{y}}_k^{\text{DL}}$, $\lambda_{:,k,n}$ is the k^{th} column in (7) and $\tilde{w}_{k,n}$ is the n^{th} element of $\tilde{\mathbf{w}}_k$.

III. SINGLE- AND MULTI-STREAM UL DETECTION

By assigning ℓ_k streams to each UE, the massive MIMO BS allocates resources with the CP-SCM waveform. The UL receiver structure is depicted in Fig. 4.

Because UL detection is performed entirely in the FD, the first step after removing the CP is to convert the N -length received vectors from (2) to the FD representation in (6) by taking the N -point FFT. As described in Section II-A, if the transmitted UL signal is spread with the circulant, unitary matrix \mathbf{Z} , which is represented as $\mathbf{\Omega}$ in the FD, then the signal is despread by multiplying by $\mathbf{\Omega}^*$. The diagonal matrix multiplication by $\mathbf{\Omega}^*$ can be replaced with a Hadamard multiply by the diagonal elements of $\mathbf{\Omega}^*$ for increased efficiency.

Detection is performed on a per-bin basis. As such, the calculations will be based upon the vector of received FD samples for the n^{th} bin, $\tilde{\mathbf{y}}_{:,n}^{\text{UL}}$, as defined in Section II-A. We next express the FD received

$$\mathbf{B}_{k,n} = \begin{bmatrix} \lambda_{:,k,n} & \omega^n \lambda_{:,k,n} & \omega^{2n} \lambda_{:,k,n} & \dots & \omega^{(\ell_k-1)n} \lambda_{:,k,n} \\ \lambda_{:,k,n+\frac{N}{L}} & \omega^{n+\frac{N}{L}} \lambda_{:,k,n+\frac{N}{L}} & \omega^{2(n+\frac{N}{L})} \lambda_{:,k,n+\frac{N}{L}} & \dots & \omega^{(\ell_k-1)(n+\frac{N}{L})} \lambda_{:,k,n+\frac{N}{L}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{:,k,n+(L-1)\frac{N}{L}} & \omega^{n+(L-1)\frac{N}{L}} \lambda_{:,k,n+(L-1)\frac{N}{L}} & \omega^{2(n+(L-1)\frac{N}{L})} \lambda_{:,k,n+(L-1)\frac{N}{L}} & \dots & \omega^{(\ell_k-1)(n+(L-1)\frac{N}{L})} \lambda_{:,k,n+(L-1)\frac{N}{L}} \end{bmatrix} \quad (12)$$

vector in terms of the transmitted FD symbol vector. We note that there are L frequency bins that have components of $\tilde{\mathbf{s}}_{k,l,n}$ based on (6) because of the structure of the tiling matrix, $\mathbf{T}_{N,L}$, and those bins are N/L bins apart. We define the composite vector of received signals with components of the n^{th} bin of $\tilde{\mathbf{s}}_{k,l}$ as $\bar{\mathbf{y}}_n^{\text{UL}} = [(\tilde{\mathbf{y}}_{:,n}^{\text{UL}})^{\text{T}} (\tilde{\mathbf{y}}_{:,n+N/L}^{\text{UL}})^{\text{T}} \dots (\tilde{\mathbf{y}}_{:,n+(L-1)N/L}^{\text{UL}})^{\text{T}}]^{\text{T}}$. With this construction, $\bar{\mathbf{y}}_n^{\text{UL}}$ has ML elements representing each virtual antenna and is only defined for N/L bins.

We next define the composite FD symbol vector corresponding to the transmitted signal based on $\tilde{\mathbf{s}}_{k,:,n}$ from Section II-A. The length of the FD symbol vector for the k^{th} UE is ℓ_k , which could vary from UE to UE. We combine the FD symbol vectors from each UE into the composite FD symbol vector, which is defined as $\bar{\mathbf{s}}_n = [\tilde{\mathbf{s}}_{1,:,n}^{\text{T}} \tilde{\mathbf{s}}_{2,:,n}^{\text{T}} \dots \tilde{\mathbf{s}}_{K,:,n}^{\text{T}}]^{\text{T}}$. The length of the composite vector is $K_v = \sum_{k=1}^K \ell_k$, which is the total number of virtual users. The order specified for $\bar{\mathbf{s}}_n$ can be chosen arbitrarily. We choose to organize the elements by data stream and then by user number. The same convention must be followed for the composite channel matrix defined below.

The composite channel matrix for per-bin processing is based upon the diagonal elements of $\mathbf{\Lambda}_{m,k}$ for each combination of m and k . The n^{th} diagonal element of $\mathbf{\Lambda}_{m,k}$ is represented as $\lambda_{m,k,n}$. We form the vector $\lambda_{:,k,n}$ over all M BS antennas, where $\lambda_{:,k,n} = [\lambda_{1,k,n} \lambda_{2,k,n} \dots \lambda_{M,k,n}]^{\text{T}}$, which is the k^{th} column of \mathbf{A}_n in (7). Next, we define $\mathbf{B}_{k,n}$ as an intermediate matrix for the k^{th} UE in (12) at the top of this page. The matrix $\mathbf{B}_{k,n}$ has dimensions $ML \times \ell_k$. All streams besides the first are affected by the linear phase term from $\Psi^{(l-1)}$ as a result of the circular shift associated with interleaving. The $\mathbf{B}_{k,n}$ matrices from each UE are combined to form the composite channel matrix

$$\bar{\mathbf{B}}_n = \frac{1}{\sqrt{L}} [\mathbf{B}_{1,n} \mathbf{B}_{2,n} \dots \mathbf{B}_{K,n}], \quad (13)$$

where we note that the order of the columns aligns with the order of the virtual users in the composite symbol vector, $\bar{\mathbf{s}}_n$. Likewise, stacking the $\lambda_{:,k,n}$ vectors from periodic bins of the spectrum, as shown in (12), aligns with the virtual antennas of the composite received signal vector, $\bar{\mathbf{y}}_n^{\text{UL}}$.

The last composite vector related to the UL received signal is the composite noise vector. It follows the same mapping as $\bar{\mathbf{y}}_n^{\text{UL}}$ and is defined as $\bar{\mathbf{w}}_n = [(\tilde{\mathbf{w}}_{:,n})^{\text{T}} (\tilde{\mathbf{w}}_{:,n+N/L})^{\text{T}} \dots (\tilde{\mathbf{w}}_{:,n+(L-1)N/L}^{\text{T}})^{\text{T}}]^{\text{T}}$, where $\tilde{\mathbf{w}}_{:,n}$ is the FD noise vector over all M antennas corresponding to the n^{th} bin.

We now express the per-bin representation of (6) after despreading as

$$\bar{\mathbf{y}}_n^{\text{UL}} = \bar{\mathbf{B}}_n \bar{\mathbf{s}}_n + \bar{\mathbf{w}}_n, \quad (14)$$

which is valid for the first N/L bins. With this construction, we can apply the MRC-MMSE detector from [11], which is given as

$$\hat{\bar{\mathbf{s}}}_n^{\text{UL}} = \alpha \left(\bar{\mathbf{B}}_n^{\text{H}} \bar{\mathbf{B}}_n + \sigma_w^2 \mathbf{I}_{K_v} \right)^{-1} \bar{\mathbf{B}}_n^{\text{H}} \bar{\mathbf{y}}_n^{\text{UL}}, \quad (15)$$

where the scale factor, α , is calculated for ML antennas and K_v users. The estimates of the FD symbols are calculated for the first N/L bins. After the estimates for the N/L bins have been calculated, the results are rearranged to form K_v length- N/L vectors corresponding to the FD estimates of the symbol vectors for each stream. The FD vectors are then converted to the TD using the inverse FFT (IFFT).

A. UL Performance Analysis

The performance of the MRC-MMSE detection is characterized by the massive MIMO processing gain. We define the processing gain as the ratio of output E_s/N_0 to input E_s/N_0 for each antenna, where E_s/N_0 is the energy per symbol divided by the noise power spectral density. We begin by examining the single-stream case where $\ell_k = 1$ for all k . At high input E_s/N_0 , we see from [11] that the value of α goes to unity. Likewise, the $\sigma_w^2 \mathbf{I}_{K_v}$ term in (15) vanishes. By examining the signal component of $\bar{\mathbf{y}}_n^{\text{UL}}$ from (14), we see that $(\bar{\mathbf{B}}_n^{\text{H}} \bar{\mathbf{B}}_n)^{-1} \bar{\mathbf{B}}_n^{\text{H}} \bar{\mathbf{B}}_n \bar{\mathbf{s}}_n^{\text{UL}} = \bar{\mathbf{s}}_n^{\text{UL}}$, resulting in an unbiased estimate of the FD symbol vector.

The gain of the noise component is scaled by $(\bar{\mathbf{B}}_n^{\text{H}} \bar{\mathbf{B}}_n)^{-1} \bar{\mathbf{B}}_n^{\text{H}}$. The power of the noise scaling evaluated over all channel realizations is expressed for each user as

$$\begin{aligned} \frac{1}{K} \mathbb{E} \left[\text{tr} \left\{ \left((\bar{\mathbf{B}}_n^{\text{H}} \bar{\mathbf{B}}_n)^{-1} \bar{\mathbf{B}}_n^{\text{H}} \right) \left((\bar{\mathbf{B}}_n^{\text{H}} \bar{\mathbf{B}}_n)^{-1} \bar{\mathbf{B}}_n^{\text{H}} \right)^{\text{H}} \right\} \right] &= \frac{L}{K} \mathbb{E} \left[\text{tr} \left\{ \left(L \bar{\mathbf{B}}_n^{\text{H}} \bar{\mathbf{B}}_n \right)^{-1} \right\} \right] \\ &= \frac{L}{ML - K}, \end{aligned} \quad (16)$$

where the coefficient of L is used inside the inverse operator to scale the entries of the constituent matrices to have unit variance. Because $\bar{\mathbf{B}}_n^{\text{H}} \bar{\mathbf{B}}_n$ is a central, complex Wishart matrix with ML degrees of freedom (i.e., $\bar{\mathbf{B}}_n^{\text{H}} \bar{\mathbf{B}}_n \sim \mathcal{W}_K(ML, \mathbf{I}_K/L)$), we leverage Lemma 6 from [27] to get the $K/(ML-K)$ result for the expectation of the trace of the inverse. Hence, the noise term associated with each UE's signal is divided by $M-K/L$ after detection. Since K_v equals K for the single-stream case, we can express the UL processing gain as

$$G^{\text{UL}} = M - \frac{K_v}{L}. \quad (17)$$

Here, we see an improvement over the standard MRC-MMSE detector whose processing gain is $M-K$, as shown in [11]. The improvement is due to the reduced data stream length of N/L and the tailored processing.

For the single-stream analysis above, we used the fact that the elements of $\bar{\mathbf{B}}_n$ are zero-mean, Gaussian random variables that are i.i.d. In the multi-stream case, the elements are no longer i.i.d. because subsequent streams are phase rotated versions of the initial stream for a given UE. As a result, the matrix $\bar{\mathbf{B}}_n^H \bar{\mathbf{B}}_n$ is not a Wishart matrix per the classical definition. However, we will show that $\bar{\mathbf{B}}_n^H \bar{\mathbf{B}}_n$ is statistically equivalent to a Wishart matrix. Accordingly, the random matrix theory result for Wishart matrices that defines the processing gain for massive MIMO can be applied to MSP.

We note that each entry of $\bar{\mathbf{B}}_n^H \bar{\mathbf{B}}_n$ is the sum of ML products. Even with modest values for M and L such that $ML = 64$, the number of terms in the summation for each element is high enough for the Central Limit Theorem (CLT) to apply. As such, each element is approximately normally distributed, meaning that it is completely defined by the first two moments. Each of the diagonals of $\bar{\mathbf{B}}_n^H \bar{\mathbf{B}}_n$ is a summation of ML squared magnitudes. There is no difference between the single- and multi-stream matrices for the diagonals because each diagonal is the inner product of a column vector with itself.

The multi-stream matrix has three types of off-diagonal elements. The first type is identical to the single-stream matrix, which are the off-diagonals of the Wishart matrix. We note that each inner product is the summation of ML products of complex, normally distributed random variables ($CN \sim (0, 1/L)$). Each product has a variance of $1/L^2$, and the resulting summation has a variance of M/L .

The second type of off-diagonal element is the inner product of a secondary stream vector with any vector corresponding to a different user. Since the channel matrix entries for each user are independent, the additional phase rotations are of no consequence, and the second type of off-diagonal has identical statistics to the first type. The first and second types of off-diagonals result from inter-user inner products.

The third type of off-diagonal is the intra-user inner products. Although each column of $\mathbf{B}_{k,n}$ is based on the same set of complex coefficients, the phase shifts applied to each section of the vectors differentiate the inner products. There are L sections with different phase shifts, as shown in (12). Depending on the phases of the two columns, L distinct phase values could result, but there could be as few as two unique phases. The number of unique phases will be even and evenly spaced around 2π . Each group of products with the same phase can be matched with another group of products with the opposite phase. In the proof that follows, we show that the variance of the summation of ML cross-products corresponding to the intra-user off-diagonals is M/L .

Proof: We first note that the expected value of the magnitude squared of a zero-mean Gaussian

1 random variable has a mean equal to the variance of the random variable [28]. The corresponding
 2 variance is the square of the variance of the random variable, which follows from the fact that such a
 3 random variable is a scaled chi-squared with a scaling factor of 0.5. The Gaussian random variables
 4 forming the columns of $\bar{\mathbf{B}}_n$ have a variance of $1/L$. Consequently, the magnitude squared of the random
 5 variable has a variance of $1/L^2$.
 6
 7
 8
 9

10 The inner product of two intra-user columns of $\bar{\mathbf{B}}_n$ is the sum of L groups of M sub-products. Each
 11 group of sub-products has the same phase, and there exists another group of sub-products with the
 12 opposite phase. By taking a pair of groups with opposite phases, we note that the mean value of the sum
 13 of the elements in the pair is zero due to an equal number of elements with equal magnitude statistics
 14 and opposite phases. Let U be a random variable representing an element from a pair of groups with
 15 opposite phases. Given that $\mathbb{E}[U] = 0$, we express the variance as $\text{Var}(U) = \mathbb{E}[U^2] - \mathbb{E}[U]^2 = \mathbb{E}[U^2]$.
 16 By squaring U , elements from both groups in the pair are mapped to the same phase, and the resulting
 17 variance for the pair of groups is identical to the variance of either of the individual groups, namely
 18 $1/L^2$. Since all of the elements are i.i.d., the variance of the sum of the pair of groups equals the sum
 19 of the constituent variances. Given that there are $2M$ elements in a pair of groups, the partial variance
 20 is $2M/L^2$. Applying the same property over the sum of the $L/2$ pairs of groups, we obtain the desired
 21 variance of M/L . ■
 22
 23
 24
 25
 26
 27
 28
 29
 30

31 Given that all diagonal and off-diagonal entries have the same statistics as the Wishart matrix of the
 32 single-stream case, we conclude that the single-stream result obtained in (16) also applies to the multi-
 33 stream case, when K is replaced by K_v . Hence, the UL processing gain for multi-stream detection is also
 34 represented by (17).
 35
 36
 37
 38
 39

40 *B. UL Simulation Results*

41 To verify the high- E_s/N_0 asymptote for multi-stream operation in (17), we simulate a single-cell
 42 massive MIMO scenario with $M = 16$ antennas and randomly selected channels. It is assumed that each
 43 UE power controls to the BS such that the average received signal power per antenna is the same for all
 44 UEs (i.e., identical average input E_s/N_0). We calculate the resulting processing gain from the simulation
 45 versus the average E_s/N_0 at each antenna port after noise addition. The number of UEs, K , is set to 8,
 46 16, and 24, and the average number of active streams per user is 2, resulting in a total of 16, 32, and
 47 48 virtual users (K_v), respectively. Note that by using an average number of streams per user, the UEs
 48 are assigned a uniformly distributed number of streams between 1 and twice the average minus one. The
 49 results are shown in Fig. 6. At low E_s/N_0 , the detection processing gain approaches the ideal array gain
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60

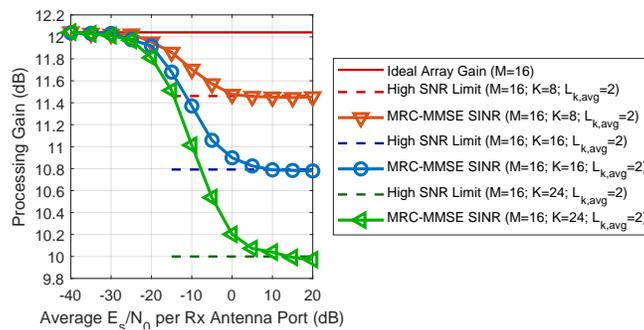


Fig. 6. Multi-stream UL simulation where $M = 16$, $L = 8$, and K is either 8, 16, or 24. The mean number of streams per UE is $\ell_k = 2$. The high- E_s/N_0 limit is also plotted in each case based on (17).

of M , as shown in [11]. We are primarily interested in the high- E_s/N_0 processing gain, which is shown to asymptotically match the expression in (17).

IV. SINGLE- AND MULTI-STREAM DL PRECODING

Single- and multi-stream precoding creates a transmit signal for the massive MIMO BS that spatially separates and delivers the targeted information to each UE. The main objectives of the MSP precoding procedure presented here are to 1) reuse the composite channel matrix and matrix inverses calculated as part of the UL detection, 2) generate a processing gain on par with UL detection, and 3) minimize the complexity at the UE terminals. Due to the TDD scenario, the reuse noted in the first objective is possible. This section details how the precoding is accomplished at the BS and the specific actions performed at each UE to extract the targeted data stream(s).

We begin by assuming that the UE stream assignments are the same as the preceding UL intervals. In [13], we showed that the same matrix inverses calculated as part of UL detection (i.e., $(\bar{\mathbf{B}}_n^H \bar{\mathbf{B}}_n + \sigma_w^2 \mathbf{I}_{K_v})^{-1}$) can be used with the regularized zero-forcing (RZF) precoding scheme for the DL. This eliminates the majority of the computational complexity associated with precoding.

The channel matrices in (7) are transposed for the DL because there are M inputs to the channel from the BS and K outputs. We likewise apply the transpose operator to the composite channel vector, $\bar{\mathbf{B}}_n$, which requires the matrix $(\bar{\mathbf{B}}_n^T \bar{\mathbf{B}}_n^* + \sigma_w^2 \mathbf{I}_{K_v})^{-1}$ for precoding. This is the complex conjugate of $(\bar{\mathbf{B}}_n^H \bar{\mathbf{B}}_n + \sigma_w^2 \mathbf{I}_{K_v})^{-1}$, which was calculated for UL detection.

The physical channel between the BS and UE does not correspond exactly to $\bar{\mathbf{B}}_n$ because $\bar{\mathbf{B}}_n$ also contains some linear phase terms used to interleave the data streams together for the UL. Consequently, the UEs are responsible for performing additional processing to accommodate the reuse of the matrix

inverse in the precoding equation. With this in mind, we recast the RZF precoder for MSP based on the RZF precoder from [13] as

$$\bar{\mathbf{x}}_n^{\text{DL}} = \beta_{\text{RZF}} \bar{\mathbf{B}}_n^* \left(\bar{\mathbf{B}}_n^T \bar{\mathbf{B}}_n^* + \sigma_w^2 \mathbf{I}_{K_v} \right)^{-1} \mathbf{Q}^{\frac{1}{2}} \bar{\mathbf{s}}_n^{\text{DL}}, \quad (18)$$

where $\bar{\mathbf{x}}_n^{\text{DL}}$ is the precoded vector for all virtual antennas in the FD, β_{RZF} is the scaling factor required to maintain the proper output power, and \mathbf{Q} is a diagonal matrix with elements that specify the optimized power for each virtual user. The ML -length vector $\bar{\mathbf{x}}_n^{\text{DL}}$ follows the structure of the submatrices of $\bar{\mathbf{B}}_n$ shown in (12), where $\bar{\mathbf{x}}_n^{\text{DL}} = [\tilde{\mathbf{x}}_{:,n}^T \tilde{\mathbf{x}}_{:,n+N/L}^T \cdots \tilde{\mathbf{x}}_{:,n+(L-1)N/L}^T]^T$. When there are multiple streams for the same physical UE, the power values in the matrix \mathbf{Q} corresponding to the same UE will be identical. As in the UL case, (18) is only valid for $n = 0, 1, \dots, N/L-1$.

The purpose of β_{RZF} is to preserve the signal power for each data stream, which is dictated by the diagonal elements of the \mathbf{Q} matrix in (18). Hence, the scale factor must compensate for the gain of the precoding by satisfying the relationship

$$\beta_{\text{RZF}} = \sqrt{\frac{K_v}{\mathbb{E}\{\text{tr}[(\bar{\mathbf{B}}_n^* (\bar{\mathbf{B}}_n^T \bar{\mathbf{B}}_n^* + \sigma_w^2 \mathbf{I}_K)^{-1})^H \bar{\mathbf{B}}_n^* (\bar{\mathbf{B}}_n^T \bar{\mathbf{B}}_n^* + \sigma_w^2 \mathbf{I}_K)^{-1}]\}}}. \quad (19)$$

Although we directly calculated β_{RZF} in [13] using numerical integration, it becomes increasingly cumbersome to perform this calculation given the larger dimensions (i.e., ML replaces M and K_v replaces K). It is more efficient to approximate the expected value in (19) through a series of random trials for each E_s/N_0 value of interest. In each trial, a random $\bar{\mathbf{B}}_n$ matrix is formed that is $ML \times K_v$ with zero-mean Gaussian entries and variance of $1/L$. The trace from the denominator of (19) is then calculated. The average value of the trace from a large number of trials will produce an estimate for the expected value. Note that β_{RZF} only needs to be calculated once for the desired E_s/N_0 operating point, so the set of values can be stored in a look-up table for future use. β_{RZF} can be approximated as $\sqrt{ML - K_v + \sigma^2} / \sqrt{M}$, which reveals the asymptotes at low and high E_s/N_0 .

Fig. 5 steps through the process to implement the precoding calculation in (18). In each transmit interval, data vectors intended for each of the K UEs are divided into streams of N/L symbols. Each of the K_v data stream vectors is converted to the FD through an N/L -length FFT. The FD vectors are reorganized to create N/L vectors with K_v entries, where each new vector takes the n^{th} element of the constituent vectors. The composite symbol vector is combined with the amplitude coefficient matrix ($\mathbf{Q}^{1/2}$), reused composite channel matrix ($\bar{\mathbf{B}}_n$), the complex conjugate of the UL matrix inverse ($(\bar{\mathbf{B}}_n^T \bar{\mathbf{B}}_n^* + \sigma_w^2 \mathbf{I}_K)^{-1}$), and the precomputed β_{RZF} value to produce the ML -length precoded vector for bin n . Note that the only inputs to (18) that change on a Tx interval basis are the composite symbol vectors, $\bar{\mathbf{s}}_n^{\text{DL}}$. All of the other inputs are calculated once per coherence period. This process is repeated for $n = 0, 1, \dots, N/L-1$, where

each bin calculation produces the M -length vectors for bins $n, n+N/L, \dots, n+(L-1)N/L$. After all bins are computed, the outputs are reorganized into M vectors of length N . As an option, each vector is spread in the FD by performing a Hadamard multiply with the diagonal elements of $\mathbf{\Omega}$. Each of the M FD vectors is then converted to the TD through an IFFT. Finally, the CP is added to each TD vector before transmission.

A. DL Power Optimization

The power optimization scheme of [13] for the conventional CP-SCM is also applicable to the MSP case with only a minor variation, which is detailed here. The power required to overcome the large-scale path loss is still p_k . Since the processing is being conducted on a stream-by-stream basis, the same power is required for each stream of a given UE. As a result, the first difference from [13] is that p_{total} includes the power of all the streams. Given that there are ℓ_k streams for the k^{th} UE, the total transmit power is expressed as

$$p_{\text{total}} = \sum_{k=1}^K \ell_k p_k. \quad (20)$$

The second implementation difference results from the $\bar{\mathbf{B}}_n$ matrix being $ML \times K_v$. Consequently, the value of K in the denominator of the equation for q_k is replaced by K_v . In addition, the computation to estimate the variance of the off-diagonal terms of $\bar{\mathbf{B}}_n^T \bar{\mathbf{B}}_n^* (\bar{\mathbf{B}}_n^T \bar{\mathbf{B}}_n^* + \sigma_w^2 \mathbf{I}_{K_v})^{-1}$ changes slightly. The variance is approximated by simulation, as was done previously. The updated calculation for q_k in the MSP context is

$$q_k = \frac{\sigma_{\text{od}}^2 p_{\text{total}} + p_k \frac{\sigma_w^2}{\beta_{\text{RZF}}^2}}{K_v \sigma_{\text{od}}^2 + \frac{\sigma_w^2}{\beta_{\text{RZF}}^2}}, \quad (21)$$

where σ_{od}^2 is the expected value of the variance of the off-diagonal elements over all channel instantiations. The K_v diagonals of \mathbf{Q} are populated by q_k according to the mapping between data streams and UEs.

B. UE Detection in the FD

The form of (18) assumes that $\bar{\mathbf{B}}_n^T$ represents the actual channels that the signals traverse. However, we know from the construction of $\bar{\mathbf{B}}_n$ in the UL case that $\bar{\mathbf{B}}_n$ includes phase modulation resulting from the interleaving of the streams. This phase modulation is not part of the physical channel, so the phase modulation portion of $\bar{\mathbf{B}}_n$ should be performed by the UE receiver. At high E_s/N_0 , the matrix inverse in (18) converges to $(\bar{\mathbf{B}}_n^T \bar{\mathbf{B}}_n^*)^{-1}$, and we see that $\bar{\mathbf{s}}_n \propto \bar{\mathbf{B}}_n^T \bar{\mathbf{x}}_n^{\text{DL}}$. This implies that $\tilde{\mathbf{s}}_{k,:,n} \propto \mathbf{B}_{k,n}^T \bar{\mathbf{x}}_n^{\text{DL}}$.

We now turn to the DL channel model in the FD provided in (11) and adapt it to the MSP framework. The composite received vector for the DL is composed of the L samples of the spectrum separated by

1
2 N/L samples, namely $\bar{\mathbf{y}}_{k,n}^{\text{DL}} = [\tilde{y}_{k,n}^{\text{DL}} \tilde{y}_{k,n+N/L}^{\text{DL}} \cdots \tilde{y}_{k,n+(L-1)N/L}^{\text{DL}}]^T$. By using (11), we represent the composite
3 received vector as

$$4 \quad \bar{\mathbf{y}}_{k,n}^{\text{DL}} = p_k^{-\frac{1}{2}} \mathbf{C}_{k,n}^T \bar{\mathbf{x}}_n^{\text{DL}} + \bar{\mathbf{w}}_{k,n}^{\text{DL}} \quad (22)$$

5
6
7 where the composite noise vector $\bar{\mathbf{w}}_{k,n}^{\text{DL}} = [\tilde{w}_{k,n} \tilde{w}_{k,n+N/L} \cdots \tilde{w}_{k,n+(L-1)N/L}]^T$, and $\mathbf{C}_{k,n}$ is the channel
8 matrix associated with the composite precoding vector. We represent this channel matrix as

$$9 \quad \mathbf{C}_{k,n} = \begin{bmatrix} 10 \quad \lambda_{:,k,n} & \mathbf{0}_{M \times 1} & \cdots & \mathbf{0}_{M \times 1} \\ 11 \quad \mathbf{0}_{M \times 1} & \lambda_{:,k,n+\frac{N}{L}} & \cdots & \mathbf{0}_{M \times 1} \\ 12 \quad \vdots & \vdots & \ddots & \vdots \\ 13 \quad \mathbf{0}_{M \times 1} & \mathbf{0}_{M \times 1} & \cdots & \lambda_{:,k,n+(L-1)\frac{N}{L}} \end{bmatrix}. \quad (23)$$

14
15
16
17 The next step is to factor $\mathbf{B}_{k,n}$ in terms of $\mathbf{C}_{k,n}$ to define the additional processing performed at the
18 UE. Upon careful inspection, one will find that $\mathbf{B}_{k,n} = \mathbf{C}_{k,n} \mathbf{D}_{k,n}$, where $\mathbf{D}_{k,n}$ is the combining matrix for
19 the k^{th} UE defined as

$$20 \quad \mathbf{D}_{k,n} = \begin{bmatrix} 21 \quad 1 & \omega^n & \cdots & \omega^{(\ell_k-1)n} \\ 22 \quad 1 & \omega^{n+\frac{N}{L}} & \cdots & \omega^{(\ell_k-1)(n+\frac{N}{L})} \\ 23 \quad \vdots & \vdots & \ddots & \vdots \\ 24 \quad 1 & \omega^{n+(L-1)\frac{N}{L}} & \cdots & \omega^{(\ell_k-1)(n+(L-1)\frac{N}{L})} \end{bmatrix}, \quad (24)$$

25
26
27 where there is one column for each of the ℓ_k data streams assigned to the k^{th} UE.

28
29
30 With the bin-specific combining matrix defined, we can address the scalar values needed to produce
31 an unbiased estimate of the FD symbols for each stream. First, the result is multiplied by $p_k^{1/2}$ to offset
32 the path loss of the channel. The result is divided by $q_k^{1/2}$ to undo the power optimization term used by
33 the transmitter, and the result is divided by β_{RZF} to undo the scale factor that was used by the BS to
34 maintain the desired transmit power. Next, the result is multiplied by α , which is the receive-side scale
35 factor that accounts for the scaling bias resulting from the regularization factor in the matrix inverse used
36 for precoding. Finally, the result is divided by \sqrt{L} , which is a factor applied when forming $\bar{\mathbf{B}}_n$ from the
37 submatrices, $\mathbf{B}_{k,n}$. The resulting vector of FD symbol estimates for the k^{th} UE is

$$38 \quad \hat{\mathbf{s}}_{k,n}^{\text{DL}} = \frac{\alpha p_k^{1/2}}{\sqrt{L} \beta_{\text{RZF}} q_k^{1/2}} \mathbf{D}_{k,n}^T \bar{\mathbf{y}}_{k,n}^{\text{DL}}. \quad (25)$$

39
40
41 The scalar values in (25) provide an exact scaling that is useful for analysis and simulation. In practice,
42 the receiver will not know the value of q_k unless it is sent by the BS ahead of time. Since the values of
43 p_k and q_k are very close, they can be dropped if the UE receiver uses an automatic gain control circuit.

44
45
46 Based on (25), we find that the FD symbol estimates corresponding to the n^{th} bin are based upon
47 taking every N/L samples of $\tilde{\mathbf{y}}_k^{\text{DL}}$, starting with the n^{th} sample. If we analyze the FD symbol estimates

for the l^{th} data stream, we find that each data stream has a different set of linear phases that follow the same pattern as the UL transmitted signal. Recalling the use of Ψ in (5), the FD symbol estimate vector for the l^{th} stream can be concisely represented as

$$\hat{\mathbf{s}}_{k,l}^{\text{DL}} = \frac{\alpha p_k^{1/2}}{\sqrt{L}\beta_{\text{RZF}}q_k^{1/2}} \mathbf{T}_{N,L}^{\text{T}} \Psi^{(l-1)} \tilde{\mathbf{y}}_k^{\text{DL}}, \quad (26)$$

where $\hat{\mathbf{s}}_{k,l}^{\text{DL}}$ is the $(N/L) \times 1$ vector of FD symbol estimates for the l^{th} stream ($l = 1$ to ℓ_k), and $\tilde{\mathbf{y}}_k^{\text{DL}}$ is defined in (10). After performing the calculation in (26) for each of the active streams for the k^{th} UE, the N/L -length FD vectors are transformed into the TD using the IFFT to obtain the final symbol estimates.

C. UE Detection in the TD

The description in the preceding subsection requires that the received signal must be converted to the FD before the processing begins. An alternate approach uses the special structure of the received signal to extract the data from the TD received signal. The operations performed by the tiling matrix and the diagonal phase matrix in (26) are very similar to the FD definition of the circular shift of the expander matrix in (4), but instead of expanding the data, the operation is combining samples to get the FD symbol estimates. In the TD, we see that the transpose of the expander matrix is used to sample a larger vector. Since the entries of the expander matrix are real, the conjugate transpose (i.e., Hermitian) can be taken in place of the transpose. Applying the Hermitian operator to both sides of (4) and pre- and post-multiplying the result by $\mathcal{F}_{N/L}$ and \mathcal{F}_N , respectively, we get

$$\mathcal{F}_{N/L} \mathbf{E}_{L(l-1)}^{\text{T}} = \frac{1}{\sqrt{L}} \mathbf{T}_{N,L}^{\text{T}} \Psi^{-(l-1)} \mathcal{F}_N,$$

where we can replace $l-1$ with $-(l-1)$ to obtain

$$\mathcal{F}_{N/L} \mathbf{E}_{L(-(l-1))}^{\text{T}} = \frac{1}{\sqrt{L}} \mathbf{T}_{N,L}^{\text{T}} \Psi^{(l-1)} \mathcal{F}_N. \quad (27)$$

We see from (27) that the FD combining operation in (26) is equivalent to a sampling operation in the TD with a shift of $-(l-1)$, where l is in the range $[1, \ell_k]$. Hence for $l > 1$, the shift is actually negative. The expression for the l^{th} data stream in the TD is

$$\hat{\mathbf{s}}_{k,l}^{\text{DL}} = \frac{\alpha p_k^{1/2}}{\beta_{\text{RZF}}q_k^{1/2}} \mathbf{E}_{L(-(l-1))}^{\text{T}} \mathbf{y}_k^{\text{DL}}. \quad (28)$$

Note that the samples available in the TD are the same as those that would be obtained by doing the extra processing in the FD. The TD approach greatly simplifies the computations performed by the UE because it is based on indexing rather than FD processing. Fig. 3 shows the simple operations performed by the UE receiver. If spreading is applied to the transmitted signal, then it is still more efficient to despread in the FD and perform the deinterleaving of the streams in the TD.

D. DL Performance Analysis

To evaluate the performance of precoding, we follow the UL analysis for the definition of the processing gain. Focusing on the high- E_s/N_0 asymptote to the processing gain, we analyze the noise component of the symbol estimates in (25). In the high- E_s/N_0 regime, α converges to unity, and the regularization term in the matrix inverses of (18) and (19) vanishes. Without loss of generality, we make the simplifying assumption that the large-scale path loss to the k^{th} UE, p_k^{-1} , is the same for all UEs (i.e., $p_k = p$). This assumption is warranted because the DL power optimization achieves the same performance as the equal path loss case, which was shown in [13]. By substituting $p_k = p$ in (20), we find that $p_{\text{total}} = pK_v$. Using these substitutions in (21), we find that $q_k = p$. Hence, the matrices \mathbf{P} and \mathbf{Q} can be replaced by the scalar value p . Based on (25), the symbol estimates are represented as

$$\begin{aligned}
\hat{\mathbf{s}}_{k,n}^{\text{DL}} &= \frac{\alpha p^{\frac{1}{2}}}{\sqrt{L}\beta_{\text{RZF}}p^{\frac{1}{2}}}\mathbf{D}_{k,n}^{\text{T}}\bar{\mathbf{y}}_{k,n}^{\text{DL}} \\
&= \frac{1}{\sqrt{L}\beta_{\text{RZF}}}\mathbf{D}_{k,n}^{\text{T}}p_k^{-\frac{1}{2}}\mathbf{C}_{k,n}^{\text{T}}\bar{\mathbf{x}}_n^{\text{DL}} + \frac{1}{\sqrt{L}\beta_{\text{RZF}}}\mathbf{D}_{k,n}^{\text{T}}\bar{\mathbf{w}}_{k,n}^{\text{DL}} \\
&= \frac{p^{-\frac{1}{2}}}{\sqrt{L}\beta_{\text{RZF}}}\mathbf{B}_{k,n}^{\text{T}}\beta_{\text{RZF}}\bar{\mathbf{B}}_n^* \left(\bar{\mathbf{B}}_n^{\text{T}}\bar{\mathbf{B}}_n^*\right)^{-1} p^{\frac{1}{2}}\bar{\mathbf{s}}_n^{\text{DL}} + \frac{1}{\sqrt{L}\beta_{\text{RZF}}}\mathbf{D}_{k,n}^{\text{T}}\bar{\mathbf{w}}_{k,n}^{\text{DL}} \\
&= \bar{\mathbf{s}}_{k,n}^{\text{DL}} + \frac{1}{\sqrt{L}\beta_{\text{RZF}}}\mathbf{D}_{k,n}^{\text{T}}\bar{\mathbf{w}}_{k,n}^{\text{DL}}, \tag{29}
\end{aligned}$$

where substitutions from (22) and (18) were made in the second and third expressions, respectively. In the third expression, $\mathbf{B}_{k,n}^{\text{T}}$ is substituted for $\mathbf{D}_{k,n}^{\text{T}}\mathbf{C}_{k,n}^{\text{T}}$. The result is an unbiased estimate of the FD symbols plus noise. The noise vector with L zero-mean elements with a variance of σ_w^2 is combined via the $\mathbf{D}_{k,n}^{\text{T}}$ matrix. All of the elements of $\mathbf{D}_{k,n}^{\text{T}}$ have unit magnitude, so the variances of the $\bar{\mathbf{w}}_{k,n}^{\text{DL}}$ elements are summed. However, due to the scale factor of $1/\sqrt{L}$, the result is still a variance of σ_w^2 prior to accounting for the β_{RZF} term. In the high- E_s/N_0 regime, the value of β_{RZF} converges to $\sqrt{M-K_v/L}$, since the denominator in (19) reduces to $\mathbb{E}\{\text{tr}[\bar{\mathbf{B}}_n^{\text{T}}\bar{\mathbf{B}}_n^*]\}$. This is the same result explained in Section III-A for the UL analysis. Hence, the noise power is divided by $M-K_v/L$, and the processing gain for the DL is

$$G^{\text{DL}} = M - \frac{K_v}{L}. \tag{30}$$

E. DL Simulation Results

The DL simulation for MSP is conducted similarly to the UL simulation with identical values for M and L . The number of users is fixed at $K = 16$, and the average number of data streams is set to 1, 2, and 3. The simplifying assumption of equal large-scale path loss, used in the analysis above, is removed. Path loss values are allowed to differ by up to 20 dB. The results plotted in Fig. 7 show that the high- E_s/N_0

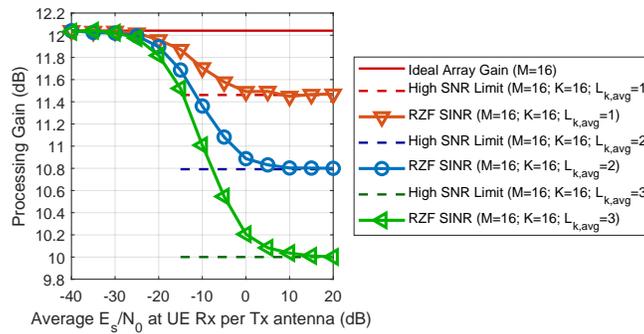


Fig. 7. Multi-stream DL simulation results where $M = 16$, $K = 16$, and $L = 8$. The average value of ℓ_k ranges from 1 to 3. The high- E_s/N_0 processing gain asymptotes from (30) are plotted for comparison.

asymptote from (30) still applies despite the power variations due to the power optimization scheme described in Section IV-A. Note that these DL results are basically identical to the UL results shown in Fig. III-B, which demonstrates that the same performance is attainable when each UE is assigned the same number of data streams (UL simulation) or when the number of data streams per UE is allowed to vary as long as the total number of data streams is maintained (DL simulation).

V. MSP BENEFITS AND USE CASES

MSP produces the following set of desirable characteristics based on the fact that it is a single carrier modulation that occupies all N frequency bins instead of a subset of those bins:

- The creation of virtual antennas through replication of the information-bearing spectrum improves the massive MIMO channel hardening effect. This means that the variance of small-scale fading is reduced due to the channel being averaged over several antennas [29]. The number of virtual antennas can enter into the range of massive MIMO (e.g., ≥ 64) with fewer physical antennas (e.g., $M = 8$ and $L = 8$). This clearly reduces deployment costs.
- Because of their spread spectrum nature, MSP signals have the potential to coexist with other signals in the same spectrum by transmitting at a low enough power to keep the interference at a manageable level.
- The spectral redundancy of the MSP signals allows for portions of the spectrum with heavy interference to be rejected through masking without causing a significant impairment.

We now introduce three different use cases that highlight some of the benefits of MSP. These use cases rely on the fact that the MSP signal is spread over a relatively wide bandwidth. For example, the MSP signal could occupy the same bandwidth as a 5G NR system (e.g., 2048 subcarriers at 15 kHz spacing).

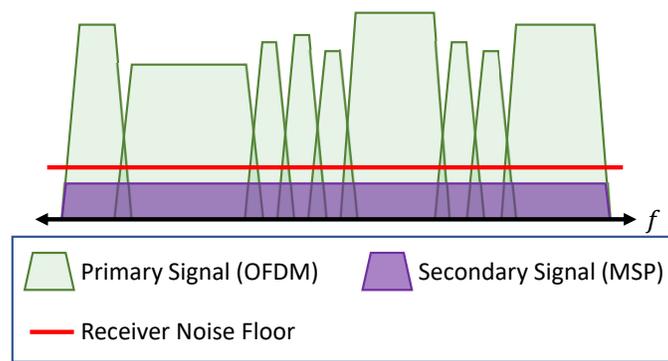


Fig. 8. Received spectrum at primary BS with a low-power MSP signal from a secondary network. Here, the power of the MSP signal is set to be below the BS receiver's noise floor to produce a negligible impairment to the primary network.

A. Low-interference Secondary Network

By transmitting at a low data rate and a low power, MSP can function as a secondary network within the same spectrum as a primary network. This is a great match to the requirements of mMTC. Because MSP is spectrally flat, the MSP secondary network causes little interference to the primary network as long as the MSP signal's power spectral density (PSD) is low enough. Part of the PSD reduction results from the data being spread in bandwidth by a factor of L . The second contributor to lower PSD is the gain associated with the multiple antennas at the MSP base station.

Another critical aspect in managing interference to the primary network is the path loss between the MSP emitters and the primary network receivers. Because the primary network UEs are mobile and distributed throughout the primary cell, managing interference to the primary network becomes simpler when focusing on the primary BS receiver instead of the primary UE receivers. Hence, as secondary users, the MSP emitters could limit transmissions to periods that are scheduled for UL transmissions from the primary UEs. In this way, the secondary users can limit the interference to the primary BS, similar to what is illustrated in Fig. 8. The interference caused by the primary UEs to the MSP receivers is addressed in the next subsection.

The flexibility of MSP provides additional options for communicating during the periods schedule for DL transmissions from the primary BS. Some of these aspects are provided in the next subsection, but the details of this low-interference secondary network requires additional attention which is beyond the scope of this paper. This subject was initially addressed in [20], but the inclusion of MSP and further details will be addressed in a future work.

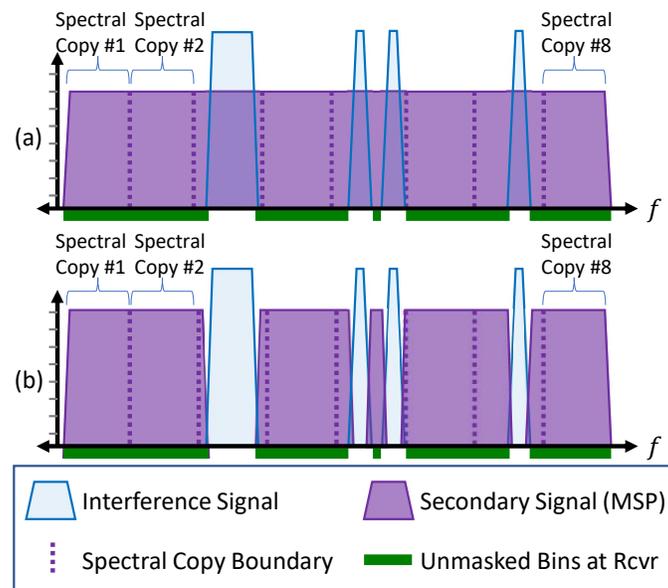


Fig. 9. Spectra illustrating interference mitigation with MSP. In spectrum (a), the MSP signal spans the targeted spectrum (i.e., no transmitter cooperation). The MSP receiver masks the spectrum in the FD to reject the interference. In other words, frequency bins containing interference are not used for detection. In spectrum (b), the transmitted MSP signal is spectrally masked to avoid interfering with other signals. The MSP receiver performs interference rejection as in case (a).

B. Interference Mitigation

The second use case addresses a scenario where MSP operates in a portion of the spectrum where there is high-power interference due to incumbent signals. The incumbent signals could be stationary in frequency or time-varying. We assume that the incumbent signals do not occupy the entirety of the targeted spectrum. The MSP receiver has the ability to sense high-power interference and then to mask out the affected frequency bins. By masking the spectrum, the MSP receiver reduces the effective number of virtual antennas. However, since the processing gain shown in (17) and (30) is primarily dependent upon the number of physical antennas, M , the impact of spectral masking is quite tolerable. Fig. 9 part (a) illustrates the spectrum as viewed by the MSP receiver. Bins with high interference are not incorporated into the detection calculations as a result of masking.

If the incumbent signals are somewhat stationary such that their spectral positions are known to both the MSP transmitter and the MSP receiver, then the MSP transmitter can mask the spectrum prior to transmission. Fig. 9 part (b) illustrates the spectrum at the receiver for this second case. Coordinated spectral masking provides two benefits. First, the transmitted signal power can be concentrated into the frequency bins that will be used by the receiver. Second, masking the transmitted signal protects the incumbent signals from interference caused by the MSP network. The details of interference avoidance

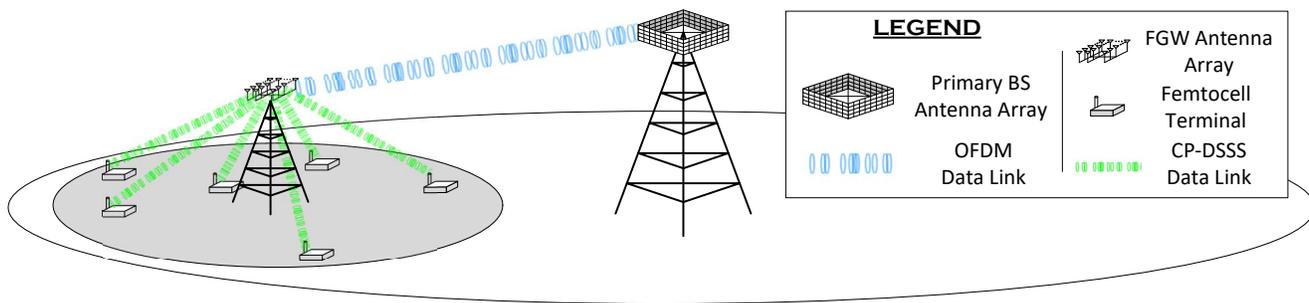


Fig. 10. Heterogeneous MSP scenario. The gray oval represents the femtocell within the primary cell. Not to scale.

is presented in [21].

C. Heterogeneous Signal Operation

The secondary network solution presented in [20] features a hub at the center of each femtocell that operates simultaneously on the primary and secondary wireless networks. As such, the hub acts as a gateway and is referred to as a femtocell gateway (FGW). The FGW communicates on the primary network using OFDM while simultaneously communicating to several FTs on a secondary network using MSP. A simplified illustration is provided in Fig. 10. The primary and secondary networks operate using the same spectrum. At the FGW, both primary and secondary signals are received during the same time interval, which is the DL interval for the primary network and the UL interval for the secondary network. Likewise, the primary and secondary signals are transmitted by the FGW during the same interval. The processing defined in [20] did not take advantage of the improved processing gain nor the resource allocation flexibility associated with MSP. These improvements are explained in the following section. In the FGW scenario, there is only one OFDM signal that occupies a set of contiguous frequency bins, which will be our focus in the next section. However, these methods are extensible to multiple OFDM signals for other applications.

VI. HETEROGENEOUS MULTI-STREAM PROCESSING

This section provides the details needed to process the OFDM signals in the MSP framework by dividing the frequency bins (i.e., subcarriers) into data streams of length N/L or less, as is done for MSP users. For simplicity and consistency with the scenario in [20], we assume that the OFDM pilot signals are sent separately from the MSP pilot signals. Since the OFDM pilot is limited to the subcarrier assignment, all frequency bins outside of the OFDM signal's bandwidth assignment are represented as zeros in the composite channel matrix that will follow. In contrast, the MSP pilots occupy the full bandwidth (i.e., N

bins). We continue to assume channel reciprocity due to TDD operation and perfect CSI in the following subsections.

A. Heterogeneous UL Detection

The MSP UL detection using the MRC-MMSE detector is defined in (15). The same structure is used for heterogeneous operation, but $\bar{\mathbf{B}}_n$ in (13) is modified to add an entry for the OFDM UE, which will be designated as the K^{th} user. In the MSP construction, each stream consists of N/L symbols and is assigned a column in the $\mathbf{B}_{k,n}$ submatrix (12). Likewise, for the OFDM user, the assigned subcarriers are divided into contiguous groups of N/L with all groups except possibly the last containing exactly N/L bin assignments. These groups will also be referred to as data streams. Since the last stream may have less than N/L bins assigned, it may not be involved in the calculation for each value of n . To simplify the discussion going forward, we assume that the number of OFDM subcarriers in the assigned RBs divided by N/L is the integer ℓ_K .

We now form the columns of $\mathbf{B}_{K,n}$ for the OFDM user. Given that each stream has N/L entries, one of the bins from each stream maps to one of the ℓ_K columns associated with $\bar{\mathbf{B}}_n$, for all n from 0 to $N/L-1$. The lowest OFDM subcarrier need not start at the beginning of an N/L boundary. Unlike the entries for MSP data streams, there are only M channel coefficients corresponding to each stream, and all of the other entries in each column are zero. Each section of M rows in the $\mathbf{B}_{K,n}$ submatrix represents the FD channel coefficients for a different part of the spectrum, equally spaced by N/L bins. Because each data stream has N/L symbols, the M non-zero entries for each column are not in the same rows as the non-zero entries from other columns in $\mathbf{B}_{K,n}$. Consequently, the columns of $\mathbf{B}_{K,n}$ are orthogonal.

Given that there are ℓ_K data streams for the OFDM UE, we designate δ as the index of the start bin of assigned RBs and note that the last bin is index $\delta + \ell_K N/L - 1$. The submatrix $\mathbf{B}_{K,n}$ will incorporate entries for all OFDM bins in the interval $[\delta, \delta + \ell_K N/L - 1]$ where the modulo- N/L result equals n . We represent these bins as $n + \epsilon N/L, n + (\epsilon + 1)N/L, \dots, n + (\epsilon + \ell_K - 1)N/L$. The M channel coefficients that compose the non-zero elements of the first column of $\mathbf{B}_{K,n}$ are denoted as $\lambda_{:,K,n+\epsilon N/L}$. The value of $\epsilon + 1$ specifies the section of M rows where $\lambda_{:,K,n+\epsilon N/L}$ is placed. For example, if $\epsilon = 0$, then $\lambda_{:,K,n}$ occupies the first M rows. The location of the M channel coefficients in subsequent columns is shifted down by

M . The general form of $\mathbf{B}_{K,n}$ for the OFDM UE is represented as

$$\mathbf{B}_{K,n} = \begin{bmatrix} \mathbf{0}_{\epsilon M \times 1} & \mathbf{0}_{\epsilon M \times 1} & \cdots & \mathbf{0}_{\epsilon M \times 1} \\ \lambda_{:,K,n+\epsilon \frac{N}{L}} & \mathbf{0}_{M \times 1} & \cdots & \mathbf{0}_{M \times 1} \\ \mathbf{0}_{M \times 1} & \lambda_{:,K,n+(\epsilon+1) \frac{N}{L}} & \cdots & \mathbf{0}_{M \times 1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{M \times 1} & \mathbf{0}_{M \times 1} & \cdots & \lambda_{:,K,n+(\epsilon+\ell_K-1) \frac{N}{L}} \\ \mathbf{0}_{(L-\epsilon-\ell_K)M \times 1} & \mathbf{0}_{(L-\epsilon-\ell_K)M \times 1} & \cdots & \mathbf{0}_{(L-\epsilon-\ell_K)M \times 1} \end{bmatrix}, \quad (31)$$

where the $(L-\epsilon-\ell_K)M$ -length zero vectors are due to there being ML rows in the matrix. Note that the value of ϵ will either be the same for all values of n , or it will be one of two consecutive values.

With the addition to the composite channel matrix noted above, heterogeneous UL detection can proceed using (15) and the signal flow in Fig. 4. However, there are three adjustments necessary for the OFDM signal. First, since the OFDM symbols are defined in the FD, they are not converted to the TD using the N/L -point IFFT. Second, the BS must circularly shift the N/L received symbols upward by δ modulo- N/L . The circular shift accounts for where δ does not line up with the start of an N/L section boundary. Third, if the optional spreading is used for the MSP UL signals, then despreading will be applied for each BS antenna in the FD using the diagonal matrix $\mathbf{\Omega}^*$ defined in Section II. Because the OFDM signal was not spread by the UE, the diagonal elements δ through $\delta+N/L-1$ of $\mathbf{\Omega}$ are Hadamard multiplied with the detected symbols of the first data stream. The diagonal elements $\delta+(i-1)N/L$ through $\delta+iN/L-1$ of $\mathbf{\Omega}$ are Hadamard multiplied with the post-detection symbols of the i^{th} data stream, for all ℓ_K data streams of the OFDM signal. This final multiplication undoes the effect of the despreading that was applied to all of the received signals.

B. Heterogeneous UL Performance Analysis

The changes to the composite channel matrix for heterogeneous operation do not significantly alter the statistics of the $\overline{\mathbf{B}}_n^H \overline{\mathbf{B}}_n$ matrix. The M non-zero entries in each column of $\mathbf{B}_{K,n}$ are zero-mean, unit-variance, complex Gaussian variables. The columns of $\overline{\mathbf{B}}_n$ corresponding to the MSP users have ML entries that are also zero-mean, complex Gaussian variables with a variance of $1/L$. The diagonals of $\overline{\mathbf{B}}_n^H \overline{\mathbf{B}}_n$ corresponding to the K^{th} UE are a summation of M non-zero squared magnitudes with unit mean, resulting in an expected value of M . This is consistent with the diagonal elements corresponding to the UEs using MSP. The expected value for the inter-user off-diagonals of $\overline{\mathbf{B}}_n^H \overline{\mathbf{B}}_n$ that involve the K^{th} user is still M/L because M non-zero values are summed, where each value has a variance of $1/L$. Hence, given a single-stream OFDM UE, $\overline{\mathbf{B}}_n^H \overline{\mathbf{B}}_n$ has the same statistics as a Wishart matrix, and the processing gain in (17) still applies.

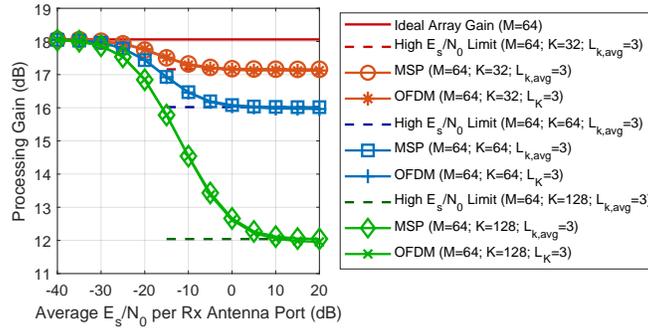


Fig. 11. Multi-stream heterogeneous UL simulation where $M = 64$, $L = 8$, and $\ell_{k,\text{avg}} = 3$ for all k . The number of users is $K = 32$, $K = 64$, and $K = 128$. Separate curves are plotted for the average processing gain for the MSP and OFDM UEs.

The main difference in the statistics of $\overline{\mathbf{B}_n^H \mathbf{B}_n}$ for a multi-stream OFDM UE is due to the intra-user off-diagonals resulting from the K^{th} user (i.e., $\ell_K > 1$). Given that the columns of $\mathbf{B}_{K,n}$ are orthogonal, the intra-user off-diagonals for the K^{th} UE are zero. This results in a lower level of inter-stream interference for the K^{th} UE's signal. However, the inter-user interference from the other $K-1$ UEs is still present. Accordingly, the high- E_s/N_0 asymptote in (17) is still a good lower bound for the processing gain in the OFDM UE's case.

C. Heterogeneous UL Simulation

Simulations were conducted to show the performance of MSP streams and OFDM streams in a heterogeneous processing scenario. There are $M = 64$ antennas, and the upsample factor is $L = 8$. The number of users is 32, 64, and 128. The average number of streams per MSP UE is three. The OFDM user was assigned three streams in each simulation run (i.e., $3N/L$ symbols). The first bin of the OFDM frequency assignment is uniformly distributed over the first $N-3N/L+1$ bins. All MSP signals use ZC spreading (i.e., CP-DSSS) to show that the despreading can be applied to the received input vectors and undone for the OFDM data streams. UL power control was implemented such that the average received power of each data stream signal was equal for all UEs.

The results are shown in Fig. 11. The processing gain for MSP and OFDM signals are plotted on the same graph to show virtually identical performance. The predicted asymptote at high SNR (17) is also plotted to show the accuracy of the analysis in the presence of heterogeneous signals.

D. Heterogeneous DL Precoding

As in Section IV, we assume that the DL assignments are identical to the UL assignments for the TDD scenario. Consequently, the DL heterogeneous precoding uses the complex conjugate of the matrix

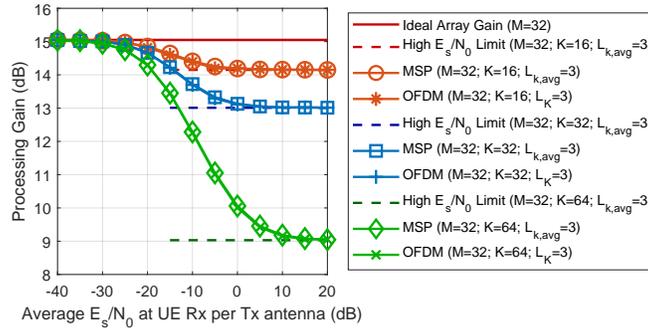


Fig. 12. Multi-stream heterogeneous DL simulation where $M = 32$, $L = 8$, and $L_{k,avg} = 3$ for all k . The number of users is $K = 16$, $K = 32$, and $K = 64$. The MSP UEs have the same processing gain as the OFDM UEs.

inverses calculated for UL heterogeneous detection. Precoding follows the process shown in Fig. 5, except that the OFDM data streams are handled differently than the MSP data streams in three ways. First, the OFDM data streams are not converted to the FD because the symbols are assigned directly to subcarriers in the FD. Second, if the optional spreading is applied by the BS precoder, then the OFDM symbols must be prescaled using the diagonal elements of the FD despreading matrix, $\mathbf{\Omega}^*$. Specifically, the elements δ through $\delta+N/L-1$ of $\mathbf{\Omega}^*$ are Hadamard multiplied with the OFDM symbols of the first data stream, and the elements $\delta+(i-1)N/L$ through $\delta+iN/L-1$ of $\mathbf{\Omega}^*$ are Hadamard multiplied with the OFDM symbols of the i^{th} data stream, for all ℓ_K data streams corresponding to the OFDM UE. Third, the data streams are circularly shifted downward by δ modulo- N/L so that the data will arrive at the UE in the expected order (i.e., no special processing is required at the OFDM UE). With these modifications to the OFDM data stream(s), the precoding process can proceed as outlined in Section IV. Based on the analysis provided in Sections IV-D and VI-B, the high- E_s/N_0 asymptote for both signal types is expected to follow (30).

E. Heterogeneous DL Simulation

DL simulations that parallel the heterogeneous UL simulations described in Section VI-C were conducted to show the average processing gain associated with the MSP and OFDM data streams. The results are plotted in Fig. 12 in relation to the high- E_s/N_0 asymptotes. As with the UL case, all MSP signals use ZC spreading to model the CP-DSSS waveform. The spreading is applied to the M output vectors in the FD using the diagonal elements of $\mathbf{\Omega}$. The OFDM data streams are prescaled with the appropriate elements of the FD despreading sequence to counteract the spreading that is applied across the outputs. The results show that MSP signals achieve the same average processing gain as the OFDM signals, consistent with the predicted high- E_s/N_0 asymptote in (30).

VII. CONCLUSION

We presented the details of processing CP-SCM waveforms in the FD with flexible resource allocation called MSP. The number of potential symbols per interval, N , was divided into L potential data streams. Each UE transmitted/received using the number of data streams dictated by the BS as part of resource allocation. We explained how the MSP technique creates L virtual antennas from each physical antenna. With the increased number of virtual antennas, the total number of simultaneous data streams can exceed the number of physical antennas. We showed an improvement in the processing gain for MSP over conventional CP-SCM for both UL and DL when a subset of N potential symbols is sent. Simulation results confirmed the results predicted in the analysis.

While it is not our intent to replace OFDM-based waveforms in future standards, we proposed some new scenarios where operation with MSP signals may be beneficial. In an mMTC scenario, we described a low-interference secondary network that uses the same spectrum as the primary network. The interference avoidance scenario was also introduced, where the MSP signals can use spectral masking to limit the effect of interference from/to incumbent signals in the spectrum. Lastly, we discussed a scenario where the multi-antenna hub simultaneously communicated using a mixture of MSP and OFDM signals. We then presented a method for detecting UL OFDM signals and precoding DL OFDM signals within the MSP framework. This heterogeneous set of signals allows MSP to coexist with traditional 4G and 5G signals, giving further flexibility to this resource allocation solution for CP-SCM signals.

REFERENCES

- [1] S. Parkvall, E. Dahlman, A. Furuskar and M. Frenne, "NR: The new 5G radio access technology," *IEEE Communications Standards Magazine*, vol. 1, no. 4, pp. 24-30, Dec 2017.
- [2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.
- [3] F. Rusek et al., "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40-60, Jan. 2013.
- [4] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742-758, Oct. 2014.
- [5] T. L. Marzetta, "Massive MIMO: An Introduction," *Bell Labs Technical Journal*, vol. 20, pp. 11-22, 2015.
- [6] J. Hoydis, S. ten Brink and M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160-171, Feb. 2013.
- [7] Thomas Marzetta, Erik Larsson, Hong Yang, and Hien Ngo, *Fundamentals of Massive MIMO*, Cambridge University Press, 1st Edition, 2016.
- [8] N. Benvenuto, R. Dinis, D. Falconer and S. Tomasin, "Single Carrier Modulation With Nonlinear Frequency Domain Equalization: An Idea Whose Time Has Come—Again," *Proceedings of the IEEE*, vol. 98, no. 1, pp. 69-96, Jan. 2010.

- 1
2 [9] Z. Mokhtari, M. Sabbaghian and R. Dinis, "Massive MIMO Downlink Based on Single Carrier Frequency Domain Processing," *IEEE*
3 *Transactions on Communications*, vol. 66, no. 3, pp. 1164-1175, Mar. 2018.
- 4 [10] Z. Mokhtari, M. Sabbaghian and R. Dinis, "Downlink Sum-Rates of SC-FDP and OFDM Massive MIMO Systems in Doubly Dispersive
5 Channels," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8065-8079, Aug. 2020.
- 6 [11] B. A. Kenney, A. J. Majid, H. Moradi and B. Farhang-Boroujeny, "Frequency Domain Detection and Precoding for Massive MIMO
7 with Single Carrier Modulation," *IEEE Transactions on Wireless Communications*, vol. 21, no. 5, pp. 3232-3248, May 2022.
- 8 [12] B. A. Kenney, A. J. Majid, H. Moradi and B. Farhang-Boroujeny, "Frequency Domain Multi-User Detection for Single Carrier
9 Modulation with Cyclic Prefix," *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1-6.
- 10 [13] B. A. Kenney, A. J. Majid, H. Moradi and B. Farhang-Boroujeny, "Efficient Precoding for Single Carrier Modulation in Multi-User
11 Massive MIMO Networks," *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1-6.
- 12 [14] B. A. Kenney, A. J. Majid, H. Moradi and B. Farhang-Boroujeny, "Resource Allocation for Single Carrier Massive MIMO Systems,"
13 *ICC 2022 - IEEE International Conference on Communications*, 2022, pp. 950-955.
- 14 [15] Y. Chen, B. Wang, Y. Han, H. Lai, Z. Safar and K. J. R. Liu, "Why Time Reversal for Future 5G Wireless? [Perspectives]," *IEEE*
15 *Signal Processing Magazine*, vol. 33, no. 2, pp. 17-26, March 2016.
- 16 [16] Y. Han, Y. Chen, B. Wang and K. J. R. Liu, "Time-Reversal Massive Multipath Effect: A Single-Antenna "Massive MIMO" Solution,"
17 *IEEE Transactions on Communications*, vol. 64, no. 8, pp. 3382-3394, Aug. 2016.
- 18 [17] Q. Xu, C. Jiang, Y. Han, B. Wang and K. J. R. Liu, "Waveforming: An Overview With Beamforming," *IEEE Communications Surveys*
19 *& Tutorials*, vol. 20, no. 1, pp. 132-149, Firstquarter 2018.
- 20 [18] A. Aminjavaheri, A. RezazadehReyhani, R. Khalona, H. Moradi and B. Farhang-Boroujeny, "Underlay Control Signaling for Ultra-
21 Reliable Low-Latency IoT Communications," *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*,
22 2018, pp. 1-6.
- 23 [19] B. A. Kenney, S. N. Jenkins, A. J. Majid, H. Moradi and B. Farhang-Boroujeny, "Cyclic Prefix Direct Sequence Spread Spectrum
24 Capacity Analysis," *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, 2020, pp. 1-6.
- 25 [20] B. A. Kenney, A. J. Majid, H. Moradi and B. Farhang-Boroujeny, "Opportunistic Secondary Communications Downlink Using Multi-
26 Antenna Femtocell Gateways," *2021 17th International Conference on Wireless and Mobile Computing, Networking and Communications*
27 *(WiMob)*, 2021, pp. 283-288.
- 28 [21] B. A. Kenney, A. J. Majid, H. Moradi and B. Farhang-Boroujeny, "Interference Mitigation for Multi-Stream Processing with Massive
29 MIMO". TechRxiv, 26-Sep-2022, doi: 10.36227/techrxiv.21137188.v1.
- 30 [22] A. Goldsmith, *Wireless Communications*, Cambridge University Press, 2005, pp. 383-392.
- 31 [23] D. Tse, P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005, pp. 95-101.
- 32 [24] Behrouz Farhang-Boroujeny, *Signal Processing Techniques for Software Radios*, 2nd Edition, Lulu Publishing House, 2010, pp.103.
- 33 [25] A. V. Oppenheim, R W. Schaffer, with J. R. Buck, *Discrete-time Signal Processing*, 2nd Edition, Prentice Hall, 1998, pp. 564-567.
- 34 [26] Robert M. Gray, "Toeplitz and Circulant Matrices: A Review," *Foundations and Trends in Communications and Information Theory*,
35 Vol. 2, No. 3, pp 155-239.
- 36 [27] A. Lozano, A. M. Tulino and S. Verdu, "Multiple-antenna capacity in the low-power regime," *IEEE Transactions on Information*
37 *Theory*, vol. 49, no. 10, pp. 2527-2544, Oct. 2003.
- 38 [28] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, Second Edition, USA: McGraw-Hill, 1984, pp. 111, 153.
- 39 [29] E. Björnson, E. G. Larsson and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," *IEEE Communications Magazine*,
40 vol. 54, no. 2, pp. 114-123, February 2016.
- 41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

University of Utah
Department of Electrical and Computer Engineering
50 S. Central Campus Drive
MEB Room 2110
Salt Lake City, UT 84112

Assigned Editor and Undisclosed Reviewers
IEEE Transactions on Communications

September 27, 2022

Dear Editor and Reviewers,

Thank you for your willingness to conduct this review. The purpose of this supplemental information is to provide some additional details on our manuscript, "Single Carrier Modulation Resource Allocation for Massive MIMO with Virtual Antennas" and to make a case for the novelty of the manuscript. We have elected to call this new method of resource allocation Multi-Stream Processing (MSP) because the resources are divided into data streams. Each user equipment (UE) is assigned one or more streams, thus creating a flexible resource allocation solution for cyclic prefixed single carrier modulation (CP-SCM). The analog to MSP data streams for existing OFDM standards is resource blocks. To our knowledge, this is the first time that flexible resource allocation has been presented for CP-SCM waveforms in a massive MIMO time domain duplexed (TDD) scenario.

As we stated in the cover letter to the editor, this paper is an expanded version of our ICC 2022 symposium paper entitled "Resource Allocation for Single Carrier Massive MIMO Systems." The 2022 ICC symposium paper was based on MSP uplink detection. This manuscript includes a refined version of the uplink solution for completeness. We also provide further analysis to explain the uplink detection results observed in simulation. This manuscript goes beyond the ICC symposium paper by also detailing the MSP precoding solution for the TDD downlink. In addition, we provide three examples of scenarios where CP-SCM may be beneficial in a massive MIMO context. One of these scenarios involves a mixture of signals (OFDM and MSP). We show how the MSP detection and precoding framework can be adapted to detect/precoding both signal types simultaneously (i.e., heterogeneous signal operation). We are confident that you will find significant contributions in this manuscript beyond the associated ICC symposium paper. We believe that this manuscript captures the essence of this new technique that provides additional flexibility to the established research area of CP-SCM waveforms.

Perhaps just as important as highlighting the contributions of this paper is identifying what it is not. Although we provide an option for a spreading sequence to be applied prior to transmission, the MSP solution should not be confused with code division multiple access (CDMA). The spreading sequence can be used to spread each symbol over a given frame, but it is assumed that each UE uses the same spreading sequence. The spreading sequence is also applied in such a way that it does not increase the signal bandwidth because symbols within a frame spread in an overlapping manner. To be clear, the multiple-access technique employed here is spatial multiplexing (i.e., MIMO). The spreading sequence could differ between cells to suppress inter-cell interference, but that is only mentioned in the

1
2
3 manuscript and is not one of the main points. We also note that even with the addition of resource
4 allocation, we are not proposing that CP-SCM should replace OFDM in all cases. The point that we are
5 trying to convey in this paper is that there are certain scenarios that could benefit from the single-carrier
6 nature of CP-SCM. We have made efforts to identify some of these cases.
7

8
9 We hope that this supplemental information will help to avoid any misconceptions about this resource
10 allocation method for CP-SCM. If there is anything that could be better explained in the manuscript,
11 please let us know. We look forward to seeing your comments.
12
13
14

15 Sincerely,

16 Brent Kenney

17 PhD candidate, University of Utah

18 Department of Electrical and Computer Engineering

19 Wireless Communications Lab

20 <https://wireless.ece.utah.edu/>
21
22

23 Co-authors:

24 Dr. Arslan Majid, Idaho National Laboratory (arslan.majid@inl.gov)

25 Dr. Hussein Moradi, Idaho National Laboratory (hussain.moradi@inl.gov)

26 Dr. Behrouz Farhang-Boroujeny, University of Utah (farhang@ece.utah.edu)
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60