IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 10, 2020, accepted September 14, 2020, date of publication September 18, 2020,
date of current version September 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3024889

The Effect of Spatial Multiplexing of the Interference on MIMO Communication Performance

YIFTACH RICHTER, (Member, IEEE), AND ITSIK BERGEL, (Senior Member, IEEE)
Faculty of Engineering, Bar-Ilan University, Ramat-Gan 5290002, Israel
Corresponding author: Yiftach Richter (richtey@biu.ac.il)

ABSTRACT This work sheds light on the effects of spatially multiplexed interference on multiple-input-multiple-output (MIMO) networks. In particular, any increase in the number of interfering data streams (while keeping the total interference power constant) is shown to degrade the quality of the interfered link. Although this statement appears very intuitive, it has yet to be proven. In this work, we first give a mathematical definition of the intuitive notion of 'increasing the number of streams' leads to proof that the achievable rate of a link decreases when any of its interferers increases its number of data streams. The achievable rate is measured by the mutual information of the link or by the spectral efficiency of the optimal linear Minimum-Mean-Square (MMSE) receiver. Correlatively, we show that the worst power allocation for an interferer is the equal power allocation for all the streams. This result highlights the importance of the optimization of the number of data streams at each transmitter in MIMO networks.

INDEX TERMS Multiple-input-multiple-output (MIMO), wireless communication.

I. INTRODUCTION

Multiple-input-multiple-output (MIMO) communication can increase the data rate in wireless networks. MIMO networks simultaneously transmit several data streams from each transmitter, using spatial multiplexing through multiple antennas. However, coordinating all the data streams in the network requires complete channel state information (CSI) in all the transmitters. Such cooperation is rarely achieved in modern communication system. Hence, it is also important to consider the interference between uncoordinated links.

MIMO technologies have been thoroughly studied and used in various applications, e.g., cellular, adhoc, heterogeneous and local area networks (LANs). In particular, MIMO is considered as one of the main enablers to achieve connectivity goals of cellular 5G and 6G networks.

While adhoc networks typically consider non-cooperating transmitters, 5G+ networks allow various levels of cooperation between different base stations (BSs) (e.g., [1], [2]). Nevertheless, BS cooperation incurs significant backhaul costs. Thus, most deployments are expected to use clusters of cooperating BSs, while between the clusters, there is no cooperation in the physical layer.

The associate editor coordinating the review of this manuscript and approving it for publication was Hayder Al-Hraishawi.

Any part of the network that has complete cooperation can be considered as a single transmitter. If this transmitter serves a single user, then the transmission of multiple data streams, together with optimal power allocation per stream enables the maximization of the link throughput (e.g., [3]–[5]). In particular, in the presence of white noise solely, a singular value decomposition (SVD) of the channel matrix together with water filling power allocation was shown to be optimal (e.g., [6], [7]).

The cooperation is more difficult between parts of the network that share CSI but do not share the user data. This setup is known as the interference channel and has been studied extensively. Approaches for the optimization of such network include for example game theory methods, SLNR, max-SINR and other interference alignment methods (see [8]–[15]).

In this work we focus on the case of interfering transmitters with no CSI sharing. We assume that each receiver can measure the channel states of all interfering transmitters. But, each transmitter only acquires data on its served users. We recall that this setup also allows us to consider sets of cooperating BSs, if they also share user data (and hence can be considered as a single transmitter). In this setup, there is no coordination between the beamforming vectors of the independent transmitters.

We consider a network with non-cooperating transmitters; e.g., MIMO cellular networks [16]–[20], or MIMO ad-hoc networks [21]–[24]. We also assume that each receiver treats the unknown interference as additional (spatially colored) noise.

However, when each receiver considers the undesired transmissions as additional noise, the selection of the number of data streams in each transmitter impacts the interference characteristics. In this case, finding the optimal number of data streams at each node is important (although quite challenging).

This work proves an intuitive but non-trivial characteristic of the interference in MIMO networks. We show that any increase in the number of data streams of any interferer will decrease the throughput of the interfered link, even if the total transmission power of each interferer remains constant.

To clarify, it is obvious that any increase in the power of any interfering data stream will degrade the performance of the link. It is also obvious that adding an additional interfering stream without changing the power of the previously interfering streams will decrease performance. We consider the case where the same interference power that was used for a given number of data streams is now divided into a larger number of data streams. This change in power allocation can increase the throughput of the interfering link in many cases, and may be advantageous for the network. Nevertheless, we show that it decreases the throughput of any interfered link.

This result appears intuitive since it seems logical that transmitting a larger number of data streams 'exploits more of the spatial dimensions of the network and hence leaves less spatial dimensions for other links'. However, this intuition does not lead to a trivial proof. Furthermore, we first need to define an 'increase in the number of data streams' mathematically in this case.

In the absence of CSI on other transmitters, it is intuitive to try to maximize the throughput of the desired transmitter. Such a maximization will often lead to a transmission of a maximal number of data streams. Yet, the throughput increase from some of these data streams may be negligible.

The importance of our work is in proving that any increase in the number of data streams will decrease the throughput of the interfered links. Thus, we need to avoid the use of additional data streams if their contribution is not significant. The definition of 'significant contribution' varies between different networks. Yet, our conclusion holds in any network: data streams that do not bring significant throughput increase will decrease the overall network throughput. This conclusion is even more important in the case that the transmitter is equipped with a large number of antennas (e.g. massive MIMO [17]). Such a transmitter can support a large number of data streams. Thus, each user must limit its number of data streams in order to allow a good network throughput.

The main novelties in this work are as follows: (1) We show that any increase in the number of interfering data streams (when the total interference power is kept constant) decreases the performance of the interfered links; (2) Equal transmission power for all the data streams of an interferer leads to worst case interference given the interference power.

The rest of this paper is organized as follows. Section II describes the structure of the analyzed MU-MIMO network. Section III presents the interference characterization in Multi-User MIMO (MU-MIMO). Section IV reports the numerical results, and Section V presents the conclusions.

II. SYSTEM MODEL

We consider a wireless network with a finite or infinite number of transmitters. All the transmitters operate in the same frequency band and at the same time. We focus on one transmitter-receiver pair (which we term *probe pair*) in the presence of other (non-cooperating) transmitters. Without loss of generality, we assume that the probe transmitter has index 0.

Denote the number of antennas at transmitter j by N_j and the number of antennas at the probe receiver by N. The received vector at the probe receiver is given by (e.g., [25], [26])

$$\mathbf{y} = \sum_j \gamma_j \mathbf{H}_j \mathbf{x}_j + \mathbf{v} \tag{1}$$

where γ_j is the mean channel gain from transmitter j to the probe receiver (also considering the path loss attenuation; see for example [27]), $\mathbf{H}_j$ is the channel matrix and $\mathbf{v}$ is the (spatially colored) additive Gaussian noise with zero mean and covariance matric $\mathbf{C}_v$. We assume throughout that all the elements of all channel matrices, $\{\mathbf{H}_j\}$, are statistically independent and identically distributed (i.i.d.) proper complex Gaussian random variables with zero mean and unit variance. This assumption is quite common, and represents Rayleigh fading channels.

The channel input at transmitter j is denoted by $\mathbf{x}_j \in \mathbb{C}^{N_j \times K_j}$ and given by:

$$\mathbf{x}_j = \mathbf{F}_j \cdot \mathrm{diag}(\sqrt{\boldsymbol{\rho}_j}) \cdot \mathbf{u}_j \tag{2}$$

where $\boldsymbol{\rho}_j$, $\mathbf{F}_j$ and $\mathbf{u}_j$ denote the power allocation vector, precoding matrix and transmitted symbol vector of transmitter j, respectively. Transmitter j transmits K_j data streams through $\mathbf{u}_j = [u_{j,1}, u_{j,2}, \ldots, u_{j,K_j}]^T \in \mathbb{C}^{K_j \times 1}$. We assume that all the data symbols, $\{u_{j,k}\}$, are i.i.d. standard complex Gaussian random variables with zero mean and unit variance. Thus, $\mathbb{E}\{u_{j,k} \cdot u_{m,\ell}\} = \delta_{j,m}\delta_{k,\ell}$ where $\delta_{k,\ell}$ is the Kronecker delta (i.e., $\delta_{k,\ell} = 1$ only if $k = \ell$, and 0 otherwise). We collect the power allocations of each user into a vector and term it the user power allocation vector, $\boldsymbol{\rho}_j = [\rho_{j,1}, \rho_{j,2}, \ldots, \rho_{j,K_j}]^T$. We also normalize all transmission powers such that $\sum_{k=1}^{K_j} \rho_{j,k} = 1$ for all j.[1]

Since the network is not cooperative, we assume throughout that the transmitters have no knowledge of the receiver environment, and in particular the channel state of

[1]This does not limit the generality, since any different value for the total transmission power can be absorbed in γ_j.

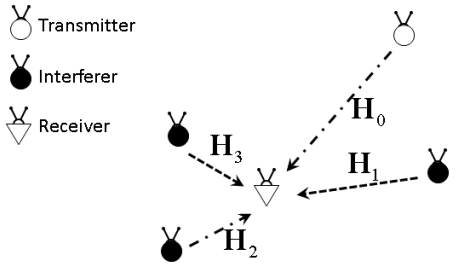

FIGURE 1. Illustration of the analyzed MU-MIMO system. Each node in the network, either a transmitter or a receiver, is equipped with multiple antennas. The channel matrix between the probe transmitter and the probe receiver is $\mathbf{H_0}$. The channel matrix between the j-th interfering transmitter and the probe receiver is $\mathbf{H}_j$.

their interferers. Thus, we assume that each transmitter selects its precoding vectors based solely on its desired channel matrix. That is, the design of the precoding matrix of the probe transmitter, $\mathbf{F}_0$, may depend on its channel, $\mathbf{H}_0$, but not on any other channel.

We further assume that the j-th precoding matrix, $\mathbf{F}_j = [\mathbf{f}_{j,1}, \mathbf{f}_{j,2}, \ldots, \mathbf{f}_{j,K_j}] \in \mathbb{C}^{N_j \times K_j}$, is composed of orthonormal precoding vectors (i.e., $\mathbf{F}_j^H \mathbf{F}_j = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix). This assumption is commonplace in most precoding approaches (see for example singular-value-decomposition (SVD) (e.g., [6], [24]) and zero-forcing beamforming (e.g., [28]).

We assume that the probe receiver is only interested in the transmission from the probe transmitter (transmitter 0) while all the other transmitters are interferers. We denote the aggregate interference vector from all the undesired transmitters by

$$\begin{aligned}\mathbf{z} &= \sum_{j\neq 0} \gamma_j \mathbf{H}_j \mathbf{x}_j \\ &= \sum_{j\neq 0} \gamma_j \sum_{k=1}^{K_j} \sqrt{\rho_{j,k}} \cdot \mathbf{H}_j \mathbf{f}_{j,k} u_{j,k}. \end{aligned} \tag{3}$$

The covariance matrix of the aggregate interference plus noise is:

$$\begin{aligned}\mathbf{C}_z &\triangleq \mathbb{E}\{\mathbf{z}\mathbf{z}^H\} + \mathbb{E}\{\mathbf{v}\mathbf{v}^H\} \\ &= \mathbb{E}\Big\{ \sum_{j\neq 0} \gamma_j^2 \mathbf{H}_j \mathbf{x}_j \mathbf{x}_j^H \mathbf{H}_j^H \Big\} + \mathbf{C}_v \\ &= \sum_{j\neq 0} \gamma_j^2 \mathbf{H}_j \mathbf{F}_j \mathrm{diag}(\boldsymbol{\rho}_j) \mathbf{F}_j^H \mathbf{H}_j^H + \mathbf{C}_v. \end{aligned} \tag{4}$$

We consider performance with either an optimal receiver or an optimal linear receiver. Thus, the performance metrics are the *mutual information* and the *achievable spectral efficiency* using the minimal mean square error (MMSE) equalizer (which we denote *MMSE spectral efficiency*).

The mutual information (MI) between input vector, $\mathbf{x}_0$, and the output, $\mathbf{y}$, represents the maximal achievable rate at a receiver that treats the interference as noise. In our system model, the MI is given by (e.g., [29]):

$$I(\mathbf{x}_0, \mathbf{y}) \triangleq \mathbb{E}\Big\{ \log_2 \Big(\det(\mathbf{I} + \mathbf{C}_z^{-1} \mathbf{H}_0 \mathbf{C}_x \mathbf{H}_0^H) \Big) \Big\} \tag{5}$$

where

$$\mathbf{C}_x \triangleq \mathbb{E}\{\mathbf{x}_0 \cdot \mathbf{x}_0^H\}. \tag{6}$$

is the covariance matrix of the desired input signal. The expectation in (5) is with respect to (w.r.t.) the channel matrices and the expectation in (6) is w.r.t. the desired symbols.

The MMSE receiver decodes the data of the k-th stream by multiplying the received vector by the optimal linear weight vector,

$$\mathbf{w}_k = \mathbf{C}_0^{-1} \mathbf{H}_0 \mathbf{f}_{0,k} \tag{7}$$

where $\mathbf{C}_0$ is the noise plus the interference covariance matrix:

$$\mathbf{C}_0 = \mathbb{E}\{\mathbf{y}\mathbf{y}^H\} = \mathbf{H}_0 \mathbf{C}_x \mathbf{H}_0^H + \mathbf{C}_z \tag{8}$$

and $\mathbf{C}_z$ is defined in (4). Here we assume that the probe receiver knows the covariance matrix of the received signal. In practice this covariance matrix is easily estimated from the received signal.

The resulting signal to noise ratio (SNR) for the decoding of the k-th stream using the MMSE weight vector is (see Appendix B):

$$\frac{1}{\frac{1}{\gamma_0^2 \rho_{0,k}} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1} \mathbf{H}_0 \mathbf{f}_{0,k})^{-1} - 1}.$$

Assuming a near optimal coding scheme with a Gaussian codebook and a long enough codeword, the achievable MMSE spectral efficiency over K_0 data streams is:

$$\begin{aligned} &R_{\mathrm{MMSE}} \\ &= \sum_{k=1}^{K_0} \mathbb{E}\Big\{ \log_2 \Big(1 + \frac{1}{\gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1} \mathbf{H}_0 \mathbf{f}_{0,k})^{-1} - 1} \Big) \Big\}. \end{aligned} \tag{9}$$

III. INTERFERENCE CHARACTERIZATION IN MU-MIMO

In this section we show that any increase in the number of data streams of even a single interferer (keeping the total interference power constant) will degrade the quality of the probe transmitter-receiver link. Since we need to keep the total power constant, increasing the number of data streams requires a power decrease in at least one of the existing data streams. Thus we first need to define what is considered an increase in the number of data streams.

Furthermore, "quantifying" the spatial multiplexing allocation depends not only on the number of data streams, but also on their power allocation. Because we are comparing power allocations with different numbers of data streams, it is convenient to zero-pad the shortest allocation to produce equal length vectors. Obviously, adding streams with zero power does not change the network, and is done solely for the convenience of notation.

Definition 1: A power allocation for user j, $\boldsymbol{\rho}_j$, is considered as an increase in the number of data streams compared to a reference power allocation, $\tilde{\boldsymbol{\rho}}_j$, if there exists a number

of streams $\tilde{K} \leq K_j$ such that $\rho_{j,k} \leq \tilde{\rho}_{j,k}$ for any $k \leq \tilde{K}$, $\rho_{j,k} > \tilde{\rho}_{j,k}$ for any $k > \tilde{K}$, and

$$\max_{k>\tilde{K}} \rho_{j,k} \leq \min_{k\leq\tilde{K}} \rho_{j,k}. \tag{10}$$

Note that Definition 1 is more general, but in most cases we will assume that the original number of active data streams was $\tilde{K}$, while K_j is used to denote the number of data streams in the new power allocation scheme. That is, in most interesting cases we will have $\tilde{\rho}_{j,k} = 0$ for any $k > \tilde{K}$. This definition is consistent with the two main power allocation approaches.

The simpler approach uses equal allocation for all active streams. Thus we switch from $\tilde{\rho}_{j,k} = 1/\tilde{K}$ for $k \leq \tilde{K}$ (e.g., $\tilde{\boldsymbol{\rho}}_j = [1/3, 1/3, 1/3, 0]$) to $\rho_{j,k} = 1/K_j$ (e.g., $\boldsymbol{\rho}_j = [1/4, 1/4, 1/4, 1/4]$) for all the data streams, which complies with Definition 1.

A smarter approach assigns larger power to better streams (e.g., water filling). In this approach, the better streams will be activated first ($k \leq \tilde{K}$). Then, when the number of streams is increased, the power for the previously active streams will be reduced, but the newly activated streams will have weaker gains than the others. Thus, these streams will have lower powers as stipulated in (10).

Using Definition 1, we can now present our main finding:

Theorem 1: If some of the interferers in the network increase their number of data streams (as in Definition 1) while all others keep the same power allocation, the **MI** and the **MMSE spectral efficiency** of the probe receiver will decrease.

Proof: We prove Theorem 1 by breaking down the change in the interferers into a sequence of changes, each involving power changes in only two streams. We next define a change in which only the powers of two data streams are altered, and these powers become closer to the average power:

Definition 2: A change in the power allocation from one power allocation vector, $\boldsymbol{\rho}_j$, to another, $\breve{\boldsymbol{\rho}}_j$, is termed an interference balancing change if there exists $\ell \neq m$ such that $\rho_{j,k} = \breve{\rho}_{j,k}$ for any $k \neq \{\ell, m\}$, $\rho_{j,m} + \rho_{j,\ell} = \breve{\rho}_{j,m} + \breve{\rho}_{j,\ell}$, and:

$$\rho_{j,m} < \breve{\rho}_{j,m} \leq \bar{\rho} \leq \breve{\rho}_{j,\ell} < \rho_{j,\ell} \tag{11}$$

where $\bar{\rho} = (\rho_{j,\ell} + \rho_{j,m})/2$.

Using Definition 2, the proof of Theorem 1 can be divided into the following three lemmas:

Lemma 1: Any increase in the number of data streams of an interferer can be written as a sequence of interference balancing changes.

Lemma 2: Any interference balancing change decreases the MI of the probe link.

Lemma 3: Any interference balancing change decreases the MMSE spectral efficiency of the probe link.

Proof of Lemma 1: By Definition 1, any stream with $k \leq \tilde{K}$ requires a power decrease, and any stream with $k > \tilde{K}$ requires a power increase. Note that all the powers in the first group are higher than all the powers in the second group, which guarantees the existence of an appropriate sequence of interference balancing changes.

To further illustrate this proof, consider the example above: the vector $\tilde{\boldsymbol{\rho}}_j = [1/3, 1/3, 1/3, 0]$, can be changed through the sequence $[1/3, 1/3, 1/4, 1/12]$, $[1/3, 1/4, 1/4, 2/12]$ and eventually $\boldsymbol{\rho}_j = [1/4, 1/4, 1/4, 1/4]$, where all changes are interference balancing. ■

Proof of Lemma 2: See Appendix A. ■

Proof of Lemma 3: See Appendix B. ■

Corollary 1: Dividing the transmission power equally over all streams of all interferers will lead to the lowest probe rate, given the power of each interferer.

Proof: The corollary immediately follows from Lemma 2, by noting that a uniform power allocation across all streams can always be obtained from any other power allocation by interference balancing changes. ■

Intuitively, increasing the number of data streams by one user can be understood as taking more of the system degrees of freedom (DoF) and hence leaving less DoF to other users. But, the above statement does not hold mathematically, as the DoF is measured when the SNR grows to infinity (and hence cannot allow any interference). In a non-cooperating network (without CSI of the interfered users) any interference from another transmitter (also with SNR growing to infinity) will reduce the number of network's DoF to zero. Thus, the DoF in such networks is bounded by the maximal DoF of a single transmitter in the network [30]. Our proof cannot characterize the total achievable network throughput. Yet, we show that even at a finite SNR, any increase in the number of data streams by a single transmitter will reduce the throughput of all interfered links.

IV. NUMERICAL RESULTS

This section presents simulation results illustrating Theorem 1. We first report the results for specific node locations, and then consider a random setup that enables us to study behavior over many spatial realizations of the node locations.

The simulated thermal noise has a proper complex Gaussian distribution with zero mean and $\mathbb{E}\{\mathbf{v}\mathbf{v}^H\} = \mathbf{I}$. The precoding vectors were chosen using eigen-beamforming, and the transmission power was divided equally between the active data streams (uniform power allocation): $\boldsymbol{\rho}_j = (1/K_j)\cdot[1, 1, \ldots, 1]^T$. In all the simulations, the channel matrices experienced Rayleigh fading as described above (i.e., the elements of all channel matrices are statistically i.i.d. proper complex Gaussian random variables with zero mean and unit variance). All the results are averaged over 10^4 network realizations.

We start with a scenario that composed solely of a single interferer, located so that the signal to interference ratio (SIR) is $\gamma_0^2/\gamma_1^2 = 20$dB. The location and transmission power of the desired transmitter are set so that its SNR $= \gamma_0^2/N = 0$dB. Both transmitters and the desired receiver have $N = N_0 = N_1 = 4$ antennas.

Fig. 2 presents the spectral efficiency of the MI, (5), and the MMSE spectral efficiency, (9), at the desired receiver,

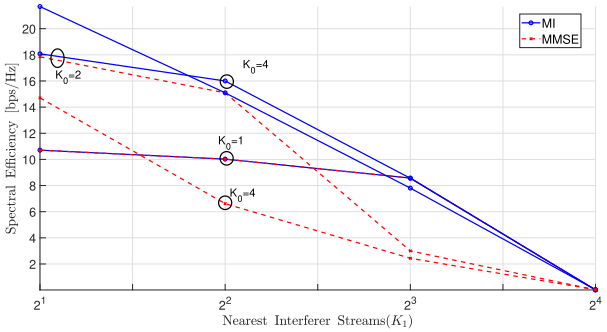

FIGURE 2. Mutual-Information (MI), (5), and the MMSE spectral efficiency, (9), as a function of K_1, the number of streams of a single interferer, where $N = 4$, SNR= 0dB and SIR= 20dB.

as a function of the number of data streams transmitted by the interfering transmitter, K_1. The figure depicts the performance when the desired transmitter transmits $K_0 = 1, 2$ and 4 data streams.

As expected by Theorem 1, Fig. 2 shows that in all cases, the spectral efficiency and the MMSE spectral efficiency decrease with the number of interferer data streams.

To capture a broader picture, we next examine the case of a random network where the locations of the nodes follow a homogenous Poisson point process (PPP) (e.g., [24], [31]) and the node density is set to $\lambda = 2.5$ [nodes/km^2] over a disk of radius $r = 6.18$km, centered at the probe receiver, and zero everywhere else (i.e., the average number of nodes in the network is $\pi\lambda r^2 = 300$). The channel gain is set to $\gamma_j^2 = d_j^{-\alpha}$ where d_j is the distance between transmitter j and the probe receiver, and α is the path loss exponent. The transmission power is set such that the signal to noise ratio at a distance of 1km is $d_0^{-\alpha}/N = 3$dB. We also use the bias correction of [32].

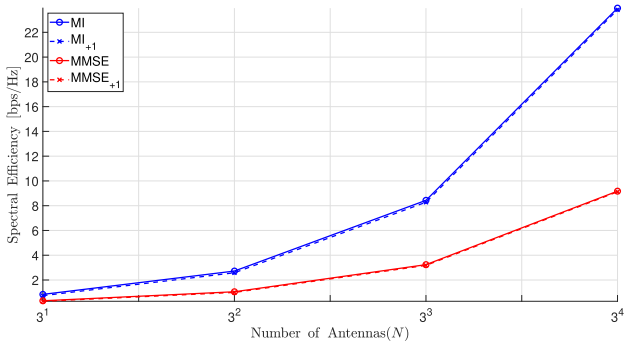

FIGURE 3. MI and MMSE spectral efficiency as a function of the number of antennas, N, for PPP nodes and a path loss exponent of $\alpha = 4$. All transmitters use $K_j = 2N/3$, except for the nearest interferer which uses $K_1 = 2N/3 + 1$ streams, but only for the curves of MI_{+1} and $MMSE_{+1}$. The plot shows that performance decreases when the nearest interferer increases its number of streams.

All the nodes in the network have the same number of antennas ($N_j = N$). Fig. 3 shows the MI and the MMSE spectral efficiency as a function of the number of antennas per node, N, for a path loss exponent of $\alpha = 4$.

The figure depicts two curves for each metric. The baseline curve (labeled MI and MMSE) used $K_j = 2N/3$ data streams for all interferers (including $K_0 = 2N/3$). The other curves depict performance in a scenario that only differs from the baseline scenarios for one additional stream transmitted by the nearest interferer (i.e., $K_1 = 2N/3 + 1$). These scenarios are labeled by MI_{+1} and $MMSE_{+1}$.

As can be seen, in all cases, even if only a single interferer increases its number of streams by a single stream, the MI and the MMSE spectral efficiency decrease. Again, this is consistent with Theorem 1.

For the simulation of Fig. 4, all the nodes were equipped with $N_j = N = 4$ antennas. Here, the desired transmitter used $K_0 = 2$ data streams, while all the other transmitters used an identical number of data streams ($K_i = K_j$ for all $i, j \geq 1$). Fig. 4 depicts the MI and the MMSE spectral efficiency at the desired receiver, as a function of the number of data streams of each interferer.

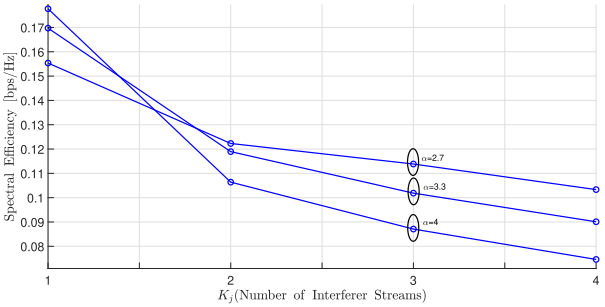

FIGURE 4. MI as a function of the number of streams transmitted by each interferer ($K_j, j \geq 1$). All nodes use $N_j = N = 4$ antennas and the desired transmitter uses $K_0 = 2$ streams.

Fig. 4 shows, for three different path loss exponents, α, that the MI decrease when the number of interferer data streams increases. The MMSE spectral efficiency is not shown here since it is indistinguishable from the MI in this scenario. Thus, all of our numerical results support the claim of Theorem 1, and demonstrate that any increase in the number of interfering data streams leads to a decrease in the rate of the desired link.

V. CONCLUSION

In this paper we characterized the effect of interferer spatial multiplexing in MIMO wireless networks. We started by providing a mathematical definition for the phrase 'increasing the number of streams' and then proved that the MI and the MMSE spectral efficiency of a link decrease whenever one or more of its interferers increases its number of data streams. This statement holds even when the total transmitted power by each interferer does not change.

As a corollary, we showed that the worst power allocation for an interferer is the equal power allocation per stream. The results are proven for MI and MMSE spectral efficiency,

and the numerical results confirmed performance for various cases.

This paper highlights the importance of the optimization of the number of data streams at each transmitter, and the need to consider the effects of this optimization on all the affected receivers in the network. Furthermore, ''quantifying'' the spatial multiplexing depends not only on the number of data streams, but also on their power allocation.

APPENDIX A
PROOF OF LEMMA 2

We assume that a single interferer makes an interference balancing change, and prove that the MI of the probe pair decreases. Without loss of generality, we assume that only interferer m makes the interference balancing change within its first two streams. That is, interferer m starts with the power allocation $\boldsymbol{\rho}_m = [\rho_{m,1}, \rho_{m,2}, \rho_{m,3}, \ldots, \rho_{m,K_m}]^T$ and changes to: $\breve{\boldsymbol{\rho}}_m = [\breve{\rho}_{m,1}, \breve{\rho}_{m,2}, \rho_{m,3}, \ldots, \rho_{m,K_m}]^T$.

In the following, it is convenient to consider both cases as a function of a single parameter. To that end, we define $\boldsymbol{\rho}(\epsilon) = [\bar{\rho}+\epsilon, \bar{\rho}-\epsilon, \rho_{m,3}, \ldots, \rho_{m,K_m}]^T$, where $\bar{\rho} = (\rho_{m,1}+\rho_{m,2})/2$. Thus, we have: $\boldsymbol{\rho}_m = \boldsymbol{\rho}\big((\rho_{m,1}-\rho_{m,2})/2\big)$ and $\breve{\boldsymbol{\rho}}_m = \boldsymbol{\rho}\big((\breve{\rho}_{m,1}-\breve{\rho}_{m,2})/2\big)$. We also define $\mathbf{C}_z(\epsilon)$ as the covariance matrix in (4), with $\boldsymbol{\rho}_m$ replaced by $\boldsymbol{\rho}(\epsilon)$, and $I(\epsilon)$ as the MI in (5) with $\mathbf{C}_z$ replaced by $\mathbf{C}_z(\epsilon)$.

Recalling that $\rho_{m,1} - \rho_{m,2} > \breve{\rho}_{m,1} - \breve{\rho}_{m,2}$, and using the ϵ notation, we can prove Lemma 2 by proving that $I(\epsilon)$ increases monotonically with ϵ for $\epsilon > 0$. We prove the monotonic increase of $I(\epsilon)$ by showing that:

$$\frac{d^2 I(\epsilon)}{d\epsilon^2} > 0 \tag{12}$$

$$\frac{dI(\epsilon)}{d\epsilon}\Big|_{\epsilon=0} = 0 \tag{13}$$

where (12) shows the convexity in ϵ, and (13) states that it has a single minimum point at $\epsilon = 0$.

We start by noting that the signal covariance matrix, $\mathbf{C}_x$ (6), is symmetric and positive definite. Using the Cholesky decomposition we can find $\mathbf{R}_x$ that satisfies:

$$\mathbf{C}_x = \mathbf{R}_x \mathbf{R}_x^H. \tag{14}$$

By also using the Weinstein-Aronszajn identity, the MI of the MIMO channel, (5), can be rewritten as:

$$I(\epsilon) = \mathbb{E}\Big\{ \log_2 \Big(\det(\mathbf{I} + \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon) \tilde{\mathbf{H}}_0) \Big) \Big\} \tag{15}$$

where $\tilde{\mathbf{H}}_0 \triangleq \mathbf{H}_0 \mathbf{R}_x$.

The first derivative of (15) w.r.t. ϵ is given by:

$$\frac{dI(\epsilon)}{d\epsilon} = \frac{1}{\log 2} \mathbb{E}\Big\{ Tr\Big((\mathbf{I} + \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon) \tilde{\mathbf{H}}_0)^{-1} \cdot \tilde{\mathbf{H}}_0^H \frac{d\mathbf{C}_z^{-1}(\epsilon)}{d\epsilon} \tilde{\mathbf{H}}_0 \Big) \Big\} \tag{16}$$

where we used the derivative rules $\frac{d}{d\epsilon} \log\Big(\det(\mathbf{X}(\epsilon))\Big) = \mathrm{Tr}\Big(\mathbf{X}^{-1}(\epsilon) \frac{d\mathbf{X}(\epsilon)}{d\epsilon}\Big)$ (e.g., [33, Equation (43)]). By also using $\frac{d}{d\epsilon}\mathbf{X}^{-1}(\epsilon) = -\mathbf{X}^{-1}(\epsilon) \frac{d\mathbf{X}(\epsilon)}{d\epsilon} \mathbf{X}^{-1}(\epsilon)$ (e.g., [33, Equation (40)]) we get:

$$\frac{dI(\epsilon)}{d\epsilon} = -\frac{1}{\log 2} \mathbb{E}\Big\{ \mathrm{Tr}\Big((\mathbf{I} + \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon) \tilde{\mathbf{H}}_0)^{-1} \cdot \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon) \frac{d\mathbf{C}_z(\epsilon)}{d\epsilon} \mathbf{C}_z^{-1}(\epsilon) \tilde{\mathbf{H}}_0 \Big) \Big\}. \tag{17}$$

We next define the random vectors $\tilde{\mathbf{f}}_{j,\ell} \triangleq \mathbf{H}_j \mathbf{f}_{j,\ell}$. To characterize the distribution of $\tilde{\mathbf{f}}_{j,\ell}$, we first consider the conditional distribution of $\tilde{\mathbf{f}}_{j,\ell}$ and $\tilde{\mathbf{f}}_{j,v}$ given $\mathbf{F}_j$. Recalling that $\mathbf{H}_j$ follows a Gaussian distribution, the conditional distribution is Gaussian with zero mean and the following cross covariance matrix:

$$\mathbb{E}\{\tilde{\mathbf{f}}_{j,\ell} \tilde{\mathbf{f}}_{j,v}^H | \mathbf{F}_j\} = \mathbb{E}\{\mathbf{H}_j \mathbf{f}_{j,\ell} \mathbf{f}_{j,v}^H \mathbf{H}_j^H | \mathbf{f}_{j,\ell}, \mathbf{f}_{j,v}\}. \tag{18}$$

Denote by $\mathbf{h}_{j,k}^H$ the k-th *row* of $\mathbf{H}_j$; then, the $[k, m]$-th element of $\mathbb{E}\{\tilde{\mathbf{f}}_{j,\ell} \tilde{\mathbf{f}}_{j,v}^H | \mathbf{f}_{j,\ell}, \mathbf{f}_{j,v}\}$ is given by:

$$\begin{aligned} \mathbb{E}\left\{ \mathbf{h}_{j,k}^H \mathbf{f}_{j,\ell} \mathbf{f}_{j,v}^H \mathbf{h}_{j,m} | \mathbf{f}_{j,\ell}, \mathbf{f}_{j,v} \right\} &= \mathrm{tr}\left(\mathbb{E}\{\mathbf{h}_{j,m} \mathbf{h}_{j,k}^H\} \mathbf{f}_{j,\ell} \mathbf{f}_{j,v}^H \right) \\ &= \mathrm{tr}(\delta_{m,k} \mathbf{I} \cdot \mathbf{f}_{j,\ell} \mathbf{f}_{j,v}^H) \\ &= \delta_{m,k} \cdot \delta_{\ell,v}. \end{aligned} \tag{19}$$

where the second line used the statistical independence of the elements of matrix $\mathbf{H}_j$, and the third line used the orthonormality assumption on $\mathbf{F}_j$.

Thus, we can write

$$\mathbb{E}\left\{ \tilde{\mathbf{f}}_{j,\ell} \tilde{\mathbf{f}}_{j,v}^H | \mathbf{F}_j \right\} = \delta_{\ell,v} \cdot \mathbf{I} \tag{20}$$

and conclude that given $\mathbf{F}_j$, vectors $\tilde{\mathbf{f}}_{j,\ell}$ and $\tilde{\mathbf{f}}_{j,v}$ are jointly Gaussian, statistically independent for $\ell \neq v$ and each have a zero mean and covariance matrix of $\mathbf{I}$. Furthermore, as this holds for any allowed $\mathbf{F}_j$, it also holds for the unconditional distribution of $\tilde{\mathbf{f}}_{j,\ell}$ and $\tilde{\mathbf{f}}_{j,v}$ (see for example, [21, Lemma 1]). Furthermore, for any $k \neq j$, $\mathbf{H}_j$ is independent of $\mathbf{H}_k$ and $\mathbf{f}_{j,\ell}$ is independent of $\mathbf{f}_{k,v}$. Hence, $\tilde{\mathbf{f}}_{j,\ell}$ is statistically independent of $\tilde{\mathbf{f}}_{k,v}$ for any $j \neq k$ and/or $\ell \neq v$.

To evaluate (17) we need the derivative of $\mathbf{C}_z(\epsilon)$ w.r.t. ϵ. This is given by:

$$\begin{aligned} \frac{d\mathbf{C}_z(\epsilon)}{d\epsilon} &= \gamma_m^2 \Big[\mathbf{H}_m \mathbf{f}_{m,1} \mathbf{f}_{m,1}^H \mathbf{H}_m^H - \mathbf{H}_m \mathbf{f}_{m,2} \mathbf{f}_{m,2}^H \mathbf{H}_m^H \Big] \\ &= \gamma_m^2 \Big[\tilde{\mathbf{f}}_{m,1} \tilde{\mathbf{f}}_{m,1}^H - \tilde{\mathbf{f}}_{m,2} \tilde{\mathbf{f}}_{m,2}^H \Big]. \end{aligned} \tag{21}$$

Substituting in (17):

$$\begin{aligned} \frac{dI(\epsilon)}{d\epsilon} = &-\frac{\gamma_m^2}{\log 2} \mathbb{E}\Big\{ \mathrm{Tr}\Big((\mathbf{I} + \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon) \tilde{\mathbf{H}}_0)^{-1} \cdot \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon) \tilde{\mathbf{f}}_{m,1} \tilde{\mathbf{f}}_{m,1}^H \mathbf{C}_z^{-1}(\epsilon) \tilde{\mathbf{H}}_0 \Big) \Big\} \\ &+ \frac{\gamma_m^2}{\log 2} \mathbb{E}\Big\{ \mathrm{Tr}\Big((\mathbf{I} + \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon) \tilde{\mathbf{H}}_0)^{-1} \cdot \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon) \tilde{\mathbf{f}}_{m,2} \tilde{\mathbf{f}}_{m,2}^H \mathbf{C}_z^{-1}(\epsilon) \tilde{\mathbf{H}}_0 \Big) \Big\}. \end{aligned} \tag{22}$$

To prove (13), note that for $\epsilon = 0$, $\mathbf{C}_z(\epsilon)$ is symmetric with respect to interchanging $\tilde{\mathbf{f}}_{m,1}$ and $\tilde{\mathbf{f}}_{m,2}$. Noting also that these two vectors are i.i.d., and statistically independent of

any other quantity in the expectation, we conclude that the two expectations in (22) are equal and hence their difference is indeed 0.

To prove (12), we consider the second derivative of $I(\epsilon)$. Taking the derivative of (16) we write:

$$\frac{d^2\, I(\epsilon)}{d\epsilon^2} = \frac{1}{\log 2}\mathbb{E}\left\{a_1(\epsilon) + a_2(\epsilon)\right\} \tag{23}$$

where

$$a_1(\epsilon) \triangleq \mathrm{Tr}\left[\frac{d}{d\epsilon}\left[(\mathbf{I} + \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\tilde{\mathbf{H}}_0)^{-1}\right] \cdot \tilde{\mathbf{H}}_0^H \frac{d\mathbf{C}_z^{-1}(\epsilon)}{d\epsilon}\tilde{\mathbf{H}}_0\right] \tag{24}$$

and

$$a_2(\epsilon) \triangleq \mathrm{Tr}\left[(\mathbf{I} + \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\tilde{\mathbf{H}}_0)^{-1} \frac{d}{d\epsilon}\left[\tilde{\mathbf{H}}_0^H \frac{d\mathbf{C}_z^{-1}(\epsilon)}{d\epsilon}\tilde{\mathbf{H}}_0\right]\right]. \tag{25}$$

Starting with $a_1(\epsilon)$, the derivative of the term within the internal square brackets in (24) is:

$$\begin{aligned}\frac{d}{d\epsilon}&\left[(\mathbf{I} + \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\tilde{\mathbf{H}}_0)^{-1}\right] \\ &= -(\mathbf{I} + \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\tilde{\mathbf{H}}_0)^{-1} \\ &\quad \cdot \tilde{\mathbf{H}}_0^H \frac{d\mathbf{C}_z^{-1}(\epsilon)}{d\epsilon}\tilde{\mathbf{H}}_0 \cdot (\mathbf{I} + \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\tilde{\mathbf{H}}_0)^{-1}.\end{aligned} \tag{26}$$

Thus, $a_1(\epsilon)$ can be written as

$$a_1(\epsilon) = -\mathrm{Tr}(\mathbf{A} \cdot \mathbf{A}) \tag{27}$$

where

$$\begin{aligned}\mathbf{A} &\triangleq (\mathbf{I} + \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\tilde{\mathbf{H}}_0)^{-1}\tilde{\mathbf{H}}_0^H \frac{d\mathbf{C}_z^{-1}(\epsilon)}{d\epsilon}\tilde{\mathbf{H}}_0 \\ &= -(\mathbf{I} + \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\tilde{\mathbf{H}}_0)^{-1}\tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\frac{d\mathbf{C}_z(\epsilon)}{d\epsilon}\mathbf{C}_z^{-1}(\epsilon)\tilde{\mathbf{H}}_0.\end{aligned} \tag{28}$$

Noting that $(\mathbf{I} + \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\tilde{\mathbf{H}}_0)^{-1}$ is positive definite, we can find a matrix, $\mathbf{R}_\mathrm{A}$, that satisfies

$$\mathbf{R}_\mathrm{A}\mathbf{R}_\mathrm{A}^H = \left(\mathbf{I} + \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\tilde{\mathbf{H}}_0\right)^{-1}. \tag{29}$$

Hence, (28) can be written as:

$$\mathbf{A} = -\mathbf{R}_\mathrm{A}\mathbf{R}_\mathrm{A}^H \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\frac{d\mathbf{C}_z(\epsilon)}{d\epsilon}\mathbf{C}_z^{-1}(\epsilon)\tilde{\mathbf{H}}_0. \tag{30}$$

Substituting in (27) and using the cyclic property of the trace:

$$\begin{aligned}a_1(\epsilon) &= -\mathrm{Tr}(\mathbf{R}_\mathrm{A}^H \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\frac{d\mathbf{C}_z(\epsilon)}{d\epsilon}\mathbf{C}_z^{-1}(\epsilon)\tilde{\mathbf{H}}_0\mathbf{R}_\mathrm{A} \\ &\quad \cdot \mathbf{R}_\mathrm{A}^H \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\frac{d\mathbf{C}_z(\epsilon)}{d\epsilon}\mathbf{C}_z^{-1}(\epsilon)\tilde{\mathbf{H}}_0\mathbf{R}_\mathrm{A}). \\ &= -\mathrm{Tr}\left(\mathbf{Q}\mathbf{R}_z(\epsilon)\tilde{\mathbf{H}}_0\mathbf{R}_\mathrm{A}\mathbf{R}_\mathrm{A}^H \tilde{\mathbf{H}}_0^H \mathbf{R}_z^H(\epsilon)\mathbf{Q}^H\right)\end{aligned} \tag{31}$$

where we defined

$$\mathbf{Q} \triangleq \mathbf{R}_\mathrm{A}^H \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\frac{d\mathbf{C}_z(\epsilon)}{d\epsilon}\mathbf{R}_z^H(\epsilon) \tag{32}$$

and $\mathbf{R}_z(\epsilon)$ is defined such that

$$\mathbf{C}_z^{-1}(\epsilon) = \mathbf{R}_z^H(\epsilon)\mathbf{R}_z(\epsilon). \tag{33}$$

Next, we simplify $a_2(\epsilon)$ by evaluating the derivative of the term in the internal square brackets in (25):

$$\begin{aligned}\frac{d}{d\epsilon}&\left[\tilde{\mathbf{H}}_0^H \frac{d\mathbf{C}_z^{-1}(\epsilon)}{d\epsilon}\tilde{\mathbf{H}}_0\right] = -\frac{d}{d\epsilon}\left[\tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\frac{d\mathbf{C}_z(\epsilon)}{d\epsilon}\mathbf{C}_z^{-1}(\epsilon)\tilde{\mathbf{H}}_0\right] \\ &= -\tilde{\mathbf{H}}_0^H \frac{d\mathbf{C}_z^{-1}(\epsilon)}{d\epsilon}\frac{d\mathbf{C}_z(\epsilon)}{d\epsilon}\mathbf{C}_z^{-1}(\epsilon)\tilde{\mathbf{H}}_0 \\ &\quad -\tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\frac{d^2\mathbf{C}_z(\epsilon)}{d\epsilon^2}\mathbf{C}_z^{-1}(\epsilon)\tilde{\mathbf{H}}_0 \\ &\quad -\tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\frac{d\mathbf{C}_z(\epsilon)}{d\epsilon}\frac{d\mathbf{C}_z^{-1}(\epsilon)}{d\epsilon}\tilde{\mathbf{H}}_0 \\ &= -2\tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\frac{d\mathbf{C}_z(\epsilon)}{d\epsilon}\frac{d\mathbf{C}_z^{-1}(\epsilon)}{d\epsilon}\tilde{\mathbf{H}}_0\end{aligned} \tag{34}$$

where the last line used $\frac{d^2\mathbf{C}_z(\epsilon)}{d\epsilon^2} = 0$ and the fact that all matrices in the product are Hermitian. Substituting (29) and (34), we have:

$$\begin{aligned}a_2(\epsilon) &= -2\mathrm{Tr}\left(\mathbf{R}_\mathrm{A}\mathbf{R}_\mathrm{A}^H \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\frac{d\mathbf{C}_z(\epsilon)}{d\epsilon}\frac{d\mathbf{C}_z^{-1}(\epsilon)}{d\epsilon}\tilde{\mathbf{H}}_0\right) \\ &= -2\mathrm{Tr}\left(\mathbf{R}_\mathrm{A}^H \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\frac{d\mathbf{C}_z(\epsilon)}{d\epsilon}\frac{d\mathbf{C}_z^{-1}(\epsilon)}{d\epsilon}\tilde{\mathbf{H}}_0\mathbf{R}_\mathrm{A}\right) \\ &= 2\mathrm{Tr}\left(\mathbf{R}_\mathrm{A}^H \tilde{\mathbf{H}}_0^H \mathbf{C}_z^{-1}(\epsilon)\frac{d\mathbf{C}_z(\epsilon)}{d\epsilon}\mathbf{C}_z^{-1}(\epsilon)\frac{d\mathbf{C}_z(\epsilon)}{d\epsilon}\right. \\ &\quad \left.\cdot \mathbf{C}_z^{-1}(\epsilon)\tilde{\mathbf{H}}_0\mathbf{R}_\mathrm{A}\right)\end{aligned} \tag{35}$$

where the second equality used the cyclic property of the trace and the third equality evaluated the derivative of $\mathbf{C}_z^{-1}(\epsilon)$. By also using the definitions of $\mathbf{Q}$ and $\mathbf{R}_z(\epsilon)$, (35) can be simplified to:

$$a_2(\epsilon) = 2\mathrm{Tr}\left(\mathbf{Q}\mathbf{Q}^H\right). \tag{36}$$

Finally, substituting $a_1(\epsilon)$ from (31) and $a_2(\epsilon)$ from (36) into (23) we have

$$\frac{d^2\, I(\epsilon)}{d\epsilon^2} = \frac{1}{\log 2}\mathbb{E}\left\{\mathrm{Tr}\left[\mathbf{Q}\Delta\mathbf{Q}^H\right]\right\} \tag{37}$$

where

$$\Delta \triangleq 2\mathbf{I} - \mathbf{R}_z(\epsilon)\tilde{\mathbf{H}}_0\mathbf{R}_\mathrm{A}\mathbf{R}_\mathrm{A}^H \tilde{\mathbf{H}}_0^H \mathbf{R}_z^H(\epsilon) \tag{38}$$

Thus, to prove that $\frac{d^2\, I(\epsilon)}{d\epsilon^2} > 0$ it is sufficient to prove that the matrix Δ is positive definite. Substituting (29) and (33):

$$\Delta = 2\mathbf{I} - \mathbf{R}_z(\epsilon)\tilde{\mathbf{H}}_0(\mathbf{I} + \tilde{\mathbf{H}}_0^H \mathbf{R}_z^H(\epsilon)\mathbf{R}_z(\epsilon)\tilde{\mathbf{H}}_0)^{-1}\tilde{\mathbf{H}}_0^H \mathbf{R}_z^H(\epsilon).$$

By also using the identity $(\mathbf{I} + \mathbf{AB})^{-1} = \mathbf{I} - \mathbf{A}(\mathbf{I} + \mathbf{BA})^{-1}\mathbf{B}$ (e.g., [33, Equation (166)]) leads to:

$$\Delta = \mathbf{I} + \left(\mathbf{I} + \mathbf{R}_z(\epsilon)\tilde{\mathbf{H}}_0\tilde{\mathbf{H}}_0^H \mathbf{R}_z^H(\epsilon)\right)^{-1} \tag{39}$$

which is obviously positive definite, which completes the proof of Lemma 2.

APPENDIX B
PROOF OF LEMMA 3

This appendix presents the proof of Lemma 3 , and hence completes the proof of Theorem 1 for the MMSE spectral efficiency, (9). The receiver extracts the decision variable of the k-th data stream by $\hat{x}_k = \mathbf{w}_k^H \mathbf{y}$, where $\mathbf{w}_k \in \mathbb{C}^{N\times 1}$ is the MMSE equalizer, (7). The total energy of the decision variable of the k-th stream is given by:

$$\begin{aligned}\mathbb{E}\{|\hat{x}_k|^2\} &= \mathbb{E}\{\mathbf{w}_k^H \mathbf{y}\mathbf{y}^H \mathbf{w}_k\} = \mathbf{w}_k^H \mathbf{C}_0 \mathbf{w}_k \\ &= \mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1} \mathbf{H}_0 \mathbf{f}_{0,k}. \end{aligned} \tag{40}$$

The energy of the desired signal in the decision variable of the k-th stream satisfies:

$$S_k = \left|\mathbf{w}_k^H (\gamma_0 \sqrt{\rho_{0,k}} \mathbf{H}_0 \mathbf{f}_{0,k})\right|^2 \mathbb{E}\left\{\left|u_{0,k}\right|^2\right\} \tag{41}$$

$$= \gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1} \mathbf{H}_0 \mathbf{f}_{0,k})^2. \tag{42}$$

Thus, the SINR of the k-th stream is:

$$\begin{aligned} SINR_k &= \frac{S_k}{\mathbb{E}\{|\hat{x}_k|^2\} - S_k} \\ &= \frac{\gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1} \mathbf{H}_0 \mathbf{f}_{0,k})^2}{\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1} \mathbf{H}_0 \mathbf{f}_{0,k} - \gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1} \mathbf{H}_0 \mathbf{f}_{0,k})^2} \\ &= \frac{1}{\frac{1}{\gamma_0^2 \rho_{0,k}} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1} \mathbf{H}_0 \mathbf{f}_{0,k})^{-1} - 1}, \end{aligned} \tag{43}$$

and the MMSE spectral efficiency of the k-th stream is $R_k = \mathbb{E}\{\log_2(1 + SINR_k)\}$.

As in the proof of Lemma 2, we consider an interference balancing change and replace R_k with $R_k(\epsilon)$, where $\mathbf{C}_0$ is replaced by $\mathbf{C}_0(\epsilon)$, and $\boldsymbol{\rho}_m$ is replaced by $\boldsymbol{\rho}(\epsilon)$. Noting that the MMSE spectral efficiency, (9), is $R = \sum_{k=1}^{K_0} R_k(\epsilon)$, we prove the monotonic increase of the MMSE spectral efficiency by showing that each stream satisfies:

$$\frac{d^2 R_k(\epsilon)}{d\epsilon^2} > 0 \tag{44}$$

$$\frac{dR_k(\epsilon)}{d\epsilon}\Big|_{\epsilon=0} = 0. \tag{45}$$

Substituting (43) into the spectral efficiency expression, the spectral efficiency of the k-th stream using MMSE equalizer is given by:

$$\begin{aligned} R_k(\epsilon) &= \mathbb{E}\left\{\log_2 \frac{\gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1}(\epsilon) \mathbf{H}_0 \mathbf{f}_{0,k})^{-1}}{\gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1}(\epsilon) \mathbf{H}_0 \mathbf{f}_{0,k})^{-1} - 1}\right\} \\ &= -\mathbb{E}\left\{\log_2\left(1 - \gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1}(\epsilon) \mathbf{H}_0 \mathbf{f}_{0,k})\right)\right\}. \end{aligned} \tag{46}$$

The first derivative of (46) is given by:

$$\begin{aligned} \frac{dR_k(\epsilon)}{d\epsilon} &= \frac{1}{\log_2} \mathbb{E}\Big\{\left(1 - \gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1}(\epsilon) \mathbf{H}_0 \mathbf{f}_{0,k})\right)^{-1} \\ &\quad \cdot \left(\gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \frac{d\mathbf{C}_0^{-1}(\epsilon)}{d\epsilon} \mathbf{H}_0 \mathbf{f}_{0,k})\right)\Big\}. \end{aligned} \tag{47}$$

Focusing on the second line, we have

$$\frac{d\mathbf{C}_0^{-1}(\epsilon)}{d\epsilon} = \mathbf{C}_0^{-1}(\epsilon) \frac{d\mathbf{C}_0(\epsilon)}{d\epsilon} \mathbf{C}_0^{-1}(\epsilon). \tag{48}$$

Recalling that $\frac{d\mathbf{C}_0(\epsilon)}{d\epsilon} = \frac{d\mathbf{C}_z(\epsilon)}{d\epsilon}$, we again have (as in (22)) $\frac{d\mathbf{C}_z(\epsilon)}{d\epsilon}\big|_{\epsilon=0}$ due to the symmetry with respect to interchanging $\tilde{\mathbf{f}}_{m,1}$ and $\mathbf{f}_{m,2}$. Thus, (45) is satisfied.

To prove (44), we consider the second derivative of $R_k(\epsilon)$. Taking the derivative of (47):

$$\frac{d^2 R_k(\epsilon)}{d\epsilon^2} = \frac{1}{\log_2} \mathbb{E}\Big\{b_1(\epsilon) + b_2(\epsilon)\Big\} \tag{49}$$

where

$$b_1(\epsilon) \triangleq \frac{d}{d\epsilon}\Big[\left(1 - \gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1}(\epsilon) \mathbf{H}_0 \mathbf{f}_{0,k})\right)^{-1}\Big] \tag{50}$$

$$\cdot\left(\gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \frac{d\mathbf{C}_0^{-1}(\epsilon)}{d\epsilon} \mathbf{H}_0 \mathbf{f}_{0,k})\right) \tag{51}$$

and

$$b_2(\epsilon) \triangleq \left(1 - \gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1}(\epsilon) \mathbf{H}_0 \mathbf{f}_{0,k})\right)^{-1} \tag{52}$$

$$\cdot \frac{d}{d\epsilon}\Big[\left(\gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \frac{d\mathbf{C}_0^{-1}(\epsilon)}{d\epsilon} \mathbf{H}_0 \mathbf{f}_{0,k})\right)\Big]. \tag{53}$$

In the following we show that $b_1(\epsilon) \geq 0$ and $b_2(\epsilon) \geq 0$. Starting with

$$b_1(\epsilon) = -\left(1 - \gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1}(\epsilon) \mathbf{H}_0 \mathbf{f}_{0,k})\right)^{-2} \tag{54}$$

$$\cdot \frac{d}{d\epsilon}\Big[\left(-\gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1}(\epsilon) \mathbf{H}_0 \mathbf{f}_{0,k})\right)\Big] \tag{55}$$

$$\cdot\left(\gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \frac{d\mathbf{C}_0^{-1}(\epsilon)}{d\epsilon} \mathbf{H}_0 \mathbf{f}_{0,k})\right) \tag{56}$$

$$= \left(1 - \gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1}(\epsilon) \mathbf{H}_0 \mathbf{f}_{0,k})\right)^{-2} \tag{57}$$

$$\cdot\left(\gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \frac{d\mathbf{C}_0^{-1}(\epsilon)}{d\epsilon} \mathbf{H}_0 \mathbf{f}_{0,k})\right)^2. \tag{58}$$

From (46) we see that $1 - \gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1}(\epsilon) \mathbf{H}_0 \mathbf{f}_{0,k}) = 1/(1 + SINR_k) \geq 0$ and hence $b_1(\epsilon) \geq 0$.

Next, considering $b_2(\epsilon)$ we have:

$$\begin{aligned} &b_2(\epsilon) \\ &= \left(1 - \gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1}(\epsilon) \mathbf{H}_0 \mathbf{f}_{0,k})\right)^{-1} \\ &\quad \cdot \frac{d}{d\epsilon}\Big[\left(-\gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1}(\epsilon) \frac{d\mathbf{C}_z(\epsilon)}{d\epsilon} \mathbf{C}_0^{-1}(\epsilon) \mathbf{H}_0 \mathbf{f}_{0,k})\right)\Big] \\ &= 2\left(1 - \gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1}(\epsilon) \mathbf{H}_0 \mathbf{f}_{0,k})\right)^{-1} \\ &\quad \cdot\Big[\Big(\gamma_0^2 \rho_{0,k} (\mathbf{f}_{0,k}^H \mathbf{H}_0^H \mathbf{C}_0^{-1}(\epsilon) \frac{d\mathbf{C}_z(\epsilon)}{d\epsilon} \\ &\qquad \cdot \mathbf{C}_0^{-1}(\epsilon) \frac{d\mathbf{C}_z(\epsilon)}{d\epsilon} \mathbf{C}_0^{-1}(\epsilon) \mathbf{H}_0 \mathbf{f}_{0,k})\Big)\Big] \geq 0. \end{aligned}$$

where we used $\frac{d^2 \mathbf{C}_z(\epsilon)}{d\epsilon^2} = 0$. This concludes the proof of Lemma 3 and Theorem 1 in the case of MMSE spectral efficiency.

REFERENCES

[1] V. W. Wong, *Key Technology for 5G wireless Systerm*. Cambridge, U.K.: Cambridge Univ. Press, 2017.
[2] I. F. Akyildiz, A. Kak, and S. Nie, "6G and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133995–134030, 2020.
[3] B. Wang, Y. Chang, and D. Yang, "On the SINR in massive MIMO networks with MMSE receivers," *IEEE Commun. Lett.*, vol. 18, no. 11, pp. 1979–1982, Nov. 2014.
[4] M. G. Khoshkholgh, K. G. Shin, K. Navaie, and V. C. M. Leung, "Coverage performance in multistream MIMO-ZFBF heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 6801–6818, Aug. 2017.
[5] M. G. Khoshkholgh and V. C. M. Leung, "Closed-form approximations for coverage probability of multistream MIMO-ZFBF receivers in HetNets," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 9862–9879, Nov. 2017.
[6] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov. 1999.
[7] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.
[8] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink multi-user MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1711–1721, May 2007.
[9] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.
[10] A. Leshem and E. Zehavi, "Cooperative game theory and the Gaussian interference channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1078–1088, Sep. 2008.
[11] Y. Richter and I. Bergel, "MMSE-SLNR precoding for multi-antenna cognitive radio," *IEEE Trans. Signal Process.*, vol. 62, no. 10, pp. 2719–2729, May 2014.
[12] N. Zhao, F. R. Yu, M. Jin, Q. Yan, and V. C. M. Leung, "Interference alignment and its applications: A survey, research issues, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1779–1803, 3rd Quart., 2016.
[13] X. Xia, K. Xu, D. Zhang, Y. Xu, and Y. Wang, "Beam-domain full-duplex massive MIMO: Realizing co-time co-frequency uplink and downlink transmission in the cellular system," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 8845–8862, Oct. 2017.
[14] K. Xu, Z. Shen, Y. Wang, X. Xia, and D. Zhang, "Hybrid time-switching and power splitting SWIPT for full-duplex massive MIMO systems: A beam-domain approach," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7257–7274, Aug. 2018.
[15] Z. Shen, K. Xu, X. Xia, W. Xie, and D. Zhang, "Spatial sparsity based secure transmission strategy for massive MIMO systems against simultaneous jamming and eavesdropping," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3760–3774, 2020.
[16] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst, and M. Haardt, "An introduction to the multi-user MIMO downlink," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 60–67, Oct. 2004.
[17] 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.
[18] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 996–1019, 3rd Quart., 2013.
[19] S. T. Veetil, K. Kuchi, and R. K. Ganti, "Performance of PZF and MMSE receivers in cellular networks with multi-user spatial multiplexing," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 4867–4878, 2015.
[20] G. George, R. K. Mungara, A. Lozano, and M. Haenggi, "Ergodic spectral efficiency in MIMO cellular networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2835–2849, 2017.
[21] N. Jindal, J. G. Andrews, and S. Weber, "Multi-antenna communication in ad hoc networks: Achieving MIMO gains with SIMO transmission," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 529–540, 2011.
[22] K. Huang, J. G. Andrews, D. Guo, R. W. Heath Jr, and R. A. Berry, "Spatial interference cancellation for multiantenna mobile ad hoc networks," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1660–1676, 2012.
[23] R. Vaze and R. W. Heath, "Transmission capacity of ad-hoc networks with multiple antennas using transmit stream adaptation and interference cancellation," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 780–792, Feb. 2012.
[24] Y. Richter and I. Bergel, "Opportunistic routing based on partial CSI in MIMO random ad-hoc networks," Tech. Rep., 2020.
[25] F. Perez-Cruz, M. R. D. Rodrigues, and S. Verdu, "MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070–1084, Mar. 2010.
[26] C. Xiao, Y. R. Zheng, and Z. Ding, "Globally optimal linear precoders for finite alphabet signals over complex vector Gaussian channels," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3301–3314, Jul. 2011.
[27] M. Franceschetti, "Stochastic rays pulse propagation," *IEEE Trans. Antennas Propag.*, vol. 52, no. 10, pp. 2742–2752, Oct. 2004.
[28] Taesa and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.
[29] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.
[30] T. Gou and S. A. Jafar, "Degrees of freedom of the K user $M \times N$ MIMO interference channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6040–6057, Dec. 2010.
[31] M. Haenggi, *Stochastic Geometry for Wireless Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[32] Y. Richter and I. Bergel, "Analysis of the simulated aggregate interference in random ad-hoc networks," in *Proc. IEEE 15th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2014, pp. 374–378.
[33] K. B. Petersen and M. S. Pedersen. (Nov. 2012). *The Matrix Cookbook*. [Online]. Available: http://www2.imm.dtu.dk/pubdb/p.php

YIFTACH RICHTER (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Bar-Ilan University, Ramat-Gan, Israel, in 2011 and 2013, respectively, where he is currently pursuing the Ph.D. degree with the Faculty of Engineering. His research interests include MIMO, routing, and stochastic geometry for wireless communication networks.

ITSIK BERGEL (Senior Member, IEEE) received the B.Sc. degree in electrical engineering and in physics from Ben Gurion University, and the M.Sc. and Ph.D. degrees in electrical engineering from Tel Aviv University. From 2001 to 2003, he was a Senior Researcher with Intel Communications Research Laboratory. In 2005, he was a Postdoctoral Researcher with the Politecnico di Torino, Italy. He is currently a Faculty Member with the Faculty of Engineering, Bar-Ilan University, Israel. His main research interests include interference mitigation in wireline and wireless communications, cooperative transmission, and cross layer optimization of random networks. He served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and currently serves as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.