Queueing Stability and CSI Probing of a TDD Wireless Network with Interference Alignment

Matha Deghel, *Student*, *IEEE*, Mohamad Assaad, *Senior*, *IEEE*,
Mérouane Debbah, *Fellow*, *IEEE*, and Anthony Ephremides, *Life Fellow, IEEE*

Abstract

This paper characterizes the performance in terms of queueing stability of a network composed of multiple MIMO transmitter-receiver pairs taking into account the dynamic traffic pattern and the probing/feedback cost. We adopt a centralized scheduling scheme that selects a number of active pairs in each time-slot. We consider that the transmitters apply interference alignment (IA) technique if two or more pairs are active, whereas in the special case where one pair is active point-to-point MIMO singular value decomposition (SVD) is used. We consider a time-division duplex (TDD) system where transmitters acquire their channel state information (CSI) by decoding the pilot sequences sent by the receivers. Since global CSI knowledge is required for IA, the transmitters have also to exchange their estimated CSIs over a backhaul of limited capacity (i.e. imperfect case). Under this setting, we characterize in this paper the stability region of the system under both the imperfect and perfect (i.e. unlimited backhaul) cases, then we examine the gap between these two resulting regions. Further, under each case we provide a centralized probing policy that achieves the max stability region. These stability regions and scheduling policies are given for the symmetric system, where all the path loss coefficients are equal to each other, as well as for the general system. For the symmetric system, we provide the conditions under which IA yields a queueing stability gain compared to SVD. Under the general system, the adopted scheduling policy is of a high computational complexity for moderate numbers of pairs, consequently we propose an approximate policy that has a reduced complexity but that achieves only a fraction of the system stability region. A characterization of this fraction is provided.

M. Deghel and M. Assaad are with Laboratoire de Signaux et Systèmes (L2S, CNRS, UMR8506) CentraleSupélec, 3 rue Joliot-Curie, 91192, Gif-sur-Yvette, cedex. France. (matha.deghel@centralesupelec.fr, mohamad.assaad@centralesupelec.fr).

M. Debbah is with the Large Systems and Networks Group (LANEAS), CentraleSupélec, Gif-sur-Yvette, France (merouane.debbah@centralesupelec.fr) and also with the Mathematical and Algorithmic Sciences Lab, Huawei Technologies Co. Ltd., France (merouane.debbah@huawei.com).

A. Ephremides is with the Department of Electrical and Computer Engineering and Institute for Systems Research University of Maryland, College Park, MD 20742. (etony@umd.edu).

Parts of this paper have been presented at the IEEE International Conference on Communications (ICC), London, U.K., 2015 [1] and at the IEEE International Symposium on Information Theory (ISIT), Hong Kong, 2015 [2].

Index Terms

MIMO channel, queueing, stability, interference alignment, singular value decomposition

I. Introduction

One of the key issues in wireless communication systems is the interference that is caused by a large number of users communicating on the same channel, resulting into severe performance degradations unless treated properly. In this regard, interference alignment (IA) was introduced [3] as an efficient interference management technique and is shown to result in higher throughputs compared to conventional interference-agnostic methods. Indeed, IA is a linear precoding technique that attempts to align interfering signals in time, frequency, or space. In multiple-input multiple-output (MIMO) networks, IA utilizes the spatial dimension offered by multiple antennas for alignment. By aligning interference at all receivers (i.e. users), IA reduces the dimension of interference, allowing users to suppress interference via linear techniques and decode their desired signals interference free. A major challenge of the above IA scheme lies in the fact that the global channel state information (CSI) must be available at each transmitter. In scenarios where the receivers quantize and send the CSI back to the transmitters, the IA scheme is explored over frequency selective channels for single-antenna users in [4] and for multiple-antenna users in [5]. Both references provide degree-of-freedom (DoF)-achieving quantization schemes and establish the required scaling of the number of feedback bits. For alignment using spatial dimensions, [6] provides the scaling of feedback bits to achieve IA in MIMO interference channel (IC). To overcome the problem of scaling codebook size, and relax the reliance on frequency selectivity for quantization, [7] proposed an analog feedback strategy for constant MIMO interference channels. In [8], the Grassmannian Manifold quantization technique was adopted to reduce the information exchange over the backhaul.

We draw the attention to the fact that IA technique can be used with a number of transmitter-receiver pairs greater than or equal to two. On the other side, for the special case where we have point-to-point MIMO system, i.e. only one pair, singular value decomposition (SVD) technique can be applied and was shown to provide very good performances [9], [10]; note that in this case other techniques can be used such as zero forcing (ZF) and matched filtering (MF).

All the above cited works, however, do not account for the dynamic traffic processes of the users, i.e. they assume users with infinite back-logged data. In practice, the users of a wireless network request data, so it is of great interest to investigate the impact of MIMO in the higher layers [11], more specifically in the media access control (MAC) layer. The cross-layer design goal here is the achievement of the entire *stability region* of the system. In broad terms, the stability region of a network is the set of arrival rate vectors such that the entire network load can be served by some service policy without an infinite blow up of any queue. The special *scheduling policy* achieving the entire stability region, called the optimal policy, is hereby of particular interest. The concept of stability-optimal operation comes originally from the control and automation theory [12]. It was applied to the wireless communication systems first in [13], and the view was extended by some bounds in [14]. Since then, this concept has been extensively investigated in the wireless framework under various traffic and network scenarios. For instance, the authors in [15] have considered the broadcast channel (BC) and proposed a technique based on ZF precoding, with a heuristic user scheduling scheme that selects users whose channel states are nearly orthogonal vectors and illustrate the stability region this policy achieves via simulations. In [16], it has been noticed that the policy resulting from the minimization of the drift of a quadratic Lyapunov function is to solve a weighted sum rate maximization problem (with weights being the queue lengths) each time-slot and they propose an iterative water-filling algorithm for this purpose. In addition, taking into account the probing cost, the authors in [17] have examined three different scheduling policies (centralized, decentralized and mixed policies) for MISO wireless downlink systems under ZF precoding technique. Further, in [18] the authors studied the impact of channel state quantization on the stability of a system using ZF precoding under a centralized scheme. Compared with the above works, in this paper we consider a system with a different physical layer based on IA, which we will present later.

As alluded earlier, taking into consideration the traffic arrivals of the users, the key performance metric we examine in this work is the queueing stability. Specifically, for a control policy, the main performance measure we adopt and characterize is the stability region this policy can achieve. Under the condition that the mean arrival-rate vectors are inside the stability region, stability implies that the mean queue length of every queue in the system is finite and, by *Little's* theorem, it also implies finite average packet-delays [19]. It is worth mentioning that for delay-sensitive systems we usually have to deal with control problems under harder delay constraints, such as minimizing the average delay or guaranteeing that the average delay is lower than a certain bound. Several approaches can be used to deal with such delay-aware problems, such as the Lyapunov stability drift approach and the Markov decision process approach; see [20]–[23] and references therein. Note that in our work here the delay is not a figure of interest in the analysis, and the main focus is on the stability of the system.

In this paper, we consider a MIMO network where multiple transmitter-receiver pairs operate in time-division duplex (TDD) mode under backhaul links of limited capacity. Each transmitter acquires its local CSI from its corresponding user by exploiting the channel reciprocity. We use (pre-assigned) orthogonal pilot sequences among the users, so the length of each one of these sequences should be proportional to the number of active users in the system. It means that after acquiring the CSI of, for example, L users, the throughput is multiplied by $1 - L\theta$, where θ is the fraction of time that takes the CSI acquisition of one user [24]. Depending on the number of scheduled pairs at a time-slot, we distinguish two cases: (i) if the number of scheduled pairs is greater than or equal to two, IA technique is applied, and each transmitter needs to send a quantized version of its local CSI to other transmitters using a fixed number of bits B per channel matrix, and (ii) if only one pair is scheduled, we apply SVD technique. It is important to focus on the *trade-off* between having a large number of active transmitter-receiver pairs (so having a high probing cost but many pairs can communicate simultaneously) and having much time of the slot dedicated to data transmission (which means getting a low probing cost but few pairs can communicate simultaneously) [17]. In order to choose the subset of active pairs at each slot, we adopt a centralized scheme where the decision of which pairs to schedule is made at a central scheduler (CS) based only on the statistics of the channels of the users and the state of their queue lengths at each slot [18]. Note that the centralized approach is used in current standards (e.g. LTE [25]), where the base station explicitly requests some users for their CSI.

In addition to increasing the CSI probing cost, scheduling more pairs requires more backhaul usage since under IA technique the active transmitters need to exchange their CSI knowledge. In some scenarios, it may be beneficial not to occupy the backhaul with this huge amount of signaling but instead exploited it more efficiently. For instance, if the backhaul is wireless, the CSI exchange process consumes a part of the total reserved bandwidth, which can be instead used in the transmission process. Hence, it is of high interest to study the system under an interference management technique for which no CSI exchange over the backhaul is required. For this purpose, we investigate the system performance under time division multiple access (TDMA) as a channel access method, meaning that there is only one active pair at a given time-slot and thus no backhaul usage occurs, and using SVD as a precoding technique. One may wonder which one between TDMA-SVD and IA outperforms the other in terms of stability. We will provide an answer to this challenging question by comparing the stability performances of these two techniques. To the best of our knowledge, among all the works (which we have cited some of them earlier, e.g. [15]–[18]) considering the impact of MIMO in the higher layers, this is the first work that considers the contrast of SVD with IA in terms of queueing stability taking into account the probing cost and the fact that the backhaul is of limited capacity.

The *main contributions* and the organization of the paper are summarized as follows.

• Section II presents the adopted system model and the interaction between the physical layer and the queueing performance.

• The average rate expressions for both the perfect (i.e. unlimited backhaul capacity) and imperfect (i.e. limited backhaul capacity) cases are derived in Section III.

• In Section IV, we present a deep stability analysis for the *symmetric* system where all the path loss coefficients are equal to each other. Specifically, for this system:

 ⋆ We provide a precise characterization of the system stability region and we propose an optimal scheduling decision to achieve this region in both the perfect and imperfect cases.

 ⋆ Under both cases, we investigate the conditions under which the use of IA (i.e. applied if the number of active pairs ≥ 2) can yield a queueing stability gain compared to SVD (i.e. applied if there is only 1 active pair).

 ⋆ We examine the maximum gap between the stability region under the imperfect case and the stability region under the perfect case. We also investigate the impact of changing the number of bits on the system stability region.

• In Section V, we present a stability analysis for the *general* system where the path loss coefficients are not necessarily equal to each other. In detail, for this system:

 ⋆ We investigate the stability performances by characterizing the system stability region and providing an optimal scheduling policy under both the imperfect and perfect cases.

 ⋆ Since the scheduling policy is of a high computational complexity, under the imperfect case we propose an approximate policy that has a reduced complexity but that achieves only a fraction of the stability region of the imperfect case; we point out that this policy relies on an average rate approximation (for the imperfect case) that we first calculate. A characterization of the fraction this policy achieves is also provided.

 ⋆ Using the average rate approximation mentioned above, we examine the gap between the stability region under the imperfect case and the stability region under the perfect case.

• Section VI is dedicated to numerical results and relevant discussions.

• Finally, Section VII concludes the paper.

Notation: Boldface uppercase symbols represent matrices and lowercases are used for vectors, unless stated otherwise. The symbol $\mathbf{I}_N$ denotes the identity matrix of size N. The operator $\otimes$ is the Kronecker product. The notation $|\cdot|$ is used to indicate the absolute value for scalars and the cardinality for sets (or subsets). $\mathbb{1}_{(\cdot)}$ is the indicator function. In addition, $\|\cdot\|_1$ and $\|\cdot\|$ are used for the norms of first and second degree, respectively. Finally, superscripts T and H over a matrix or vector denote its transpose and conjugate transpose, respectively.

II. System Model

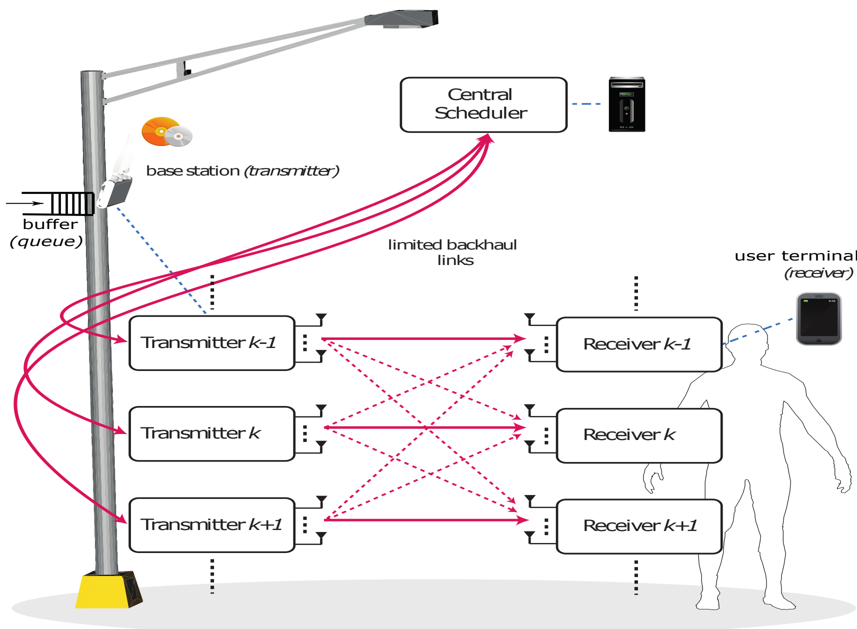

Figure 1: A sketch of an N-user MIMO interference channel with limited backhaul.

We consider the MIMO interference channel with N transmitter-receiver pairs shown in Figure 1. For simplicity of exposition, we consider a homogeneous network where all transmitters are equipped with N_t antennas and all receivers (users) with N_r antennas. We assume that time is slotted. As will be explained later, only a subset $\mathcal{L}(t)$, of cardinality $L(t)$, of pairs is active at time-slot t, with $L(t) \leq N$. Transmitter k has $d_k \leq \min\left(N_\mathrm{t}, N_\mathrm{r}\right)$ independent data streams to transmit to its intended user k. For the cases where $L(t) \geq 2$, while each transmitter communicates with its intended receiver, it also creates interference to other $L(t) - 1$ unintended receivers.

Given this channel model, the received signal at active user k $(\in \mathcal{L}(t))$ can be expressed as

$$\mathbf{y}_k = \sum_{i \in \mathcal{L}(t)} \sqrt{\frac{\zeta_{ki} P}{d_i}} \mathbf{H}_{ki} \sum_{j=1}^{d_i} \mathbf{v}_i^{(j)} x_i^{(j)} + \mathbf{z}_k, \tag{1}$$

where $\mathbf{y}_k$ is the $N_r \times 1$ received signal vector, $\mathbf{z}_k$ is the additive white complex Gaussian noise with zero mean and covariance matrix $\sigma^2 \mathbf{I}_{N_\mathrm{r}}$, $\mathbf{H}_{ki}$ is the $N_\mathrm{r} \times N_\mathrm{t}$ channel matrix between transmitter i and receiver k with independent and identically distributed (i.i.d.) zero mean and unit variance complex Gaussian entries, ζ_{ki} represents the path loss of channel $\mathbf{H}_{ki}$, $x_i^{(j)}$ represents the j-th data stream from transmitter i, P is the total power at each transmitting node, which is equally allocated among its data streams, and $\mathbf{v}_i^{(j)}$ is the corresponding $N_\mathrm{t} \times 1$ precoding vector of unit norm. For the rest of the paper we denote by α_{ki} the fraction $\frac{\zeta_{ki} P}{d_i}$, i.e. $\alpha_{ki} = \frac{\zeta_{ki} P}{d_i}$.

The scheduling rule to select the subset of active pairs at each time-slot will be discussed later in the paper. Depending on the number of scheduled pairs $L(t)$, two cases are to consider:

- If $L(t) = 1$, i.e. only one pair is active at time-slot t: in this case we use SVD (singular value decomposition) technique, which was shown to provide very good performances for point-to-point MIMO systems [9]. Note that other techniques can be considered, such as ZF (zero forcing) and MF (matched filtering).
- If $L(t) \geq 2$: in this case we perform IA (interference alignment) technique, which was shown to provide good performances for multipoint-to-multipoint MIMO systems [3].

For clarity of exposition, in this section we present the IA scheme and in the next Section (Section III, in which we also derive the average rates expressions) we provide the SVD scheme.

A. *Interference Alignment Technique*

IA is an efficient linear precoding technique that often achieves the full DoF supported by MIMO interference channels. In cases where the full DoF cannot be guaranteed, IA has been shown to provide significant gains in high signal-to-noise ratio (SNR) sum-rate. To investigate IA in our model, we start by examining the effective channels created after precoding and combining. For tractability, we restrict ourselves to a per-stream zero-forcing receiver. Recall that in the high (but finite) SNR regime, in which IA is most useful, gains from more involved receiver designs are limited [26]. Under the adopted system, receiver k uses the $N_\mathrm{r} \times 1$ *combining* vector $\mathbf{u}_k^m$ of unit norm to detect its m-th stream, such as

$$\begin{aligned}\hat{x}_k^{(m)} &= \left(\mathbf{u}_k^{(m)}\right)^H \mathbf{y}_k \\ &= \overbrace{\sqrt{\alpha}_{kk}\left(\mathbf{u}_k^{(m)}\right)^H \mathbf{H}_{kk}\mathbf{v}_k^{(m)} x_k^{(m)}}^{\text{desired signal}} + \overbrace{\sqrt{\alpha}_{kk}\sum_{j=1, j\neq m}^{d_k}\left(\mathbf{u}_k^{(m)}\right)^H \mathbf{H}_{kk}\mathbf{v}_k^{(j)} x_k^{(j)}}^{\text{inter-stream interference (ISI)}} \\ &\quad + \overbrace{\sum_{i\in\mathcal{L}(t), i\neq k}\sqrt{\alpha}_{ki}\sum_{j=1}^{d_i}\left(\mathbf{u}_k^{(m)}\right)^H \mathbf{H}_{ki}\mathbf{v}_i^{(j)} x_i^{(j)}}^{\text{inter-user interference (IUI)}} + \overbrace{\left(\mathbf{u}_k^{(m)}\right)^H \mathbf{z}_k}^{\text{noise}}, \end{aligned} \tag{2}$$

where the first term at the right-hand-side of this expression is the desired signal, the second one is the inter-stream interference (ISI) caused by the same transmitter, and the third one is the inter-user interference (IUI) resulting from the other transmitters. In order to mitigate these interferences and improve the system performances, IA is performed accordingly, that is designing the set of combining and precoding vectors such that

$$\left(\mathbf{u}_k^{(m)}\right)^H \mathbf{H}_{ki}\mathbf{v}_i^{(j)} = 0, \qquad \forall (i,j) \neq (k,m), \text{with}\, i,k \in \mathcal{L}(t). \tag{3}$$

In the following we state some additional assumptions on the design of these vectors [7]: (a) vector $\mathbf{v}_i^{(j)}$ is a function of all the cross channels $\mathbf{H}_{kl}$ $\forall k, l$, $k \neq l$ only, (b) vector $\mathbf{u}_k^{(m)}$ is a function of vectors $\mathbf{H}_{ki}\mathbf{v}_i^{(j)}$ $\forall i \neq k$, $\forall j$ and $\mathbf{H}_{kk}\mathbf{v}_k^{(j)}$, $\forall j \neq m$, and (c) matrix $\mathbf{V}_k = \left[\mathbf{v}_k^{(1)}, \ldots, \mathbf{v}_k^{(d_k)}\right]$ is unitary. From these properties/assumptions we can deduce that $\mathbf{u}_k^{(m)}$ is independent of $\mathbf{H}_{kk}\mathbf{v}_k^{(m)}$. Note that the conditions in (3) are those of a *perfect interference alignment*. In other words, suppose that all the transmitting nodes have perfect global CSI and each receiver obtains a perfect version of its corresponding combining vectors, ISI and IUI can be suppressed completely. However, obtaining the perfect global CSI at the transmitters is not always practical due to the fact that backhaul links, which connect transmitters to each other, are of limited capacity. The CSI sharing mechanism is detailed in the next subsection.

Finally, some remarks are in order. We note that the approach to design the IA vectors is not a figure of interest in our contribution, however our analysis holds for every approach that produces IA vectors with simultaneously all the properties given before [7]; such an approach was the subject of investigation in several works, as for instance [3], [27]. In addition, we assume that each active receiver obtains a perfect version of its corresponding combining vectors. The cost of this latter process is not considered in our analysis. Moreover, we do not perform power control for our system. This is done to further simplify the transmission scheme.

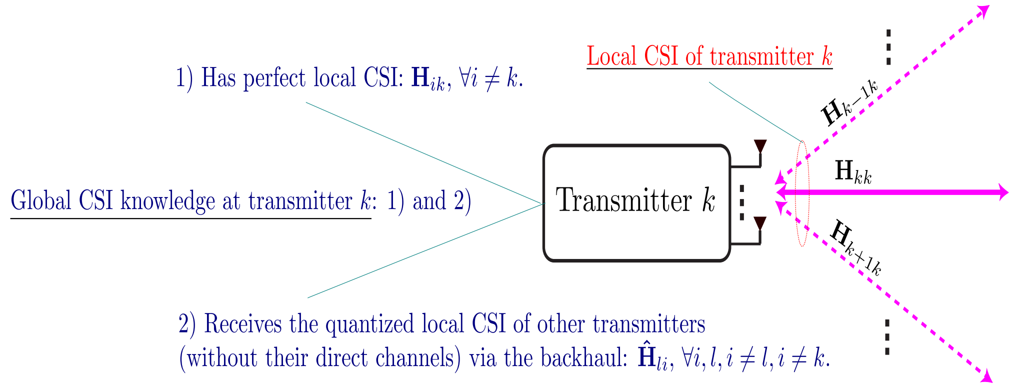

Figure 2: Global CSI knowledge at transmitter k under the imperfect case.

B. CSIT Sharing Over Limited Capacity Backhaul Links

The process of CSI sharing is restricted to the scheduled pairs (represented by subset $\mathcal{L}(t)$). Thus, here, even if we did not mention it, when we write "transmitter" (resp., "user") we mean "active transmitter" (resp., "active user").

We adopt a scenario where each transmitter receives all the required CSI and independently computes the IA vectors. As alluded earlier, global CSI is required at each transmitting node in order to design the IA vectors that satisfy (3). As shown in Figure 1, we suppose that all the transmitters are connected to a CS via their limited backhaul links, meaning that this CS serves as a way for connecting the transmitters to each other; as we will see later on, this scheduler decides which pairs to schedule at each time-slot. We assume a TDD transmission strategy, which enables the transmitters to estimate their channels toward different users by exploiting the reciprocity of the wireless channel. We consider throughout this paper that there are no errors in the channel estimation. Under the adopted strategy, the users send their training sequences in the uplink phase, allowing each transmitter to estimate (perfectly) its *local* CSI, meaning that the k-th transmitter estimates perfectly the channels $\mathbf{H}_{ik}$, $\forall i \in \mathcal{L}(t)$. However, the local CSI of other transmitters, excluding the direct channels (i.e. for the k-th transmitter the direct channel is given by matrix $\mathbf{H}_{kk}$), are obtained via backhaul links of limited capacity. In such limited backhaul conditions, a codebook-based quantization technique needs to be adopted to reduce the huge amount of information exchange used for CSI sharing, which we detail as follows. Let $\mathbf{h}_{ik}$ denote the vectorization of the channel matrix $\mathbf{H}_{ik}$. Then, for each $i \neq k$, transmitter k selects the index n_o that corresponds to the optimal codeword in a predetermined codebook

$\mathcal{CB} = \left[\hat{\mathbf{h}}_{ik}^{(1)}, ..., \hat{\mathbf{h}}_{ik}^{(2^B)}\right]$ according to

$$n_o = \arg \max_{1 \leq n \leq 2^B} \left| \left(\tilde{\mathbf{h}}_{ik}\right)^H \hat{\mathbf{h}}_{ik}^{(n)} \right|^2, \tag{4}$$

in which B is the number of bits used to quantize $\mathbf{H}_{ik}$ and $\tilde{\mathbf{h}}_{ik} = \mathbf{h}_{ik} \left\|\mathbf{h}_{ik}\right\|^{-1}$ is the channel direction vector. After quantizing all the matrices of its local CSI, we assume that transmitter k sends the corresponding optimal indexes to all other active transmitters, which share the same codebook, allowing these transmitters to reconstruct the quantized local knowledge of transmitter k. We point out that each transmitter quantizes its local CSI excluding the direct channel since, as noted earlier, it is not required when computing the IA vectors of other transmitters. The global CSI knowledge at transmitter k is summarized in Figure 2.

Let us now define the quantization error as $e_{ik} = 1 - \left(|\,\hat{\mathbf{h}}_{ik}^H \mathbf{h}_{ik}|^2 (\|\mathbf{h}_{ki}\|^2)^{-1}\right)$, where $\hat{\mathbf{h}}_{ik}$ is the quantization of $\mathbf{h}_{ik}$, and adopt a similar model to the one used in [18], [28, Lemma 6], which relies on the theory of *quantization cell approximation*. Let $Q = N_{\mathrm{t}} N_{\mathrm{r}} - 1$. The cumulative distribution function (CDF) of e_{ik} is then given by the following

$$\mathbb{P}\left\{e_{ik} \leq \varepsilon\right\} = \begin{cases} 2^B \varepsilon^Q, & \text{for } 0 \leq \varepsilon \leq 2^{-\frac{B}{Q}} \\ 1, & \text{for } \varepsilon > 2^{-\frac{B}{Q}} \end{cases} \tag{5}$$

As a final remark, we note that the number of bits B is assumed to be fixed and does not change with the number of scheduled pairs $L(t)$. This assumption is made to simplify the analysis, because otherwise we should consider the relation between these two numbers, which would add a considerable amount of complexity to the analysis.

C. Rate Model and Impact of Training

Before proceeding with the description, we define the *perfect case* as the case where the backhaul has an infinite capacity, which leads to perfect global CSI knowledge at the transmitters; so no quantization is needed. Further, we call *imperfect case* the model described previously, where a quantization is performed over the backhaul of limited capacity and a *fixed* number of bits B is used to quantize each channel matrix.

As explained in the previous subsection, for the perfect case the IA constraints null the ISI and the IUI, and no residual interference exists. On the other hand, for the imperfect case we recall that each (active) transmitter quantizes its local CSI, excluding the direct channel, and sends it to all other (active) transmitters. At the same time, this transmitter receives the quantized local CSI (except the direct channel) of each other transmitter. Let us denote $\hat{\mathbf{H}}_{ki}$ as the quantization version of $\mathbf{H}_{ki}$, which can be formed by reshaping vector $\hat{\mathbf{h}}_{ki}$. Then, e.g., transmitter k designs its IA vectors based on: (a) $\mathbf{H}_{kk}$, i.e. perfect direct channel, (b) $\hat{\mathbf{H}}_{ik}$ $\forall i \neq k$, i.e. quantized version of the channels of its local CSI (without the direct channel), and (c) $\hat{\mathbf{H}}_{li}$ $\forall i, l, i \neq l, i \neq k$, i.e. the imperfect knowledge it receives from other transmitters. It can be noticed that although transmitter k has a perfect version of the $\mathbf{H}_{ik}$ $\forall i \neq k$, this transmitter uses the quantized version of these channels in the design of its IA vectors. This is considered so that each transmitter can compute the (same) precoding vectors of other transmitters, where we recall that these vectors are used in the computation of the combining vector of this transmitter. Based on the above, under the imperfect case the IA technique is able to completely cancel the ISI but not the IUI. In other words, under this case we have

$$\begin{aligned} &\left(\hat{\mathbf{u}}_k^{(m)}\right)^H \hat{\mathbf{H}}_{ki}\hat{\mathbf{v}}_i^{(j)} = 0, \qquad \left(\hat{\mathbf{u}}_k^{(m)}\right)^H \mathbf{H}_{ki}\hat{\mathbf{v}}_i^{(j)} \neq 0, \qquad \forall i \neq k, \forall m, \forall j, \text{and}\, k, i \in \mathcal{L}(t), \\ &\left(\hat{\mathbf{u}}_k^{(m)}\right)^H \mathbf{H}_{kk}\hat{\mathbf{v}}_k^{(j)} = 0, \qquad \forall j \neq m, \forall k, \text{and}\, k \in \mathcal{L}(t), \end{aligned} \tag{6}$$

where $\hat{\mathbf{v}}_k^{(m)}$ and $\hat{\mathbf{u}}_k^{(m)}$, $\forall k, m$, are the precoding and combining vectors, respectively, designed under the imperfect case. Similarly to the perfect case, for the imperfect case the following properties hold [7]: (a) vector $\hat{\mathbf{v}}_i^{(j)}$ is a function of all the cross channels $\hat{\mathbf{H}}_{kl}$ $\forall k, l,$ $k \neq l$ *only*, (b) vector $\hat{\mathbf{u}}_k^{(m)}$ is a function of vectors $\hat{\mathbf{H}}_{ki}\hat{\mathbf{v}}_i^{(j)}$ $\forall i \neq k$, $\forall j$ and $\mathbf{H}_{kk}\hat{\mathbf{v}}_k^{(j)}$, $\forall j \neq m$, and (c) matrix $\hat{\mathbf{V}}_k = \left[\hat{\mathbf{v}}_k^{(1)}, \ldots, \hat{\mathbf{v}}_k^{(d_k)}\right]$ is unitary. Thus, we can deduce that $\hat{\mathbf{u}}_k^{(m)}$ is independent of $\mathbf{H}_{kk}\hat{\mathbf{v}}_k^{(m)}$. Using the above, the SINR/SNR for stream m at active receiver k can be written as

$$\gamma_k^{(m)} = \begin{cases} \dfrac{\alpha_{kk}\left|\left(\hat{\mathbf{u}}_k^{(m)}\right)^H \mathbf{H}_{kk}\hat{\mathbf{v}}_k^{(m)}\right|^2}{\sigma^2 + \sum\limits_{i \in \mathcal{L}(t), i \neq k} \alpha_{ki} \sum\limits_{j=1}^{d_i} \left|\left(\hat{\mathbf{u}}_k^{(m)}\right)^H \mathbf{H}_{ki}\hat{\mathbf{v}}_i^{(j)}\right|^2}, & \text{imperfect case} \\[2ex] \dfrac{\alpha_{kk}\left|\left(\mathbf{u}_k^{(m)}\right)^H \mathbf{H}_{kk}\mathbf{v}_k^{(m)}\right|^2}{\sigma^2}, & \text{perfect case} \end{cases} \tag{7}$$

where we recall that $\mathbf{v}_k^{(m)}$ and $\mathbf{u}_k^{(m)}$, $\forall k, m$, are designed under the perfect case (i.e. using perfect global CSI). As alluded earlier, only a subset $\mathcal{L}(t)$ of pairs is scheduled at a time, where we recall that $L(t) = |\mathcal{L}(t)|$. For notational convenience, we will use signal-to-interference-plus-noise ratio (SINR) as a general notation to denote SNR for the perfect case and SINR for the imperfect case, unless stated otherwise.

We now explain some useful points that are adopted in the rate model. At a given time-slot, the instantaneous rate R of stream m of pair k is transmitted successfully if the SINR $\gamma_k^{(m)}$ of receiver (i.e. user) k is higher than or equal to a given threshold τ; otherwise, the transmission is not successful and the instantaneous bit rate is 0. Obviously, the choice of τ depends on R and vice-versa. The relation (mainly point to point relation) between the SINR threshold and the bit rate has been studied widely in the literature and therefore lies out of the scope of this paper. In fact, our analysis is valid for any point to point relation between τ and R. Our assumption is simply that a given rate R is transmitted if the SINR is above a given threshold τ. Let us denote by $\tilde{R}_k(t)$ the assigned rate (in units of bits/slot) for user k at time-slot t, thus $\tilde{R}_k(t)$ is the sum of the assigned rates for all the streams of user k at time-slot t. In other words, we have

$$\tilde{R}_k(t) = \sum_{m=1}^{d_k} R\,\mathbb{1}_{\left(\gamma_k^{(m)}(t) \geq \tau\right)}, \tag{8}$$

where $\mathbb{1}_{(\cdot)}$ is the indicator function. For this model, channel acquisition cost is not negligible and should be considered. As mentioned earlier, we consider a system under TDD mode where users send training sequences in the uplink so that the transmitters can estimate their channels. This scheme uses orthogonal sequences among the users, so their lengths are proportional to the number of active users in the system. We assume that acquiring the CSI of one user takes fraction θ of the slot. Thus, since we have $L(t)$ active users, the *actual* rate for transmission to active user k at time-slot t is $(1 - L(t)\theta)\tilde{R}_k(t)$. We denote this rate by $D_k(t)$, i.e.

$$D_k(t) = (1 - L(t)\theta)\tilde{R}_k(t) = (1 - L(t)\theta) \sum_{m=1}^{d_k} R\,\mathbb{1}_{\left(\gamma_k^{(m)}(t) \geq \tau\right)}. \tag{9}$$

Note that $D_k(t)$ is set to 0 if pair k is not active at time-slot t.

Under the above setting, the *average rate* for active user k can be written in function of the transmission *success probability* (i.e. the probability that the corresponding SINR is greater than or equal to a certain threshold) conditioned on the subset of active pairs as

$$\mathbb{E}\left\{D_k(t) \mid \mathcal{L}(t)\right\} = (1 - L(t)\theta) \sum_{m=1}^{d_k} R\,\mathbb{P}\left\{\gamma_k^{(m)}(t) \geq \tau \mid \mathcal{L}(t)\right\}. \tag{10}$$

It can be noticed that the feedback overhead $(1 - L(t)\theta)$ scales with the number of active pairs, meaning that when $L(t)$ is large there will be little time left to transmit in the time-slot before the channels change again. We point out that if $L(t) = 1$, then $\gamma_k^{(m)}(t)$ is the SINR obtained using the SVD scheme, which will be presented in Section III.

D. Queue Dynamics, Stability and Scheduling Policy

For each user, we assume that the incoming data is stored in a respective queue (i.e. buffer) until transmission, and we denote by $\mathbf{q}(t) = (q_1(t), ..., q_N(t))$ the queue length vector. We designate by $\mathbf{A}(t) = (A_1(t), ..., A_N(t))$ the vector of number of bits arriving in the buffers in time-slot t, which is an i.i.d. in time process, independent across users and with $A_k(t) < A_{\max}$. The mean arrival rate (in units of bits/slot) for user k is denoted by $a_k = \mathbb{E}\{A_k(t)\}$. We recall that, e.g., user k will get $D_k(t)$ served bits per slot if it gets scheduled and zero otherwise. Note that $D_k(t)$ is finite because R is finite, so we can define a finite positive constant $D_{\max}$ such that $D_k(t) < D_{\max}$, for $k = 1, \ldots, N$.

At each time-slot, the CS selects the pairs to schedule based on the queue lengths and the average rates in the system. To this end, we suppose that: (i) this scheduler has full knowledge of average rate values under different combinations of choosing active pairs, which can be provided offline since an average rate is time-independent, (ii) at each time-slot, each transmitter sends its queue length to the CS so that it can obtain all the queue dynamics of the system, and (iii) the cost of providing such knowledge to the scheduler will not be taken into account in our analysis. After selecting the set of pairs to be scheduled (represented by $\mathcal{L}(t)$), the CS broadcasts this information so that the selected transmitter-user pairs activate themselves, and then the active users send their pilots in the uplink so that the (active) transmitters can estimate the CSI. It is worth noting that, as alluded previously, if we select a large number of pairs ($L(t)$) for transmission, many pairs can communicate but a high CSI acquisition cost is needed (i.e. this will leave a small fraction of time for transmission). On the other hand, a small $L(t)$ requires a low acquisition cost, but, at the same time, it allows a few number of simultaneous transmissions. The decision of selecting active pairs is referred simply as the *scheduling policy*. At the t-th slot, this policy can be represented by an indicator vector $\mathbf{s}(t) \in \mathcal{S} := \{0, 1\}^N$, where the k-th component of $\mathbf{s}(t)$, denoted $s_k(t)$, is equal to 1 if the k-th queue (pair) is scheduled or otherwise equal to 0. It can be seen that the cardinality of set $\mathcal{S}$ is equal to $|\mathcal{S}| = 2^N$. Remark that, in terms of notation, $\mathbf{s}(t)$ and $\mathcal{L}(t)$ are used to represent the same thing, that is, the scheduled pairs at time-slot t, but they illustrate it differently. Specifically, using $\mathbf{s}(t)$ the active pairs correspond to the non-zero coordinates (equal to 1), whereas $\mathcal{L}(t)$ contains the indexes (i.e. positions) of these pairs. Let $\mathfrak{L}$ be the set of all possible $\mathcal{L}(t)$.

Now, using the definition of $D_k(t)$, which was provided earlier, the queueing *dynamics* (i.e. how the queue lengths evolve over time) can be described by the following

$$q_k(t+1) = \max\{q_k(t) - D_k(t), 0\} + A_k(t), \qquad \forall k \in \{1, \ldots, N\}, \forall t \in \{0, 1, \ldots\}, \tag{11}$$

where we note that $D_k(t)$ depends on the scheduling policy.

In this work, the focus will be mainly on the stability of the system. Thus, in the following we provide the definitions of "Strong Stability" and "Stability region" of a system.

Definition 1 (Strong Stability)**.** *The condition for strong stability of the system can be expressed as* $\limsup\limits_{T\to\infty}\frac{1}{T}\sum\limits_{t=0}^{T-1}\mathbb{E}\left\{q_k(t)\right\}<\infty,\ \forall k\in\{1,...,N\}$.

From this definition, stability implies that the mean queue length of every queue in the system is finite, further implying finite delays in single hop systems.

Definition 2 (Stability Region)**.** *The stability region of the aforementioned model is defined as the set of vectors of mean arrival rates for which the system is strongly stable. Furthermore, a scheduling policy that achieves this region is called throughput optimal.*

When describing and characterizing stability regions, we implicitly mean that the system is strongly stable in the *interior* of the characterized region. Normally, for the boundary points the system has at least a weaker form of stability called "mean rate stability". Note that in the remainder of the manuscript "stable" will imply "strongly stable", unless stated otherwise.

Max-Weight Scheduling Policy: If the arrival rates are known, the queue stability can be achieved by pre-defined time-sharing between scheduling different subsets of queues [18]. However, in practice, the arrival rates are usually unknown. For this case, the queues can be stabilized using a policy that considers the knowledge of average rates and queue lengths. Such a policy is called Max-Weight [18], [29], i.e. it maximizes a weighted sum, and is described below.

$$\Delta^{*}:\mathbf{s}(t)=\arg\max_{\mathbf{s}\in\mathcal{S}}\left\{\mathbf{r}(\mathbf{s})\cdot\mathbf{q}(t)\right\}, \tag{12}$$

where "$\cdot$" is the scalar (dot) product, and $\mathbf{r}(\mathbf{s})$ is constructed by replacing the non-zero coordinates of $\mathbf{s}$, which represent the selected pairs, with their corresponding average rate values. Recall that $\mathcal{L}$ represents the positions (indexes) of the non-zero coordinates of $\mathbf{s}$; e.g. if $\mathbf{s}=\{0,0,1,1,1,0\}$, then $\mathcal{L}=\{3,4,5\}$, meaning that pairs 3, 4 and 5 are active. Hence, vector $\mathbf{r}(\mathbf{s})$ contains $\mathbb{E}\{D_k\mid\mathcal{L}\}$ at position k if the k-th coordinate of $\mathbf{s}$ is '1' and 0 if this coordinate is '0'. The following lemma states the optimality of the above policy, where Λ denotes the stability region.

Lemma 1. *Under the adopted system, the scheduling policy* Δ^{*} *is throughput optimal, meaning that it can stabilize the system for every mean arrival rate vector in* Λ.

Proof. We show that policy Δ^* stabilizes the system for all $\mathbf{a} \in \Lambda$ by proving that the Markov chain of the corresponding system is positive recurrent. For this purpose, we use Foster's theorem. Such a proof is standard in the literature and is thus omitted for sake of brevity. □

It is worth noting that, in the above lemma, the stability region Λ is the result of the adopted system model. More in detail, this region depends mainly, but not only, on the following points: (i) the use of IA, for $L \geq 2$, and SVD, for $L = 1$, as interference management techniques, (ii) accounting for the probing cost, (iii) the scheduling is done in a centralized manner, i.e. in each slot the CS schedules the set of active pairs (before receiving any feedback) that must send their pilots, and (iv) the quantization process over the backhaul (for the imperfect case).

Computational Complexity of Δ^*: For such an optimal policy, an important factor to investigate is the computational complexity (CC), which we derive next. Because what we are looking for is the maximum over 2^N possible values, due to 2^N combinations, thus it takes $\mathcal{O}(2^N)$ after computing all values $\mathbf{r}(\mathbf{s}) \cdot \mathbf{q}(t)$ to find the maximum value (resp., the corresponding argument). Note that for two fixed vectors we can compute this product in time $\mathcal{O}(N)$. Thus we would have $\mathcal{O}(N2^N)$ ignoring computing $\mathbf{r}(\mathbf{s})$, which can be done offline. We can notice that this computational complexity increases considerably with the maximum number of pairs N.

III. Derivation of Success Probabilities and Average Rates

In this section, we give the expression for the success probability and, subsequently, the expression for the average transmission rate for each of the imperfect and perfect cases. Then, we present the SVD scheme (used when $L(t) = 1$) and derive its average rate expression.

A. Average Rate Expression for IA

For the calculation of the average rate, we next provide a proposition in which we calculate the success probabilities under the considered setting.

Proposition 1. *The probability that the received SINR corresponding to stream m of active user k exceeds a threshold τ given that $\mathcal{L}(t)$ is the set of scheduled pairs (including pair k) can be given by*

$$\mathbb{P}\left\{\gamma_k^{(m)}(t) \geq \tau \mid \mathcal{L}(t)\right\} = \begin{cases} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}} \, MGF_{RI_k^{(m)}}\left(-\frac{\tau}{\alpha_{kk}}\right), & \textit{imperfect case} \\ e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}, & \textit{perfect case} \end{cases} \tag{13}$$

where $RI_k^{(m)} = \sum_{i \in \mathcal{L}(t), i \neq k} \alpha_{ki} \sum_{j=1}^{d_i} \left| \left(\hat{\mathbf{u}}_k^{(m)} \right)^H \mathbf{H}_{ki} \hat{\mathbf{v}}_i^{(j)} \right|^2$ *is the residual interference, which appears in the denominator of* $\gamma_k^{(m)}$ *in the imperfect case, and* $MGF_{RI_k^{(m)}}(\cdot)$ *stands for the moment-generating function (MGF) of* $RI_k^{(m)}$.

Proof. Please refer to [1] for the proof. □

In the above result, the success probability expression in the imperfect case is given in function of the MGF of the leakage interference $RI_k^{(m)}$. It is noteworthy to mention that the explicit expression of this MGF will be given afterwards during the average rate calculations. But first, let us focus on the expression $RI_k^{(m)}$. Indeed, we have

$$\begin{aligned} RI_k^{(m)} &= \sum_{i \in \mathcal{L}(t), i \neq k} \alpha_{ki} \sum_{j=1}^{d_i} \left| \left(\hat{\mathbf{u}}_k^{(m)} \right)^H \mathbf{H}_{ki} \hat{\mathbf{v}}_i^{(j)} \right|^2 = \sum_{i \in \mathcal{L}(t), i \neq k} \alpha_{ki} \sum_{j=1}^{d_i} \left| \mathbf{h}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2 \\ &= \sum_{i \in \mathcal{L}(t), i \neq k} \alpha_{ki} \left\| \mathbf{h}_{ki} \right\|^2 \sum_{j=1}^{d_i} \left| \tilde{\mathbf{h}}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2, \end{aligned} \tag{14}$$

in which $\mathbf{T}_{k,i}^{(m,j)} = \hat{\mathbf{v}}_i^{(j)} \otimes ((\hat{\mathbf{u}}_k^{(m)})^H)^T$ (where $\otimes$ is the Kronecker product) and $\tilde{\mathbf{h}}_{ki}$ is the normalized vector of channel $\mathbf{h}_{ki}$, i.e. $\tilde{\mathbf{h}}_{ki} = \frac{\mathbf{h}_{ki}}{\|\mathbf{h}_{ki}\|}$. Note that $((\hat{\mathbf{u}}_k^{(m)})^H)^T$ is nothing but the conjugate of $\hat{\mathbf{u}}_k^{(m)}$. Following the model used in [30], the channel direction $\tilde{\mathbf{h}}_{ki}$ can be written as follows

$$\tilde{\mathbf{h}}_{ki} = \sqrt{1 - e_{ki}}\, \hat{\mathbf{h}}_{ki} + \sqrt{e_{ki}}\, \mathbf{w}_{ki}, \tag{15}$$

where $\hat{\mathbf{h}}_{ki}$ is the channel quantization vector of $\mathbf{h}_{ki}$ and $\mathbf{w}_{ki}$ is a unit norm vector isotropically distributed in the null space of $\hat{\mathbf{h}}_{ki}$, with $\mathbf{w}_{ki}$ independent of e_{ki}. Since IA is performed based on the quantized CSI $\hat{\mathbf{h}}_{ki}$, we get

$$\left| \tilde{\mathbf{h}}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2 = \left| \sqrt{1 - e_{ki}}\, \hat{\mathbf{h}}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} + \sqrt{e_{ki}}\, \mathbf{w}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2 = e_{ki} \left| \mathbf{w}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2. \tag{16}$$

Therefore, $RI_k^{(m)}$ can be rewritten as

$$RI_k^{(m)} = \sum_{i \in \mathcal{L}(t), i \neq k} \alpha_{ki} \left\| \mathbf{h}_{ki} \right\|^2 e_{ki} \sum_{j=1}^{d_i} \left| \mathbf{w}_{ki}^H \mathbf{T}_{k,i}^{(m,j)} \right|^2. \tag{17}$$

Based on the above results, we now have all the required materials to derive the average rate expressions for both the perfect and imperfect cases. We recall that if $\mathcal{L}(t)$ is the subset of scheduled pairs, the general formula of the average rate of active user k can be given as

$$\mathbb{E}\left\{ D_k(t) \mid \mathcal{L}(t) \right\} = (1 - L(t)\theta) \sum_{m=1}^{d_k} R\, \mathbb{P}\left\{ \gamma_k^{(m)}(t) \geq \tau \mid \mathcal{L}(t) \right\}. \tag{18}$$

The explicit rate expressions are provided in the following theorem.

Theorem 1. *Given a subset of scheduled pairs, $\mathcal{L}(t)$, the average rate of user k ($\in \mathcal{L}(t)$) is:*

- *For the imperfect case:*

$$(1 - L(t)\theta)d_k R\, e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} MGF_{RI_k^{(m)}}\left(-\frac{\tau}{\alpha_{kk}}\right), \tag{19}$$

in which the MGF can be written as

$$MGF_{RI_k^{(m)}}\left(-\frac{\tau}{\alpha_{kk}}\right) = \prod_{i \in \mathcal{L}(t), i \neq k} \left(\frac{\alpha_{ki}\tau d_i}{\alpha_{kk} 2^{\frac{B}{Q}}} + 1\right)^{-Q} {}_2F_1(c_{2i}, Q; c_{1i} + c_{2i}; \frac{1}{\frac{\alpha_{kk} 2^{\frac{B}{Q}}}{\alpha_{ki}\tau d_i} + 1}). \tag{20}$$

In the above equation, ${}_2F_1$ represents the Hypergeometric function, $c_{1i} = (Q+1)Q^{-1}d_i - Q^{-1}$ and $c_{2i} = (Q-1)c_{1i}$, with $Q = N_t N_r - 1$.

- *For the perfect case:*

$$(1 - L(t)\theta)d_k R\, e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}}. \tag{21}$$

Proof. The proof can be found in [1]. □

B. SVD Scheme and its Average Rate Expression

1) SVD Scheme: In the case where $L(t) = 1$, there is only one active pair, which we denote by index k, and the system is reduced to a point-to-point MIMO system. We recall that receiver k sends its pilot sequence, then transmitter k estimates perfectly the channel matrix; clearly, here the probing cost is $(1-\theta)$. With one active pair, the only source of interference is the ISI caused by the transmitter itself. To manage this problem, we use SVD technique. Specifically, by the singular value decomposition theorem we have

$$\mathbf{H}_{kk} = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^H, \tag{22}$$

where $\mathbf{U}_k$ and $\mathbf{V}_k$ are $N_\mathrm{r} \times N_\mathrm{r}$ and $N_\mathrm{t} \times N_\mathrm{t}$ unitary matrices, respectively. $\mathbf{\Sigma}_k$ is an $N_\mathrm{r} \times N_\mathrm{t}$ diagonal matrix with the singular values of $\mathbf{H}_{kk}$ in diagonal. These singular values are denoted by $\sqrt{\lambda_k^{(m)}}$. Note that here we consider the same assumptions and parameters that we have used for IA. We assume that active transmitter has d_k ($\leq \min(N_\mathrm{t}, N_\mathrm{r})$) data streams to transmit to its receiver. We also assume that $\mathbf{H}_{kk}$ is full rank, meaning that its rank is given by $\min(N_\mathrm{t}, N_\mathrm{r})$; d_k should be less than or equal to the rank of matrix $\mathbf{H}_{kk}$, which is satisfied under our setting. Further, we assume that the power P is equally allocated among the d_k data streams.

We define $\mathbf{x}_k$ to be the following $N_{\text{t}} \times 1$ vector: $\mathbf{x}_k = (x_k^{(1)}, \ldots, x_k^{(d_k)}, 0 \ldots, 0)^T$, where we recall that $x_k^{(m)}$ represents stream m of pair k. Under SVD, the transmitter sends vector $\mathbf{V}_k \mathbf{x}_k$ instead of $\mathbf{x}_k$, thus the received signal, which we denote by $\mathbf{y}_k$, can be written as

$$\mathbf{y}_k = \sqrt{\frac{\zeta_{kk} P}{d_k}} \mathbf{H}_{kk} \mathbf{V}_k \mathbf{x}_k + \mathbf{z}_k, \tag{23}$$

where we recall that $\mathbf{H}_{kk}$ denotes the $N_{\text{r}} \times N_{\text{t}}$ channel matrix with i.i.d. zero mean and unit variance complex Gaussian entries, $\mathbf{z}_k$ is the additive white complex Gaussian noise vector with zero mean and covariance matrix $\sigma^2 \mathbf{I}_{N_{\text{r}}}$, and ζ_{kk} stands for the path loss. Then, at the receiver we multiply the corresponding received signal by $\mathbf{U}_k^H$ to detect the desired signal. Hence, after multiplying by $\mathbf{U}_k^H$, we obtain

$$\mathbf{U}_k^H \mathbf{y}_k = \sqrt{\frac{\zeta_{kk} P}{d_k}} \mathbf{U}_k^H \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^H \mathbf{V}_k \mathbf{x}_k + \mathbf{U}_k^H \mathbf{z}_k = \sqrt{\frac{\zeta_{kk} P}{d_k}} \mathbf{\Sigma}_k \mathbf{x}_k + \mathbf{U}_k^H \mathbf{z}_k. \tag{24}$$

Recall that $\mathbf{U}_k$ and $\mathbf{V}_k$ are unitary matrices, so $\mathbf{U}_k^H \mathbf{z}_k$ and $\mathbf{V}_k \mathbf{x}_k$ have the same distribution as $\mathbf{z}_k$ and $\mathbf{x}_k$, respectively. Based on the above, it can be noticed that the ISI is completely canceled.

2) Average Rate for SVD: The equivalent MIMO system can be seen as d_k uncoupled parallel sub-channels. The SNR for stream m at time-slot t can be written as the following

$$\gamma_k^{(m)}(t) = \frac{\zeta_{kk} P}{d_k \, \sigma^2} \lambda_k^{(m)}(t), \qquad 1 \leq m \leq d_k. \tag{25}$$

Let $m_1 = \max(N_{\text{t}}, N_{\text{r}})$ and $m_2 = \min(N_{\text{t}}, N_{\text{r}})$. It was shown in [9] that the distribution of any one of the unordered eigenvalues, which we denote by λ, is given by

$$p(\lambda) = \frac{1}{m_2} \sum_{n=0}^{m_2 - 1} \frac{n!}{(n + m_1 - m_2)!} [L_n^{m_1 - m_2}(\lambda)]^2 \lambda^{m_1 - m_2} e^{-\lambda}, \qquad \lambda \geq 0, \tag{26}$$

where $L_n^{m_1 - m_2}(x)$ is the associated Laguerre polynomial of degree (order) n and is given by

$$L_n^{m_1 - m_2}(\lambda) = \sum_{l=0}^{n} (-1)^l \frac{(n + m_1 - m_2)!}{(n - l)! \, (m_1 - m_2 + l)!} \frac{\lambda^l}{l!}. \tag{27}$$

Adopting the same rate model as for IA, the average rate of the active user can be written as

$$(1 - \theta) d_k R \, \mathbb{P} \left\{ \gamma_k^{(m)}(t) \geq \tau \right\}. \tag{28}$$

Based on the above, the average rate expression for SVD if pair k is the active pair, which we denote by $r_{\text{svd},k}$, is provided in the following statement.

Proposition 2. *Under SVD technique, the average rate for the active pair is given by*

$$r_{\text{svd},k} = (1 - \theta) d_k R \sum_{n=0}^{m_2 - 1} \Omega_n \sum_{j=0}^{2n} \kappa_j \, \Gamma(j + m_1 - m_2 + 1, \frac{d_k \sigma^2 \tau}{\zeta_{kk} P}), \tag{29}$$

where $\Gamma(\cdot,\cdot)$ *is the upper incomplete Gamma function,* $\Omega_n = n!\,(m_2(n+m_1-m_2)!\,)^{-1}$*,* $\kappa_j = \sum_{i=0}^{j} \omega_i\,\omega_{j-i}$*,* $\omega_l = (-1)^l\,(n+m_1-m_2)!\;((n-l)!\,(m_1-m_2+l)!\,l\,!\,)^{-1}$*, with* $\omega_l = 0$ *if* $l > n$*.*

Proof. Please refer to Appendix A for the proof. □

IV. Stability Analysis for the Symmetric Case

In this section, we consider a symmetric system in which the path loss coefficients have the same value, namely $\zeta = \zeta_{ki}$, $\forall k, i$, and all the pairs have equal number of data streams, namely $d = d_k$, $\forall k$; note that we still assume different average arrival rates. Under this system, the *feasibility condition* of IA, given in [31], becomes $N_{\mathrm{t}} + N_{\mathrm{r}} \geq (L+1)d$, which we assume is satisfied here. We recall that at each time-slot, for the selected pairs, rate R can be supported if the corresponding SINR is greater than or equal to a given threshold τ; otherwise, the packets are not received correctly and the instantaneous bit-rate is considered to be equal to 0. Let $\alpha = \frac{P\zeta}{d}$. For notational convenience, in the remainder of the paper we sometimes drop the notation for dependence of $\mathcal{L}$, L and $\gamma_k^{(m)}$ on t.

A. Average Rate Expressions

In the following, we provide the average rate expressions under the above assumptions for the different cases in the system.

1) If the number of scheduled pairs $L \geq 2$*:* The IA technique is applied, and thus the SINR of stream m at user k can be given by

$$\gamma_k^{(m)} = \begin{cases} \dfrac{\alpha \left| \left(\hat{\mathbf{u}}_k^{(m)}\right)^H \mathbf{H}_{kk} \hat{\mathbf{v}}_k^{(m)} \right|^2}{\sigma^2 + \sum\limits_{i \in \mathcal{L}, i \neq k} \alpha \left\|\mathbf{h}_{ki}\right\|^2 e_{ki} \sum\limits_{j=1}^{d} \left| \mathbf{w}_{ki}^H \, \mathbf{T}_{k,i}^{(m,j)} \right|^2}, & \text{imperfect case} \\[2ex] \dfrac{\alpha \left| \left(\mathbf{u}_k^{(m)}\right)^H \mathbf{H}_{kk} \mathbf{v}_k^{(m)} \right|^2}{\sigma^2}, & \text{perfect case} \end{cases} \tag{30}$$

where the sum in the denominator of the first expression results from equation (17). As explained in the previous sections, if $\mathcal{L}$ is the subset of scheduled pairs, the average transmission rate per active user is given by $(1 - L\theta)dR\,\mathbb{P}\left\{\gamma_k^{(m)} \geq \tau \mid \mathcal{L}\right\}$. Relying on the average rate expressions in Theorem 1, we get the following results.

a) Imperfect Case: The *average transmission rate* for an active user $k \in \mathcal{L}$ can be given by

$$(1 - L\theta)dR\, e^{-\frac{\sigma^2 \tau}{\alpha}} \left(\left(d\tau\, 2^{-\frac{B}{Q}} + 1 \right)^{-Q} {}_2F_1\left(c_2, Q; c_1 + c_2; (2^{\frac{B}{Q}} (d\tau)^{-1} + 1)^{-1}\right) \right)^{L-1}, \tag{31}$$

where ${}_2F_1$ is the Hypergeometric function, $c_1 = (Q+1)Q^{-1}d - Q^{-1}$ and $c_2 = (Q-1)c_1$. It can be noticed that this average rate is independent of the identity of active user k and the $L-1$ other active pairs, yet depends on the cardinality L of subset $\mathcal{L}$. By denoting this rate as $r(L)$, the expression in (31) can be re-written as

$$r(L) = (1 - L\theta)dR\, e^{-\frac{\sigma^2\tau}{\alpha}} F^{L-1}, \tag{32}$$

in which $F = \left(d\tau\, 2^{-\frac{B}{Q}} + 1\right)^{-Q} {}_2F_1\left(c_2, Q; c_1 + c_2; (2^{\frac{B}{Q}}(d\tau)^{-1} + 1)^{-1}\right)$. Consequently, the *total average transmission rate* of the system is given by

$$r_{\mathrm{T}}(L) = L(1 - L\theta)dR\, e^{-\frac{\sigma^2\tau}{\alpha}} F^{L-1}. \tag{33}$$

Studying the variation of these rate functions w.r.t. the number of active pairs L is essential for the stability analysis and is thus described by the following lemma.

Lemma 2. *Given a number of users to be scheduled, L, the average transmission rate is a decreasing function with L, whereas the total average transmission rate is increasing from 0 to L_0 and decreasing from L_0 to $\frac{1}{\theta}$, meaning that r_{T} reaches its maximum at L_0, where $L_0 < \frac{1}{2\theta}$ and is given by*

$$L_0 = \frac{\frac{1}{\theta} - \frac{2}{\log F} - \sqrt{\left(\frac{2}{\log F} - \frac{1}{\theta}\right)^2 + \frac{4}{\theta \log F}}}{2}. \tag{34}$$

Proof. The proof is provided in Appendix B. □

From (34) we can notice that L_0 is in general a real value. But, since it represents a number of users, we need to find the best and nearest integer to L_0, i.e. best in terms of maximizing the total average rate function. We denote this integer by L_{I} and we assume without lost of generality that $2 \leq L_{\mathrm{I}} \leq N$.

b) Perfect Case: In this case no residual interference exists and the corresponding SINR expression is given in (30). Using Theorem 1, the *average* and *total average transmission rate* expressions can be given, respectively, by

$$\mu(L) = (1 - L\theta)dR\, e^{-\frac{\sigma^2\tau}{\alpha}}, \tag{35}$$

$$\mu_{\mathrm{T}}(L) = L(1 - L\theta)dR\, e^{-\frac{\sigma^2\tau}{\alpha}}. \tag{36}$$

A similar observation to that given in the first case can be made here, that is, the rate functions depend only on the cardinality L of $\mathcal{L}$ and not on the subset itself. Notice that $\mu(L)$ is a decreasing function with L, while $\mu_{\mathrm{T}}(L)$ is concave at $\frac{1}{2\theta}$. Since $\frac{1}{2\theta}$ represents a number of pairs, we can use a similar procedure to that proposed for the imperfect case to find the best and nearest integer to $\frac{1}{2\theta}$. For the remainder of this paper, we denote this integer by L_{P} .

2) If the number of scheduled pairs $L = 1$: The SVD technique is applied, and thus the SINR of stream m at user k becomes

$$\gamma_k^{(m)} = \frac{\zeta P}{d\,\sigma^2}\lambda_k^{(m)}. \tag{37}$$

Based on Proposition 2, the average rate (which is also the total average rate) can be written as

$$r_{\mathrm{svd},k} = (1-\theta)dR\sum_{n=0}^{m_2-1}\Omega_n\sum_{j=0}^{2n}\kappa_j\,\Gamma(j+m_1-m_2+1,\frac{d\sigma^2\tau}{\zeta P}). \tag{38}$$

Note that this expression is independent of the identity of the active pair. For the rest of the paper, we use r_{svd} to denote this expression. Obviously, here the average rate is independent of the case (i.e. perfect or imperfect) we adopt since anyway the backhaul is not used.

B. Stability Regions and Scheduling Policies for the Imperfect and Perfect Cases

After presenting results on the average rate functions, we now provide a precise characterization of the stability region of the adopted system under both the imperfect and perfect cases.

1) Imperfect Case: We first define subset $S_L = \{\mathrm{s} \in \mathcal{S} : \|\mathrm{s}\|_1 = L\}$, where we recall that $\mathrm{s} \in \mathbb{Z}^N$ is the scheduling vector whose coordinates take values 0 or 1 (see Section II); note that S_L is the subset of scheduling decision vectors for which the number of active pairs is equal to L. Given a number $L \geq 2$, the subset of average rate vectors is defined as $I_L = \{r(L)\mathrm{s} : \mathrm{s} \in S_L\}$. For $L = 1$, the subset of average rate vectors is defined as $I_1 = \{r_{\mathrm{svd}}\mathrm{s} : \mathrm{s} \in S_1\}$. We define set $\mathcal{I} = \{\mathbf{0}, I_1, I_2, ..., I_{L_{\mathrm{I}}}\}$, i.e. it contains the origin point $\mathbf{0}$ and the set of average rate vectors when the number of active pairs L is between 1 and L_{I}. We also define set $\bar{\mathcal{I}} = \{I_{L_{\mathrm{I}}+1}, ..., I_N\}$, i.e. it contains the set of average rate vectors for which L is between $L_{\mathrm{I}} + 1$ and N. Note that in terms of cardinality we have $|\mathcal{I}| + |\bar{\mathcal{I}}| = |\mathcal{S}|$. Using these definitions, we can state the following lemma which will be useful to characterize the stability region of the system.

Lemma 3. *Each point in the set $\bar{\mathcal{I}}$ is inside the convex hull of $\mathcal{I}$. Consequently, this hull will also contain any point in the convex hull of $\bar{\mathcal{I}}$.*

Proof. We provide the following lemma that will help us in the proof of Lemma 3.

Lemma 4. *For any point $\mathbf{s}_{i,L+1} \in S_{L+1}$, there exists a point on the convex hull of S_L that is in the same direction toward the origin as $\mathbf{s}_{i,L+1}$. Furthermore, $\mathbf{s}_{i,L+1}$ can be written as $\frac{L+1}{L}\times$ its corresponding point on the convex hull of S_L.*

Please refer to Appendix C for the complete proof of Lemma 3, which includes the proof of Lemma 4. □

Lemma 3 means that by increasing the number of active pairs L beyond L_{I}, the set of achievable average rates will not increase. Based on the above, we are now able to characterize the stability region of the considered system. We recall that this region is defined as the set of all mean arrival rate vectors for which the system is stable. Here, this region is given by the following theorem.

Theorem 2. *The stability region of the system in the symmetric case with limited backhaul can be characterized as*

$$\Lambda_{\mathrm{I}} = \mathcal{CH}\left\{\mathcal{I}\right\} = \mathcal{CH}\left\{\mathbf{0}, I_1, I_2, ..., I_{L_{\mathrm{I}}}\right\}, \tag{39}$$

where $\mathcal{CH}$ represents the convex hull.

Proof. Please refer to Appendix D for the proof. □

Unlike classical results in which the stability region is given by the convex hull over all possible decisions, here the characterization is more precise and is defined by the decision subsets S_L for all $L \leq L_{\mathrm{I}}$. In addition, this theorem provides an exact specification of the corner points (i.e. vertices) of the stability region, meaning that this region is characterized by the set $\mathcal{I}$ and not by the whole set $\mathcal{I} \cup \bar{\mathcal{I}}$. An additional point to note is that Λ_{I} is a convex polytope in the N-dimensional space $\mathbb{R}_+^N$.

In order to choose the active pairs at each time-slot, we use the Max-Weight scheduling policy defined earlier (see (12)). Under the symmetric and imperfect case, this policy, which we denote as Δ_{I}^*, selects decision vector $\mathbf{s}(t)$ that yields the following max

$$\Delta_{\mathrm{I}}^* : \max\left\{\max_{\mathbf{s}\in S_L,\, 2\leq L\leq N}\left\{r(\|\mathbf{s}\|_1)\,\mathbf{s}\cdot\mathbf{q}(t)\right\}, \max_{\mathbf{s}\in S_1}\left\{r_{\mathrm{svd}}\,\mathbf{s}\cdot\mathbf{q}(t)\right\}\right\}, \tag{40}$$

where $\|\mathbf{s}\|_1$ gives the number of '1' coordinates in $\mathbf{s}$ (or equivalently, the number of active pairs L). Recall that these non-zero coordinates indicate which pairs to schedule. The above policy chooses the subset of pairs that should be active at time-slot t, which is represented by vector

$\mathbf{s}(t)$. As mentioned earlier, if only one pair is selected to be active, then we use SVD technique, otherwise we use IA technique. For the proposed policy, the following proposition holds.

Proposition 3. *The scheduling policy Δ_{I}^{*} is throughput optimal. In other words, Δ_{I}^{*} stabilizes the system (under the imperfect case) for every arrival rate vector $\mathbf{a} \in \Lambda_{\mathrm{I}}$.*

Proof. The proof can be done in the same way as the proof of Lemma 1 and is thus omitted to avoid repetition. □

Based on the analysis done at the end of Section II, it was shown that applying policy Δ^{*} will result in a computational complexity (CC) of $\mathcal{O}(N2^{N})$. The same holds here for policy Δ_{I}^{*}. Consequently, a moderately large N will lead to considerably high CC. Recall that this analysis is for the classical implementation of the Max-Weight algorithm, that is finding the maximum over 2^{N} products of two vectors. However, in our case the implementation of this algorithm does not require all this complexity. This is due to the fact that all the active users have the same average transmission rate. This structural property allows us to propose an equivalent reduced CC implementation of Δ_{I}^{*}, which we provide in Algorithm 1.

Algorithm 1 : A Reduced Computational Complexity Implementation of Δ_{I}^{*}

Sort the queues in a descending order; break ties arbitrarily.
Initialize $L_{\mathrm{s}} = 1$ and $prod = 0$.
Set $r(1) = r_{\mathrm{svd}}$. For $L \geq 2$, let $l = L$ and $r(l) = r(L)$.
for $l = 1:1:N$ **do**
  Consider sum_{l} = sum of the lengths of the first l queues (i.e. sum of the l biggest queue lengths).
  if $r(l)\, sum_{l} > prod$ **then**
    Put $L_{\mathrm{s}} = l$ and $prod = r(L_{\mathrm{s}}) sum_{L_{\mathrm{s}}}$
  end if
end for
Schedule pairs corresponding to the first L_{s} queues.

The proposed implementation depends essentially on two steps: the "sorting algorithm" and the "for loop". We assume that we use a sorting algorithm of complexity $\mathcal{O}(N^{2})$, such as the "bubble sort" algorithm. For the "for loop", the (worst case) complexity is also $\mathcal{O}(N^{2})$ since this loop is executed N times (i.e. iterations) and every iteration has another dependency to N. Therefore, the computational complexity of the proposed implementation is $\mathcal{O}(N^{2} + N^{2})$, or equivalently $\mathcal{O}(N^{2})$, which is very small compared to $\mathcal{O}(N2^{N})$, especially for large N.

2) Perfect Case: A similar study to that done for the imperfect case can be adopted here. To begin with, for $L \geq 2$ we define $P_L = \{\mu(L)\mathbf{s} : \mathbf{s} \in S_L\}$; i.e. P_L is the subset of average rate vectors where the number of active pairs is L. Recall that $S_L = \{\mathbf{s} \in \mathcal{S} : \|\mathbf{s}\|_1 = L\}$. For $L = 1$ we define $P_1 = \{r_{\text{svd}}\mathbf{s} : \mathbf{s} \in S_1\}$. As seen earlier, the total average rate for IA given in (36) reaches its maximum at $\frac{1}{2\theta}$ for which the best and nearest integer is denoted by L_{P}, where we assume without lost of generality that $2 \leq L_{\text{P}} \leq N$. In addition, the average rate $\mu(L)$ decreases with L. Under these observations, the stability region can be characterized as follows.

Theorem 3. *For the symmetric system with unlimited backhaul, the stability region can be defined as the following*

$$\Lambda_{\text{P}} = \mathcal{CH}\left\{\mathbf{0}, P_1, P_2, ..., P_{L_{\text{P}}}\right\}. \tag{41}$$

Proof. The proof is very similar to that of Theorem 2, just consider the average rate functions $\mu(L)$ and $\mu_{\text{T}}(L)$ instead of $r(L)$ and $r_{\text{T}}(L)$; so we omit this proof to avoid redundancy. □

To achieve the stability region that is characterized in the above, we use the Max-Weight policy, which we denote here as Δ_{P}^*. Specifically, this *optimal* policy (i.e. it can achieve Λ_{P}) selects decision vector $\mathbf{s}(t)$ that yields the following max

$$\Delta_{\text{P}}^* : \max\left\{\max_{\mathbf{s}\in S_L,\, 2\leq L\leq N}\left\{\mu(\|\mathbf{s}\|_1)\,\mathbf{s}\cdot\mathbf{q}(t)\right\}, \max_{\mathbf{s}\in S_1}\left\{r_{\text{svd}}\,\mathbf{s}\cdot\mathbf{q}(t)\right\}\right\}. \tag{42}$$

As observed in the imperfect case, applying this policy using its classical implementation will result in a CC of $\mathcal{O}(N2^N)$. Hence, to avoid a high complexity for large N, and since the structural properties of this policy and those of policy Δ_{I}^* are similar, the equivalent implementation proposed for the imperfect case can be applied here but after replacing $r(L)$ with $\mu(L)$. Consequently, we get a reduced complexity of $\mathcal{O}(N^2)$.

C. Conditions under which IA Provides a Gain in terms of Queueing Stability

In this subsection, we investigate the conditions under which the use of IA can provide a queueing stability gain. We provide this investigation under the imperfect case, while noting that a similar analysis can be done under the perfect case.

Based on Theorem 2, we can notice that in the characterization of the stability region the vertices that correspond to IA are given by the subsets $I_2, I_3, ..., I_{L_{\text{I}}}$. On the other hand, the vertices that correspond to SVD can be found in I_1. In order for IA to provide a queueing stability gain, at least one of its vertices should be outside the part of the stability region that is yielded by SVD, where this part is nothing but the convex hull $\mathcal{CH}\{\mathbf{0}, I_1\}$. In the following we provide a simple example for which we illustrate the different shapes of the stability region.

Example: Let $N = 2$. Here there is only one vertex of IA, given by point $(r(2), r(2))$. The vertices of SVD are $(r_{\text{svd}}, 0)$ and $(0, r_{\text{svd}})$. Hence, the part of the stability region resulting from SVD is $\mathcal{CH}\{\mathbf{0}, I_1\} = \mathcal{CH}\{\mathbf{0}, (r_{\text{svd}}, 0), (0, r_{\text{svd}})\}$. Depending on whether vertex $(r(2), r(2))$ is inside this convex hull, two stability region shapes are possible. These shapes are illustrated in Figure 3. We point out that the illustrations in Figure 3 are not the results of a specific setting and are only provided as general illustrations in order to help understanding the analysis.

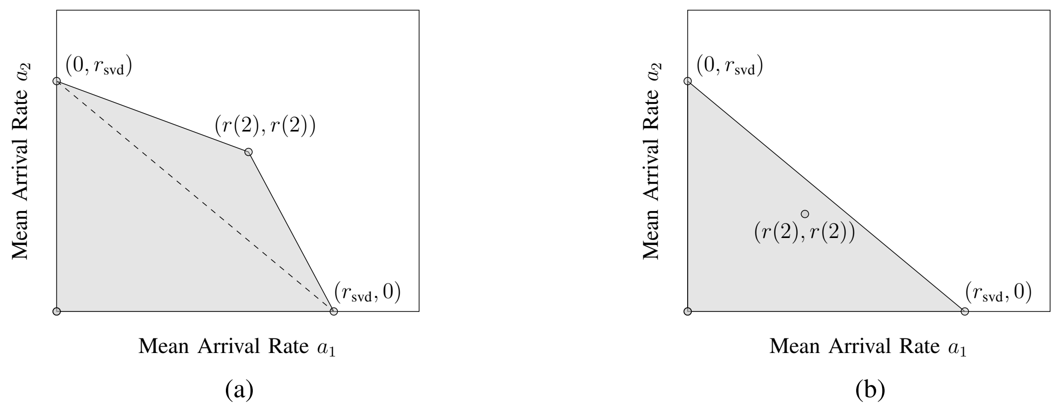

Figure 3: Stability region (in gray) for the symmetric system under the imperfect case, where $N = 2$. (a) IA provides a queueing stability gain and (b) IA does not provide a queueing stability gain.

Based on these observations, we now provide the conditions under which IA yields a queueing stability gain. These conditions are given (for the perfect and imperfect cases) as follows.

Proposition 4. *For the symmetric system under the imperfect case, IA provides a queueing stability gain iff there exists a number L such that $Lr(L) > r_{\text{svd}}$, with $2 \leq L \leq L_{\text{I}}$. For the same system but under the perfect case, we get a similar result, that is, IA yields a queueing stability gain iff there exists a number L such that $L\mu(L) > r_{\text{svd}}$, with $2 \leq L \leq L_{\text{P}}$.*

Proof. The proof is provided in Appendix E. □

Based on the above, it can be noticed that if there exists an L that satisfies $Lr(L) > r_{\text{svd}}$ and $2 \leq L \leq L_{\text{I}}$, the characterization of the stability region Λ_{I} can be more precise since the points in $I_2, \ldots, I_{L-1}$ are inside the stability region part resulting from SVD, and the only vertices of IA that are outside this part are given by the subsets $I_L, \ldots, I_{L_{\text{I}}}$. In addition, if we cannot find an L that satisfies the above conditions, then all the vertices of IA will be inside the region part corresponding to SVD. These observations lead us to the following remark; recall that $P_1 = I_1$.

Remark 1. *For the imperfect case, if there exists an L such that $Lr(L) > r_{\text{svd}}$, with $2 \leq L \leq L_{\text{I}}$, a more precise characterization of Λ_{I} can be given as*

$$\Lambda_{\text{I}} = \mathcal{CH}\left\{\mathbf{0}, I_1, I_L, ..., I_{L_{\text{I}}}\right\}. \tag{43}$$

If such an L does not exist, then the characterization reduces to: $\Lambda_{\text{I}} = \mathcal{CH}\left\{\mathbf{0}, I_1\right\}$.
For the perfect case, if we can find an L such that $L\mu(L) > r_{\text{svd}}$, with $2 \leq L \leq L_{\text{P}}$, we get

$$\Lambda_{\text{P}} = \mathcal{CH}\left\{\mathbf{0}, P_1, P_L, ..., P_{L_{\text{P}}}\right\}. \tag{44}$$

Otherwise, the characterization of Λ_{P} reduces to $\Lambda_{\text{P}} = \mathcal{CH}\left\{\mathbf{0}, P_1\right\}$.

D. Compare the Imperfect and Perfect Cases in terms of Stability

After having characterized the stability region for both the perfect and imperfect cases, we now investigate the maximum *gap* between these two regions. Clearly, here the investigation is reserved for the scenario where *IA provides queueing stability gains* in the perfect case; because otherwise IA will also not provide queueing stability gains in the imperfect case, and thus the stability region for the imperfect case will be the same as that for the perfect case, and can be given by $\mathcal{CH}\left\{\mathbf{0}, I_1\right\} = \mathcal{CH}\left\{\mathbf{0}, P_1\right\}$, which is the stability region of SVD. The gap can be interpreted as the impact of having limited backhaul, and thus quantization, on the stability region. It is straightforward to notice that the quantization process will result in shrinking the stability region compared with that of the perfect case. To capture this shrinkage, we find that the imperfect case achieves a (guaranteed) fraction of the stability region achieved in the perfect case, which we refer to as *minimum fraction* in the sequel; i.e. the term "minimum fraction" is justified by the fact that the stability region in the imperfect case achieves at least this fraction of the stability region in the perfect case. To be more precise, if we denote this fraction by β (≤ 1), then under the imperfect case the queues are stable for any mean arrival rate lying inside $\beta\Lambda_{\text{P}}$ and it may be possible to achieve stability for some mean arrival rates outside $\beta\Lambda_{\text{P}}$. We highlight that this fraction represents the maximum gap between the two regions.

To begin with, we first draw the attention to the fact that in addition to having $\mu(L) > r(L)$, we have $L_{\text{P}} \geq L_{\text{I}}$. In order to provide some insights into how we will derive the minimum achievable fraction, in the following we give a simple example for which the stability region shapes are illustrated.

Example: Let $N = 2$. Depending on whether IA provides queueing stability gains under the imperfect case, two scenarios are to consider. In Figure 4 we depict the general shapes of the two (i.e. perfect and imperfect cases) stability regions under these two scenarios; note that these depicted regions are not the result of a specific setting and are only given to help understanding the analysis. From this figure, we can observe that we have different gaps over different directions.

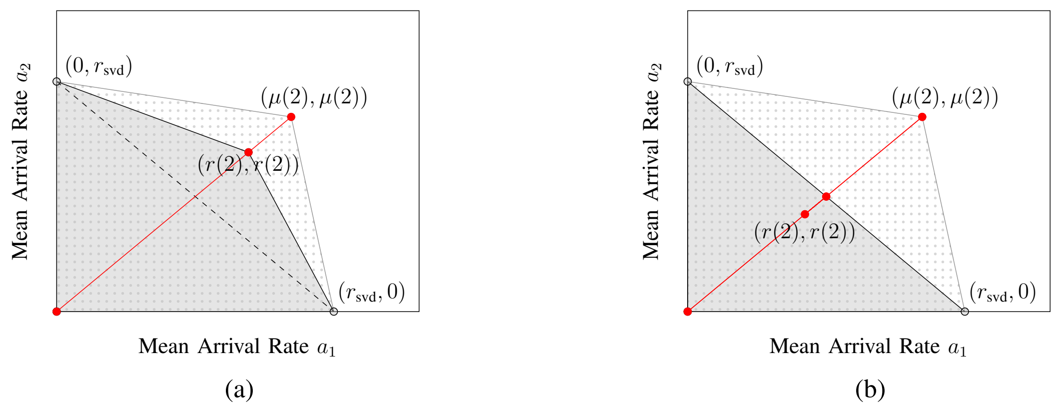

Figure 4: Stability regions of the perfect (dotted region) and imperfect (gray region) cases for the symmetric system, where $N = 2$. (a) IA provides a queueing stability gain and (b) IA does not provide a queueing stability gain.

To find the minimum fraction (i.e. maximum gap), we adopt the following approach. We take any point from subset $P_{L_{\mathrm{P}}}$, and then we try to see how far is this point from the convex hull Λ_{I} in the direction toward the origin. We adopt such an approach mainly for three reasons: (i) $I_{L_{\mathrm{I}}}$ is a subset of the vertices that characterize the convex hull of the imperfect case (Λ_{I}), (ii) $P_{L_{\mathrm{P}}}$ is the subset that contains points (vertices) on the convex hull of the perfect case, and (iii) the points in $P_{L_{\mathrm{P}}}$ are the farthest from Λ_{I}. Using this approach, we can state the following theorem that characterizes the minimum achievable fraction.

Theorem 4. *For the symmetric system, in general the stability region in the imperfect case achieves at least a fraction* $\frac{r_{\mathrm{T}}(L_{\mathrm{I}})}{\mu_{\mathrm{T}}(L_{\mathrm{P}})} = \frac{L_{\mathrm{I}} r(L_{\mathrm{I}})}{L_{\mathrm{P}} \mu(L_{\mathrm{P}})}$ (< 1) *of the stability region achieved in the perfect case. In other words, the region* Λ_{I} *can be bounded as*

$$\frac{r_{\mathrm{T}}(L_{\mathrm{I}})}{\mu_{\mathrm{T}}(L_{\mathrm{P}})} \Lambda_{\mathrm{P}} \subseteq \Lambda_{\mathrm{I}} \subseteq \Lambda_{\mathrm{P}}. \tag{45}$$

In the special case where IA delivers no gain under the imperfect case, the above fraction becomes $\frac{r_{\text{svd}}}{\mu_{\mathrm{T}}(L_{\mathrm{P}})} = \frac{r_{\text{svd}}}{L_{\mathrm{P}} \mu(L_{\mathrm{P}})}$.

Proof. We provide the following result that will help us in the proof of this theorem.

Lemma 5. *Each point in S_{L+1} can be written as $\frac{L+1}{L+1-n}\times$ some point on the convex hull of S_{L+1-n}, for $1 \leq n \leq L$.*

Please refer to Appendix F for the complete proof of Theorem 4, which includes the proof of Lemma 5. □

It should be noted that the fraction given in the above theorem represents the maximum impact of limited backhaul on the stability region.

E. Impact of the Number of Bits B on the System Stability Region

Here the analysis is restricted for the imperfect case of IA, where the backhaul is of finite capacity. We recall that under the adopted system the number of bits is B. Since the stability analysis depends essentially on this parameter, it is important to study the impact of changing this parameter on the stability performance of the system. To this end, we next investigate the impact of reducing B to B'. In this investigation, we consider the scenario where IA always yields a queueing stability gain, as this is the most likely scenario.

To begin with, let Δ_B^* and $\Delta_{B'}^*$ denote the same algorithm, that is, the Max-Weight policy, but the first one considers the case where the number of bits is equal to B and for the second one this number is B' (with $B' < B$). Further, let $\mathcal{L}_B$ and $\mathcal{L}_{B'}$ denote the subsets of pairs selected by Δ_B^* and $\Delta_{B'}^*$, respectively. Also, we denote by Λ_B and $\Lambda_{B'}$ the stability regions achieved by Δ_B^* and $\Delta_{B'}^*$, respectively. In addition, we define $r(L, B)$ as the average rate $r(L)$ with a number of bits B. Equivalently, $r(L, B')$ is the average rate function $r(L)$ in which we replace B by B'. For this model, we can state the following theorem.

Theorem 5. *For the same (symmetric) system, where the maximum number of pairs is N, if we decrease the number of bits from B to B', the stability region $\Lambda_{B'}$ can be bounded as*

$$\frac{r(N, B')}{r(N, B)}\Lambda_B \subseteq \Lambda_{B'} \subseteq \Lambda_B. \tag{46}$$

Proof. Please refer to Appendix G for the proof. □

V. Algorithmic Design and Performance Analysis for the General Case

We now consider a more general model where, unlike the symmetric case, the path loss coefficients are not necessarily equal to each other. However, for the sake of simplicity, and without loss of generality, we keep the same assumption on the number of streams, that is, all the pairs have equal number of data streams, namely d. Also, as in the symmetric case, we suppose at each time-slot that rate R can be supported if the corresponding SINR is greater than or equal to τ; otherwise, the packets are not received correctly and the instantaneous bit-rate is considered to be equal to 0. We recall that $\mathcal{L}$ stands for the subset of active pairs, with $|\mathcal{L}| = L$. We also recall that $\alpha_{ki} = \frac{P\zeta_{ki}}{d}$ and $\alpha_{kk} = \frac{P\zeta_{kk}}{d}$.

A. Average Rate Expressions

1) If the number of scheduled pairs $L \geq 2$: The IA technique is performed, and thus the SINR of stream m at active user k can be given by (see (7))

$$\gamma_k^{(m)} = \begin{cases} \dfrac{\alpha_{kk}\left|\left(\hat{\mathbf{u}}_k^{(m)}\right)^H \mathbf{H}_{kk}\hat{\mathbf{v}}_k^{(m)}\right|^2}{\sigma^2 + \sum\limits_{i\in\mathcal{L}, i\neq k} \alpha_{ki}\left\|\mathbf{h}_{ki}\right\|^2 e_{ki} \sum\limits_{j=1}^{d}\left|\mathbf{w}_{ki}^H\,\mathbf{T}_{k,i}^{(m,j)}\right|^2}, & \text{imperfect case} \\ \dfrac{\alpha_{kk}\left|\left(\mathbf{u}_k^{(m)}\right)^H \mathbf{H}_{kk}\mathbf{v}_k^{(m)}\right|^2}{\sigma^2}, & \text{perfect case} \end{cases} \tag{47}$$

a) Imperfect Case: Using Theorem 1 and the fact that $\frac{\alpha_{ki}}{\alpha_{kk}} = \frac{\zeta_{ki}}{\zeta_{kk}}$, the average rate of user k ($\in \mathcal{L}$) can be written as $\quad r_k =$

$$(1-L\theta)dR\,e^{-\frac{\sigma^2\tau}{\alpha_{kk}}} \prod_{i\in\mathcal{L}, i\neq k}\left(\zeta_{ki}\tau d\,(\zeta_{kk}2^{\frac{B}{Q}})^{-1}+1\right)^{-Q} {}_2F_1(c_2, Q; c_1+c_2; (\zeta_{kk}2^{\frac{B}{Q}}(\zeta_{ki}\tau d\,)^{-1}+1)^{-1}), \tag{48}$$

where we recall that $c_1 = (Q+1)Q^{-1}d - Q^{-1}$ and $c_2 = (Q-1)c_1$.

b) Perfect Case: Based on Theorem 1, the average rate of active user k can be expressed as

$$\mu_k = (1-L\theta)dR\,e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}. \tag{49}$$

2) If the number of scheduled pairs $L = 1$: The SVD technique is applied, and the SINR of stream m of the active user, which is denoted by index k, can be given by (see (25))

$$\gamma_k^{(m)} = \frac{\zeta_{kk}P}{d\,\sigma^2}\lambda_k^{(m)}. \tag{50}$$

Using (29), the average rate for the active user (i.e. user k) is

$$r_{\text{svd},k} = (1-\theta)dR\sum_{n=0}^{m_2-1}\Omega_n\sum_{j=0}^{2n}\kappa_j\,\Gamma(j+m_1-m_2+1, \frac{d\sigma^2\tau}{\zeta_{kk}P}). \tag{51}$$

B. Stability Regions and Scheduling Policies for the Imperfect and Perfect cases

1) Imperfect case: Let $\mathbf{r}$ be a vector that contains at position k the average rate of user k if pair k is active and 0 otherwise; for ex, let $N = 2$: if only pair 1 is active then $\mathbf{r} = (r_{\text{svd},1}, 0)$, whereas if both of pairs 1 and 2 are active, then $\mathbf{r} = (r_1, r_2)$. As mentioned earlier (see Section II, for instance), $\mathbf{s}$ and $\mathcal{L}$ are two different representations for the (same) set of active pairs, so we will use $\mathbf{r}(\mathbf{s})$ to represent the fact that $\mathbf{r}$ results from decision vector $\mathbf{s}$. Notice that, in contrast to the symmetric case, here the average rate expression depends on the identity of the active pairs; consider the same example as before: if $\mathbf{s} = (1, 1)$, i.e. pairs 1 and 2 are scheduled, then $\mathbf{r}(\mathbf{s}) = (r_1, r_2)$ where r_1 is different from r_2 (in general). This lack of symmetry will make us incapable of finding the exact set of vertices of the corresponding stability region. However, we can still provide a characterization of this stability region, denoted Λ_{GI}, by considering all the possible decisions of scheduling the pairs, as follows

$$\Lambda_{\text{GI}} = \mathcal{CH}\left\{\mathbf{0}, GI_1, GI_2, ..., GI_N\right\}, \tag{52}$$

where $GI_L = \{\mathbf{r}(\mathbf{s}) : \mathbf{s} \in S_L\}$, i.e. GI_L is the set of average rate vectors when the number of active pairs is L. To achieve this stability region we can apply the Max-Weight rule, which is an optimal scheduling policy, denoted by Δ^*_{GI}, such as

$$\Delta^*_{\text{GI}} : \mathbf{s}(t) = \arg\max_{\mathbf{s} \in \mathcal{S}} \left\{\mathbf{r}(\mathbf{s}) \cdot \mathbf{q}(t)\right\}. \tag{53}$$

where we recall that $\mathcal{S}$ is the set of all possible decision vectors $\mathbf{s}$.

2) Perfect Case: For this case, we denote by $\boldsymbol{\mu}$ the average rate vector that contains at position k the average rate of pair k if pair k is scheduled and 0 otherwise. Also, let $\boldsymbol{\mu}(\mathbf{s})$ be the rate vector under decision vector $\mathbf{s}$; as a simple example, let $N = 2$: if $\mathbf{s} = (0, 1)$, then $\boldsymbol{\mu}(\mathbf{s}) = (0, r_{\text{svd},2})$, whereas if $\mathbf{s} = (1, 1)$, then $\boldsymbol{\mu}(\mathbf{s}) = (\mu_1, \mu_2)$. Let Λ_{GP} be the system stability region under this case. This region can be represented as

$$\Lambda_{\text{GP}} = \mathcal{CH}\left\{\mathbf{0}, GP_1, GP_2, ..., GP_N\right\}, \tag{54}$$

where $GP_L = \{\boldsymbol{\mu}(\mathbf{s}) : \mathbf{s} \in S_L\}$. The (optimal) policy that schedules the pairs and achieves this above region is denoted by Δ^*_{GP} and can be given by

$$\Delta^*_{\text{GP}} : \mathbf{s}(t) = \arg\max_{\mathbf{s} \in \mathcal{S}} \left\{\boldsymbol{\mu}(\mathbf{s}) \cdot \mathbf{q}(t)\right\}. \tag{55}$$

C. β_{A}-Approximate Policy and its Corresponding Achievable Stability Region

As detailed earlier, the classical implementation of the Max-Weight policy, in both the perfect and imperfect cases, has a computational complexity of $\mathcal{O}(N2^N)$. Whereas for the symmetric case some structural properties allowed us to find a low computational complexity implementation of this policy, here no such properties exist. To deal with this problem, we propose an alternative policy that has a reduced computational complexity, so that we can apply it instead of the optimal policy, and that achieves an important fraction of the system stability region. Specifically, here we are interested in finding this alternative policy under the imperfect case, which can be considered as the hardest case to analyze since the average rate expressions are very complicated compared with those under the perfect case. The alternative policy in this case is denoted by Δ^{A} and termed as β_{A}-*approximate policy*, where this latter expression is justified by the fact that this policy approximates Δ^*_{GI} to a fraction of β_{A} (≤ 1). More specifically, for every queue length vector $\mathbf{q}$, the following holds ([19])

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta^{\text{A}})} \geq \beta_{\text{A}} (\mathbf{r} \cdot \mathbf{q})^{(\Delta^*_{\text{GI}})}, \tag{56}$$

where we recall that $\mathbf{q}$ is the queue lengths vector, and where $(\mathbf{r} \cdot \mathbf{q})^{(\Delta^{\text{A}})}$ (resp., $(\mathbf{r} \cdot \mathbf{q})^{(\Delta^*_{\text{GI}})}$) implies that $\mathbf{r}$ results from the scheduling decisions of policy Δ^{A} (resp., Δ^*_{GI}).

For the rest of the paper, for notational conciseness, we will use the term "approximate policy" instead of "β_{A}-approximate policy" unless stated otherwise. A key step in the investigation is to determine a specific approximation of the average rate expression r_k, more specifically an approximation that possesses a set of structural features that let us define the approximate policy. Indeed, we will derive such an approximation and prove that it is very accurate if the fractions $\frac{\zeta_{kk} 2^{\frac{B}{Q}}}{\zeta_{ki} \tau d}$ (or equivalently, $\frac{\alpha_{kk} 2^{\frac{B}{Q}}}{\alpha_{ki} \tau d}$) are sufficiently high (i.e. ≥ 10), for $i \neq k$. For fixed τ and d, these conditions correspond to a scenario where the number of quantization bits is high and/or the cross channels have small path loss coefficients in comparison with the direct channels (i.e. low interference scenario). It should be noted that here the approximation is only provided for the average rates r_k, meaning that rate $r_{\text{svd},k}$ is not approximated; recall that if $L \geq 2$, r_k stands for the average rate for active user k, whereas if $L = 1$, i.e. one active pair, $r_{\text{svd},k}$ is the corresponding average rate. Under the aforementioned assumptions/conditions, the approximation of r_k is given by the following proposition.

Rate Approximation 1. *Under the general system and the imperfect case, and given a subset of active pairs, $\mathcal{L}$, with cardinality $L \geq 2$, if we have a relatively low interference scenario, the average rate of active user k ($\in \mathcal{L}$) can be accurately approximated as*

$$r_k \approx (1 - L\theta)\, dR\, e^{-\frac{\sigma^2\tau}{\alpha_{kk}}} \prod_{i\in\mathcal{L},\, i\neq k} (1 - g_{ki}), \tag{57}$$

where $g_{ki} = \left(\alpha_{kk} 2^{\frac{B}{Q}} (\alpha_{ki}\tau d)^{-1} + 1\right)^{-1} = \left(\zeta_{kk} 2^{\frac{B}{Q}} (\zeta_{ki}\tau d)^{-1} + 1\right)^{-1}$.

Proof. The derivation is provided in Appendix H. □

It can be noted that the low-interference condition, i.e. $\zeta_{kk} 2^{\frac{B}{Q}} (\zeta_{ki}\tau d)^{-1}$ is sufficiently high, implies that all the g_{ki}, $\forall i \neq k$, are sufficiently small. To proceed further with the analysis, let $\bar{g}_k$ be the average value of all the g_{ki}, $\forall i \neq k$, for the same number of active pairs L. More specifically, for a fixed cardinality L we take all the possible subsets (i.e. scheduling decisions) in which user k is active. For each of these subsets, there are $L-1$ values of g_{ki}. Hence, $\bar{g}_k$ is the average of these g_{ki} values over all the considered decisions. Using the average value $\bar{g}_k$ and the approximate expression of r_k (given in (57)), we define function $\phi_k(L)$, as the following

$$\phi_k(L) = \begin{cases} (1 - L\theta) dR\, e^{-\frac{\sigma^2\tau}{\alpha_{kk}}} (1 - \bar{g}_k)^{L-1}, & \text{if } L \geq 2 \\ r_{\text{svd},k}, & \text{if } L = 1 \end{cases} \tag{58}$$

where the expression of $r_{\text{svd},k}$ is given in (51). Le $\boldsymbol{\phi}$ be the vector containing $\phi_k(L)$ at position k if pair k is scheduled (with $L-1$ other pairs); otherwise, we put 0 at this position. Under this setting, we define the approximate policy Δ^{A} as

$$\Delta^{\text{A}} : \mathbf{s}(t) = \arg\max_{\mathbf{s}\in\mathcal{S}} \left\{ \boldsymbol{\phi}(\mathbf{s}) \cdot \mathbf{q}(t) \right\},$$

where $\boldsymbol{\phi}(\mathbf{s})$ results from decision vector $\mathbf{s}$; as a simple example, set $N = 2$: if $\mathbf{s} = (1, 0)$, then $\boldsymbol{\phi}(\mathbf{s}) = (r_{\text{svd},1}, 0)$, whereas if $\mathbf{s} = (1, 1)$, then $\boldsymbol{\phi}(\mathbf{s}) = (\phi_1(2), \phi_2(2))$. It is noteworthy to mention that, for $L \geq 2$, although we use $\phi_k(L)$ to make the scheduling decision under Δ^{A}, the *actual* average rate of user k is still r_k. Also, remark that Δ^{A} follows the Max-Weight rule, thus, as was shown earlier, implementing Δ^{A} as a classical maximization problem over all the possible decisions s needs a CC of $\mathcal{O}(N2^N)$. However, in contrast to Δ^*_{GI}, policy Δ^{A} has a structural property that will allow us to propose an equivalent reduced CC implementation instead of the classical one. This property results from to the fact that $\phi_k(L)$ is independent of the $L-1$ other active users, and only depends on pair k and the cardinality L. The proposed implementation of policy Δ^{A} is given by Algorithm 2.

Algorithm 2 : A Reduced Computational Complexity Implementation of Δ^{A}

1: Initialize $L_{\mathrm{g}} = 0$ and $ws_{L_{\mathrm{g}}} = 0$.
2: **for** $l = 1 : 1 : N$ **do**
3: Sort the pairs in a descending order with respect to the product $pro_k = \phi_k(l)\, q_k$; break ties arbitrarily.
4: Let ws_l = sum of the first l biggest pro_k values.
5: Save sq_l as the subset of l pairs that yields ws_l.
6: **if** $ws_l > ws_{L_{\mathrm{g}}}$ **then**
7: Put $ws_{L_{\mathrm{g}}} = ws_l$ and $L_{\mathrm{g}} = l$.
8: **end if**
9: **end for**
10: Schedule the pairs given by $sq_{L_{\mathrm{g}}}$.

Computational Complexity of the proposed implementation of Δ^{A}: To compare with the classical implementation, we now focus on the computational complexity of the proposed implementation, which depends essentially on a "for loop" of N iterations, each of which contains: (i) a "bubble sorting algorithm", which needs $\mathcal{O}(N^2)$, (ii) a sum of l terms in iteration l, and (iii) other steps of small CC compared with those mentioned before. Thus, by neglecting the CC of the steps in (iii) and noticing that the summing steps (in (ii)) over all the iterations need $\mathcal{O}(\frac{N(N+1)}{2}) = \mathcal{O}(N^2)$, the CC of the proposed implementation is roughly $\mathcal{O}(N^2 N + N^2) = \mathcal{O}(N^3)$, which is very small compared with $\mathcal{O}(N2^N)$ for large N.

In general, an approximate policy comes with the disadvantage of reducing the achievable stability region compared with the optimal policy. Indeed, we will show that policy Δ^{A} guarantees to achieve only a fraction of the stability region achieved by policy Δ^*_{GI}. Let us now provide some remarks an definitions that will be useful for the analysis. We recall that $g_{ki} = \left(\zeta_{kk} 2^{\frac{B}{Q}} (\zeta_{ki}\tau d)^{-1} + 1\right)^{-1}$. In addition, we define $\mathcal{L}_{\mathrm{A}}$ as the subset of pairs chosen by Δ^{A}, and we let the cardinality of this subset be L_{A}. As for Δ^*_{GI} we keep the original notation, that is, $\mathcal{L}$ is the scheduled subset, with $L = |\mathcal{L}|$. Recall that $\mathfrak{L}$ stands for the set of all possible decision subsets, so $\mathcal{L}_{\mathrm{A}}$ and $\mathcal{L}$ are elements of $\mathfrak{L}$. In the following we provide a result that is essential for the characterization of the fraction that Δ^{A} can achieve.

Rate Approximation 2. *If the values of g_{ki}, $\forall i \neq k$, are relatively close to $\bar{g}_k$, the approximation* *of r_k in (57) can in its turn be accurately approximated as*

$$\begin{aligned} r_k &\approx (1 - L\theta)\, dR\, e^{-\frac{\sigma^2\tau}{\alpha_{kk}}} \prod_{i\in\mathcal{L},\, i\neq k} (1 - g_{ki}) \\ &\approx (1 - L\theta)\, dR\, e^{-\frac{\sigma^2\tau}{\alpha_{kk}}} \left[(1 - \bar{g}_k)^{L-1} - (1 - \bar{g}_k)^{L-2} \sum_{i\in\mathcal{L},\, i\neq k} (g_{ki} - \bar{g}_k) \right]. \end{aligned} \tag{59}$$

Proof. The derivation is provided in Appendix I. □

Let us define Λ_{GI} as the stability region of the general system under the imperfect case, which can be achieved by Δ^*_{GI}. We also define Λ_{A} to be the stability region that the approximate policy Δ^{A} can achieve. Concerning the fraction that policy Δ^{A} can achieve w.r.t. the stability region Λ_{GI}, and under all the above-mentioned conditions, we have the following result.

Theorem 6. *The approximate policy Δ^{A} achieves at least a fraction β_{A} (≤ 1) of the stability region achieved by the optimal policy Δ^*_{GI}, meaning that Λ_{A} can be bounded as*

$$\beta_{\mathrm{A}}\Lambda_{\mathrm{GI}} \subseteq \Lambda_{\mathrm{A}} \subseteq \Lambda_{\mathrm{GI}}, \tag{60}$$

where β_{A} is given by

$$\beta_{\mathrm{A}} = \frac{1 + \min\limits_{\mathcal{L}_{\mathrm{A}}\in\boldsymbol{\mathfrak{L}}} \left\{ \min\limits_{k\in\mathcal{L}_{\mathrm{A}}} \left\{ -(1-\bar{g}_k)^{-1} \sum\limits_{i\in\mathcal{L}_{\mathrm{A}}, i\neq k} (g_{ki} - \bar{g}_k) \right\} \right\}}{1 + \max\limits_{\mathcal{L}\in\boldsymbol{\mathfrak{L}}} \left\{ \max\limits_{k\in\mathcal{L}} \left\{ -(1-\bar{g}_k)^{-1} \sum\limits_{i\in\mathcal{L}, i\neq k} (g_{ki} - \bar{g}_k) \right\} \right\}}. \tag{61}$$

Proof. Please refer to Appendix J for the proof. □

We point out that in the above theorem we suppose, without lost of generality, that the $\min$ term in the numerator takes values in the interval $]-1, 0[$, whereas the $\max$ term in the denominator is greater than 0. More details about these assumptions can be found in the proof.

D. Compare the Imperfect Case with the Perfect Case in Terms of Stability

At the very beginning of this section, we showed that policy Δ^*_{GP} achieves the (general) system stability region under the perfect case. Let us denote by $\mathcal{L}_{\mathrm{P}}$ the subset of scheduled pairs using Δ^*_{GP} and by L_{P} the cardinality of this subset. Also, we define Λ_{GP} to be the stability region of the perfect case. On the other hand, for the imperfect case we adopt the same notation as before, i.e. the subset of scheduled users and its cardinality are represented by $\mathcal{L}$ and L, respectively, and the stability region is denoted by Λ_{GI}. Under our system, an essential parameter to investigate is the fraction the stability region the imperfect case achieves compared with the stability region of the perfect case. This fraction is captured in the following theorem.

Theorem 7. *The stability region of the imperfect case reaches at least a fraction β_{P} of the stability region achieved in the perfect case, meaning that Λ_{GI} can be bounded as*

$$\beta_{\text{P}}\Lambda_{\text{GP}} \subseteq \Lambda_{\text{GI}} \subseteq \Lambda_{\text{GP}}, \tag{62}$$

in which β_{P} is given as

$$\beta_{\text{P}} = \min_{\mathcal{L}_{\text{P}} \in \mathfrak{L}} \left\{ \min_{k \in \mathcal{L}_{\text{P}}} \left\{ \prod_{i \in \mathcal{L}_{\text{P}}, i \neq k} (1 - g_{ki}) \right\} \right\}, \tag{63}$$

where $g_{ki} = \left(\zeta_{kk} 2^{\frac{B}{Q}} (\zeta_{ki} \tau d)^{-1} + 1 \right)^{-1}$.

Proof. Please refer to Appendix K for the proof. □

We draw the attention to the fact that the proof of the above result relies on the (first) approximation of r_k given in Rate Approximation 1, meaning that the above theorem holds under the low-interference condition (i.e. all the g_{ki}, with $i \neq k$, are sufficiently small).

Relation between β_{P} and B: An important factor on which fraction β_{P} depends is the number of quantization bits B, so it is essential to compute the number of bits that can guarantee this fraction. Finding the explicit relation that gives the number of bits in function of β_{P} is a difficult task, however we can obtain the required result numerically. In detail, using the expression of β_{P} given in the above theorem, we start from a small value of B for which we calculate the corresponding fraction, then we keep increasing B until the desired value of β_{P} is obtained. Although computing a closed form relation of B in function of β_{P} is hard to achieve, we can still find a relation that gives a rough idea of the required number of bits. Specifically, we know that $1 - g_{ki} = (1 + 2^{-\frac{B}{Q}} c_{ki})^{-1}$, where $c_{ki} = \frac{\zeta_{ki} \tau d}{\zeta_{kk}}$, then, after selecting the subset $\mathcal{L}_{\text{P}}$ (of cardinality L_{P}) and k ($\in \mathcal{L}_{\text{P}}$) that yield β_{P}, we find a parameter c such that $c = \min_{i \in \mathcal{L}_{\text{P}}, i \neq k} c_{ki}$. Thus, we get

$$\boxed{\beta_{\text{P}} \leq \left(1 + 2^{-\frac{B}{Q}} c \right)^{-(L_{\text{P}}-1)}} \tag{64}$$

or equivalently we obtain

$$\boxed{B \geq Q \log_2 \left(c \left(\beta_{\text{P}}^{-(L_{\text{P}}-1)^{-1}} - 1 \right)^{-1} \right)} \tag{65}$$

Therefore, it suffices to use a number of quantization bits equals to the lower bound in the above inequality to guarantee the fraction β_{P}. Note that the exact number of bits, given by the numerical method, is less than the calculated lower bound.

VI. Numerical Results

In this section we present our numerical results. We consider a system where the number of antennas $N_{\mathrm{t}} = N_{\mathrm{r}} = 7$, $P = 10$, $\sigma = 1$, $d = 2$, $\theta = 0.01$. We take $N = 6$, which satisfies the condition $N_{\mathrm{t}} + N_{\mathrm{r}} \geq (N+1)d$. In addition, we assume that all the users have Poisson incoming traffic with the same average arrival rates as $a_k = a$. Also, we assume a scheme with a rate of $\log_2(1+\tau)$ bits per channel use if the SINR of a scheduled user exceeds τ. We set the slot duration to be $T_{\mathrm{s}} = 1000$ channel uses. Even though in practice all the path loss coefficients are different, we consider in this section a very special case that simplifies the simulations and can still provide insights on the comparison between IA and TDMA-SVD. In detail, we assume that all the direct links have a path loss coefficient of 1 and all the cross links have a path loss coefficient of ζ_{c} (with $\zeta_{\mathrm{c}} \leq 1$). This setting allows us to examine, with respect to parameter ζ_{c}, the impact of the cross channels (or equivalently, the impact of interference) on the system stability performances and it let us detect when IA technique can provide a queueing stability gain. To show the stability performance of the system, we plot the total average queue length given by $\frac{1}{M_{\mathrm{s}}}\sum_{t=0}^{M_{\mathrm{s}}-1}\sum_{k=1}^{N} q_k(t)$ for different values of a, where each simulation lasts M_{s} time-slots. We set $M_{\mathrm{s}} = 10^4$. Note that the point where the total average queue length function increases very steeply is the point at which the system becomes unstable.

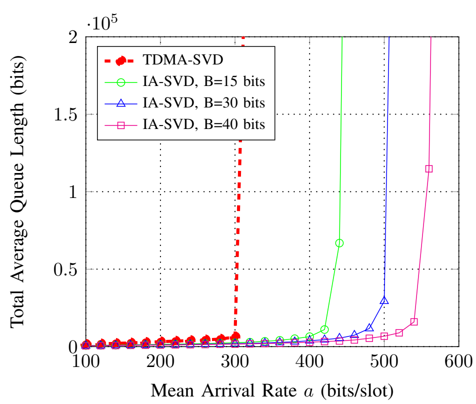

Figure 5: Total average queue length vs. mean arrival rate a. Here $\zeta_{\mathrm{c}} = 0.2$ and $\tau = 1$.

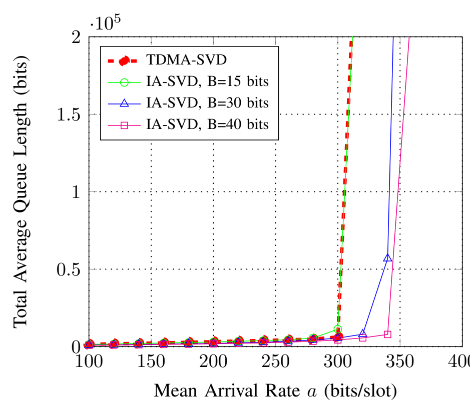

Figure 6: Total average queue length vs. mean arrival rate a. Here $\zeta_{\mathrm{c}} = 0.4$ and $\tau = 1$.

Figures 5 and 6 depicts the variations of the total average queue length w.r.t. the mean arrival rate a for two systems: (i) our adopted system, i.e. IA and SVD are used, where we consider

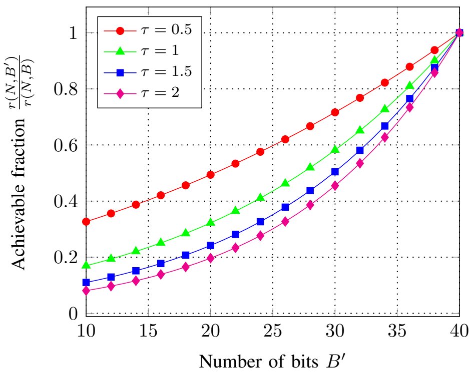

Figure 7: Achievable fraction $\frac{r(N,B')}{r(N,B)}$ vs. Number of bits B'. Here $\zeta_c = 1$ and $B = 40$ bits.

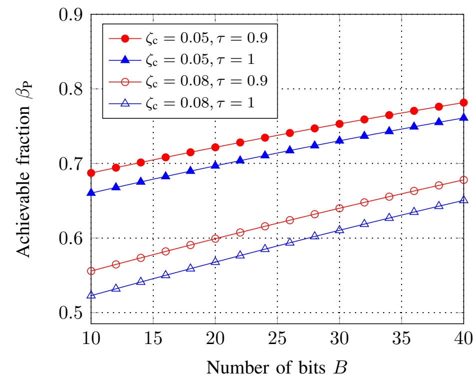

Figure 8: Achievable fraction β_P vs. Number of bits B.

several values of the number of quantization bits B, and (ii) a TDMA-SVD system, i.e. there is always only one active pair and where SVD is applied. In Figure 5, where a relatively low-interference scenario ($\zeta_c = 0.2$) is considered, we can see that IA can provide a queueing stability gain and this gain increases with the number of quantization bits. This is due to the fact that the more the quantization is precise, the more we achieve higher rates which implies better stability performances. On the other side, from Figure 6, where a relatively high-interference scenario ($\zeta_c = 0.4$) is considered, we can notice that for small B (e.g. $B = 15$ bits) IA does not provide any additional stability gain to that of SVD. However, when we increase the number of bits (e.g. $B = 30$ and 40 bits), IA becomes capable of yielding a stability gain. This results from the fact that in high interference scenarios, IA needs better CSI knowledge in order to maintain a good alignment of interference, and this can be provided by using a sufficiently large number of bits in the quantization process. It is worth noting that, although we considered the impact of B and ζ_c, there exist other parameters that may affect the system performance, such as the number of antennas, the threshold τ, the number of data streams, etc. Further, we point out that for low values of a the average queue lengths appear to be 0 in Figure 5 and 6, however these length values are not 0 but just very small compared to the maximum length value (of order 10^5).

Figure 7 depicts the variation of the fraction $\frac{r(N,B')}{r(N,B)}$ with the number of bits B', for different values of τ and for a fixed reference number of bits $B = 40$ bits; here, we set $\zeta_c = 1$ since $\frac{r(N,B')}{r(N,B)}$ is defined for the symmetric system. It is clear from this figure that increasing the number of quantization bits and/or decreasing the threshold τ result in higher achievable fractions. Also, one can notice that changing the number of quantization bits has a higher impact on the achievable fraction for greater values of τ, meaning that the more the threshold is high, the more the fraction $\frac{r(N,B')}{r(N,B)}$ is sensitive to the variation of the number of bits.

In Figure 8 we illustrate the variation of fraction β_{P} with the number of bits B, for different values of τ and ζ_{c}. The plots in this figure confirm the expectation that the stability region in the imperfect case gets bigger, i.e. the fraction this stability region achieves with respect to the stability region in the perfect case is bigger, for greater B and/or lower ζ_{c}.

VII. Conclusion

In this paper, we considered a MIMO interference network under TDD mode with limited backhaul capacity and taking into account the probing cost, and where we adopt a centralized scheduling scheme to select the active pairs in each time-slot. For the case where only one pair is active we apply the SVD technique, whereas if this number is greater than or equal to two we apply the IA technique. Under this setting, we have characterized the stability region and proposed a scheduling policy to achieve this region for the perfect case as well as for the imperfect case. Then, we have captured the maximum gap between these two resulting stability regions. These stability regions, scheduling rules and maximum gaps are provided for the symmetric system (i.e. equal path loss coefficients) as well as for the general system. In addition, for the symmetric system we have provided the conditions under which IA can deliver a queueing stability gain compared to SVD. Moreover, under the general system, since the scheduling policy is of a high computational complexity, we propose an approximate policy that has a reduced complexity but that achieves only a fraction of the system stability region. Also, a characterization of this achievable fraction is provided.

Important extensions can be addressing the stability analysis when we adopt decentralized or even mixed (centralized + decentralized) methods for feedback and scheduling.

Appendix A
Proof of Proposition 2

To begin with, let $\omega_l = (-1)^l\,(n+m_1-m_2)!\,\left((n-l)!\,(m_1-m_2+l)!\,l!\,\right)^{-1}$ and $\Omega_n = n!\,\left(m_2(n+m_1-m_2)!\right)^{-1}$. Then, we have
$L_n^{m_1-m_2}(\lambda) = \sum_{l=0}^{n} \omega_l \lambda_m^l$ and $p(\lambda_m) = \sum_{n=0}^{m_2-1} \Omega_n [L_n^{m_1-m_2}(\lambda_m)]^2 \lambda_m^{m_1-m_2} e^{-\lambda_m}$. For the Laguerre polynomial, we can write

$$[L_n^{m_1-m_2}(\lambda_m)]^2 = \sum_{j=0}^{2n} \kappa_j \lambda_m^j, \tag{66}$$

where $\kappa_j = \sum_{i=0}^{j} \omega_i \omega_{j-i}$, with $\omega_s = 0$ if $s > n$. On the other side, since $\gamma_k^{(m)} = \frac{\zeta P}{d\,\sigma^2}\lambda_m$, the corresponding success probability can be written as

$$\begin{aligned}\mathbb{P}\left\{\gamma_k^{(m)} \geq \tau\right\} &= \mathbb{P}\left\{\lambda_m \geq \frac{d\,\sigma^2\tau}{\zeta P}\right\} \\ &= \sum_{n=0}^{m_2-1} \Omega_n \sum_{j=0}^{2n} \kappa_j \int_{\frac{d\,\sigma^2\tau}{\zeta P}}^{\infty} \lambda_m^{j+m_1-m_2} e^{-\lambda_m} d\lambda_m \\ &= \sum_{n=0}^{m_2-1} \Omega_n \sum_{j=0}^{2n} \kappa_j \Gamma(j+m_1-m_2+1, \frac{d\,\sigma^2\tau}{\zeta P}),\end{aligned} \tag{67}$$

where $\Gamma(\cdot,\cdot)$ stands for the upper incomplete Gamma function. Hence, the desired result follows.

APPENDIX B
PROOF OF LEMMA 2

We start the proof by first showing that $r(L)$ decreases with L. The first derivative of this rate function is given by

$$\frac{dr}{dL} = dRe^{-\frac{\sigma^2\tau}{\alpha}} \left(-\theta + (1-L\theta)\log F\right) F^{L-1}. \tag{68}$$

Since we have $L < \frac{1}{\theta}$ and $\log F < 0$, the first derivative is negative and so r decreases with L.

To study the variation of $r_{\mathrm{T}}(L)$ (w.r.t. L) we need to first compute its first derivatives, which will help us determine the optimal number of pairs. The first derivative can be written as

$$\frac{dr_{\mathrm{T}}}{dL} = dRe^{-\frac{\sigma^2\tau}{\alpha}} \left(-L^2\theta\log F + L(-2\theta + \log F) + 1\right) F^{L-1}. \tag{69}$$

To study the sign of this derivative w.r.t. L, we first calculate its zeros and investigate if they are feasible (i.e. they satisfy $L\theta < 1$). Setting $\frac{dr_{\mathrm{T}}}{dL} = 0$ yields

$$-L^2\theta\log F + L(-2\theta + \log F) + 1 = 0, \tag{70}$$

or equivalently

$$L^2 + L\left(\frac{2}{\log F} - \frac{1}{\theta}\right) - \frac{1}{\theta \log F} = 0, \tag{71}$$

We can easily show that the only zeros of $\frac{dr_{\mathrm{T}}}{dL}$ are at

$$L_0 = \frac{\frac{1}{\theta} - \frac{2}{\log F} - \sqrt{\left(\frac{2}{\log F} - \frac{1}{\theta}\right)^2 + \frac{4}{\theta\log F}}}{2}, \tag{72}$$

$$L_1 = \frac{\frac{1}{\theta} - \frac{2}{\log F} + \sqrt{\left(\frac{2}{\log F} - \frac{1}{\theta}\right)^2 + \frac{4}{\theta\log F}}}{2}. \tag{73}$$

Note that $\log F < 0$ and $\left(\frac{2}{\log F} - \frac{1}{\theta}\right)^2 + \frac{4}{\theta \log F} = \frac{1}{\theta^2} + \frac{4}{(\log F)^2}$. Let us now examine the feasibility of L_0 and L_1. Indeed, under our setting a number L is feasible if it satisfies $0 < L < \frac{1}{\theta}$, since $L\theta$ should be < 1. For L_0 we have

$$L_0 = \frac{\frac{1}{\theta} - \frac{2}{\log F} - \sqrt{\frac{1}{\theta^2} + \frac{4}{(\log F)^2}}}{2} < \frac{\frac{1}{\theta} - \frac{2}{\log F} - \frac{2}{|\log F|}}{2} = \frac{1}{2\theta}, \tag{74}$$

where the inequality results from the fact that $\frac{2}{|\log F|} < \sqrt{\frac{1}{\theta^2} + \frac{4}{(\log F)^2}}$. We can also observe that

$$L_0 = \frac{\frac{1}{\theta} - \frac{2}{\log F} - \sqrt{\frac{1}{\theta^2} + \frac{4}{(\log F)^2}}}{2} > \frac{\frac{1}{\theta} - \frac{2}{\log F} - \frac{1}{\theta} - \frac{2}{|\log F|}}{2} = 0. \tag{75}$$

Thus, L_0 is a feasible solution since $0 < L_0 < \frac{1}{\theta}$. On the other hand, for L_1 we can notice that

$$L_1 = \frac{\frac{1}{\theta} - \frac{2}{\log F} + \sqrt{\frac{1}{\theta^2} + \frac{4}{(\log F)^2}}}{2} > \frac{\frac{1}{\theta} + \sqrt{\frac{1}{\theta^2}}}{2} = \frac{1}{\theta}. \tag{76}$$

Hence, L_1 is not a feasible solution because $L_1 > \frac{1}{\theta}$. To complete the proof it suffices to show that $r_{\mathrm{T}}(L)$ reaches its maximum at L_0. To this end, we note that $r_{\mathrm{T}}(0) = 0$, $r_{\mathrm{T}}(\frac{1}{\theta}) = 0$ and $\frac{dr_{\mathrm{T}}}{dL}|_{L=\frac{1}{2\theta}} < 0$, and we recall that $0 < L_0 < \frac{1}{2\theta} < \frac{1}{\theta}$. In addition, one can easily notice that r_{T} and its first derivative ($\frac{dr_{\mathrm{T}}}{dL}$) are continuous over $\left[0, \frac{1}{\theta}\right]$. Based on these observations, the variation of r_{T} over $\left[0, \frac{1}{\theta}\right]$ can be described as follows: r_{T} is increasing from 0 to L_0 and decreasing from L_0 to $\frac{1}{\theta}$. This concludes the proof.

APPENDIX C
PROOF OF LEMMA 3

We first provide the proof of Lemma 4 that will help us in the proof of Lemma 3.

Step 1 (*Proof of Lemma 4*): We start the proof by first defining $\mathcal{E}_{i,L}$ as the set containing the points (vectors) that only have L '1' (the other coordinate values are '0') and where the positions (indexes) of these '1' are the same as those of L '1' coordinates of $\mathbf{s}_{i,L+1}$. Note that the points in $\mathcal{E}_{i,L}$ are all different from each other. The cardinality of $\mathcal{E}_{i,L}$, which is denoted by $|\mathcal{E}_{i,L}|$, is nothing but the result of the combination of $L+1$ elements taken L at a time without repetition, and it can be computed as the following

$$|\mathcal{E}_{i,L}| = \binom{L+1}{L} = \frac{(L+1)!}{L!(L+1-L)!} = L+1. \tag{77}$$

Thus, we have $L+1$ elements from S_L that if we take them in a specific convex combination, we get a point on the same line (from the origin) as that of $\mathbf{s}_{i,L+1}$. This can be represented by

$$\sum_{j \in \mathcal{E}_{i,L}} \delta_j \mathbf{s}_{j,L} \equiv \mathbf{s}_{i,L+1}, \tag{78}$$

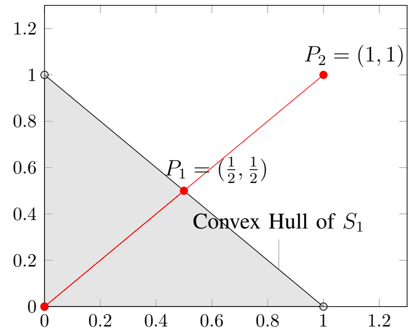

Figure 9: Example that illustrates the result of Lemma 4.

where $\equiv$ is a notation used to represent the fact that these two points are on the same line from the origin, and where $\sum_{j\in\mathcal{E}_{i,L}} \delta_j = 1$ and $\delta_j \geq 0$. Let us suppose that all the coefficients $\delta_j = \frac{1}{L+1}$. This assumption satisfies the above constraints, namely $\sum_{j\in\mathcal{E}_{i,L}} \delta_j = 1$ and $\delta_j \geq 0$. By replacing these coefficients in the term at the left-hand-side of (78), we get

$$\sum_{j\in\mathcal{E}_{i,L}} \delta_j \mathbf{s}_{j,L} = \frac{1}{L+1} \sum_{j\in\mathcal{E}_{i,L}} \mathbf{s}_{j,L} = \frac{L}{L+1} \mathbf{s}_{i,L+1}, \tag{79}$$

where the second equality holds since we have $L+1$ elements to sum (due to the fact that $|\mathcal{E}_{i,L}| = L+1$), each of which contains L '1' at the same positions as L '1' coordinates of $\mathbf{s}_{i,L+1}$, and (these elements) differ from each other in the position of one '1' (and consequently of one '0'); for instance, suppose that $N = 5$, $L = 2$ and $\mathbf{s}_{i,L+1} = (1,1,1,0,0)$, then the points $\mathbf{s}_{j,L}$ are given by the subset $\mathcal{E}_{i,L} = \{(1,1,0,0,0); (1,0,1,0,0); (0,1,1,0,0)\}$. The sum corresponding to each coordinate is then equal to L. To complete the proof, it remains to show that $\frac{1}{L+1}\sum_{j\in\mathcal{E}_{i,L}} \mathbf{s}_{j,L}$ is on the convex hull of S_L. To this end, note that all the points in S_L are on the same hyperplane (in $\mathbb{R}_+^N$), which is described by the equation $\sum_{k=1}^N \nu_k - L = 0$; ν_k represents the k-th coordinate. Hence, a point on the convex hull of S_L is also on this hyperplane. If we compute $\sum_{i=k}^N \nu_k$ for point $\frac{1}{L+1}\sum_{j\in\mathcal{E}_{i,L}} \mathbf{s}_{j,L}$, it yields $\frac{(L+1)L}{L+1} = L$ due to the definition of $\mathcal{E}_{i,L}$, thus this point is on the defined hyperplane and consequently on the convex hull of S_L.

Example: In order to better understand the result of this lemma, we provide a simple example for which the geometric illustration is in Figure 9. In this example, we take $N = 2$, $S_1 = \{(1,0); (0,1)\}$ and $S_2 = \{(1,1)\}$. In addition, we define points $P_2 = (1,1)$ and $P_1 = (\frac{1}{2}, \frac{1}{2})$. Note that $P_2 \in S_2$ and P_1 is on the convex hull of S_1. We can express P_2 as $P_2 = \frac{2}{1}[\frac{1}{2}(1,0) + \frac{1}{2}(0,1)] = 2(\frac{1}{2}, \frac{1}{2}) = \frac{2}{1}P_1$. Thus, P_2 equals $\frac{2}{1}\times$ its corresponding point (P_1) on the convex hull of S_1. This completes the proof of Lemma 4.

Step 2: Now, using the above lemma, a point $\mathbf{s}_{i,L+1}$ in S_{L+1} can be expressed in function of $L+1$ specific points in S_L as $\mathbf{s}_{i,L+1} = \frac{L+1}{L}\sum_{j\in\mathcal{E}_{i,L}}\delta_j\mathbf{s}_{j,L}$, which implies that

$$r(L+1)\mathbf{s}_{i,L+1} = r(L+1)\frac{L+1}{L}\sum_{j\in\mathcal{E}_{i,L}}\delta_j\mathbf{s}_{j,L}, \tag{80}$$

where the definition of $\mathcal{E}_{i,L}$ can be found in the proof of Lemma 4. By Lemma 2, we have $(L+1)r(L+1) < Lr(L)$ for $L \geq L_{\mathrm{I}}$. We thus get

$$r(L+1)\frac{L+1}{L}\sum_{j\in\mathcal{E}_{i,L}}\delta_j\mathbf{s}_{j,L} \prec r(L)\frac{L}{L}\sum_{j\in\mathcal{E}_{i,L}}\delta_j\mathbf{s}_{j,L} = r(L)\sum_{j\in\mathcal{E}_{i,L}}\delta_j\mathbf{s}_{j,L}. \tag{81}$$

Note that the inequality operator $\prec$ in (81) is component-wise. Therefore, each point in I_{L+1} is in the convex hull of I_L, for $L \geq L_{\mathrm{I}}$, since $r(L+1)\mathbf{s}_{i,L+1} \in I_{L+1}$ and $r(L)\sum_{j\in\mathcal{E}_{i,L}}\delta_j\mathbf{s}_{j,L}$ is in the convex hull of I_L. Consequently, all the points in I_{L+1} for $L \geq L_{\mathrm{I}}$ (i.e. these points form $\bar{\mathcal{I}}$) are in the convex hull of $I_{L_{\mathrm{I}}}$, which is a subset of $\mathcal{I}$. Therefore, the desired result holds.

Appendix D
Proof of Theorem 2

The proof consists of two steps. First we will prove that the region in the statement of the Theorem is indeed achievable. We then have to prove the converse, that is, if there exists a centralized policy that stabilizes the system for a mean arrival rate vector $\mathbf{a}$, then $\mathbf{a} \in \Lambda_{\mathrm{I}}$.

Step 1: Using the fact that a queue is stable if the arrival rate is strictly lower than the departure rate, it is sufficient to show that for each point in the stability region there exists a scheduling policy that achieves this point. Indeed, a point $\mathbf{r}_{\mathrm{I}}$ in Λ_{I} can be written as the convex combination of the points in $\mathcal{I}$ as $\mathbf{r}_{\mathrm{I}} = \sum_{i=1}^{|\mathcal{I}|}p_i\mathbf{r}_i$, where $\mathbf{r}_i$ represents a point in $\mathcal{I}$, $p_i \geq 0$ and $\sum_{i=1}^{|\mathcal{I}|}p_i = 1$. Note that each point $\mathbf{r}_i$ represents a different scheduled subset of pairs and that the probability of choosing point $\mathbf{0}$ is equal to 0. To achieve $\mathbf{r}_{\mathrm{I}}$ it suffices to use a randomized policy that at the beginning of each time-slot selects (decision) $\mathbf{r}_i$ with probability p_i. Since $\mathbf{r}_{\mathrm{I}}$ is an arbitrary point in Λ_{I}, we can claim that this region is achievable.

Step 2: Assume the system is stable for a mean arrival rate vector $\mathbf{a}$. As explained earlier, the scheduling decision (i.e. subset $\mathcal{L}$) under the centralized policy depends on the queues only, so we show this dependency by $\mathcal{L}(\mathbf{q})$. Let us denote by $\mathbf{r}_{\mathrm{s}}$ the mean service rate vector. In addition, we denote by $\mathbf{r}(\mathcal{L}(\mathbf{q}))$ the average rate vector if the queues state is $\mathbf{q}$ and the selected subset of pairs is $\mathcal{L}(\mathbf{q})$. It is obvious that the set of all possible values of $\mathbf{r}(\mathcal{L}(\mathbf{q}))$ is nothing but $\mathcal{I}\cup\bar{\mathcal{I}}$. Under the adopted model, the system can be described as a discrete time Markov chain on a countable state space with a single communicating class (i.e. irreducible) [14]. Since we assume strong stability, then we have that this chain is positive recurrent and the mean service rates greater than the mean arrival rates [14]. In addition, we can deduce that the Markov chain has a unique stationary distribution (because the chain is irreducible and positive recurrent), which we denote by $\pi(\mathbf{q})$. Thus, the following holds for the mean service rate vector

$$\mathbf{r}_{\mathrm{s}} = \sum_{\mathbf{q}\in\mathbb{Z}_{+}^{N}} \pi(\mathbf{q})\mathbf{r}(\mathcal{L}(\mathbf{q})) = \sum_{\mathcal{L}\in\mathfrak{L}} \mathbf{r}(\mathcal{L}) \sum_{\mathbf{q}\in\mathbb{Z}_{+}^{N}:\mathcal{L}(\mathbf{q})=\mathcal{L}} \pi(\mathbf{q}) \succ \mathbf{a}, \tag{82}$$

where the operator $\succ$ is component-wise. By setting $p(\mathcal{L}) = \sum_{\mathbf{q}\in\mathbb{Z}_{+}^{N}:\mathcal{L}(\mathbf{q})=\mathcal{L}} \pi(\mathbf{q})$ and noticing that the set of all possible values of $\mathbf{r}(\mathcal{L})$ is the same as $\mathbf{r}(\mathcal{L}(\mathbf{q}))$, that is, $\mathcal{I}\cup\bar{\mathcal{I}}$, the mean service rate can be re-written as

$$\mathbf{r}_{\mathrm{s}} = \sum_{j=1}^{\left|\mathcal{I}\cup\bar{\mathcal{I}}\right|} p_j \mathbf{r}_j, \tag{83}$$

in which j is used to denote decision $\mathcal{L}$, meaning that $p_j = p(\mathcal{L})$, and $\mathbf{r}_j$ represents a point in set $\mathcal{I}\cup\bar{\mathcal{I}}$, and where $\left|\mathcal{I}\cup\bar{\mathcal{I}}\right|$ represents the cardinality of this set. Hence, we can state that $\mathbf{r}_{\mathrm{s}}$ is in the convex hull of $\mathcal{I}\cup\bar{\mathcal{I}}$. But, since we have demonstrated that $\bar{\mathcal{I}}$ is in the convex hull of $\mathcal{I}$ (see Lemma 3), we have $\mathbf{r}_{\mathrm{s}} \in \Lambda_{\mathrm{I}}$ and consequently $\mathbf{a} \in \Lambda_{\mathrm{I}}$. This completes the proof.

APPENDIX E
PROOF OF PROPOSITION 4

The proof consists of two steps. In the first step we prove that if there exists an L such that $Lr(L) > r_{\mathrm{svd}}$, then IA provides a queueing stability gain compared to SVD. In the second step, we show that if for all the L we have $Lr(L) \leq r_{\mathrm{svd}}$, then IA does not provide any gain.

Step 1: Here we assume that there exists an L such that $Lr(L) > r_{\mathrm{svd}}$, with $2 \leq L \leq L_{\mathrm{I}}$, where we recall that $r(L)$ is the average rate (per user) when L pairs are active. To prove that here IA can provide a queueing stability gain, we show that IA combined with SVD is capable of achieving points that are outside the stability region of SVD. This is proven as follows.

It is well known that any point in a stability region characterized by its vertices can be written as a convex combination of these vertices. Let $\mathbf{p}_{\mathrm{d}}$ represent any point in the stability region resulting from SVD, thus this point can be expressed as a convex combination of the vertices of this region, and we can write

$$\mathbf{p}_{\mathrm{d}} = \delta_1(r_{\mathrm{svd}}, 0, \ldots, 0) + \delta_2(0, r_{\mathrm{svd}}, 0, \ldots, 0) + \ldots + \delta_N(0, \ldots, 0, r_{\mathrm{svd}}) = (\delta_1 r_{\mathrm{svd}}, \ldots, \delta_N r_{\mathrm{svd}}), \tag{84}$$

where $\delta_i \geq 0$ and $\sum_i \delta_i = 1$. Let us define a specific scheduling policy that at each time-slot schedules a subset of l pairs and where pair i is selected with a probability $\pi_i^{(l)}$. Here we assume that l can be 1 or L. Clearly, for $l = 1$ we use SVD, whereas for $l = L$ we use IA. For $l = 1$ we choose $\pi_i^{(1)}$ such that $\pi_i^{(1)} \leq \delta_i$, $\forall i$, while for $l = L$ we set $\pi_i^{(L)} = (\delta_i - \pi_i^{(1)})L$; it can be noticed that $\sum_i \pi_i^{(1)} + L^{-1} \sum_i \pi_i^{(L)} = \sum_i \delta_i = 1$. Then, the points in the stability region achieved by combining IA and SVD can be written as

$$\begin{aligned}\mathbf{p}_{\text{ad}} &= (\pi_1^{(1)} r_{\text{svd}} + \pi_1^{(L)} r(L), \ldots, \pi_N^{(1)} r_{\text{svd}} + \pi_N^{(L)} r(L)) \\ &= (\pi_1^{(1)} r_{\text{svd}} + (\delta_1 - \pi_1^{(1)}) L r(L), \ldots, \pi_N^{(1)} r_{\text{svd}} + (\delta_N - \pi_N^{(1)}) L r(L)). \end{aligned} \tag{85}$$

Since here we have $Lr(L) > r_{\text{svd}}$, then it can be deduced that

$$\begin{aligned}\mathbf{p}_{\text{ad}} &= (\pi_1^{(1)} r_{\text{svd}} + (\delta_1 - \pi_1^{(1)}) L r(L), \ldots, \pi_N^{(1)} r_{\text{svd}} + (\delta_N - \pi_N^{(1)}) L r(L)) \\ &\succ (\pi_1^{(1)} r_{\text{svd}} + (\delta_1 - \pi_1^{(1)}) r_{\text{svd}}, \ldots, \pi_N^{(1)} r_{\text{svd}} + (\delta_N - \pi_N^{(1)}) r_{\text{svd}} = (\delta_1 r_{\text{svd}}, \ldots, \delta_N r_{\text{svd}}) = \mathbf{p}_{\text{d}}, \end{aligned} \tag{86}$$

where the operator $\succ$ is component-wise. Hence, we can claim that the proposed policy achieves points that are outside the stability region of SVD.

Step 2: In this step, we assume that the condition $Lr(L) \leq r_{\text{svd}}$ holds true for any L such that $2 \leq L \leq L_{\text{I}}$. To prove that here IA cannot yield a queueing stability gain, we show that SVD is capable of achieving any point in the stability region of IA combined with SVD. This is detailed as follows. Any point in the stability region resulting from combining IA and SVD can be achieved by a scheduling policy that at each time-slot schedules a subset of pairs where pair i is selected with a probability $\pi_i^{(L)}$, such that $\sum_i \sum_L L^{-1} \pi_i^{(L)} = 1$. So this point, denoted by $\mathbf{p}_{\text{ad}}$, can be written as the following

$$\mathbf{p}_{\text{ad}} = (\pi_1^{(1)} r_{\text{svd}} + \sum_{L:L\geq 2} \pi_1^{(L)} r(L), \ldots, \pi_N^{(1)} r_{\text{svd}} + \sum_{L:L\geq 2} \pi_N^{(L)} r(L)). \tag{87}$$

Let us define a scheduling policy under SVD (i.e. one pair is active) that at each time-slot selects pair i with a probability $\delta_i = \sum_L L^{-1} \pi_i^{(L)}$, where it is clear that $\sum_i \delta_i = \sum_i \sum_L L^{-1} \pi_i^{(L)} = 1$. Hence, this policy can achieve a point, denoted by $\mathbf{p}_{\text{d}}$, such as

$$\mathbf{p}_{\text{d}} = (\delta_1 r_{\text{svd}}, \ldots, \delta_N r_{\text{svd}}) = (\sum_L L^{-1} \pi_1^{(L)} r_{\text{svd}}, \ldots, \sum_L L^{-1} \pi_N^{(L)} r_{\text{svd}}). \tag{88}$$

The condition $Lr(L) \leq r_{\text{svd}}$ yields

$$\begin{aligned}
\mathbf{p}_{\text{d}} &= (\sum_{L} L^{-1}\pi_1^{(L)} r_{\text{svd}}, \ldots, \sum_{L} L^{-1}\pi_N^{(L)} r_{\text{svd}}) \\
&= (\pi_1^{(1)} r_{\text{svd}} + \sum_{L:L\geq 2} L^{-1}\pi_1^{(L)} r_{\text{svd}}, \ldots, \pi_N^{(1)} r_{\text{svd}} + \sum_{L:L\geq 2} L^{-1}\pi_N^{(L)} r_{\text{svd}}) \\
&\succeq (\pi_1^{(1)} r_{\text{svd}} + \sum_{L:L\geq 2} L^{-1}\pi_1^{(L)} Lr(L), \ldots, \pi_N^{(1)} r_{\text{svd}} + \sum_{L:L\geq 2} L^{-1}\pi_N^{(L)} Lr(L)) = \mathbf{p}_{\text{ad}}, \qquad (89)
\end{aligned}$$

where the equality is achieved for some judicious choice of the $\pi_i^{(L)}$. This shows that the policy using SVD can achieve any point in the stability region of IA combined with SVD.

A similar proof can be done for the perfect case by replacing $r(L)$ by $\mu(L)$, we thus omit this proof to avoid repetition. Therefore, the desired result holds.

APPENDIX F
PROOF OF THEOREM 4

We first provide the proof of Lemma 5 that will help us in the proof of the theorem.

Step 1 (*Proof of Lemma 5*): From Lemma 4, a point in S_{L+1} can be expressed in function of some subset of points, represented by $\mathcal{E}_{i,L}$, in S_L as $\mathbf{s}_{i,L+1} = \frac{L+1}{L}\sum_{i_1\in\mathcal{E}_{i,L}} \delta_{i_1,L}\mathbf{s}_{i_1,L}$. More specifically, we found that the coefficients $\delta_{i_1,L}$ are all equal to $\frac{1}{L+1}$, and thus $\mathbf{s}_{i,L+1} = \frac{L+1}{L}\sum_{i_1\in\mathcal{E}_{i,L}} \frac{1}{L+1}\mathbf{s}_{i_1,L} = \frac{1}{L}\sum_{i_1\in\mathcal{E}_{i,L}} \mathbf{s}_{i_1,L}$. Similarly, each point $\mathbf{s}_{i_1,L}$ ($\in S_L$) can be written in function of some specific subset of points, denoted $\mathcal{E}_{i_1,L-1}$, in S_{L-1} as $\mathbf{s}_{i_1,L} = \frac{1}{L-1}\sum_{i_2\in\mathcal{E}_{i_1,L-1}} \mathbf{s}_{i_2,L-1}$. Following this reasoning until index $L-(n-1)$, we can express $\mathbf{s}_{i,L+1}$ as

$$\mathbf{s}_{i,L+1} = \frac{1}{L(L-1)\ldots(L-(n-1))}\sum_{i_1\in\mathcal{E}_{i,L}}\sum_{i_2\in\mathcal{E}_{i_1,L-1}}\cdots\sum_{i_n\in\mathcal{E}_{i_{n-1},L-(n-1)}} \mathbf{s}_{i_n,L-(n-1)}. \qquad (90)$$

We denote by $\mathcal{E}_{n-1,L-(n-1)}$ the subset containing the points (vectors) that only have $L-n$ '1' (the other coordinate values are '0') and where the positions (indexes) of these '1' coordinates are the same as the positions of $L-n$ '1' coordinates of $\mathbf{s}_{i,L+1}$. It is important to point out that the elements in $\mathcal{E}_{n-1,L-(n-1)}$ are all different from each other. It can be easily noticed that each $\mathcal{E}_{i_{n-1},L-(n-1)}$ (in (90)) is a subset of $\mathcal{E}_{n-1,L-(n-1)}$. The cardinality of this latter set is the result of the combination of $L+1$ elements taken $L-(n-1)$ at a time without repetition, thus we get the following

$$\left|\mathcal{E}_{n-1,L-(n-1)}\right| = \binom{L+1}{L-(n-1)} = \frac{(L+1)!}{(L-(n-1))!\,n!}. \qquad (91)$$

Let us now examine the nested summation in the expression in (90). We can remark that the summands are the elements of $\mathcal{E}_{n-1,L-(n-1)}$. Hence, the result of this nested summation is nothing but a simple sum of the vectors in $\mathcal{E}_{n-1,L-(n-1)}$, each of which multiplied by the number of times it appears in the summation. For each vector, this number is the result of the number of possible orders in which we can remove $(L+1-(L-(n-1)))$ particular '1' coordinates from $\mathbf{s}_{i,L+1}$. It follows that the required numbers are all equal to each other and given by $(L+1-(L-(n-1)))! = n!$. From the above and the fact that $(L(L-1)\dots(L-(n-1)))^{-1} = (L-n)!(L!)^{-1}$, the expression in (90) can be rewritten as

$$\begin{aligned}\mathbf{s}_{i,L+1} &= \frac{(L-n)!}{L!}\sum_{j\in\mathcal{E}_{n-1,L-(n-1)}} n!\,\mathbf{s}_{j,L-(n-1)}\\ &= \frac{(L-n)!\,n!}{L!}\frac{(L+1)!}{(L-(n-1))!\,n!}\sum_{j\in\mathcal{E}_{n-1,L-(n-1)}}\frac{(L-(n-1))!\,n!}{(L+1)!}\mathbf{s}_{j,L-(n-1)},\end{aligned} \tag{92}$$

where the second equality is due to multiplying and dividing by $\left|\mathcal{E}_{n-1,L-(n-1)}\right|$. By noticing that the factor that multiplies the summation is equal to $\frac{L+1}{L-(n-1)}$, (92) can be re-expressed as

$$\mathbf{s}_{i,L+1} = \frac{L+1}{L-(n-1)}\sum_{j\in\mathcal{E}_{n-1,L-(n-1)}}\frac{(L-(n-1))!\,n!}{(L+1)!}\mathbf{s}_{j,L-(n-1)}. \tag{93}$$

From the proof of Lemma 4, we can claim that the point formed by the convex combination $\sum_{j\in\mathcal{E}_{n-1,L-(n-1)}}\left|\mathcal{E}_{n-1,L-(n-1)}\right|^{-1}\mathbf{s}_{j,L-(n-1)}$ is on the convex hull of $S_{L-(n-1)}$ and in the same direction from the origin as $\mathbf{s}_{i,L+1}$; this combination is convex since we have $\left|\mathcal{E}_{n-1,L-(n-1)}\right|^{-1} > 0$ and $\sum_{j\in\mathcal{E}_{n-1,L-(n-1)}}\left|\mathcal{E}_{n-1,L-(n-1)}\right|^{-1} = 1$, meaning that the coefficients of this combination are non-negative and sum to 1.

Example: In order to clarify the result of this lemma, we provide a simple example in which we set $N=4$ and $L+1=4$. Under this example, we know that S_4 will contain one point, namely $\mathbf{s}_{i,4} = (1,1,1,1)$. For this point, we want to find its corresponding point on the convex hull of S_2; this implies that $n=2$. From the tree in Figure 10, it can be seen that

$$\begin{aligned}\mathbf{s}_{i,4} &= \frac{1}{3}((1,1,1,0)+(1,1,0,1)+(1,0,1,1)+(0,1,1,1))\\ &= \frac{1}{3}((1,1,0,0)+(1,0,1,0)+(0,1,1,0)+(1,0,0,1)+(0,1,0,1)+(0,0,1,1)).\end{aligned} \tag{94}$$

Remark that the 6 different vectors in the second equality form the set $\mathcal{E}_{n-1,L-(n-1)} = \mathcal{E}_{1,2}$, thus $|\mathcal{E}_{1,2}| = 6$. Using $\mathcal{E}_{1,2}$, the point that corresponds to $\mathbf{s}_{i,4}$ and that lies on the convex hull of S_2 is given by

$$\frac{1}{6}(1,1,0,0)+\frac{1}{6}(1,0,1,0)+\frac{1}{6}(0,1,1,0)+\frac{1}{6}(1,0,0,1)+\frac{1}{6}(0,1,0,1)+\frac{1}{6}(0,0,1,1). \tag{95}$$

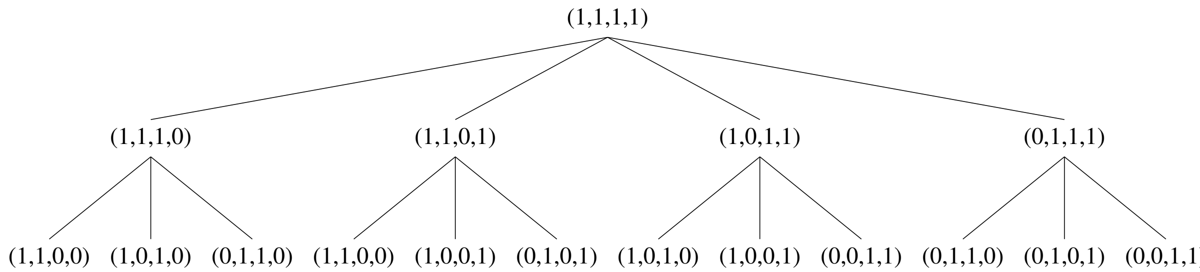

Figure 10: A tree that shows the vectors in S_2 that yield $\mathbf{s}_{i,4} = (1,1,1,1)$. Here, $N = 4$ and $n = 2$.

We can obtain $\mathbf{s}_{i,4}$ by just multiplying this convex combination by a factor of 2, which verifies the general formula provided in (93). This completes the proof of Lemma 5.

Step 2: To find the minimum achievable fraction between the stability region of the imperfect case (Λ_{I}) and the stability region of the perfect case (Λ_{P}), we examine the gap between each vertex that contributes in the characterization of Λ_{P}, where the set of these vertices is given by $\{P_1, \ldots, P_{L_{\mathrm{P}}}\}$, and the convex hull of Λ_{I}. To begin with, using the above lemma, we recall that a point in $S_{L_{\mathrm{P}}}$ can be written in function of some point that lies on the convex hull of $S_{L_{\mathrm{I}}}$, where these two points are in the same direction toward the origin. Furthermore, the gap between these two points can be captured using the fraction $\frac{L_{\mathrm{I}}}{L_{\mathrm{P}}}$. Since any point in $P_{L_{\mathrm{P}}}$ can be written as $\mu(L_{\mathrm{P}})$ times its corresponding point in $S_{L_{\mathrm{P}}}$ and a point on the convex hull of $I_{L_{\mathrm{I}}}$ is $r(L_{\mathrm{I}})$ times its corresponding point on the convex hull of $S_{L_{\mathrm{I}}}$, we can claim that the fraction between any point in $P_{L_{\mathrm{P}}}$ and its corresponding point on the convex hull of $I_{L_{\mathrm{I}}}$, and thus on the convex hull of Λ_{I}, can be given by the following

$$\frac{L_{\mathrm{I}} r(L_{\mathrm{I}})}{L_{\mathrm{P}} \mu(L_{\mathrm{P}})} = \frac{r_{\mathrm{T}}(L_{\mathrm{I}})}{\mu_{\mathrm{T}}(L_{\mathrm{P}})}. \tag{96}$$

More generally, using the above approach, we can show that the fraction between any point (vertex) in P_L, for $L_{\mathrm{I}} \leq L \leq L_{\mathrm{P}}$, and the convex hull of Λ_{I} is equal to $\frac{L_{\mathrm{I}} r(L_{\mathrm{I}})}{L \mu(L)} = \frac{r_{\mathrm{T}}(L_{\mathrm{I}})}{\mu_{\mathrm{T}}(L)}$. For these fractions, since $\mu_{\mathrm{T}}(L)$ increases for $L \leq L_{\mathrm{P}}$, the following holds

$$\frac{r_{\mathrm{T}}(L_{\mathrm{I}})}{\mu_{\mathrm{T}}(L_{\mathrm{P}})} < \frac{r_{\mathrm{T}}(L_{\mathrm{I}})}{\mu_{\mathrm{T}}(L_{\mathrm{P}} - 1)} < \ldots < \frac{r_{\mathrm{T}}(L_{\mathrm{I}})}{\mu_{\mathrm{T}}(L_{\mathrm{I}})}. \tag{97}$$

On the other side, the fraction between any vertex in P_L, for $2 \leq L \leq L_{\mathrm{I}}$, and its corresponding point on the convex hull of Λ_{I} is given by $\frac{L r(L)}{L \mu(L)} = \frac{r_{\mathrm{T}}(L)}{\mu_{\mathrm{T}}(L)}$. This is due to the fact that the point on the convex hull of $\Lambda_{\rm I}$ and that corresponds to a specific vertex in P_L, for $2 \le L \le L_{\rm I}$, is nothing but a vertex in I_L. For the fractions in this case, it is obvious that

$$\frac{r_{\rm T}(L_{\rm I})}{\mu_{\rm T}(L_{\rm I})} < \frac{r_{\rm T}(L_{\rm I}-1)}{\mu_{\rm T}(L_{\rm I}-1)} < \ldots < \frac{r_{\rm T}(2)}{\mu_{\rm T}(2)}. \tag{98}$$

Hence, using the inequalities in (97) and (98), the minimum achievable fraction is given by $\frac{r_{\rm T}(L_{\rm I})}{\mu_{\rm T}(L_{\rm P})}$. Therefore, the desired result holds.

APPENDIX G
PROOF OF THEOREM 5

The proof consists of three steps. We first show that $(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_{B'})} \ge \frac{r(L_B,B')}{r(L_B,B)}(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_B)}$. We then find the minimum fraction $\frac{r(L_B,B')}{r(L_B,B)}$ under the condition that the number of active pairs, L_B, can be less than or equal to N; we get $\frac{r(N,B')}{r(N,B)}$ as a minimum fraction. Finally, we show that the stability region $\Lambda_{B'}$ achieves at least a fraction $\frac{r(N,B')}{r(N,B)}$ of the stability region Λ_B and we conclude that $\Lambda_{B'}$ can be bounded as given in (46).

Step 1: Recall that under the symmetric case all the active pairs have the same average rate, which we denote here by $r(L_B,B)$. Thus, we can write $(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_B)} = r(L_B,B)\sum_{k\in\mathcal{L}_B} q_k$, where $L_B = |\mathcal{L}_B|$. Similarly, we get $(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_{B'})} = r(L_{B'},B')\sum_{k\in\mathcal{L}_{B'}} q_k$, with $L_{B'} = |\mathcal{L}_{B'}|$. Note that $r(L_B,B') \le r(L_B,B)$, or equivalently $\frac{r(L_B,B')}{r(L_B,B)} \le 1$, if $B' \le B$.

Depending on the subset of scheduled pairs under each of $\Delta^*_{B'}$ and Δ^*_B, four cases are to consider:

1) If the Number of Scheduled Pairs is 1 *under both of* $\Delta^*_{B'}$ *and* Δ^*_B*:* Let us denote the active pair by i. It is clear that in this case the average rate expression is independent of the number of bits and thus we have $r(1,B) = r(1,B') = r_{\text{svd}}$. Based on this, it can be seen that

$$(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_B)} = (\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_{B'})} = r_{\text{svd}}\, q_i. \tag{99}$$

Hence, for any positive fraction $\beta \le 1$, we have $(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_{B'})} \ge \beta(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_B)}$. We can therefore write $(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_{B'})} \ge \frac{r(L_B,B')}{r(L_B,B)}(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_B)}$.

2) If the Number of Scheduled Pairs is 1 *under* Δ^*_B *and strictly greater than* 1 *under* $\Delta^*_{B'}$*:* It can be observed that this case cannot take place. We prove this claim using the concept of proof by contradiction. Let i denote the index of the scheduled pair under Δ^*_B. Here we have $(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_B)} = r_{\text{svd}}\, q_i$ and $(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_{B'})} = r(L_{B'},B')\sum_{k\in\mathcal{L}_{B'}} q_k$. Since $\Delta^*_{B'}$ maximizes the product $(\mathbf{r}\cdot\mathbf{q})$ for the case where B' is the number of bits, we get

$$r(L_{B'},B')\sum_{k\in\mathcal{L}_{B'}} q_k \ge r_{\text{svd}}\, q_i. \tag{100}$$

On the other side, based on the definition of Δ_B^* and the fact that $r(L_B, B') < r(L_B, B)$, we get

$$r_{\text{svd}}\, q_i \geq r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k > r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k. \tag{101}$$

The above inequality holds for any $\mathcal{L}_B$ and in particular for $\mathcal{L}_B = \mathcal{L}_{B'}$, we thus have

$$r_{\text{svd}}\, q_i \geq r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k > r(L_{B'}, B') \sum_{k \in \mathcal{L}_{B'}} q_k. \tag{102}$$

Based on equations (101) and (102), we obtain $r_{\text{svd}}\, q_i > r_{\text{svd}}\, q_i$, which is incorrect.

3) If the Number of Scheduled Pairs is 1 *under* $\Delta_{B'}^*$ *and strictly greater than* 1 *under* Δ_B^**:* If we denote by i the index of the scheduled pair under $\Delta_{B'}^*$, we can write $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{B'}^*)} = r_{\text{svd}}\, q_i$. In addition, we have $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_B^*)} = r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k$. Since $\Delta_{B'}^*$ maximizes the product $(\mathbf{r} \cdot \mathbf{q})$ for the case where the number of bits is B', the following holds

$$r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k \leq r_{\text{svd}}\, q_i. \tag{103}$$

In addition, using the definition of Δ_B^* and the fact that $r(L_B, B') \leq r(L_B, B)$, we can write

$$r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k \leq r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k. \tag{104}$$

In order to obtain $r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k \geq \beta\, r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k$, for some $\beta \leq 1$, it suffices to take $\beta \leq \frac{r(L_B, B')}{r(L_B, B)}$. Setting $\beta = \frac{r(L_B, B')}{r(L_B, B)}$, we get $r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k = \beta\, r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k$. Combining this equality with the inequality in (103) yields

$$r_{\text{svd}}\, q_i \geq \frac{r(L_B, B')}{r(L_B, B)} r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k. \tag{105}$$

Hence, we can deduce that $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{B'}^*)} \geq \beta (\mathbf{r} \cdot \mathbf{q})^{(\Delta_B^*)}$, i.e. $(\mathbf{r} \cdot \mathbf{q})^{(\Delta_{B'}^*)} \geq \frac{r(L_B, B')}{r(L_B, B)} (\mathbf{r} \cdot \mathbf{q})^{(\Delta_B^*)}$.

4) If the Number of Scheduled Pairs is strictly greater than 1 *under both of* $\Delta_{B'}^*$ *and* Δ_B^**:* Here, the analysis is similar to the third case. In detail, since $\Delta_{B'}^*$ maximizes the product $(\mathbf{r} \cdot \mathbf{q})$ for the case where the number of bits is B', it follows that

$$r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k \leq r(L_{B'}, B') \sum_{k \in \mathcal{L}_{B'}} q_k. \tag{106}$$

Also, using the definition of Δ_B^* and the fact that $r(L_B, B') \leq r(L_B, B)$, we can write

$$r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k \leq r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k. \tag{107}$$

In order to get $r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k \geq \beta \, r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k$, for some $\beta \leq 1$, it suffices to take $\beta \leq \frac{r(L_B, B')}{r(L_B, B)}$. Setting $\beta = \frac{r(L_B, B')}{r(L_B, B)}$, we get $r(L_B, B') \sum_{k \in \mathcal{L}_B} q_k = \beta \, r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k$. Combining this equality with the inequality in (106) yields

$$r(L_{B'}, B') \sum_{k \in \mathcal{L}_{B'}} q_k \geq \frac{r(L_B, B')}{r(L_B, B)} r(L_B, B) \sum_{k \in \mathcal{L}_B} q_k. \tag{108}$$

Therefore, we can deduce that $(\mathbf{r} \cdot \mathbf{q})^{(\Delta^*_{B'})} \geq \frac{r(L_B, B')}{r(L_B, B)} (\mathbf{r} \cdot \mathbf{q})^{(\Delta^*_B)}$.

Step 2: In Step 1 we have proven that, in the four considered cases, the following holds: $(\mathbf{r} \cdot \mathbf{q})^{(\Delta^*_{B'})} \geq \beta (\mathbf{r} \cdot \mathbf{q})^{(\Delta^*_B)}$, with $\beta = \frac{r(L_B, B')}{r(L_B, B)}$.

We now want to find the minimum fraction $\frac{r(L_B, B')}{r(L_B, B)}$ w.r.t. L_B, such as

$$\underset{L_B}{\text{minimize}} \quad \frac{r(L_B, B')}{r(L_B, B)} \tag{109}$$

$$\text{subject to} \quad L_B \leq N \tag{110}$$

To solve this problem, we show that the objective function to minimize in (109) is a decreasing function w.r.t. L_B. Indeed, using (32), we have

$$\frac{r(L_B, B')}{r(L_B, B)} = \frac{(1 - L_B\theta) dR e^{-\frac{\sigma^2 \tau}{\alpha}} \left(F(B')\right)^{L_B - 1}}{(1 - L_B\theta) dR e^{-\frac{\sigma^2 \tau}{\alpha}} \left(F(B)\right)^{L_B - 1}} = \left(\frac{F\left(B'\right)}{F(B)}\right)^{L_B - 1}, \tag{111}$$

in which function F was already defined for equation (32). It is clear that $F\left(B'\right) < F\left(B\right)$ because $B' < B$, which implies that $\left(\frac{F(B')}{F(B)}\right)^{L_B - 1}$ decreases with L_B. Since $L_B \leq N$, the optimization problem reaches its minimum at $L_B = N$. Based on the above, the minimum fraction can be given by $\frac{r(N, B')}{r(N, B)}$. For the rest of the proof, we define $\beta_{\text{m}} = \frac{r(N, B')}{r(N, B)}$.

Step 3: Using the minimum fraction derived before, we now want to examine the stability region achieved by $\Delta^*_{B'}$. To this end, we define the quadratic *Lyapunov function* as

$$Ly(\mathbf{q}(t)) \triangleq \frac{1}{2} \left(\mathbf{q}(t) \cdot \mathbf{q}(t)\right) = \frac{1}{2} \sum_{k=1}^{N} q_k(t)^2. \tag{112}$$

From the evolution equation for the queue lengths (see (11)) we have

$$\begin{aligned} Ly(\mathbf{q}(t+1)) - Ly(\mathbf{q}(t)) &= \frac{1}{2} \sum_{k=1}^{N} \left[q_k(t+1)^2 - q_k(t)^2\right] \\ &= \frac{1}{2} \sum_{k=1}^{N} \left[\max\left\{q_k(t) - D_k(t), 0\right\}^2 + A_k(t)^2 - q_k(t)^2\right] \\ &\leq \sum_{k=1}^{N} \frac{\left[A_k(t)^2 + D_k(t)^2\right]}{2} + \sum_{k=1}^{N} q_k(t) \left[A_k(t) - D_k(t)\right], \end{aligned} \tag{113}$$

where in the final inequality we have used the fact that for any $q \geq 0$, $A \geq 0$, $D \geq 0$, we have

$$(\max\{q - B, 0\} + A)^2 \leq q^2 + A^2 + D^2 + 2q(A - B).$$

Now define $Dr(\mathbf{q}(t))$ as the *conditional Lyapunov drift* for time-slot t

$$Dr(\mathbf{q}(t)) \triangleq \mathbb{E}\left\{Ly(\mathbf{q}(t+1)) - Ly(\mathbf{q}(t)) \mid \mathbf{q}(t)\right\}. \tag{114}$$

From (113), we have that $Dr(\mathbf{q}(t))$ for a general scheduling policy satisfies

$$Dr(\mathbf{q}(t)) \leq \mathbb{E}\left\{\sum_{k=1}^{N} \frac{A_k(t)^2 + D_k(t)^2}{2} \mid \mathbf{q}(t)\right\} + \sum_{k=1}^{N} q_k(t) a_k - \mathbb{E}\left\{\sum_{k=1}^{N} q_k(t) D_k(t) \mid \mathbf{q}(t)\right\}, \tag{115}$$

where we have used the fact that arrivals are i.i.d. over slots and hence independent of current queue backlogs, so that $\mathbb{E}\left\{A_k(t) \mid \mathbf{q}(t)\right\} = \mathbb{E}\left\{A_k(t)\right\} = a_k$. Now define E as a finite positive constant that bounds the first term on the right-hand-side of the above drift inequality, so that for all t, all possible $q_k(t)$, and all possible control decisions that can be taken, we have

$$\mathbb{E}\left\{\sum_{k=1}^{N} \frac{A_k(t)^2 + D_k(t)^2}{2} \mid \mathbf{q}(t)\right\} \leq E. \tag{116}$$

Note that E exists since $A_k(t) < A_{\text{max}}$ and $D_k(t) < D_{\text{max}}$. Using the expression in (115) yields

$$Dr(\mathbf{q}(t)) \leq E + \sum_{k=1}^{N} q_k(t) a_k - \mathbb{E}\left\{\sum_{k=1}^{N} q_k(t) D_k(t) \mid \mathbf{q}(t)\right\}. \tag{117}$$

The conditional expectation at the right-hand-side of the above inequality is with respect to the randomly observed channel states. Thus, the drift under $\Delta^*_{B'}$ can be expressed as

$$Dr^{(\Delta^*_{B'})}(\mathbf{q}(t)) \leq E - \sum_{k=1}^{N} q_k(t) \left[\mathbb{E}\left\{D_k^{(\Delta^*_{B'})}(t) \mid \mathbf{q}(t)\right\} - a_k\right], \tag{118}$$

Note that here we have $\mathbb{E}\left\{D_k^{(\Delta^*_{B'})}(t) \mid \mathbf{q}(t)\right\} = r(L_{B'}, B')$, where the expectation at the left-hand-side of this latter equality is over the randomly observed channel state. Similarly, we have $\mathbb{E}\left\{D_k^{(\Delta^*_{B})}(t) \mid \mathbf{q}(t)\right\} = r(L_B, B)$. Hence, using (108) and the fact that the minimum fraction is $\beta_{\text{m}} = \frac{r(N,B')}{r(N,B)}$, we can write

$$\sum_{k=1}^{N} q_k(t) \mathbb{E}\left\{D_k^{(\Delta^*_{B'})}(t) \mid \mathbf{q}(t)\right\} \geq \sum_{k=1}^{N} q_k(t) \beta_{\text{m}} \mathbb{E}\left\{D_k^{(\Delta^*_{B})}(t) \mid \mathbf{q}(t)\right\}. \tag{119}$$

Plugging this directly into (118) yields

$$Dr^{(\Delta^*_{B'})}(\mathbf{q}(t)) \leq E - \sum_{k=1}^{N} q_k(t) \left[\beta_{\text{m}} \mathbb{E}\left\{D_k^{(\Delta^*_{B})}(t) \mid \mathbf{q}(t)\right\} - a_k\right]. \tag{120}$$

The above expression can be re-expressed as

$$Dr^{(\Delta^*_{B'})}(\mathbf{q}(t)) \leq E - \beta_{\mathrm{m}} \sum_{k=1}^{N} q_k(t) \left[\mathbb{E}\left\{ D_k^{(\Delta^*_B)}(t) \mid \mathbf{q}(t) \right\} - \beta_{\mathrm{m}}^{-1} a_k \right]. \tag{121}$$

Because Δ^*_B maximizes the weighted sum $\sum_{k=1}^{N} q_k(t)\mathbb{E}\left\{D_k(t) \mid \mathbf{q}(t)\right\}$ over all alternative decisions, the following holds

$$\sum_{k=1}^{N} q_k(t)\mathbb{E}\left\{ D_k^{(\Delta^*_B)}(t) \mid \mathbf{q}(t) \right\} \geq \sum_{k=1}^{N} q_k(t)\mathbb{E}\left\{ D_k^{(\Delta)}(t) \mid \mathbf{q}(t) \right\}. \tag{122}$$

where Δ represents any alternative (possibly randomized) scheduling decision that can stabilize the system. Plugging the above directly into (121) yields

$$Dr^{(\Delta^*_{B'})}(\mathbf{q}(t)) \leq E - \beta_{\mathrm{m}} \sum_{k=1}^{N} q_k(t) \left[\mathbb{E}\left\{ D_k^{(\Delta)}(t) \mid \mathbf{q}(t) \right\} - \beta_{\mathrm{m}}^{-1} a_k \right]. \tag{123}$$

Let us suppose that the mean arrival rate vector $\mathbf{a}$ is interior to fraction β_{m} of the stability region Λ_B. Thus, there exists an $\epsilon_{\max}$ such that

$$(a_1 + \epsilon_{\max}, \ldots, a_N \epsilon_{\max}) \in \beta_{\mathrm{m}} \Lambda_B, \tag{124}$$

or equivalently we have

$$(\beta_{\mathrm{m}}^{-1} a_1 + \beta_{\mathrm{m}}^{-1} \epsilon_{\max}, \ldots, \beta_{\mathrm{m}}^{-1} a_N + \beta_{\mathrm{m}}^{-1} \epsilon_{\max}) \in \Lambda_B. \tag{125}$$

Based on the above and considering a particular policy Δ that depends only on the states of the channels, we can write

$$\mathbb{E}\left\{ D_k^{(\Delta)}(t) \mid \mathbf{q}(t) \right\} = \mathbb{E}\left\{ D_k^{(\Delta)}(t) \right\} \geq \beta_{\mathrm{m}}^{-1} a_k + \beta_{\mathrm{m}}^{-1} \epsilon_{\max}, \quad \forall k \in \{1, \ldots, N\}. \tag{126}$$

Plugging the above in (123) yields

$$Dr^{(\Delta^*_{B'})}(\mathbf{q}(t)) \leq E - \epsilon_{\max} \sum_{k=1}^{N} q_k(t). \tag{127}$$

Taking an expectation of $Dr^{(\Delta^*_{B'})}$ over the randomness of the queue lengths and summing over $t \in \{0, 1, \ldots, T-1\}$ for some integer $T > 0$ we get

$$\mathbb{E}\left\{ Ly(\mathbf{q}(T)) \right\} - \mathbb{E}\left\{ Ly(\mathbf{q}(0)) \right\} \leq ET - \epsilon_{\max} \sum_{t=0}^{T-1} \sum_{k=1}^{N} \mathbb{E}\left\{ q_k(t) \right\}. \tag{128}$$

Rearranging terms, dividing by $\epsilon_{\max} T$, and taking a $\limsup$ we eventually obtain

$$\limsup_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \sum_{k=1}^{N} \mathbb{E}\left\{ q_k(t) \right\} \leq \frac{E}{\epsilon_{\max}}. \tag{129}$$

Based on the above inequality and the definition of strong stability (see Definition 1), it follows that $\Delta^*_{B'}$ stabilizes the system for any arrivals such that the mean arrival rate vector is interior to fraction β_{m} of the stability region of Δ^*_B, meaning that $\Delta^*_{B'}$ achieves up to $\Lambda_{B'} = \frac{r(N,B')}{r(N,B)}\Lambda_B$. Note that this achievable region corresponds to the worst case, that is, when the fraction is $\frac{r(N,B')}{r(N,B)}$. Hence, since the fraction is greater than or equal to $\frac{r(N,B')}{r(N,B)}$, we get

$$\frac{r(N,B')}{r(N,B)}\Lambda_B \subseteq \Lambda_{B'} \subseteq \Lambda_B, \tag{130}$$

meaning that $\Lambda_{B'}$ achieves *at least* a fraction $\frac{r(N,B')}{r(N,B)}$ of Λ_B. This completes the proof.

APPENDIX H
DERIVATION OF RATE APPROXIMATION 1

To begin with, we note that the expression of r_k given in (48) can be re-expressed as

$$r_k = (1-L\theta)dRe^{-\frac{\sigma^2\tau}{\alpha_{kk}}} \prod_{i\in\mathcal{L}, i\neq k} (1-g_{ki})^Q \,{}_2F_1(c_2, Q; c_1+c_2; g_{ki}), \tag{131}$$

which follows since $1 - g_{ki} = \left(\zeta_{ki}\tau d\,(\zeta_{kk}2^{\frac{B}{Q}})^{-1}+1\right)^{-1}$. We recall that we work under the assumption/condition that the g_{ki} are sufficiently small.

We focus on the term $(1-g_{ki})^Q\,{}_2F_1(c_2, Q; c_1+c_2; g_{ki})$. Using linear transformations (of variable) properties for the Hypergeometric function, we have the relation [32, Page 559]

$$(1-g_{ki})^Q\,{}_2F_1(c_2, Q; c_1+c_2; g_{ki}) = {}_2F_1(c_1, Q; c_1+c_2; \frac{g_{ki}}{g_{ki}-1}). \tag{132}$$

For sufficiently small g_{ki} values, we have the approximation: $\frac{g_{ki}}{g_{ki}-1} \approx -g_{ki}$. We can (numerically) verify that for $g_{ki} < 0.1$ the following accurate approximation holds

$${}_2F_1(c_1, Q; c_1+c_2; \frac{g_{ki}}{g_{ki}-1}) \approx {}_2F_1(c_1, Q; c_1+c_2; -g_{ki}). \tag{133}$$

We recall that $c_1 = (Q+1)Q^{-1}d - Q^{-1}$, $c_2 = (Q-1)c_1$ and $Q = N_{\mathrm{t}}N_{\mathrm{r}} - 1$. For sufficiently large Q, and since $d \leq \min(N_{\mathrm{t}}, N_{\mathrm{r}})$ (this implies that Q is sufficiently larger than d), we can easily see that $c_1 \approx d$, $c_2 \approx Qd - d$ and $c_1 + c_2 \approx Qd$. Now, using the Maclaurin expansion to the second order we can write

$${}_2F_1(c_1,Q; c_1+c_2; -g_{ki}) \approx 1 - \frac{c_1 Q}{c_1+c_2}g_{ki} + \frac{1}{2}\frac{c_1 Q}{c_1+c_2}\frac{(c_1+1)(Q+1)}{c_1+c_2+1}g_{ki}^2 + \mathcal{O}(g_{ki}^2) \approx 1 - g_{ki}. \tag{134}$$

In this approximation we have used the facts that $\frac{1}{2}g_{ki}^2 \ll g_{ki}$, $\frac{c_1Q}{c_1+c_2} = \frac{dQ}{Qd} = 1$, d is sufficiently high, and $\frac{(c_1+1)(Q+1)}{c_1+c_2+1} = \frac{(d+1)(Q+1)}{Qd+1} \approx 1$. In addition, $\mathcal{O}(g_{ki}^2)$ can be removed since it is negligible compared with $1 - g_{ki}$. This latter property follows from the fact that the Maclaurin expansion to higher orders (greater than two) will add terms in $g_{ki}^3, g_{ki}^4, \ldots$, which are, as g_{ki}^2, very small with respect to 1 and to the term in g_{ki}; this is due to the condition that g_{ki} is sufficiently small (i.e. $g_{ki} < 0.1$). Hence, by replacing the above approximation in the expression of r_k given in (131), we obtain the following

$$r_k \approx (1 - L\theta) dR e^{-\frac{\sigma^2 \tau}{\alpha_{kk}}} \prod_{i \in \mathcal{L}, i \neq k} (1 - g_{ki}). \tag{135}$$

APPENDIX I
DERIVATION OF RATE APPROXIMATION 2

Let us first recall that here we suppose that all the g_{ki} (with $i \neq k$) are relatively close to $\bar{g}_k$. We focus on the product $\prod_{i \in \mathcal{L}, i \neq k} (1 - g_{ki})$. Our goal here is to show that this product can be accurately approximated by the following expression

$$(1 - \bar{g}_k)^{L-1} - (1 - \bar{g}_k)^{L-2} \sum_{i \in \mathcal{L}, i \neq k} (g_{ki} - \bar{g}_k). \tag{136}$$

To prove this latter result, we start by a simple example and then we provide the general result. We consider a simple example where the product function is $f(g_{k1}, g_{k2}) = (1 - g_{k1})(1 - g_{k2})$, i.e. it can be seen as an example where $L = 3$ and $k = 3$. For this function, the Taylor expansion of order 2 around the point $(\bar{g}_k, \bar{g}_k)$ can be given as the following

$$\begin{aligned}
f(g_{k1}, g_{k2}) = {} & f(\bar{g}_k, \bar{g}_k) + (g_{k1} - \bar{g}_k) \frac{\partial f}{\partial g_{k1}}|_{(\bar{g}_k, \bar{g}_k)} + (g_{k2} - \bar{g}_k) \frac{\partial f}{\partial g_{k2}}|_{(\bar{g}_k, \bar{g}_k)} \\
& + \frac{1}{2}(g_{k1} - \bar{g}_k)^2 \frac{\partial^2 f}{\partial g_{k1}^2}|_{(\bar{g}_k, \bar{g}_k)} + \frac{1}{2}(g_{k2} - \bar{g}_k)^2 \frac{\partial^2 f}{\partial g_{k2}^2}|_{(\bar{g}_k, \bar{g}_k)} \\
& + (g_{k1} - \bar{g}_k)(g_{k2} - \bar{g}_k) \frac{\partial f}{\partial g_{k1}}|_{(\bar{g}_k, \bar{g}_k)} \frac{\partial f}{\partial g_{k2}}|_{(\bar{g}_k, \bar{g}_k)} \\
= {} & (1 - \bar{g}_k)(1 - \bar{g}_k) - (g_{k1} - \bar{g}_k)(1 - \bar{g}_k) - (g_{k2} - \bar{g}_k)(1 - \bar{g}_k) + (g_{k1} - \bar{g}_k)(g_{k2} - \bar{g}_k) \\
= {} & (1 - \bar{g}_k)^{3-1} - (1 - \bar{g}_k)^{3-2}(g_{k1} - \bar{g}_k + g_{k2} - \bar{g}_k) + (g_{k1} - \bar{g}_k)(g_{k2} - \bar{g}_k).
\end{aligned} \tag{137}$$

Note that in the first line of the above equation we use operator $=$ (and not $\approx$) since, as it can be easily noticed, we have all the expansions of f for order ≥ 3 are equal to the one for order 2. This latter statement results from the fact that $\frac{\partial^\alpha f}{\partial g_{ki}^\alpha} = 0$, $\forall \alpha \geq 2$. Since we work under the condition that all the g_{ki} are close to $\bar{g}_k$, we can claim that the term $(g_{k1} - \bar{g}_k)(g_{k2} - \bar{g}_k)$ (in (137)) is sufficiently small compared with the other terms. Hence, we get

$$f(g_{k1}, g_{k2}) \approx (1 - \bar{g}_k)^{3-1} - (1 - \bar{g}_k)^{3-2}(g_{k1} - \bar{g}_k + g_{k2} - \bar{g}_k). \tag{138}$$

The obtained result can be easily generalized, and thus we can write

$$\prod_{i\in\mathcal{L},i\neq k}(1-g_{ki})\approx(1-\bar{g}_k)^{L-1}-(1-\bar{g}_k)^{L-2}\sum_{i\in\mathcal{L},i\neq k}(g_{ki}-\bar{g}_k). \tag{139}$$

By recalling that the approximated average rate expression is $(1-L\theta)dRe^{-\frac{\sigma^2\tau}{\alpha_{kk}}}\prod_{i\in\mathcal{L},i\neq k}(1-g_{ki})$ and replacing $\prod_{i\in\mathcal{L},i\neq k}(1-g_{ki})$ with its approximation given in (139), we eventually obtain

$$r_k\approx(1-L\theta)\,dRe^{-\frac{\sigma^2\tau}{\alpha_{kk}}}\left[(1-\bar{g}_k)^{L-1}-(1-\bar{g}_k)^{L-2}\sum_{i\in\mathcal{L},i\neq k}(g_{ki}-\bar{g}_k)\right]. \tag{140}$$

APPENDIX J
PROOF OF THEOREM 6

The proof consists of two parts. The first part provides the explicit expression of fraction β_{A}. Then, in the second part, we show that the approximate policy Δ^{A} achieves at least a fraction β_{A} of the stability region achieved by Δ^{*}_{GI}.

Step 1: As in the first part of the proof of Theorem 5, depending on the number of scheduled pairs under each policy, we consider (the same) four cases.

1) If the Number of Scheduled Pairs is 1 under both of Δ^{A} and Δ^{*}_{GI}: We recall that in this case the scheduled pair is denoted by j and its average rate is given by $r_{\text{svd},j}$, independently of whether we consider the perfect or imperfect case. Thus, the dot product $\mathbf{r}\cdot\mathbf{q}$ can be written as $r_{\text{svd},j}\,q_j$ under Δ^{A} as well as under Δ^{*}_{GI}. Hence, we have $(\mathbf{r}\cdot\mathbf{q})^{(\Delta^{\text{A}})}=(\mathbf{r}\cdot\mathbf{q})^{(\Delta^{*}_{\text{GI}})}$, and for any positive constant $\beta_{\text{A}}\leq 1$ we can deduce that the following inequality holds: $(\mathbf{r}\cdot\mathbf{q})^{(\Delta^{\text{A}})}\geq\beta_{\text{A}}(\mathbf{r}\cdot\mathbf{q})^{(\Delta^{*}_{\text{GI}})}$.

2) If the Number of Scheduled Pairs is strictly greater than 1 under both of Δ^{A} and Δ^{*}_{GI}: Using Rate Approximation 2, under policy Δ^{A} the dot product $\mathbf{r}\cdot\mathbf{q}$ can be expressed as

$$(1-L_{\text{A}}\theta)dR\left[\sum_{k\in\mathcal{L}_{\text{A}}}e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L_{\text{A}}-1}q_k-\sum_{k\in\mathcal{L}_{\text{A}}}e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L_{\text{A}}-2}q_k\sum_{i\in\mathcal{L}_{\text{A}},i\neq k}(g_{ki}-\bar{g}_k)\right], \tag{141}$$

whereas under Δ^{*}_{GI} this dot product is given by

$$(1-L\theta)dR\left[\sum_{k\in\mathcal{L}}e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L-1}q_k-\sum_{k\in\mathcal{L}}e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L-2}q_k\sum_{i\in\mathcal{L},i\neq k}(g_{ki}-\bar{g}_k)\right]. \tag{142}$$

Since the approximate policy Δ^{A} schedules the subset $\mathcal{L}_{\text{A}}$ that maximizes the dot product $\boldsymbol{\phi}\cdot\mathbf{q}$, and recalling that $\phi_k(l)=(1-l\theta)dRe^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{l-1}$, it yields

$$(1-L_{\text{A}}\theta)dR\sum_{k\in\mathcal{L}_{\text{A}}}e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L_{\text{A}}-1}q_k\geq(1-L\theta)dR\sum_{k\in\mathcal{L}}e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L-1}q_k. \tag{143}$$

Similarly, using the definition of the optimal policy Δ^*_{GI} under which the dot product $\mathbf{r}\cdot\mathbf{q}$ is maximized, the following inequality holds

$$(1-L\theta)dR\left[\sum_{k\in\mathcal{L}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L-1}q_k - \sum_{k\in\mathcal{L}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L-2}q_k \sum_{i\in\mathcal{L},i\neq k}(g_{ki}-\bar{g}_k)\right] \geq$$
$$(1-L_{\text{A}}\theta)dR\left[\sum_{k\in\mathcal{L}_{\text{A}}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L_{\text{A}}-1}q_k - \sum_{k\in\mathcal{L}_{\text{A}}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L_{\text{A}}-2}q_k \sum_{i\in\mathcal{L}_{\text{A}},i\neq k}(g_{ki}-\bar{g}_k)\right]. \tag{144}$$

These last two inequalities, in (143) and (144), lead us to the simple observation

$$-(1-L\theta)dR\left[\sum_{k\in\mathcal{L}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L-2}q_k \sum_{i\in\mathcal{L},i\neq k}(g_{ki}-\bar{g}_k)\right] \geq$$
$$-(1-L_{\text{A}}\theta)dR\left[\sum_{k\in\mathcal{L}_{\text{A}}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L_{\text{A}}-2}q_k \sum_{i\in\mathcal{L}_{\text{A}},i\neq k}(g_{ki}-\bar{g}_k)\right]. \tag{145}$$

For the rest of this proof, we define

$$o_1 = (1-L_{\text{A}}\theta)dR\sum_{k\in\mathcal{L}_{\text{A}}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L_{\text{A}}-1}q_k,$$

$$p_1 = -(1-L_{\text{A}}\theta)dR\left[\sum_{k\in\mathcal{L}_{\text{A}}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L_{\text{A}}-2}q_k \sum_{i\in\mathcal{L}_{\text{A}},i\neq k}(g_{ki}-\bar{g}_k)\right],$$

$$o_2 = (1-L\theta)dR\sum_{k\in\mathcal{L}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L-1}q_k,$$

$$p_2 = -(1-L\theta)dR\left[\sum_{k\in\mathcal{L}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L-2}q_k \sum_{i\in\mathcal{L},i\neq k}(g_{ki}-\bar{g}_k)\right].$$

We can easily notice that p_1 and p_2 can be rewritten, respectively, as

$$p_1 = -(1-L_{\text{A}}\theta)dR\left[\sum_{k\in\mathcal{L}_{\text{A}}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{-1}(1-\bar{g}_k)^{L_{\text{A}}-1}q_k \sum_{i\in\mathcal{L}_{\text{A}},i\neq k}(g_{ki}-\bar{g}_k)\right], \tag{146}$$

$$p_2 = -(1-L\theta)dR\left[\sum_{k\in\mathcal{L}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{-1}(1-\bar{g}_k)^{L-1}q_k \sum_{i\in\mathcal{L},i\neq k}(g_{ki}-\bar{g}_k)\right]. \tag{147}$$

We next point out two simple but important remarks that will help us complete the proof.

- For any policy Δ_2 that approximates any policy Δ_1 to a fraction β (≤ 1), we have

$$(\mathbf{r}\cdot\mathbf{q})^{(\Delta_2)} \geq \beta(\mathbf{r}\cdot\mathbf{q})^{(\Delta_1)}.$$

If w.r.t. the approximate policy (Δ_2) there exists a scheduling policy Δ_{22} such that $(\mathbf{r}\cdot\mathbf{q})^{(\Delta_{22})} \leq (\mathbf{r}\cdot\mathbf{q})^{(\Delta_2)}$, then we can derive a fraction based on $(\mathbf{r}\cdot\mathbf{q})^{(\Delta_{22})}$ instead of $(\mathbf{r}\cdot\mathbf{q})^{(\Delta_2)}$. We can easily notice that this fraction is lower than or equal to β, therefore, w.r.t. the stability region achieved by Δ_1, Δ_2 reaches a fraction larger than that achieved by Δ_{22}.

- If w.r.t. the approximated policy (Δ_1) there exists a scheduling policy Δ_{11} such that

$$(\mathbf{r}\cdot\mathbf{q})^{(\Delta_1)} \leq (\mathbf{r}\cdot\mathbf{q})^{(\Delta_{11})},$$

then we can derive an achievable fraction based on $(\mathbf{r}\cdot\mathbf{q})^{(\Delta_{11})}$, and this fraction will be lower than or equal to β. The key idea here is that sometimes it is easier to find the fraction using Δ_{22} (resp., Δ_{11}) instead of Δ_2 (resp., Δ_1), but this will be to the detriment of finding an achievable fraction that is, in general, lower than the exact solution.

To proceed further, we consider the extreme case that corresponds to define

$$p_{1\mathrm{e}} = \min_{\mathcal{L}_\mathrm{A}\in\mathfrak{L}} \left\{ \min_{k\in\mathcal{L}_\mathrm{A}} \left\{ -(1-\bar{g}_k)^{-1} \sum_{i\in\mathcal{L}_\mathrm{A}, i\neq k} (g_{ki}-\bar{g}_k) \right\} \right\} (1-L_\mathrm{A}\theta) dR \sum_{k\in\mathcal{L}_\mathrm{A}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}} (1-\bar{g}_k)^{L_\mathrm{A}-1} q_k,$$

$$p_{2\mathrm{e}} = \max_{\mathcal{L}\in\mathfrak{L}} \left\{ \max_{k\in\mathcal{L}} \left\{ -(1-\bar{g}_k)^{-1} \sum_{i\in\mathcal{L}, i\neq k} (g_{ki}-\bar{g}_k) \right\} \right\} (1-L\theta) dR \sum_{k\in\mathcal{L}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}} (1-\bar{g}_k)^{L-1} q_k.$$

It is obvious that $p_1 \geq p_{1\mathrm{e}}$ and $p_2 \leq p_{2\mathrm{e}}$. Let us define m_1 and m_2 as

$$m_1 = \min_{\mathcal{L}_\mathrm{A}\in\mathfrak{L}} \left\{ \min_{k\in\mathcal{L}_\mathrm{A}} \left\{ -(1-\bar{g}_k)^{-1} \sum_{i\in\mathcal{L}_\mathrm{A}, i\neq k} (g_{ki}-\bar{g}_k) \right\} \right\},$$

$$m_2 = \max_{\mathcal{L}\in\mathfrak{L}} \left\{ \max_{k\in\mathcal{L}} \left\{ -(1-\bar{g}_k)^{-1} \sum_{i\in\mathcal{L}, i\neq k} (g_{ki}-\bar{g}_k) \right\} \right\}.$$

Then, it is easy to see that $p_{1\mathrm{e}} = m_1 o_1$ and $p_{2\mathrm{e}} = m_2 o_2$. This yields the following

$$(\mathbf{r}\cdot\mathbf{q})^{(\Delta^\mathrm{A})} = o_1 + p_1 \geq o_1 + p_{1e} = o_1 + m_1 o_1, \tag{148}$$

$$(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_\mathrm{GI})} = o_2 + p_2 \leq o_2 + p_{2e} = o_2 + m_2 o_2. \tag{149}$$

As mentioned earlier, Δ^A approximates Δ^*_GI to a fraction β if the following inequality holds

$$(\mathbf{r}\cdot\mathbf{q})^{(\Delta^\mathrm{A})} \geq \beta(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_\mathrm{GI})}. \tag{150}$$

In our case, it is difficult to derive β, however we can compute a fraction $\beta_\mathrm{A} \leq \beta$. In detail, based on the two properties about the achievable fraction given in the above paragraph, and combining (148) with (149), the problem turns out to find β_A such that

$$o_1(1+m_1) \geq \beta_{\mathrm{A}} o_2(1+m_2). \tag{151}$$

Using the fact that $o_2 \leq o_1$, which was shown at the beginning of this proof, it suffices to have $\beta_{\mathrm{A}} \leq \frac{1+m_1}{1+m_2}$, to satisfy the inequality in (151). Let us consider the upper bound $\beta_{\mathrm{A}} = \frac{1+m_1}{1+m_2}$. Therefore, we can deduce that $(\mathbf{r} \cdot \mathbf{q})^{(\Delta^{\mathrm{A}})} \geq \beta_{\mathrm{A}} (\mathbf{r} \cdot \mathbf{q})^{(\Delta^*_{\mathrm{GI}})}$.

3) If the Number of Scheduled Pairs is 1 under Δ^{A} and strictly greater than 1 under Δ^*_{GI}: Denoting by j the index of the scheduled pair under Δ^{A}, we can write $(\mathbf{r} \cdot \mathbf{q})^{(\Delta^{\mathrm{A}})} = r_{\mathrm{svd},j}\, q_j$. On the other side, as in the second case, $(\mathbf{r} \cdot \mathbf{q})^{(\Delta^*_{\mathrm{GI}})}$ can be expressed as

$$(1-L\theta)dR\left[\sum_{k\in\mathcal{L}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L-1}q_k - \sum_{k\in\mathcal{L}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L-2}q_k \sum_{i\in\mathcal{L}, i\neq k}(g_{ki}-\bar{g}_k)\right]. \tag{152}$$

Using the expression of m_2 that was given in the second case, it can be shown that

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta^{\mathrm{A}})} \geq \frac{1}{1+m_2}(\mathbf{r} \cdot \mathbf{q})^{(\Delta^*_{\mathrm{GI}})}. \tag{153}$$

The proof is similar to the proof of the second case and thus is omitted for the sake of brevity.

4) If the Number of Scheduled Pairs is 1 under Δ^*_{GI} and strictly greater than 1 under Δ^{A}: Denoting by j the index of the scheduled pair under Δ^*_{GI}, we have $(\mathbf{r} \cdot \mathbf{q})^{(\Delta^*_{\mathrm{GI}})} = r_{\mathrm{svd},j}\, q_j$. On the other hand, as in the second case, $(\mathbf{r} \cdot \mathbf{q})^{(\Delta^{\mathrm{A}})}$ can be expressed as

$$(1-L_{\mathrm{A}}\theta)dR\left[\sum_{k\in\mathcal{L}_{\mathrm{A}}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L_{\mathrm{A}}-1}q_k - \sum_{k\in\mathcal{L}_{\mathrm{A}}} e^{-\frac{\sigma^2\tau}{\alpha_{kk}}}(1-\bar{g}_k)^{L_{\mathrm{A}}-2}q_k \sum_{i\in\mathcal{L}_{\mathrm{A}}, i\neq k}(g_{ki}-\bar{g}_k)\right]. \tag{154}$$

Using the expression of m_1 that was provided in the second case, it can be shown that

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta^{\mathrm{A}})} \geq (1+m_1)(\mathbf{r} \cdot \mathbf{q})^{(\Delta^*_{\mathrm{GI}})}. \tag{155}$$

The proof here is also similar to the proof of the second case and thus is omitted.

Step 2: We first recall the assumptions $-1 < m_1 < 0$ and $m_2 > 0$. Thus, it can be easily seen that $\frac{1+m_1}{1+m_2} < 1+m_1$ and $\frac{1+m_1}{1+m_2} < \frac{1}{1+m_2}$. Hence, based on these inequalities and on Step 1, and recalling that $\beta_{\mathrm{A}} = \frac{1+m_1}{1+m_2}$, under the four cases we can write

$$(\mathbf{r} \cdot \mathbf{q})^{(\Delta^{\mathrm{A}})} \geq \beta_{\mathrm{A}}(\mathbf{r} \cdot \mathbf{q})^{(\Delta^*_{\mathrm{GI}})}. \tag{156}$$

Now, to complete the proof, we use a similar approach to that used in Step 3 of the proof for Theorem 5. Specifically, the drift under Δ^{A} can be expressed as

$$Dr^{(\Delta^{\mathrm{A}})}(\mathbf{q}(t)) \leq E - \sum_{k=1}^{N} q_k(t)\left[\mathbb{E}\left\{D_k^{(\Delta^{\mathrm{A}})}(t) \mid \mathbf{q}(t)\right\} - a_k\right], \tag{157}$$

for some finite constant E. Using equation (156), we can write

$$\sum_{k=1}^{N} q_k(t)\mathbb{E}\left\{D_k^{(\Delta^{\mathrm{A}})}(t) \mid \mathbf{q}(t)\right\} \geq \sum_{k=1}^{N} q_k(t)\beta_{\mathrm{A}}\,\mathbb{E}\left\{D_k^{(\Delta^*_{\mathrm{GI}})}(t) \mid \mathbf{q}(t)\right\}. \tag{158}$$

Plugging this directly into (157) yields

$$Dr^{(\Delta^{\mathrm{A}})}(\mathbf{q}(t)) \leq E - \beta_{\mathrm{A}} \sum_{k=1}^{N} q_k(t)\left[\mathbb{E}\left\{D_k^{(\Delta^*_{\mathrm{GI}})}(t) \mid \mathbf{q}(t)\right\} - \beta_{\mathrm{A}}^{-1} a_k\right], \tag{159}$$

Following similar steps to those used at the end of the proof of Theorem 5, we eventually obtain

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \sum_{k=1}^{N} \mathbb{E}\left\{q_k(t)\right\} \leq \frac{E}{\epsilon_{\max}}, \tag{160}$$

for some $\epsilon_{\max}$ and T. Based on the above, it follows that Δ^{A} stabilizes any arrivals such that the mean arrival rate vector is interior to fraction β_{A} of the stability region of Δ^*_{GI}. Note that the term *at least* in the theorem is justified by the fact that β_{A} is lower than or equal to the exact solution (β). Therefore, the desired statement follows.

Appendix K
Proof of Theorem 7

As in the first part of the proof of Theorem 5, depending on the number of scheduled pairs under each policy, we distinguish (the same) four cases.

1) If the Number of Scheduled Pairs is equal to 1 *under both of* Δ^*_{GI} *and* Δ^*_{GP}*:* We recall that in this case the scheduled pair is denoted by j and its average rate is given by $r_{\mathrm{svd},j}$, independently of whether we consider the perfect or imperfect case. Hence, the dot product $\mathbf{r}\cdot\mathbf{q}$ can be written as $r_{\mathrm{svd},j}\, q_j$ under Δ^*_{GP} as well as under Δ^*_{GI}. Hence, we have $(\boldsymbol{\mu}\cdot\mathbf{q})^{(\Delta^*_{\mathrm{GP}})} = (\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_{\mathrm{GI}})}$, and for any (positive constant) $\beta_{\mathrm{P}} \leq 1$ the following inequality holds $(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_{\mathrm{GI}})} \geq \beta_{\mathrm{P}}(\boldsymbol{\mu}\cdot\mathbf{q})^{(\Delta^*_{\mathrm{GP}})}$.

2) If the Number of Scheduled Pairs is strictly greater than 1 *under both of* Δ^*_{GI} *and* Δ^*_{GP}*:* Under policy Δ^*_{GI} and using the approximate expression of r_k given in (57), the product $(\mathbf{r}\cdot\mathbf{q})^{\Delta^*_{\mathrm{GI}}}$ can be written as

$$(1-L\theta)\left[\sum_{k\in\mathcal{L}} dRe^{-\frac{\sigma^2\tau}{\alpha_{kk}}}\, q_k \prod_{i\in\mathcal{L}, i\neq k}(1-g_{ki})\right]. \tag{161}$$

On the other hand, using the definition of Δ^*_{GP}, the product $(\boldsymbol{\mu}\cdot\mathbf{q})^{(\Delta^*_{\mathrm{GP}})}$ can be expressed as

$$(1-L_{\mathrm{P}}\theta)\sum_{k\in\mathcal{L}_{\mathrm{P}}} dRe^{-\frac{\sigma^2\tau}{\alpha_{kk}}} q_k. \tag{162}$$

One can easily remark that this latter expression has the following equivalent representation, which results from multiplying and dividing by the same term $\prod_{i\in\mathcal{L}_{\mathrm{P}},i\neq k}(1-g_{ki})$,

$$(1-L_{\mathrm{P}}\theta)\sum_{k\in\mathcal{L}_{\mathrm{P}}} dRe^{-\frac{\sigma^2\tau}{\alpha_{kk}}}q_k\left[\prod_{i\in\mathcal{L}_{\mathrm{P}},i\neq k}(1-g_{ki})\right]\left[\prod_{i\in\mathcal{L}_{\mathrm{P}},i\neq k}(1-g_{ki})\right]^{-1}. \tag{163}$$

The extreme case of $(\boldsymbol{\mu}\cdot\mathbf{q})^{(\Delta^*_{\mathrm{GP}})}$ corresponds to

$$m_3(1-L_{\mathrm{P}}\theta)\left[\sum_{k\in\mathcal{L}_{\mathrm{P}}} dRe^{-\frac{\sigma^2\tau}{\alpha_{kk}}}q_k\prod_{i\in\mathcal{L}_{\mathrm{P}},i\neq k}(1-g_{ki})\right], \tag{164}$$

where $m_3^{-1}=\min_{\mathcal{L}_{\mathrm{P}}\in\mathfrak{L}}\left\{\min_{k\in\mathcal{L}_{\mathrm{P}}}\left\{\prod_{i\in\mathcal{L}_{\mathrm{P}},i\neq k}(1-g_{ki})\right\}\right\}$. Since, by definition, policy Δ^*_{GI} produces the subset $\mathcal{L}$ and maximizes the product $(\mathbf{r}\cdot\mathbf{q})$, it yields

$$(1-L\theta)\left[\sum_{k\in\mathcal{L}} dRe^{-\frac{\sigma^2\tau}{\alpha_{kk}}}q_k\prod_{i\in\mathcal{L},i\neq k}(1-g_{ki})\right]\geq(1-L_{\mathrm{P}}\theta)\left[\sum_{k\in\mathcal{L}_{\mathrm{P}}} dRe^{-\frac{\sigma^2\tau}{\alpha_{kk}}}q_k\prod_{i\in\mathcal{L}_{\mathrm{P}},i\neq k}(1-g_{ki})\right]. \tag{165}$$

As explained earlier, the stability region achieved by Δ^*_{GI} approximates the one achieved by Δ^*_{GP} to a fraction β if $(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_{\mathrm{GI}})}\geq\beta\,(\boldsymbol{\mu}\cdot\mathbf{q})^{(\Delta^*_{\mathrm{GP}})}$. It is hard to find β based on the product $(\boldsymbol{\mu}\cdot\mathbf{q})^{(\Delta^*_{\mathrm{GP}})}$, however, using a similar observation to that provided at the end of the proof of Theorem 6, we can compute a fraction $\beta_{\mathrm{P}}\leq\beta$ based on an upper bound on this product. In detail, using (164), which represents this upper bound, our problem turns out to find β_{P} such that

$$\begin{aligned}m_3(1-L_{\mathrm{P}}\theta)&\left[\sum_{k\in\mathcal{L}_{\mathrm{P}}} dRe^{-\frac{\sigma^2\tau}{\alpha_{kk}}}q_k\prod_{i\in\mathcal{L}_{\mathrm{P}},i\neq k}(1-g_{ki})\right]\leq\\ &\beta_{\mathrm{P}}^{-1}(1-L\theta)\left[\sum_{k\in\mathcal{L}} dRe^{-\frac{\sigma^2\tau}{\alpha_{kk}}}q_k\prod_{i\in\mathcal{L},i\neq k}(1-g_{ki})\right].\end{aligned} \tag{166}$$

It suffices to take $\beta_{\mathrm{P}}^{-1}\geq m_3$, or equivalently $\beta_{\mathrm{P}}\leq m_3^{-1}$, to satisfy the above inequality. By considering $\beta_{\mathrm{P}}=m_3^{-1}$, the inequality in (166) becomes an equality. Combining the above with equation (165), we therefore get $(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_{\mathrm{GI}})}\geq\beta_{\mathrm{P}}(\boldsymbol{\mu}\cdot\mathbf{q})^{(\Delta^*_{\mathrm{GP}})}$.

3) If the Number of Scheduled Pairs is 1 *under* Δ^*_{GI} *and strictly greater than* 1 *under* Δ^*_{GP}*:* Denoting by j the index of the scheduled pair under Δ^*_{GI}, we can write $(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_{\mathrm{GI}})}=r_{\mathrm{svd},j}\,q_j$. On the other hand, as in the second case, $(\boldsymbol{\mu}\cdot\mathbf{q})^{(\Delta^*_{\mathrm{GP}})}$ can be expressed as

$$(1-L_{\mathrm{P}}\theta)\sum_{k\in\mathcal{L}_{\mathrm{P}}} dRe^{-\frac{\sigma^2\tau}{\alpha_{kk}}}q_k. \tag{167}$$

Using the expression of m_3 that was given in the second case, it can be shown that

$$(\mathbf{r}\cdot\mathbf{q})^{(\Delta^*_{\mathrm{GI}})}\geq m_3^{-1}(\boldsymbol{\mu}\cdot\mathbf{q})^{(\Delta^*_{\mathrm{GP}})}. \tag{168}$$

The proof is similar to the proof of the second case and thus is omitted for the sake of brevity.

4) If the Number of Scheduled Pairs is 1 *under* Δ^*_{GP} *and strictly greater than* 1 *under* Δ^*_{GI}*:* It can be seen that such a case cannot arise. This can be shown using the concept of proof by contradiction. The proof is simple and will not be provided for the sake of brevity.

Based on the different cases given above, we can deduce that $(\mathbf{r} \cdot \mathbf{q})^{(\Delta^*_{\text{GI}})} \geq \beta_{\text{P}} \, (\boldsymbol{\mu} \cdot \mathbf{q})^{(\Delta^*_{\text{GP}})}$, where $\beta_{\text{P}} = m_3^{-1}$. Finally, we note that the rest of the proof can be done in a similar way as in the proofs of Theorem 5 and Theorem 6, so we omit this part to avoid repetition.

REFERENCES

[1] M. Deghel, M. Assaad, and M. Debbah, "System performance of interference alignment under TDD mode with limited backhaul capacity," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 1673–1678.

[2] ——, "Queueing stability and CSI probing of a TDD wireless network with interference alignment," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 794–798.

[3] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3425–3441, August 2008.

[4] J. Thukral and H. Bölcskei, "Interference alignment with limited feedback," *CoRR*, vol. abs/0905.0374, 2009. [Online]. Available: http://arxiv.org/abs/0905.0374

[5] R. Krishnamachari and M. Varanasi, "Interference alignment under limited feedback for MIMO interference channels," in *2010 IEEE International Symposium on Information Theory Proceedings (ISIT)*, June 2010, pp. 619–623.

[6] M. Rezaee and M. Guillaud, "Limited feedback for interference alignment in the K-user MIMO interference channel," in *Information Theory Workshop (ITW), 2012 IEEE*, Sept 2012, pp. 667–671.

[7] O. E. Ayach and R. W. Heath, "Interference alignment with analog channel state feedback," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 626–636, February 2012.

[8] M. Rezaee, M. Guillaud, and F. Lindqvist, "CSIT sharing over finite capacity backhaul for spatial interference alignment," in *IEEE International Symposium on Information Theory Proceedings (ISIT'13)*. IEEE, July 2013, pp. 569–573.

[9] E. T. Ar and I. E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, 1999.

[10] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of mimo channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, June 2003.

[11] H. Boche and M. Wiczanowski, "The interplay of link layer and physical layer under MIMO enhancement: benefits and challenges," *Wireless Communications, IEEE*, vol. 13, no. 4, pp. 48–55, Aug 2006.

[12] N. McKeown, A. Mekkittikul, V. Anantharam, and J. Walrand, "Achieving 100 percent throughput in an input-queued switch," *IEEE Transactions on Communications*, vol. 47, no. 8, pp. 1260–1267, Aug 1999.

[13] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Transactions on Automatic Control*, vol. 37, no. 12, pp. 1936–1948, Dec 1992.

[14] M. Neely, E. Modiano, and C. Rohrs, "Power allocation and routing in multibeam satellites with time-varying channels," *IEEE/ACM Transactions on Networking*, vol. 11, no. 1, pp. 138–152, Feb 2003.

[15] C. Swannack, E. Uysal-biyikoglu, and G. Wornell, "Low complexity multiuser scheduling for maximizing throughput in the MIMO broadcast channel," in *Proc. Allerton Conf. Commun., Contr., and Computing*, 2004.

[16] M. Kobayashi and G. Caire, "An iterative water-filling algorithm for maximum weighted sum-rate of gaussian MIMO-BC," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1640–1646, Aug 2006.

[17] A. Destounis, M. Assaad, M. Debbah, and B. Sayadi, "Traffic-aware training and scheduling for MISO wireless downlink systems," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2574–2599, May 2015.

[18] K. Huang and V. Lau, "Stability and delay of zero-forcing SDMA with limited feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6499–6514, Oct 2012.

[19] M. J. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synthesis Lectures on Communication Networks*, vol. 3, no. 1, pp. 1–211, 2010.

[20] Y. Cui, V. K. N. Lau, R. Wang, H. Huang, and S. Zhang, "A survey on delay-aware resource control for wireless systems – large deviation theory, stochastic lyapunov drift, and distributed stochastic learning," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1677–1701, March 2012.

[21] Y. Cui, V. K. N. Lau, and H. Huang, "Dynamic partial cooperative MIMO system for delay-sensitive applications with limited backhaul capacity," *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5889–5895, November 2013.

[22] J. Chen and V. K. N. Lau, "Large deviation delay analysis of queue-aware multi-user MIMO systems with two-timescale mobile-driven feedback," *IEEE Trans. Signal Processing*, vol. 61, no. 16, pp. 4067–4076, 2013.

[23] V. K. N. Lau, F. Zhang, and Y. Cui, "Low complexity delay-constrained beamforming for multi-user MIMO systems with imperfect CSIT," *IEEE Trans. Signal Processing*, vol. 61, no. 16, pp. 4090–4099, 2013.

[24] P. Chaporkar, A. Proutiere, H. Asnani, and A. Karandikar, "Scheduling with limited information in wireless systems," in *Proceedings of the Tenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc '09. New York, NY, USA: ACM, 2009, pp. 75–84. [Online]. Available: http://doi.acm.org/10.1145/1530748.1530759

[25] M. B. Stefania Sesia, Issam Toufik, "LTE, the UMTS long term evolution: From theory to practice." Wiley, 2009.

[26] O. El Ayach, A. Lozano, and R. Heath, "On the overhead of interference alignment: Training, feedback, and cooperation," *IEEE Transactions on Wireless Communications*, vol. 11, no. 11, pp. 4192–4203, November 2012.

[27] S. W. Peters and R. W. Heath, "Interference alignment via alternating minimization," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2009, pp. 2445–2448.

[28] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5045–5060, Nov 2006.

[29] L. Georgiadis, M. J, and R. Tassiulas, "Resource allocation and cross-layer control in wireless networks," in *Foundations and Trends in Networking*, 2006, pp. 1–149.

[30] X. Chen and C. Yuen, "Performance analysis and optimization for interference alignment over MIMO interference channels with limited feedback," *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1785–1795, April 2014.

[31] C. M. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran, "On feasibility of interference alignment in MIMO interference networks," *IEEE Trans. Signal Processing*, vol. 58, no. 9, pp. 4771–4782, September 2010.

[32] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover, 1964.