H-Infinity Control Based Scheduler for the Deployment of Small Cell Networks

Subhash Lakshminarayana,[1,2,*], Mohamad Assaad[1,*], Merouane Debbah[2,*]

Abstract

In this work, we address the joint problem of traffic scheduling and interference management related to the deployment of Small Cell Networks (SCNs). The base stations of the SCNs (which we will refer to as Micro Base Stations, MBSs) are low power devices with limited buffer size. They are connected to a Central Scheduler (CS) with limited capacity backhaul links. In this scenario, traffic has to be scheduled from the network to the MBS queues in such a way that the queue-length at MBS remains as close as possible to a given target queue-length. The challenge is to design a scheduler which is oblivious to the wireless link between the MBSs and the User Terminals (UTs). For the traffic arriving at the MBS, we need to efficiently transmit it over the wireless channel to the UTs in an interference limited environment. Additionally, real time centralized interference management techniques will not be feasible. In this paper, we decouple the joint scheduling and interference management into two separate parts. For the scheduling problem, we propose a H^{∞} control based scheduler which regulates the arrival rates to the queues at the MBS. For the problem of interference management over the wireless channel, we propose a multi-cell beamforming technique and formulate a decentralized algorithm using tools from the field

*Corresponding author.

Email addresses: subhash.lakshminarayana@supelec.fr (Subhash Lakshminarayana,), mohamad.assaad@supelec.fr (Mohamad Assaad), merouane.debbah@supelec.fr (Merouane Debbah)

[1]Department of Telecommunications, SUPELEC, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France

[2]Alcatel-Lucent Chair on Flexible Radio, SUPELEC, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France

of random matrix theory. The beamforming vectors are designed to optimize two performance metrics of interest namely downlink power minimization and weighted sum rate maximization. Our simulation results show that the H^{∞} based queue length control algorithm stabilizes the queue-lengths at the MBS and keeps the variation of the queue-length around the target to a minimum.

Keywords: Small cell netwoks, Random matrix theory, Scheduler, Beamforming, H^{∞} control

1. Introduction

According to the observation made by Martin Cooper of Arraycomm, during the last century, the capacity of wireless networks has doubled every 30 months. It has been observed that the biggest gains in the efficiency of spectrum utilization has been due to network densification, i.e., shrinking of cell sizes [1]. Hence the next paradigm shift in the field of cellular networks has been identified as shrinkage of cell sizes resulting in *Small Cells Networks,* (SCN) [2]. The reduction in the cell sizes implies reducing the distance between the transmitter and the receiver, creating the dual benefit of higher link qualities and more spatial reuse. It also implies reduction in the amount of power transmitted. In this document, we will refer to the SCN base stations as Micro Base Stations (MBSs). Typically the MBSs of the small cells are of the form of data access points with limited battery and buffer sizes installed by the users to provide better data and voice coverage over a small area.

However, the reduction of cell sizes imposes numerous technical challenges in the design of such SCNs. Some of the technical issues faced in the implementation of SCNs are the problem of interference management, traffic scheduling, optimal cell size planning, call handover issues, security, backhaul infrastructure etc. For a detailed survey on the SCN deployment and design challenges in SCNs, please refer to [2],[3],[4].

In this work, we address two of the issues associated with the deployment SCNs namely interference management and traffic flow regulation.

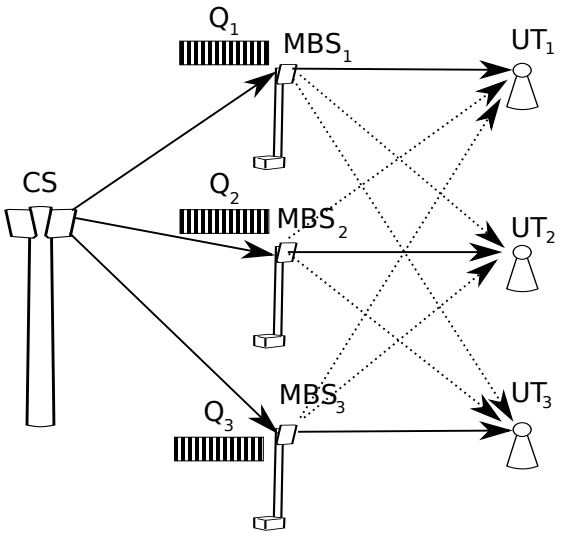

Figure 1: System Model

Consider the scenario in Figure 1. The set up consists of small cells with MBS serving the UTs over the wireless channel. The MBSs are equipped with multiple antennas and the UTs have a single antenna each. Inter-cell interference is a major bottleneck in achieving good data rates for the UTs, specially the UTs at the cell edge. Hence the MBSs perform joint multi-cell beamforming in order to manage interference. However, the MBSs must perform beamforming in a decentralized manner. To this end, we use joint multi-cell beamforming technique developed in [5] using tools from random matrix theory [6]. We consider two performance metrics of interest namely downlink power minimization and weighted sum rate maximization.

The MBSs are in turn connected to a Central Station (CS) which forwards the traffic arriving from the backbone network to the MBSs through wired backhaul link. We assume that the CS is an infinite reservoir of packets. Since the buffer sizes at the MBS is limited, we need an efficient flow controller which regulates the traffic from the CS to the MBS in such a way that the queue-length at the MBS remains as close as possible to a given target queue-length (usually the capacity of the buffer at the MBS). The main challenge in the design of the

flow controller is that it must regulate the traffic flow while being oblivious to the wireless channel conditions between the MBS and the UTs. This is a very practical assumption, since the CS is typically connected to a large number of MBSs and hence the cannot keep track of the channel conditions of all the MBSs.

We model the queue-length evolution as a linear dynamic system and develop a robust queue-length controller/regulator based on H^∞ control [7]. The H^∞ type robust controller controls the trajectory of a linear dynamical system under unknown disturbances (noise) without the knowledge of the statistical distribution of the noise process. The advantage of such a control algorithm in our setting is that the MBS do not need to feedback the CSI of the UTs to the CS. The task of the H^∞ controller is to specify an instantaneous arrival rate every time slot (from the CS to the queues present at the MBS) such that it minimizes the variation of queue length size around the target queue length.

1.1. Comparison with Cross-Layer Optimization in Wireless Networks

The problem handled in this paper is multi-disciplinary and spans the fields scheduling and power allocation in wireless networks and decentralized interference management techniques in the context of cellular networks. The problem set up described above has a flavor of cross-layer design to it. Cross-layer optimization has been a well studied topic in the context of single-hop and multi-hop wireless networks [8], [9], [10]. In these works, the problem framework naturally leads to decomposition of the problem into two parts. A fairness based flow controller to regulate the traffic flow into the queues and a max-weight based scheduler to select a set of flows whose packets can be scheduled for transmission at each time slot.

Our work is loosely connected to the cross layer framework of the above works. In our work, the scheduling aspect of cross-layer design is replaced by beamforming on the wireless channel. Equipped with multiple antennas, the MBSs obtain spatial degrees of freedom for interference management. The MBSs can simultaneously serve multiple UTs by beamforming on the wireless

channel. The flow controller part is replaced by an H^∞ based flow controller. However, unlike the fairness based congestion controllers in previous works, the objective of the H^∞ based flow controller is to minimize the variation of the queue-length around a given target queue-length. The application of H^∞ control for queue-length stabilization in the context of wireless network has been used in [11],[12]. In particular, [12] addresses a dynamic resource allocation problem in multi-service downlink OFDMA systems. However, to the authors' knowledge, the joint problem of scheduling using H^∞ control theory, power allocation and beamforming has not been handled before in the context of cellular networks.

Throughout this work, we use boldface lowercase and uppercase letters to designate column vectors and matrices, respectively. For a matrix $\mathbf{X}$, $\mathbf{X}_{i,j}$ denotes the (i,j) entry of $\mathbf{X}$. $\mathbf{X}^T$ and $\mathbf{X}^H$ denote the transpose and complex conjugate transpose of matrix $\mathbf{X}$. We denote an identity matrix of size M as $\mathbf{I}_M$ and $\text{diag}(x_1, ..., x_M)$ is a diagonal matrix of size M with the elements x_i on its main diagonal. We use $\mathbf{x} \sim \mathcal{CN}(\mathbf{m}, \mathbf{R})$ to state that the vector $\mathbf{x}$ has a complex Gaussian distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{R}$. We will use the notation $(x)^+ = \max(x, 0)$. We will also use the notation $||\mathbf{x}||^2_{\mathbf{W}}$ to denote the weighted second order norm of the vector $\mathbf{x}$ given by $\mathbf{x}^H\mathbf{W}\mathbf{x}$. We use $\mathbf{E}[Y]$ to denote the expected value of a random variable Y.

The rest of the paper is organized as follows. In section 2, we provide the system model and introduce the notations. We introduce the problem formuation in two scenarios namely downlink power minimization and weighted sum rate maximization. In section 3, we provide the algorithm description for the decentralized beamforming design in the two scenarios described above. In section 4, we provide the details of the flow control algorithm based on H$^\infty$ control. In section 5, we compare the performance of the H$^\infty$ against standard LQG based flow control. The simulation results are provided in section 6. We also provide a brief description of the details of H$^\infty$ based control and LQG control in Appendix A and B respectively.

2. System Model

The system model is shown in Figure 1. It can broken down into two parts. The frontend consisting of the MBSs and the UTs and the backend consisting of the CS and the backhaul links.

We focus on the frontend first. We consider a SCN scenario consisting N cells and K UTs per cell. The UTs in each cell are served by their MBS which is equipped with N_t antennas and each UT has a single antenna. The notation $\mathrm{UT}_{i,j}$ denotes the j-th UT present in the i-th cell. The MBS of each cell serves only the UTs present only in its cell. Let $\mathbf{h}_{i,j,k}^t \in \mathbb{C}^{N_t}$ denote the channel from the MBS i to the $\mathrm{UT}_{k,j}$ during the coherence interval t. We assume that the elements of the channel vector are Gaussian distributed, i.e., $\mathbf{h}_{i,j,k}^t \sim \mathcal{CN}(0, (\sigma_{i,j,k}^2/N_t)\mathbf{I}_{N_t})$, [3] the variance of which depends upon the path loss model between BS i and $\mathrm{UT}_{j,k}$. The channel variance has been scaled by the factor N_t to maintain the per antenna power constraint at each base station. We denote the coherence interval of the channel by the notation T_c.

We consider that a discrete time system in which the time slots are indexed by t. Before proceeding further, we will clarify the notion of time slot t in our context. We assume that arrivals and departures from the queues present at the MBS happen once every coherence interval of the channel. Hence the notation t indexes the coherence periods of the channel.

The backend part consists of the CS and the backhaul links connecting the CS to the MBS. The task of the CS is to schedule the traffic arriving from the underlying wired backbone into the queues present at the MBS. We assume that the backhaul links have just enough capacity to deliver all the traffic scheduled by the CS during a given coherence.

Let us denote the queue length at MBS i for the $\mathrm{UT}_{i,j}$ in its cell during the time slot t by the notation $q_{i,j}^t$. We denote the target buffer length at the MBS

[3] Note that the variance of the channel is constant across all time slots since we consider that the UTs do not move.

by the notation $\bar{q}_{i,j}$. We denote the traffic arrival rate into the queue of the MBS i for $\mathrm{UT}_{i,j}$ during the time slot t by the notation $a^t_{i,j}$. We also denote the rate at which packets are transmitted out of the queue (i,j) into the wireless link, (i.e. the service rate) during the time slot t by the notation $\mu^t_{i,j}$.

We now describe the design parameters of the problem. Let $\mathbf{w}^t_{i,j} \in \mathbb{C}^{N_t}$ denote the transmit downlink beamforming vector for the $\mathrm{UT}_{i,j}$ during time slot t. Likewise, let $\Gamma^t_{i,j}$ denote the received SINR for $\mathrm{UT}_{i,j}$ and $\gamma^t_{i,j}$ the corresponding target SINR during the time slot t. The received signal $y^t_{i,j} \in \mathbb{C}$ for the $\mathrm{UT}_{i,j}$ during the time slot t is given by

$$y^t_{i,j} = \sum_{l=1}^{K} {\mathbf{h}^t}^H_{i,i,j} \mathbf{w}^t_{i,l} x^t_{i,l} + \sum_{\substack{m=1 \\ m \neq i}}^{N} \sum_{n=1}^{K} {\mathbf{h}^t}^H_{m,i,j} \mathbf{w}^t_{m,n} x^t_{m,n} + z^t_{i,j}$$

where $x^t_{i,j} \in \mathbb{C}$ represents the information signal for the $\mathrm{UT}_{i,j}$ during the time slot t and $z^t_{i,j} \sim \mathcal{CN}(0, \sigma^2)$ is the corresponding additive white Gaussian complex noise.

We now provide two different problem formulations depending on the metric of interest namely the downlink power minimization and the weighted sum rate maximization.

Downlink Power Minimization - $\mathbf{P}_{\mathrm{MIN}}$: The downlink power minimization and scheduler design problem can be cast into the following optimization problem given by

$$\begin{aligned} \mathbf{P}_{\mathrm{MIN}}: \quad \min_{\mathbf{w}^t_{i,j}} \quad & \sum_{i,j} {\mathbf{w}^t}^H_{i,j} \mathbf{w}^t_{i,j}, \qquad \forall t = 1 \ldots T \\ s.t. \quad & \Gamma^t_{i,j} \geq \gamma^t_{i,j}, \qquad t = 1 \ldots T, i = 1 \ldots N, \, j = 1 \ldots K \\ & \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} q^t_{i,j} = \bar{q}_{i,j} \end{aligned} \tag{1}$$

The objective function of the optimization problem in (1) attempts to minimize the total power transmitted in the downlink during each time slot t. The first constraint equation is the target SINR constraint for each UT. The second constraint equation denotes the scheduler design constraint which tries to maintain the long term average of the queue-lengths (across multiple coherence

intervals) as close as possible to the target queue-lengths. The received SINR in the downlink is given by the following expression

$$\Gamma_{i,j}^{t} = \frac{|\mathbf{w}^{t\,H}_{i,j}\mathbf{h}^{t}_{i,i,j}|^2}{\sum_{l\neq j}|\mathbf{w}^{t\,H}_{i,l}\mathbf{h}^{t}_{i,i,j}|^2 + \sum_{m\neq i,n}|\mathbf{w}^{t\,H}_{m,n}\mathbf{h}^{t}_{m,i,j}|^2 + \sigma^2} \tag{2}$$

The numerator term is the useful signal. The denominator terms represent the intra-cell interference, inter-cell interference and the thermal noise (in order as they appear in the denominator).

Weighted Sum Rate Maximization - $\mathbf{R}_{\mathrm{MAX}}$: We now consider the problem of maximizing the weighted sum rate subject to max-power constraint on each MBS. The optimization problem is given by

$$\begin{aligned} \mathbf{R}_{\mathrm{MAX}}: \quad \max_{\mathbf{w}^{t}_{i,j}} \quad & \sum_{i,j}\beta_{i,j}\log(1+\Gamma^{t}_{i,j}), \qquad \forall t = 1\ldots T \\ s.t. \quad & \sum_{j}\mathbf{w}^{t\,H}_{i,j}\mathbf{w}^{t}_{i,j} \leq \mathrm{P}_{max}, \qquad t = 1\ldots T,\ i = 1\ldots N \\ & \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} q^{t}_{i,j} = \bar{q}_{i,j} \end{aligned} \tag{3}$$

where the downlink SINR $\Gamma^{t}_{i,j}$ is given as in equation (2). Here, $\beta_{i,j}$ are scalar weights for the downlink rate of $\mathrm{UT}_{i,j}$ and P_{max} is the max-power constraint per MBS.

3. Algorithm Design

The basic optimization problem in (1) and (3) are difficult to solve due to the interdependencies of the various parameters involved. Hence, in this work we decouple the problem into two stages. The first stage is the scheduling or the rate control problem for the queues of the MBS. The design objective of the scheduler is to minimize the variance of the queue-length around a given target queue-length. The second stage is the interference management problem.

The parameters connecting the two problems are the SINR and the service rates of the queues. The relationship between the service rate of the queues at

the MBS, μ and the SINR, γ, is given by the Shannon-Hartley theorem [4]

$$\mu = B \log(1 + \gamma) \tag{4}$$

where B is the total number of channel uses available during the coherence interval T_c. Accordingly, we denote the service rate corresponding to the target SINR ($\bar{\gamma}_{i,j}$) for $\text{UT}_{i,j}$ by $\bar{\mu}_{i,j}$ and the service rate corresponding to the achieved SINR in the downlink ($\Gamma^t_{i,j}$) during each time slot t by $\mu^t_{i,j}$. We will refer to $\bar{\mu}_{i,j}$ as the target service rate for the queues.

The decoupled algorithm can be summarized as follows:

- The UTs request a target SINR during each time slot depending on their requirement.
- The beamformer design algorithm computes the optimal design parameters to serve the data to the UTs.
- During each time slot, the scheduler regulates the amount of traffic arriving into the queues of the MBS such that the queues are stable.

In the subsequent part of this section, we will describe the decentralized multicell beamforming algorithm in detail. We will defer the flow regulation part to section 4. The problem of decentralized multicell beamforming has received considerable attention in recent years [13], [14], [15], [16]. [13] suggests beamforming design based on maximizing the *virtual signal to interference and noise ratio* (VSINR) metric which is shown to attain optimal rate points. [14] proposes a distributed beamforming approach based on Kalman smoothing involving message passing between the BSs. [15], [16] propose decentralized beamforming technique based on localized message passing, hence eliminating the need for a central processor. The decentralized algorithm proposed in this work is based on random matrix theory based approximations. In our algorithm, the MBSs

[4] We assume that the coherence interval of the channel is long enough for the transmitter to achieve the channel capacity.

would need to exchange only the channel statistics between themselves and not the instantaneous CSI, hence reducing the information exchange. Such algorithms are asymptotically optimal (when the system dimensions become very large, to be explained later in this section) and provide good approximations for practical network dimensions. We now describe the algorithm in detail.

3.1. Beamformer Design and Power Allocation: The Downlink Power Minimization Case

We first provide the a distributed algorithm to compute the optimal beamforming vectors and downlink power minimization problem in (1). We take up the beamformer design problem during a given time slot t. Hence we drop the superscript t in subsequent equations related to the above problem. The optimization problem for the beamforming design is given by

$$\begin{aligned} \min_{\mathbf{w}_{i,j}} \quad & \sum_{i,j} \mathbf{w}_{i,j}^H \mathbf{w}_{i,j} \\ s.t. \quad & \Gamma_{i,j} \geq \gamma_{i,j}, \qquad i = 1 \ldots N,\, j = 1 \ldots K \end{aligned} \tag{5}$$

The beamformer design problem is inspired from our previous work in [5].

The downlink power minimization problem in equation (5) is more difficult to solve since the computation of the beamforming vector of a given UT in turn depends on the beamforming vectors of the other UTs as well. Hence, using an approach similar to the one in [17], we convert this problem into a dual uplink problem. The uplink problem is much easier to be solved. The dual problem in the uplink is given as

$$\begin{aligned} \max \quad & \sum_{i,j} \lambda_{i,j} \sigma^2 \\ s.t. \quad & \Lambda_{i,j} \geq \gamma_{i,j}, \qquad i = 1, \ldots, N,\, j = 1, \ldots, K \end{aligned} \tag{6}$$

where the uplink SINR, $\Lambda_{i,j}$ is given by

$$\Lambda_{i,j} = \frac{\lambda_{i,j} |\hat{\mathbf{w}}_{i,j}^H \mathbf{h}_{i,i,j}|^2}{\sum_{(m,l)\neq(i,j)} \lambda_{m,l} |\hat{\mathbf{w}}_{i,j}^H \mathbf{h}_{i,m,l}|^2 + \alpha_i ||\hat{\mathbf{w}}_{i,j}||_2^2}$$

where $\hat{\mathbf{w}}$ denotes the corresponding uplink beamforming vector and $\lambda_{i,j}$ represents the dual variable associated with the optimization problem in (5). The $\lambda_{i,j}$ can be viewed as the dual uplink power.

Before stating the optimal beamforming algorithm, we define the following matrices. Let $\mathbf{H}_i = [\mathbf{h}_{i,1,1}\mathbf{h}_{i,1,2}\cdots\mathbf{h}_{i,m,n}\cdots\mathbf{h}_{i,N,K}]$ be the matrix whose columns are formed by the channel vectors from BS i to all the UTs across all the cells. Similarly, $\mathbf{\Lambda} = \text{diag}[\lambda_{1,1}\lambda_{1,2}\cdots\lambda_{m,n}\cdots\lambda_{N,K}]$ be a diagonal matrix with diagonal elements being the uplink power allocations. We also define the matrix $\mathbf{\Sigma}_i$ as

$$\mathbf{\Sigma}_i = \mathbf{H}_i\mathbf{\Lambda}\mathbf{H}_i^H \tag{7}$$

We define the Stieltjes transform ([6]) of the empirical eigen value distribution of the matrix $\mathbf{\Sigma}_i$ by the notation $m_{\mathbf{\Sigma}_i}$. When the number of antennas N_t on the MBS and the number of UTs per cell K, become very large, $(N_t, K \to \infty)$ [5]

- The iterative function for evaluating the optimal uplink power allocation $\lambda_{i,j}$ is given by

$$\lambda_{i,j} = \Big((1+\frac{1}{\gamma_{i,j}})(\frac{\sigma_{i,i,j}^2 m_{\mathbf{\Sigma}_i}(-\alpha_i)}{1+\sigma_{i,i,j}^2\lambda_{i,j}m_{\mathbf{\Sigma}_i}(-\alpha_i)})\Big)^{-1} \tag{8}$$

- The optimal receive uplink beamformers can be evaluated as

$$\hat{\mathbf{w}}_{i,j} = \Big(\sum_{m,l}\lambda_{m,l}\mathbf{h}_{i,m,l}\mathbf{h}_{i,m,l}^H + \alpha_i I\Big)^{-1}\mathbf{h}_{i,i,j} \tag{9}$$

- The optimal transmit downlink beamformers are computed using

$$\mathbf{w}_{i,j} = \sqrt{\delta_{i,j}}\hat{\mathbf{w}}_{i,j} \tag{10}$$

---

[5] The details of the algorithm can be found in [5].

The parameter $\delta_{i,j}$ is a scaling factor between the uplink and the downlink beamforming vectors. It can be computed as in [5].

The effectiveness of our algorithm lies in the fact that the computation of the uplink power parameter λ depends only on the channel statistics and not upon the actual channel realization. Hence in order to compute the beamforming vectors, MBSs only need to exchange the channel statistics (of their UTs) between themselves. This tremendously reduces the computational complexity and the information to be exchanged between between the BSs especially in a fast fading scenario in which the channel realization changes rapidly where as the channel statistics remain constant over fairly long periods of time (assuming low mobility scenario).

Note however, that the algorithm is optimal in the asymptotic setting implying that the SINR constraints are perfectly satisfied only when the number of antennas on the MBS and the number of UTs tend to infinity. However, in a practical scenario, when the network dimensions are finite, the authors in [5] show with the help of simulation results that the actual achieved SINR fluctuates around the target SINR. The fluctuations around the target SINR depends upon the channel realizations across the multiple coherence intervals.

3.1.1. The Problem of Feasibility

We now include a practical constraint in the beamformer design problem namely the max-power constraint on each MBS. We rewrite the optimization problem of Equation (5) incorporating the max-power constraint.

$$\begin{aligned} \min_{\mathbf{w}_{i,j}} \quad & \sum_{i,j} \mathbf{w}_{i,j}^H \mathbf{w}_{i,j} \\ s.t. \quad & \Gamma_{i,j} \geq \gamma_{i,j}, \qquad i = 1 \ldots N,\ j = 1 \ldots K \\ & \sum_{j} \mathbf{w}_{i,j}^H \mathbf{w}_{i,j} \leq P_{max} \qquad i = 1 \ldots N \end{aligned} \tag{11}$$

The optimization problem in (5) has a solution only if the max-power constraints for each MBS is satisfied. In other words, if the problem is infeasible, it implies that given the max-power constraints on the MBSs, the targeted SINR lie outside the capacity region of the system. In this work, we address this problem

by selectively reducing the target SINR of a few UTs. We provide a heuristic approach to select the UTs whose target SINRs can be reduced. For the cells in which $\sum_j \mathbf{w}_{i,j}^H \mathbf{w}_{i,j} \geq P_{max}$, do the following steps.

1. Denote $P_{i,j}$ as the downlink power for the $\text{UT}_{i,j}$ given by $P_{i,j} = \mathbf{w}_{i,j}^H \mathbf{w}_{i,j}$. Among the UTs belonging to the cells for which power constraint is violated, select the set of UT(s) which has the maximum ratio $\delta_{i,j} = \frac{P_{i,j}}{q_{i,j}}$. We define the $\text{UT}_{i,j'} = \arg\max_j \delta_{i,j}$.
2. If the target SINR $\gamma_{i,j'} > \tau$, decrease the target SINR for $\text{UT}_{i,j'}$ by a small amount Δ. Else, choose the UT(s) for which the parameter $\delta_{i,j'}$ is the next highest.
3. Solve the optimization problem in Equation (5) with the new set of target SINR constraints.
4. Repeat the steps $1-4$ until the problem becomes feasible.

The intuition behind the algorithm is as follows. During each time slot t, if the max-power constraints for the BSs is not met, we start by choosing the set of UTs for which $\delta_{i,j}$ is maximum. The numerator term is the power consumed by the UT. If the UT is consuming a lot of power, then the UT probably has a bad channel realization and hence we reduce its target SINR. The denominator term is the queue-length at time slot t. Minimum queue-length implies that the UT(s) has lesser amount of information bits to be served and hence we can afford to reduce the data rate of that UT. The parameter Δ is a design parameter which can be chosen suitably depending on the simulation scenario. The parameter τ is a small quantity which ensures that each UT gets at least a minimum rate. No UT goes unserved.

Therefore, the actual achieved SINR in the downlink fluctuates around the target SINR both due to the asymptotic nature of the decentralized beamformer design algorithm and also the feasibility issue. These fluctuations of the SINR result in a variation of queue-length around a given target queue-length. Moreover, the fluctuations depend on the actual channel realization during each time slot of which the CS is oblivious. Hence, we need an effective queue-length con-

trol algorithm which accounts for the fluctuations of the SINR. We will defer the description of the H^{∞} based flow controller to Section 4.

3.2. Beamformer Design and Power Allocation: The Weighted Sum-rate Maximization Case

We now take up the weighted sum rate maximization case in (3). As in the case of downlink power minimization problem discussed in the previous subsection, we decompose the problem in (3) into two parts namely the sum-rate maximization and queue-length stabilization. The problem of weighted sum rate maximization is given by

$$\begin{aligned} \max_{\mathbf{w}_{i,j}^t} \quad & \sum_{i,j} \beta_{i,j} \log(1+\Gamma_{i,j}^t), \qquad \forall t = 1 \dots T \\ s.t. \quad & \sum_{j} \mathbf{w}_j^{t^H} \mathbf{w}_{i,j}^t \leq \mathrm{P}_{max}, \qquad t = 1 \dots T,\ i = 1 \dots N \end{aligned} \tag{12}$$

Optimizing over the beamforming vector in the case of MIMO downlink channel is a particularly challenging problem even in the centralized scenario (MIMO broadcast channel) which has been addressed in the past literature [18],[19]. Noting that the main objective of the paper is to bring out the effectiveness of the H-Infinity queue length controller, we avoid the complications associated with the above optimization problem by fixing the direction of the beamforming vector and optimize only over the power allocation associated with the UTs. We choose the MMSE beamforming vector to fix the direction of beamforming vector ($\mathbf{d}_{i,j}$) given by

$$\mathbf{d}_{i,j} = \frac{\left(\sum_{m,l} \mathbf{h}_{i,m,l}\mathbf{h}_{i,m,l}^H + \mathbf{I}\right)^{-1} \mathbf{h}_{i,i,j}}{||\left(\sum_{m,l} \mathbf{h}_{i,m,l}\mathbf{h}_{i,m,l}^H + \alpha_i \mathbf{I}\right)^{-1} \mathbf{h}_{i,i,j}||_2} \tag{13}$$

and the beamforming vector $\mathbf{w}_{i,j}$ is given by

$$\mathbf{w}_{i,j} = P_{i,j}\mathbf{d}_{i,j}$$

where $P_{i,j}$ is the power allocation to the $\mathrm{UT}_{i,j}$. As shown in the iterative algorithm in [18],[19], the MMSE beamforming direction is a reasonable choice for

optimizing the weighted sum rate. Hence the modified optimization problem is given by

$$\max_{P^t_{i,j}} \quad \sum_{i,j} \beta_{i,j} \log\left(1 + \frac{P^t_{i,j}|\mathbf{d}^{t\,H}_{i,j}\mathbf{h}^t_{i,i,j}|^2}{\sum_{l\neq j} P^t_{i,l}|\mathbf{d}^{t\,H}_{i,l}\mathbf{h}^t_{i,i,j}|^2 + \sum_{m\neq i,n} P^t_{m,n}|\mathbf{d}^{t\,H}_{m,n}\mathbf{h}^t_{m,i,j}|^2 + \sigma^2}\right) \quad (14)$$
$$s.t. \quad \sum_j P^t_{i,j} \leq \mathrm{P}_{max}, \qquad t = 1\ldots T,\ i = 1\ldots N$$

Once again, as in the previous case, we assume that the BSs are only allowed exchange the channel statistics and not the instantaneous CSI values. Hence, the MBSs optimize the power allocation based on the asymptotic approximations of the associated quantities which are given by

$$|\mathbf{d}^{t\,H}_{i,j}\mathbf{h}^t_{i,i,j}|^2 - \left(\frac{\sigma^2_{i,i,j}\bar{m}_{\mathbf{\Sigma}_i}}{1+\sigma^2_{i,i,j}\bar{m}_{\mathbf{\Sigma}_i}}\right)^2 \xrightarrow[N_t,K\to\infty]{a.s.} 0 \quad (15)$$

$$|\mathbf{d}^{t\,H}_{i,l}\mathbf{h}^t_{i,i,j}|^2 - \frac{(\sigma^2_{i,i,j}\sigma^2_{i,i,l}/N_t)\bar{m}'_{\mathbf{\Sigma}_i}}{\left(1+\sigma^2_{i,i,j}\bar{m}_{\mathbf{\Sigma}_i}\right)^2\left(1+\sigma^2_{i,i,l}\bar{m}_{\mathbf{\Sigma}_i}\right)^2} \xrightarrow[N_t,K\to\infty]{a.s.} 0 \quad (16)$$

$$|\mathbf{d}^{t\,H}_{m,n}\mathbf{h}^t_{m,i,j}|^2 - \frac{(\sigma^2_{m,i,j}\sigma^2_{m,m,n}/N_t)\bar{m}'_{\mathbf{\Sigma}_m}}{\left(1+\sigma^2_{m,i,j}\bar{m}_{\mathbf{\Sigma}_m}\right)^2\left(1+\sigma^2_{m,m,n}\bar{m}_{\mathbf{\Sigma}_m}\right)^2} \xrightarrow[N_t,K\to\infty]{a.s.} 0 \quad (17)$$

Here $\bar{m}_{\mathbf{\Sigma}_i}$ is the Stieltjes transform of the matrix $\mathbf{\Sigma}_i = \left(\sum_{m,l}\mathbf{h}_{i,m,l}\mathbf{h}^H_{i,m,l} + \mathbf{I}\right)$. The reader can refer to [5] for the derivations using random matrix theory. The details have been omitted in this paper. In order to simplify notations, let us introduce

$$\bar{G}_{i,i,j} \triangleq \frac{\sigma^2_{i,i,j}}{\left(1+\sigma^2_{i,i,j}\bar{m}_{\mathbf{\Sigma}_i}\right)^2} \qquad \bar{G}_{i,i,l} \triangleq \frac{\sigma^2_{i,i,l}}{\left(1+\sigma^2_{i,i,l}\bar{m}_{\mathbf{\Sigma}_i}\right)^2}$$
$$\bar{G}_{m,i,j} \triangleq \frac{\sigma^2_{m,i,j}}{\left(1+\sigma^2_{m,i,j}\bar{m}_{\mathbf{\Sigma}_m}\right)^2} \qquad \bar{G}_{m,m,n} \triangleq \frac{\sigma^2_{m,m,n}}{\left(1+\sigma^2_{m,m,n}\bar{m}_{\mathbf{\Sigma}_m}\right)^2}$$

Let us denote the downlink SINR obtained by replacing the numerator and denominator terms by their asymptotic equivalents by $\bar{\Gamma}_{i,j}$,

$$\bar{\Gamma}_{i,j} = \frac{P_{i,j}\sigma^2_{i,i,j}\bar{G}_{i,i,j}\bar{m}_{\mathbf{\Sigma}_i}}{\sum_{l\neq j} P_{i,l}\bar{G}_{i,i,j}\bar{G}_{i,i,l}\bar{m}'_{\mathbf{\Sigma}_i} + \sum_{m\neq i,n} P_{m,n}\bar{G}_{m,i,j}\bar{G}_{m,m,n}\bar{m}'_{\mathbf{\Sigma}_m} + \sigma^2} \quad (18)$$

Note that with the asymptotic approximation, the downlink SINR $\bar{\Gamma}_{i,j}$ is the same across the time slots (since $\bar{\Gamma}_{i,j}$ only depends upon the channel statistics which is constant across time slots). We re frame the problem with asymptotic approximations given by

$$\mathbf{R}_{\mathrm{MAX}}^{\mathrm{asymp}}: \quad \max_{P_{i,j}} \quad \sum_{i,j} \beta_{i,j} \log\left(1+\bar{\Gamma}_{i,j}\right) \tag{19}$$
$$s.t. \quad \sum_{j} P_{i,j} \leq \mathrm{P}_{max}, \qquad i = 1 \ldots N$$

The MBSs can now solve the optimization problem $R_{\mathrm{MAX}}^{\mathrm{asymp}}$ by exchanging only the channel statistics between themselves. We make the high SINR approximation and use $\log(1+x) \approx \log(x)$, for large x. However, the optimization problem in $\log(\bar{\Gamma}_{i,j})$ is not concave in terms of $\mathbf{P}$. Hence, we proceed by using geometric programming approach [20] and make the variable change to $\tilde{P}_{i,j} = \log(P_{i,j})$.

$$\mathbf{R}_{\mathrm{MAX}}^{\mathrm{asymp}}: \qquad \max_{\tilde{P}_{i,j}} \log\left(\exp(\tilde{P}_{i,j})\sigma_{i,i,j}^2 \bar{G}_{i,i,j}\bar{m}_{\mathbf{\Sigma}_i}\right) \tag{20}$$
$$- \quad \log\left(\sum_{l\neq j} \exp(\tilde{P}_{i,l})\bar{G}_{i,i,j}\bar{G}_{i,i,l}\bar{m}'_{\mathbf{\Sigma}_i} + \sum_{m\neq i,n} \exp(\tilde{P}_{m,n})\bar{G}_{m,i,j}\bar{G}_{m,m,n}\bar{m}'_{\mathbf{\Sigma}_m} + \sigma^2\right)$$
$$s.t. \quad \sum_{j} \exp(\tilde{P}_{i,j}) \leq \mathrm{P}_{max}, \qquad i = 1 \ldots N$$

The problem is now a convex optimization problem in terms of the variable $\tilde{P}_{i,j}$ (refer to [20] for details). We Introduce the Lagrange variable δ_i for the max-power constraint and formulate the Lagrangian as

$$\mathcal{L}(\tilde{\mathbf{P}}, \boldsymbol{\delta}) = \sum_{i,j} \beta_{i,j} \log\left(\exp(\tilde{P}_{i,j})\sigma_{i,i,j}^2 \bar{G}_{i,i,j}\bar{m}_{\mathbf{\Sigma}_i}\right) \tag{21}$$
$$- \log\left(\sum_{l\neq j} \exp(\tilde{P}_{i,l})\bar{G}_{i,i,j}\bar{G}_{i,i,l}\bar{m}'_{\mathbf{\Sigma}_i} + \sum_{m\neq i,n} \exp(\tilde{P}_{m,n})\bar{G}_{m,i,j}\bar{G}_{m,m,n}\bar{m}'_{\mathbf{\Sigma}_m} + \sigma^2\right) \tag{22}$$
$$- \sum_{i} \delta_i \left(\sum_{i,j} \exp(\tilde{P}_{i,j}) - \mathrm{P}_{max}\right) \tag{23}$$

Let us also denote the terms inside log in equation (22) by the notation $\bar{\mathcal{I}}_{i,j}$. We also define $\tilde{\mathbf{P}} = \left[\tilde{P}_{1,1}, \tilde{P}_{1,2}, \ldots, \tilde{P}_{1,K}, \ldots, \tilde{P}_{N,K}\right]$. The solution to the optimiza-

tion problem can now be given by

$$\min_{\boldsymbol{\delta}} \max_{\tilde{\mathbf{P}}} \ \mathcal{L}(\tilde{\mathbf{P}}, \boldsymbol{\delta}) \tag{24}$$

We will use the gradient based method to solve the dual problem [21]. At the iteration count τ for the gradient based algorithm, for a given $\boldsymbol{\delta} = \boldsymbol{\delta}(\tau)$, let us define

$$\tilde{\mathbf{P}}(\tau) = \arg\max_{\tilde{\mathbf{P}}} \ \mathcal{L}(\tilde{\mathbf{P}}, \boldsymbol{\delta}(\tau)) \tag{25}$$

The update equation for the dual variable is given by

$$\delta_i(\tau+1) = \delta_i(\tau) - \alpha \left(\sum_j \exp(\tilde{P}_{i,j}(\tau)) - \mathrm{P}_{max} \right) \qquad \forall i = 1, \ldots, N \tag{26}$$

In order to solve the optimization problem in (25), we once again use the gradient based method. The gradient of $\mathcal{L}(\tilde{\mathbf{P}}, \boldsymbol{\delta}(\tau))$ with respect to $\tilde{P}_{i,j}$ at iteration ζ is given by

$$\Delta_{i,j}(\zeta, \tau) = \frac{\partial \mathcal{L}(\tilde{\mathbf{P}}, \boldsymbol{\delta}(\tau))}{\partial \tilde{P}_{i,j}} \Big|_{\tilde{P}_{i,j} = \tilde{P}_{i,j}(\zeta,\tau)}$$

In particular, we have

$$\Delta_{i,j}(\zeta, \tau) = \beta_{i,j} - \sum_{l \neq j} \frac{\beta_{i,l} \exp(\tilde{P}_{i,l}(\zeta,\tau)) \bar{G}_{i,i,l} \bar{G}_{i,i,j} \bar{m}'_{\boldsymbol{\Sigma}_i}}{\bar{\mathcal{I}}_{i,l}}$$
$$- \sum_{m \neq i, n} \frac{\beta_{m,n} \exp(\tilde{P}_{m,n}(\zeta,\tau)) \bar{G}_{i,m,n} \bar{G}_{i,i,j} \bar{m}'_{\boldsymbol{\Sigma}_m}}{\bar{\mathcal{I}}_{m,n}} - \delta_i(\tau) \exp(\tilde{P}_{i,j}(\zeta,\tau))$$

We iteratively update the power allocation until convergence

$$\tilde{P}_{i,j}(\zeta+1, \tau) = \tilde{P}_{i,j}(\zeta, \tau) + \kappa \Delta_{i,j}(\zeta, \tau) \qquad i = 1, \ldots, N, j = 1, \ldots, K$$

where κ is a small step size. The value $\tilde{P}_{i,j}(\tau)$ is the value of the gradient algorithm at the convergence point given by,

$$\tilde{P}_{i,j}(\tau) = \lim_{\zeta \to \infty} \tilde{P}_{i,j}(\zeta, \tau)$$

Now, let use denote the solution of the optimization problem $\mathbf{R}_{\mathrm{MAX}}^{\mathrm{asymp}}$ as

$$\bar{\mu}_{i,j} = \max_{P_{i,j}} \mathbf{R}_{\mathrm{MAX}}^{\mathrm{asymp}}$$

$$\bar{P}_{i,j} = \arg\max_{P_{i,j}} \mathbf{R}_{\mathrm{MAX}}^{\mathrm{asymp}}$$

The MBSs precompute the power allocation based on the channel statistics to $\bar{P}_{i,j}$ and assume a precomputed service rate $\bar{\mu}_{i,j}$. The precomputed service rate can be viewed as an estimate of the service rate offered to the queues at the CS. However, in reality, the actual service rate offered to the queues depends on the channel realization and fluctuates around the target service rate. The service rate obtained per time slot is given by

$$\mu_{i,j}^{t} = \log\left(1 + \frac{\bar{P}_{i,j}|\mathbf{d}^{t\,H}_{i,j}\mathbf{h}^{t}_{i,i,j}|^2}{\sum_{l\neq j}\bar{P}_{i,l}|\mathbf{d}^{t\,H}_{i,l}\mathbf{h}^{t}_{i,i,j}|^2 + \sum_{m\neq i,n}\bar{P}_{m,n}|\mathbf{d}^{t\,H}_{m,n}\mathbf{h}^{t}_{m,i,j}|^2 + \sigma^2}\right)$$

The CS has only an estimate of the service rate provided to the UTs given by $\bar{\mu}_{i,j}$. However, the CS being oblivious to the actual channel realization does not have the knowledge of the actual service rates in the downlink.

Hence, it is clear that in both the optimization problems (downlink power minimization and weighted sum rate maximization) there is a mismatch between the estimated service rate and actual service rate obtained to the queues. We model the mismatch between the targeted service rate and the actual service rate as unknown noise and apply H^{∞} based queue length controller described in section 4 to regulate the arrival rates into the queues.

In what follows, we provide a flow control algorithm based on H^{∞} control.

4. Scheduler Design and Queue-Length Stability at the MBS

We now describe the scheduler design. The scheduler has to regulate the flow into the queues without the knowledge of the wireless channel conditions. (which implies that the scheduler does not have the knowledge of the actual received SINR in the downlink, $\mu_{i,j}^{t}$).

Let us now focus on the queue length dynamics at the MBS. The equation for the queue-length evolution at the MBS is given by

$$q_{i,j}^{t+1} = \left(q_{i,j}^{t} + a_{i,j}^{t} - \mu_{i,j}^{t}\right)^{+} \qquad i = 1 \dots N, j = 1 \dots K \tag{27}$$

It must be recalled that we assumed the CS to be an infinte reservoir of packets. The flow controller at the CS (to be described later in this section) specifies $a_{i,j}^{t}$, the number of packets that should flow into the queue at the MBS from the CS at each time slot t and $\mu_{i,j}^{t}$ is the number of packets departing the queue (based on the downlink SINR corresponding to the beamforming vector design). Let us subtract the target queue-length $\bar{q}_{i,j}$ from the LHS and the RHS of (27). We also add and subtract the target service rate of each queue, $\bar{\mu}_{i,j}$ on the RHS of the equation (27). Hence equation (27) after rearranging becomes

$$q_{i,j}^{t+1} - \bar{q}_{i,j} = \left([q_{i,j}^{t} - \bar{q}_{i,j}] + [a_{i,j}^{t} - \bar{\mu}_{i,j}] + [\bar{\mu}_{i,j} - \mu_{i,j}^{t}]\right)^{+} \tag{28}$$

We perform a change of variables and define $\psi_{i,j}^{t} = q_{i,j}^{t} - \bar{q}_{i,j}$ and the vectorized version by the notation $\boldsymbol{\psi}^{t} = [\psi_{1,1}^{t},\ \psi_{1,2}^{t}, \dots,\ \psi_{1,K}^{t}, \dots,\ \psi_{N,K}^{t}]^{T}$. We also define $u_{i,j}^{t} = a_{i,j}^{t} - \bar{\mu}_{i,j}$ and its vectorized notation by $\mathbf{u}^{t} = [u_{1,1}^{t},\ u_{1,2}^{t}, \dots,\ u_{1,K}^{t}, \dots,\ u_{N,K}^{t}]^{T}$. Similarly $\zeta^{t} = (\bar{\mu}_{i,j} - \mu_{i,j}^{t})$ and its vectorized notation $\boldsymbol{\zeta}^{t} = [\zeta_{1,1}^{t},\ \zeta_{1,2}^{t}, \dots,\ \zeta_{1,K}^{t}, \dots,\ \zeta_{N,K}^{t}]^{T}$. Therefore, Equation (28) in its vectorized form becomes

$$\boldsymbol{\psi}^{t+1} = \boldsymbol{\psi}^{t} + \mathbf{u}^{t} + \boldsymbol{\zeta}^{t} \tag{29}$$

Note that we have removed the $(.)^{+}$ operation assuming that the queue-length remains positive at all time instants. The CS makes an observation on the queue-length in order to regulate the traffic arrival into the queues. The observation equation is given by

$$\mathbf{y}^{t} = \boldsymbol{\psi}^{t} + \boldsymbol{\eta}^{t} \tag{30}$$

where $\boldsymbol{\eta}^{t}$ is a noise parameter. The noise parameter is representative of the fact that the CS only an estimation of queue-length. Moreover, in practical scenario this observation may even be an outdated observation. We model the

observation errors as noise. Note that we do not assume any particular statistical distribution for the noise parameter.

The objective of the scheduler design problem is to minimize the cost function given by

$$\mathbf{J} = \lim_{T\to\infty} \frac{1}{T}\big[\sum_{t=1}^{T} \big(||\boldsymbol{\psi}^t||^2_{\mathbf{W}} + ||\mathbf{u}^t||^2_{\mathbf{Q}}\big)\big] \tag{31}$$

where $\mathbf{W}$ and $\mathbf{Q}$ are weight matrices. In our algorithm, both the matrices are assumed to be identity matrices. The first term of the cost function in Equation (31) denotes the penalty for deviating from the target queue length. The second term tries to maintain the arrival rate at each queue as close as possible to the target service rate $\bar{\gamma}_{i,j}$. Note that the scheduler is unaware of the error term $\boldsymbol{\zeta}^t$, since it does not have the knowledge of the actual received SINR in the downlink.

Equation (32) represents a linear dynamic system with the state variable denoted by the notation $\boldsymbol{\psi}^t$, a control parameter denoted by the notation $\mathbf{u}^t$ and an unknown process noise denoted by $\boldsymbol{\zeta}^t$. In particular, Equation (32) denotes the evolution of the state variable and (31) the quadratic cost to be minimized. We use a controller based on H^∞ control algorithm [7] to solve the above problem. The task of the controller is to specify an arrival rate into the queue by calculating the control parameter $\mathbf{u}^t$ at each time slot t.

The reason why H^∞ filter makes controlling decisions without making any assumptions of the noise processes is because H^∞ control assumes the worst case noise. That is, it tries to minimize the cost function assuming the worst case noise. For this reason, the H^∞ control is called mini-max control. Additionally, the scheduler is trying to minimize the long term cost function. The problem is a dynamic programming problem involving stochastic processes i.e., during each time slot the CS has to take control decisions such that the long term cost function is minimized (and not the instantaneous cost function). For details of the H^∞ control design please refer to the Appendix A.

We would like to state that in practice, we can add a scaling factor α as a weight to the control parameter. We define a matrix $\mathbf{S} = \text{diag}(\frac{1}{\sqrt{\alpha}} \dots \frac{1}{\sqrt{\alpha}})$. The

equation defining the evolution of the state variable becomes

$$\boldsymbol{\psi}^{t+1} = \boldsymbol{\psi}^{t} + \mathbf{S}\mathbf{u}^{t} + \boldsymbol{\zeta}^{t} \tag{32}$$

And hence, the correspondingly, the weight matrix for the control variable becomes $\mathbf{V} = \text{diag}(\alpha \dots \alpha)$. The parameter α can be varied to adjust the convergence rate of the H^{∞} control.

5. LQG Control Vs H-Infinity

In this section, we compare the effectiveness of our H-infinity controller in minimizing the fluctuations of the queue-length as opposed to a standard LQG controller. It must be noted that standard LQG control assumes that the process noise is Gaussian in nature and minimizes the expected value of the cost function. In general, the nature of the fluctuations of the service rate to the queues around the target service rate would depend on the fluctuations of the wireless channel which is unknown to the CS. In this section, we assume a Gaussian approximation to model the fluctuations and perform simulations with standard LQG control (which is known to be optimal under Gaussian noise). We provide a brief description of a LQG control in Appendix B.

The simulation results (refer to section 6, Figure 7) show that the H-infinity flow controller performs better than the LQG controller in terms of keeping the fluctuations of the queue-length to minimum. The above observation shows that the Gaussian approximation is not a very accurate representation of the process noise. Hence, using a robust controller such as H^{∞} control keeps the queue-length fluctuations to a lower value and ensures stability.

6. Numerical Results

In this section we provide some numerical results to show the effectiveness of our algorithm. We consider a hexagonal cellular system with a cluster of 3 cells as shown in Figure 2. Each cell has a MBS which serves only the UTs in its cell. The number of transmit antennas on each MBS scales with the number

of UTs in the cell such that their ratio is a finite constant. In particular, we assume $N_t = K$. The MBS are connected to the CS. We assume that the target queue-length of the buffers at all the UTs to be 20kBs. We also assume the number of channel uses per coherence interval of the channel (B in equation (4)) as one-third the target queue-length. The channel realizations are assumed to be i.i.d. across the coherence intervals.

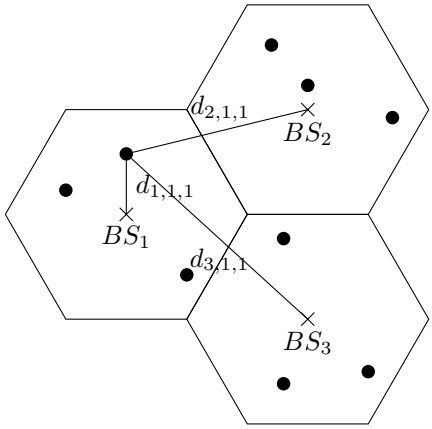

Figure 2: Example of a network with 3 Cells. The crosses represent the location of the BSs and the dots represent the location of the UTs randomly scattered inside the cells. The distances of a UT from the three BSs are also provided.

We consider a distance dependent path loss model in which the UTs are assumed to be arbitrarily scattered inside each cell. In this case, the path loss factor from MBS i to $\text{UT}_{j,k}$ is given by $\sigma^2_{i,j,k} = \left(\frac{1}{d_{i,j,k}}\right)^{\beta}$ where $d_{i,j,k}$ is the distance between MBS i to $\text{UT}_{j,k}$, normalized to the maximum distance within a cell, and β is the path loss exponent which lies usually in the range from 2 to 5 dependent on the radio environment. We normalize the variance of the total received noise to $\sigma^2 = 1$. We also assume that no user terminal is within a normalized distance of 0.1 from the closest BS.

We perform our simulations for the case of downlink power minimization (Section 3.1) and assume a target SINR of 9 dB for all the UTs during each time slot. First let us focus on the case with no max-power constraint on the

MBSs ($P_{max} = \infty$). We run our simulations for $T = 100$ time slots. We first plot the variation of queue-length values (averaged across the UTs) against the number of UTs per cell in Figure 3. The bubbles represent queue-length values at each time slot for a given number of UTs (overlapped on the same point) and the horizontal line represents the target queue-length.

In order to quantify the variation in the queue-lengths around the target queue size, we define the Normalized Mean Square Error (NMSE) of the queue-lengths given by

$$\text{NMSE} = \frac{1}{NKT} \sum_{t=1}^{T} \sum_{i,j} \frac{(q_{i,j}^t - \bar{q}_{i,j})^2}{\bar{q}_{i,j}}$$

The NMSE for the arrival rates also follows a similar definition. We plot the NMSE of the queue-lengths against the number of UTs in Figure 4.

It can be seen that in both the Figures 3 and 4 (in this case the curve corresponding to $P_{max} = \infty$), as the number of UTs increases, the variance of the queue-length becomes lesser. This is in accordance with our beamforming algorithm. Recall that as the number of UTs become large, the achieved SINR in the downlink gets very close to the target SINR for every given channel realization. Hence, the variation of the queue-length around the target also reduces.

We also plot the NMSE of the arrival rates into the queues specified by the scheduler in Figure 5. The NMSE of the arrival rates also reduce as the number of UTs increase. In fact, when the number of UTs goes to infinity, effectively, we do not require queue-length control anymore. The arrivals into the queue is then deterministic equal the the target queue-length at all the time slots.

We now introduce the max-power constraint on each MBS. When the max power constraint of a particular MBS is not satisfied, we use the approach developed in Section 3.1.1 to redesign the system parameters. In Figure 4 and 5 (curves corresponding to $P_{max} = 30, 40$), we plot the NMSE of the queue-lengths and the arrival rates with power constraints. Please note that P_{max} indicated in the plots is normalized by K, the number of UTs. It can be seen that as the P_{max} per MBS reduces, the NMSE parameter increases. This goes

well with the intuition since with the max-power constraint, the error between the target SINR and actual achieved SINR in the downlink increases because of redesigning of system parameters. The max-power plots will be useful in determining the excess buffer capacity beyond the target queue-length that the system designer will have to provide.

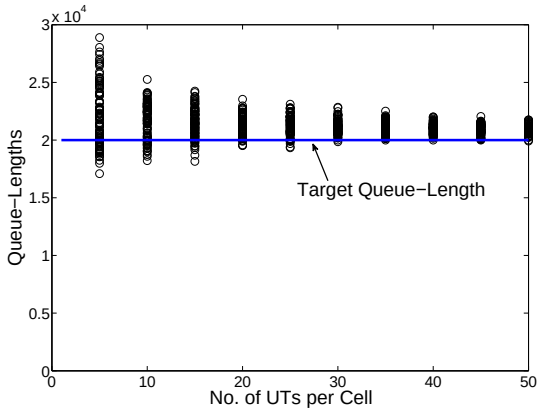

Figure 3: Scatter Plot showing the variation of queue-length around the target, Target SINR = 9dB, Target Buffer Length = 20kBs, $\beta = 3.6$, T =100

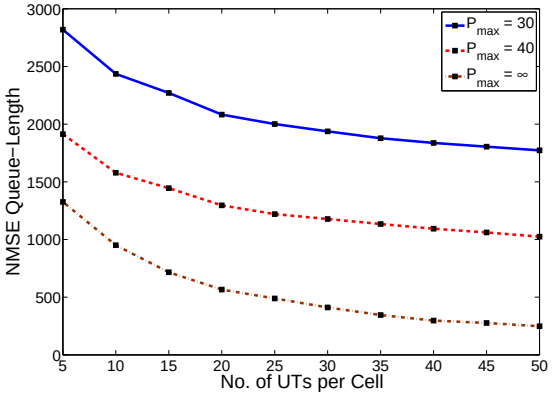

Figure 4: NMSE of Queue-Length Vs The No. of UTs. Target SINR = 9dB, Target Buffer Length = 20kBs, $\beta = 3.6$, T = 100

In Figure 6, we plot the average downlink transmitted power per UT and the

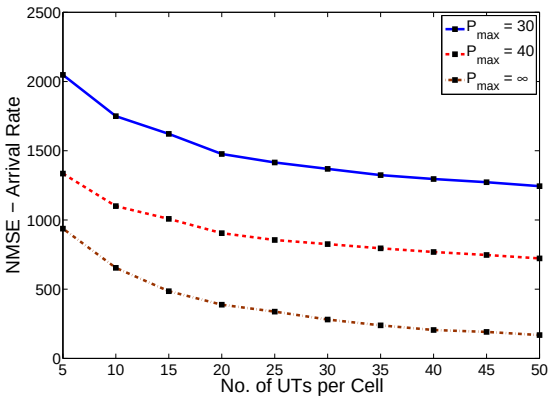

Figure 5: NMSE of Arrival Rates Vs The No. of UTs. Target SINR = 9dB, Target Buffer Length = 20kBs, $\beta = 3.6$, T = 100

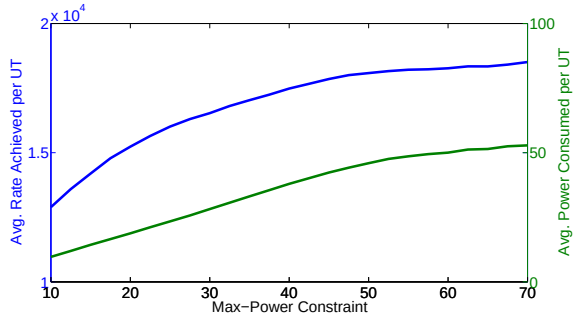

Figure 6: Mean Achieved Rate per UT (left y-axis) and Mean Power Consumed (Right y-axis) per UT Vs Max-Power of the BSs. Target SINR = 9dB, Target Buffer Length = 20kBs, $\beta = 3.6$, T =100, 10 UTs per cell

average rate achieved per UT against the given max-power constraint (P_{max}) in the same plot with 10 UTs per cell and 100 channel realizations. It can be seen that more the power consumed, more is the average rate achieved per UT . This implies that with more power, the MBSs can meet the targeted SINR without having to reset them to lower values. Hence, the error between requested SINR (target) and actual achieved SINR is lower. Also, for low levels of P_{max}, it can be seen that the increase in total downlink transmitted power is almost linear as a function of P_{max}. This is due to the fact that at low P_{max}, all the MBSs are transmitting nearly at their peak in order to provide the best possible data rates for the UTs (with the power constraint). However, as P_{max} increases beyond certain value, the system performance is close to the case of $P_{max} = \infty$. This is due to the fact that with higher powers, the targeted SINRs lie inside the capacity region.

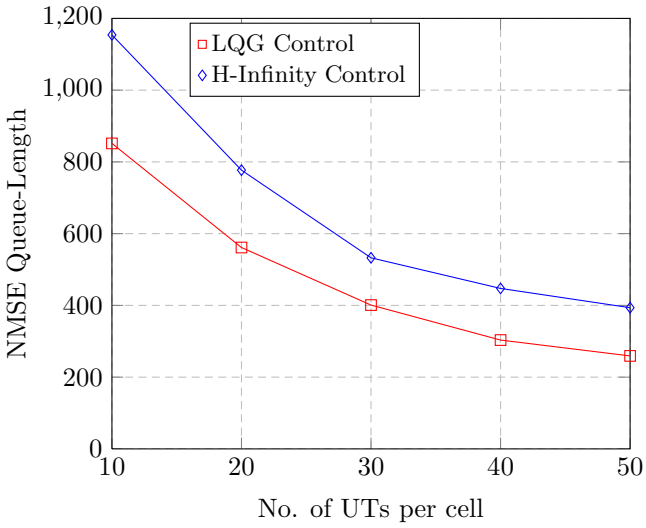

Figure 7: Comparison of H^{∞} control Vs LQG control, Target SINR = 9dB, Target Buffer Length = 20kBs, $\beta = 3.6$, T = 100

We finally plot the comparison of H^{∞} control in comparison with LQG control for the flow controller in Figure 7. The H-infinity controller provides better performance in terms of minimizing the fluctuations of the queue-length

as discusses in Section 5.

7. Conclusions

In this work, we addressed the joint problem of traffic scheduling using H^∞ control theory and decentralized beamforming design using tools from random matrix theory in the context of cellular networks. The task of the H^∞ based flow controller is to regulate the arrival process to the queues at the MBS from the CS. The H^∞ based flow controller stabilizes the queue-lengths at the MBS without the knowledge of the wireless channel between the MBS and the UTs. The H^∞ controller keeps the fluctuations around the target queue-length to a minimum as compared to LQG based control.

In this work we decoupled the two problems of traffic scheduling and beamforming design. A natural extension of this work would be to jointly address the two problems of flow control and beamforming design and develop a cross-layer model to address the same.

Appendix A: H-Infinity Control

In this section, we provide a brief overview on the solution to the discrete-time system dynamic with unpredictable noise. (Please note that in order to maintain consistency with the previous literature in this field, we deviate from the guidelines set for notations in the introduction section this paper). We consider the following state space equation:

$$\mathbf{x}^{t+1} = \mathbf{A}\mathbf{x}^t + \mathbf{B}\mathbf{u}^t + \mathbf{D}_1\Gamma^t \tag{33}$$

where state vector $\mathbf{X}$, control vector $\mathbf{U}$ and noise vector Γ take values respectively in $\mathbb{R}^n$, $\mathbb{R}^p$ and $\mathbb{R}^m$. Moreover, we define the following measurement equation for the state (i.e. the state is not known exactly but estimated via an observation),

$$\mathbf{y}^t = \mathbf{C}\mathbf{x}^t + \mathbf{D}_2\Gamma^t \tag{34}$$

observable state vector Y takes values in $\mathbb{R}^n$. The objective is to design a controller that minimizes the following cost function:

$$\overline{\mathbf{L}} = \lim_{T\to\infty} \Big\{ \frac{1}{T} \sum_{t=0}^{T} \Big(||\mathbf{x}^t||^2_{\mathbf{W}} + ||\mathbf{u}^t||^2_{\mathbf{Q}} \Big) \Big\} \tag{35}$$

where $\mathbf{W}$ and $\mathbf{Q}$ are the norm matrices. The problem to solve is then a linear control problem with quadratic cost and unpredictable noise. Please note that when the noise is Gaussian, the solution to this problem is to use LQG (Linear Quadratic Gaussian) technique. When the noise is unpredictable (we don't have information on the distribution of the noise), an efficient way to solve the aforementioned problem is to use H^∞ technique. The H^∞ controller adopted in this section is the robust controller discussed in [7] and based on zero sum game theory. To solve our problem using H^∞, we introduce the following cost function:

$$J_\pi = \lim_{T\to\infty} \Big\{ \frac{1}{T} \sum_{t=0}^{T} \Big(||\mathbf{x}^t||^2_{\mathbf{w}} + ||\mathbf{u}^t||^2_{\mathbf{Q}} - \pi^2 ||\Gamma^t||^2 \Big) \Big\} \tag{36}$$

where, π is the level of attenuation. This is a minimax optimization problem where the cost function is minimized over maximum unknown disturbance. This problem is a zero sum game with two players. The cost J_π is minimized by player 1 and maximized by player 2 using vectors U and Γ. One can refer to [7] for more general class of discrete-time zero-sum games, with various information patterns, where sufficient conditions for the existence of a saddle point are provided when the information pattern is perfect and imperfect state.

The H^∞ controller can be obtained according to the following proposition [7];

There exists a state feedback H^∞ controller $\mathbf{U}^t$ such that,

$$\mathbf{u}^t = \mathbf{M}(\mathbf{I} + (1 - \frac{1}{\pi^2})\Sigma)^{-1} \hat{\mathbf{x}}^t \tag{37}$$

where, $\overline{\rho}(\Sigma\mathbf{M}) < \pi^2$. Moreover,

1) $\mathbf{M}$ is a minimal non-negative definite solution to the following algebraic Riccati equation,

$$\mathbf{M} = \mathbf{I} + \mathbf{M}(\mathbf{I} + (1 - \frac{1}{\pi^2})\mathbf{M})^{-1} \tag{38}$$

where,

$$\pi^2\mathbf{I} - \mathbf{M} > 0 \tag{39}$$

2) Σ is a minimal non-negative definite solution to the following algebraic Riccati equation,

$$\Sigma = \mathbf{I} + \Sigma(\mathbf{I} + (1 - \frac{1}{\pi^2})\Sigma)^{-1} \tag{40}$$

where,

$$\pi^2\mathbf{I} - \Sigma > 0 \tag{41}$$

3) and the state estimate $\hat{\mathbf{X}}^t$ is generated by,

$$\hat{\mathbf{x}}^t = (\mathbf{I} + \Sigma - \pi^{-2}\Sigma\mathbf{M})^{-1}(\tilde{\mathbf{x}}^t + \Sigma\mathbf{y}^t) \tag{42}$$

and

$$\tilde{\mathbf{x}}^{t+1} = \tilde{\mathbf{x}}^t - \mathbf{u}^t + \Sigma(\mathbf{I} + (1 - \frac{1}{\pi^2})\Sigma)^{-1}(\mathbf{y}^t - (1 - \frac{1}{\pi^2})\tilde{\mathbf{x}}^t) \tag{43}$$

The above proposition allows us to find the optimal controller for given π. Therefore, to obtain the solution, we have to find the optimal value of π^2 which satisfies all the above conditions. For that, we use the following lemma: We use the following algorithm to find the optimal π [[12], [7], [22] [23]].

1. Start with a small value of $\pi^2 \geq 0$.
2. Increment the value of π^2 by a small step size δ.
3. Formulate the two Hamiltonian matrices corresponding to the Ricatti Equations
$$\mathbf{H}_1 = \mathbf{H}_2 = \begin{bmatrix} \mathbf{I} & -\big(1 - \frac{1}{\pi^2}\big)\mathbf{I} \\ -\mathbf{I} & -\mathbf{I} \end{bmatrix}$$
For our particular problem, the two Hamiltonian matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ are identical.
4. Find the eigen values of $\mathbf{H}_1$ and $\mathbf{H}_2$.
5. If the eigen value of $\mathbf{H}_1$ or $\mathbf{H}_2$ has an imaginary part, then go to step 2. ELSE for this value of π^2 solve the Ricatti Equations in (38) and (40) to find the values of $\mathbf{M}$ and $\mathbf{\Sigma}$.
6. If $\mathbf{M} < 0$ or $\mathbf{\Sigma} < 0$, go to step 2 ELSE check conditions (39) and (41).

7. If conditions (39) and (41) are not satisfied, go to step 2. Else check the spectral radius of $\mathbf{\Sigma M}$.
8. If $\rho(\mathbf{\Sigma M}) < \pi^2$, go to step 2, ELSE π^2 is the optimal value and terminate the algorithm.

Appendix B: LQG Control

We now provide a brief description of standard LQG control algorithm. Consider the following state space equation and the observation equations given by

$$\mathbf{x}^{t+1} = \mathbf{A}\mathbf{x}^t + \mathbf{B}\mathbf{u}^t + \mathbf{w}^t$$
$$\mathbf{y}^t = \mathbf{C}\mathbf{x}^t + \mathbf{v}^t$$

where $\mathbf{x}^t$ is the state variable, $\mathbf{u}^t$ the control and $\mathbf{w}^t$ and $\mathbf{v}^t$ being the process and the observation noise respectively. The sequences $\mathbf{v}^t$ and $\mathbf{w}^t$ are Gaussian distributed with zero mean and covariance matrix given by

$$\mathbf{E}\left[\mathbf{w}^{t^T}\mathbf{w}^s\right] = \mathbf{R}_1\boldsymbol{\delta}^{t,s}$$
$$\mathbf{E}\left[\mathbf{w}^{t^T}\mathbf{v}^s\right] = \mathbf{R}_2\boldsymbol{\delta}^{t,s}$$
$$\mathbf{E}\left[\mathbf{w}^{t^T}\mathbf{v}^s\right] = 0$$

where $\boldsymbol{\delta}^{t,s}$ are Dirac functions. The cost function to be minimized is given by

$$J = \lim_{T\to\infty}\frac{1}{T}\mathbf{E}\left[\sum_{t=0}^{T-1}||\mathbf{x}^t||^2_{\mathbf{W}} + ||\mathbf{u}^t||^2_{\mathbf{Q}}\right]$$

The LQG control with imperfect state observations is summarized as follows [24]. Starting with the initial state estimate $\hat{\mathbf{x}}^t\big|_{t=0} = 0$, solve the following Ricatti equations given by

$$\mathbf{S} = \mathbf{A}^T\mathbf{S}\mathbf{A} - \mathbf{A}^T\mathbf{S}\mathbf{B}\left(\mathbf{B}^T\mathbf{S}\mathbf{B} + \mathbf{Q}\right)^{-1}\mathbf{B}^T\mathbf{S}\mathbf{A} + \mathbf{W}$$
$$\mathbf{P} = \mathbf{A}(\mathbf{P} - \mathbf{P}\mathbf{C}(\mathbf{C}\mathbf{P}\mathbf{C}^T + \mathbf{R}_2)^{-1}\mathbf{C}\mathbf{P})\mathbf{A}^T + \mathbf{R}_1$$

The optimal control strategy which is a linear control law in the state estimate is given by

$$\mathbf{u}^t = -\mathbf{L}\hat{\mathbf{x}}^t$$

where the matrix $\mathbf{L}$ is given by

$$\mathbf{L} = \left(\mathbf{B}^T\mathbf{S}\mathbf{B} + \mathbf{Q}\right)^{-1}\mathbf{B}^T\mathbf{S}\mathbf{A}$$

The state estimate at time $t+1$ is given as a function of the state estimate at t and observation $\mathbf{y}^t$ is given according to the following equation

$$\hat{\mathbf{x}}^{t+1} = \mathbf{A}\hat{\mathbf{x}}^t + \mathbf{B}\mathbf{u}^t + \mathbf{K}(\mathbf{y}^t - \mathbf{C}\hat{\mathbf{x}}^t)$$

where the matrix $\mathbf{K}$ is given as

$$\mathbf{K} = \mathbf{A}\mathbf{P}\mathbf{C}^T(\mathbf{C}\mathbf{P}\mathbf{C}^T + \mathbf{R}_2)^{-1}$$

[1] M.-S. Alouini and A. Goldsmith, "Area spectral efficiency of cellular mobile radio systems," *IEEE Transactions on Vehicular Technology*, vol. 48, pp. 1047 –66, 1999.

[2] J. Hoydis, M. Kobayashi, and M. Debbah, "Green Small-Cell Networks," *IEEE Vehicular Technology Magazine*, vol. 6, no. 1, pp. 37 –43, march 2011.

[3] [Online]. Available: http://ecoscells.twn-bl.homeip.net/

[4] [Online]. Available: http://www.willcom-inc.com/en/service/why/index.html

[5] S. Lakshminarayana, J. Hoydis, M. Debbah, and M. Assaad, "Asymptotic Analysis of Distributed Multi-cell Beamforming," in *PIMRC, Istanbul, Turkey*, Sep, 2010.

[6] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge University Press, 2011.

[7] T. Basar and P. Bernhard, *H-∞ Optimal Control and Relaxed Minimax Design Problems: A Dynamic Game Approach*. Birkhauser, Boston, MA, 1991.

[8] A. Eryilmaz and R. Srikant, "Fair Resource Allocation in Wireless Networks using Queue-Length-based Scheduling and Congestion Control," in

24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE INFOCOM 2005, vol. 3, march 2005, pp. 1794 – 1803 vol. 3.

[9] M. Neely, E. Modiano, and C.-P. L. ;., "Fairness and Optimal Stochastic Control for Heterogeneous Networks," in *24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE INFOCOM 2005*, vol. 3, march 2005, pp. 1723 – 1734 vol. 3.

[10] X. Lin, N. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1452 –1463, aug. 2006.

[11] E. Altman and T. Basar, "Optimal rate control for high speed telecommunication networks," in *Proceedings of the 34th IEEE Conference on Decision and Control, 1995.*, vol. 2, dec 1995, pp. 1389 –1394 vol.2.

[12] N. Hassan and M. Assaad, "Dynamic Resource Allocation in Multi-service OFDMA Systems with Dynamic Queue Control," *IEEE Transactions on Communications*, vol. 59, no. 6, pp. 1664–1674, june 2011.

[13] E. Björnson, R. Zakhour, D. Gesbert, and B. Ottersten, "Cooperative Multicell Precoding: Rate Region Characterization and Distributed Strategies With Instantaneous and Statistical CSI," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4298 –4310, aug. 2010.

[14] B. L. Ng, J. S. Evans, S. V. Hanly, and D. Aktas, "Distributed Downlink Beamforming With Cooperative Base Stations," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5491–5499, 2008.

[15] R. Zhang and S. Cui, "Cooperative Interference Management With MISO Beamforming," *IEEE Transactions on Signal Processing,*, vol. 58, no. 10, 2010.

[16] J. Qiu, R. Zhang, Z.-Q. Luo, and S. Cui, "Optimal Distributed Beamforming for MISO Interference Channels," *IEEE Transactions on Signal Processing*, 2010, submitted. [Online]. Available: http://arxiv.org/abs/1011.3890

[17] H. Dahrouj and W. Yu, "Coordinated beamforming for the multi-cell multi-antenna wireless system," in *42nd Annual Conference on Information Sciences and Systems, CISS 2008, Princeton, NJ, USA, 19-21.*

[18] S. Shi, M. Schubert, and H. Boche, "Rate Optimization for Multiuser MIMO Systems With Linear Processing," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 4020 –4030, aug. 2008.

[19] S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, "Weighted Sum-rate Maximization using Weighted MMSE for MIMO-BC Beamforming Design," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4792 –4799, december 2008.

[20] M. Chiang, *Geometric programming for communication systems.* Foundations and Trends in Communications and Information Theory, vol. 2, no. 1-2, pp. 1-154, August 2005.

[21] S.Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, 2004.

[22] B. Anderson, "An algebraic solution to the spectral factorization problem," *IEEE Transactions on Automatic Control*, vol. 12, no. 4, aug 1967.

[23] S. Boyd, V. Balakrishnan, and P. Kabamba, "On Computing the H Infinity Norm of a Transfer Matrix," *Math. Contr. Signals Syst.*, 1988.

[24] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation.* Englewood Cliffs, NJ: Prentice Hall, 2000.