Joint UL/DL Mode Selection and Transceiver Design for Dynamic TDD Systems

Antti Tölli, Jarkko Kaleva, Ganesh Venkatraman
Center for Wireless Communications (CWC),
P.O. Box 4500, FI-90014 University of Oulu, Finland
Email: firstname.lastname@ee.oulu.fi

David Gesbert
EURECOM, Campus SophiaTech,
450 Route des Chappes, 06410 Biot Sophia Antipolis, France
Email: david.gesbert@eurecom.fr

***Abstract*—Weighted queue minimization is considered in Dynamic TDD system where the available resources per cell and per TDD frame can be freely allocated to both uplink (UL) and downlink (DL) transmission depending on traffic demand and user distribution within the network. Given the traffic state, joint UL/DL cell mode selection, beamformer design and resource allocation among available time/frequency/space dimensions is carried out. A centralized solution is found by dividing the combinatorial and non-convex problem into subproblems in which relaxed versions of the original intractable optimization problem are formulated and solved in an efficient manner. The numerical examples demonstrate significant gains from dynamic UL/DL mode selection as compared to the conventional TDD with fixed mode assignment.**

I. Introduction

In small cell scenarios, the amount of instantaneous uplink (UL) and downlink (DL) traffic may vary significantly with time and among the adjacent cells [1]. In such cases, *Dynamic TDD* allows for resources to be dynamically adapted between the UL and DL by dynamically changing the amount of resources allocated to each direction at each time instant, providing vastly improved overall resource utilization.

The dynamic allocation variation will change, however, the interference seen by neighboring cells, considerably complicating the interference management. In addition to the normal UL-to-UL and DL-to-DL interference, there are two additional types of interference associated with Dynamic TDD in multi-cell operation, namely UL-to-DL and DL-to-UL interference. Allocating resources in such a system requires balancing the gains of flexible DL/UL allocation (i.e. in the form of reduced packet delivery delay) with the potential losses (in the form of excess interference). In order to mitigate or avoid the cross-link interference, the system may employ coordinated resource allocation and beamforming, where the transmissions within a coordinating set of cells are jointly designed.

Numerous studies have been carried out in the literature that mostly consider time-slot allocation algorithms to minimize the cross-link interference in Dynamic TDD, see e.g., [2]–[4], and the references therein. Fully dynamic or flexible TDD is an essential 5G research item, e.g., in 3GPP standardization [8], [9]. With some limitations in adapting the DL/UL allocation, a form of dynamic TDD named as "enhanced Interference Mitigation and Traffic Adaptation" (eIMTA) has been already introduced 3GPP in Long Term Evolution Advanced (LTE-A) [5]. However, such evolutions of LTE-A will not trigger major latency reductions due to the restrictions to ensure backwards compatibility. Consequently, there is a need for a new 5G air interface to enable the required low physical layer latencies. A popular candidate for frame structure is one without any switching point restrictions so that any slot can be either uplink or downlink, and in addition, serve a direct device-to-device link or provide self-backhauling [5]–[7].

In this paper, a fully dynamic traffic aware TDD transmission and reception is considered in multi-cell network, where the load variation between adjacent small cells can be significant. Different traffic types are assumed, for example, data traffic consisting of short intermittent bursts and backlogged traffic corresponding to large file transfers (e.g., video). A standard approach in such settings is to consider a myopic optimization problem in each time-slot, where the objective varies from slot-to-slot with the queue-sizes [10]. In [11], a weighted queue deviation minimization was considered for designing transmit precoders and receive beamformers in a conventional synchronous TDD multiple-input multiple-output (MIMO)-orthogonal frequency division multiplexing (OFDM) multi-cell network to minimize the number of backlogged packets at the base stations (BSs).

The weighted queue minimization approach from [11] is extended in this paper to the Dynamic TDD setting where the available resources per cell and per TDD frame can be freely allocated to both UL and DL transmission depending on given traffic demand and user distribution within the network. Given the traffic state, the design problem is to jointly assign UL/DL mode selection per cell, resource allocation among available time/frequency/space dimensions and design beamformers for each data stream. This highly challenging problem is solved by dividing it into subproblems in which relaxed versions of the optimization problems are formulated, solved, and related to the original problem, e.g., integer valued UL/DL mode selection variables are replaced by continuous variables while a binary outcome is enforced using an appropriate regularization function in the objective.

II. System Model

We consider a dynamic TDD scenario for a MIMO–OFDM transmission with N sub-channels and N_B half-duplex BSs each equipped with N_T transmit antennas, serving in total K half-duplex users each with N_R receive antennas. The set of users associated with BS b is denoted by $\mathcal{U}_b$ and the set $\mathcal{U}$ represents all users in the system, i.e., $\mathcal{U} = \cup_{b\in\mathcal{B}}\mathcal{U}_b$, where $\mathcal{B}$ is the set of indices of all coordinating BSs. Data for user k is transmitted from only one BS which is denoted by $b_k \in \mathcal{B}$. Furthermore, $\mathcal{U}^{\mathrm{UL}}$ and $\mathcal{U}^{\mathrm{DL}}$ denote the set of uplink and downlink users active at given time instant, respectively. Let $\mathcal{N} = \{1, 2, \ldots, N\}$ be the set of all sub-channel indices available in the system. In a given time, a subset of base stations $\mathcal{B}_U \subseteq \mathcal{B}$ serves the uplink traffic in all sub-channels simultaneously and the rest of the base stations $\mathcal{B}_D \subseteq \mathcal{B}$ are serving the DL traffic.

We adopt linear transmit beamforming technique at BSs and UTs. Specifically for DL case, the data symbols $d_{l,k,n}$ for user k on the lth spatial stream over sub-channel n is multiplied with beamformer $\mathbf{m}_{l,k,n} \in \mathbb{C}^{N_T\times 1}$ before being transmitted. In order to detect multiple spatial streams at the user terminal, receive beamforming vector $\mathbf{w}_{l,k,n}$ is employed for each user. Similarly for uplink case, at the user terminal the data symbols $\bar{d}_{l,k,n}$ for user k is multiplied with beamformer $\bar{\mathbf{m}}_{l,k,n} \in \mathbb{C}^{N_R \times 1}$ before being transmitted. Then receive beamforming vector $\bar{\mathbf{w}}_{l,k,n}$ is employed at the BS. Consequently, the received data estimate corresponding to the lth spatial stream over sub-channel n at DL user k is given by

$$\begin{aligned} x_{l,k,n} = & \mathbf{w}_{l,k,n}^{\mathrm{H}} \mathbf{H}_{b_k,k,n} \mathbf{m}_{l,k,n} d_{l,k,n} \\ & + \mathbf{w}_{l,k,n}^{\mathrm{H}} \Big(\sum_{i \in \mathcal{U} \setminus \{k\}} \mathbf{H}_{b_i,k,n} \sum_{j=1}^{L} \mathbf{m}_{j,i,n} d_{j,i,n} \\ & + \sum_{i \in \mathcal{U} \setminus \mathcal{U}_{b_k}} \tilde{\mathbf{H}}_{i,k,n} \sum_{j=1}^{L} \bar{\mathbf{m}}_{j,i,n} \bar{d}_{j,i,n} \Big) + \mathbf{w}_{l,k,n}^{\mathrm{H}} \mathbf{n}_{l,k,n} \end{aligned} \tag{1}$$

where $\mathbf{H}_{b,k,n} \in \mathbb{C}^{N_R \times N_T}$ is the channel between BS b and user k on sub-channel n and $\tilde{\mathbf{H}}_{i,k,n} \in \mathbb{C}^{N_R \times N_R}$ is the UT-UT interference channel between users i and k on sub-channel n. $L = \min(N_T, N_R)$ is the maximum number of spatial streams per user. Consequently, the received data estimate corresponding to the lth spatial stream over sub-channel n for user k at the serving UL BS b_k is given by

$$\begin{aligned} \bar{x}_{l,k,n} = & \bar{\mathbf{w}}_{l,k,n}^{\mathrm{H}} \mathbf{H}_{b_k,k,n}^{\mathrm{T}} \bar{\mathbf{m}}_{l,k,n} \bar{d}_{l,k,n} \\ & + \bar{\mathbf{w}}_{l,k,n}^{\mathrm{H}} \Big(\sum_{i \in \mathcal{U} \setminus \{k\}} \mathbf{H}_{b_k,i,n}^{\mathrm{T}} \sum_{j=1}^{L} \bar{\mathbf{m}}_{j,i,n} \bar{d}_{j,i,n} \\ & + \sum_{i \in \mathcal{U} \setminus \mathcal{U}_{b_k}} \hat{\mathbf{H}}_{b_i,b_k,n} \sum_{j=1}^{L} \mathbf{m}_{j,i,n} d_{j,i,n} \Big) + \bar{\mathbf{w}}_{l,k,n}^{\mathrm{H}} \bar{\mathbf{n}}_{l,k,n} \end{aligned} \tag{2}$$

where $\hat{\mathbf{H}}_{b_i,b_k,n} \in \mathbb{C}^{N_T \times N_T}$ is the BS-BS interference channel, and $\bar{\mathbf{n}}_{l,k,n} \sim \mathcal{CN}(0, N_0)$ is the additive noise vector for user k on the nth sub-channel and lth spatial stream. Assuming independent detection of data streams, we can write the signal-to-interference-plus-noise ratio (SINR) of DL user as in (3) where $\widetilde{N}_0 = N_0 \operatorname{tr}(\mathbf{w}_{l,k,n} \mathbf{w}_{l,k,n}^{\mathrm{H}})$ denotes the equivalent noise variance. Similarly, SINR of UL user can be written as in (4).[1]

Let Q_k and $\bar{Q}_k$ be the number of backlogged DL and UL packets destined for user k at a given scheduling instant, respectively. The queue dynamics of user k are modeled using the Poisson arrival process with the average number of packet arrivals of $A_k = \mathbf{E}_i\{\lambda_k\}$, $\bar{A}_k = \mathbf{E}_i\{\bar{\lambda}_k\}$ packets or bits, where $\lambda_k(i) \sim \mathrm{Pois}(A_k)$, $\bar{\lambda}_k(i) \sim \mathrm{Pois}(\bar{A}_k)$ represents the instantaneous number of DL and UL packets arriving for user k at the ith time instant.[2] For example, the total number of downlink (similarly in uplink) queued packets at the $(i+1)$th instant for user k is given by

$$Q_k(i+1) = \Big[Q_k(i) - t_k(i) \Big]^+ + \lambda_k(i) \tag{5}$$

where $[x]^+ \equiv \max\{x, 0\}$ and t_k denotes the number of transmitted DL ($\bar{t}_k(i)$ in UL) packets or bits for user k. At the ith instant, the transmission rate of user k is given by

$$t_k(i) = \sum_{n=1}^{N} \sum_{l=1}^{L} t_{l,k,n}(i) \tag{6}$$

where $t_{l,k,n}$ denotes the number of transmitted packets over the lth spatial stream on the nth sub-channel. The maximum rate achieved over the space-frequency resource (l, n) for DL and UL is given by $t_{l,k,n} \le \log_2(1 + \gamma_{l,k,n})$ and $\bar{t}_{l,k,n} \le \log_2(1 + \bar{\gamma}_{l,k,n})$, respectively. Note that t_k and Q_k are represented by the same units, *i.e.*, in bits defined per channel use.

III. Problem Formulation

In order to minimize the total number of backlogged packets, we define an objective $\sum_{k \in \mathcal{U}} \alpha_k |v_k|^q + \beta_k |u_k|^q$ where queue deviation metrics are defined as

$$v_k = Q_k - t_k = Q_k - \sum_{n=1}^{N} \sum_{l=1}^{L} \log_2(1 + \gamma_{l,k,n}) \tag{7}$$

$$u_k = \bar{Q}_k - \bar{t}_k = \bar{Q}_k - \sum_{n=1}^{N} \sum_{l=1}^{L} \log_2(1 + \bar{\gamma}_{l,k,n}) \tag{8}$$

and where $\gamma_{l,k,n}$, $\bar{\gamma}_{l,k,n}$ are given by (3), (4), respectively, and the optimization variables are transmit precoders $\mathbf{m}_{l,k,n}$, $\bar{\mathbf{m}}_{l,k,n}$ and receivers $\mathbf{w}_{l,k,n}$,$\bar{\mathbf{w}}_{l,k,n}$.

Since we assume all the network nodes to be half duplex, any cell can be either in UL or DL mode at any given time instant. New BS specific binary variables $x_b \in \{0,1\}$ and $\bar{x}_b \in \{0,1\}$ are introduced to indicate whether cell b is in downlink or uplink mode, respectively, such that $x_b + \bar{x}_b = 1 \; \forall \; b$. By relaxing the SINR expressions as inequality constraints and introducing new variables $\gamma_{l,k,n}$ and $\bar{\gamma}_{l,k,n}$, the joint UL/DL mode selection and traffic aware beamforming problem formulation becomes

$$\min. \quad \|\tilde{\mathbf{v}}\|_q + \|\tilde{\mathbf{u}}\|_q \tag{9a}$$

$$\text{s. t.} \quad \gamma_{l,k,n} \le \frac{\left| \mathbf{w}_{l,k,n}^{\mathrm{H}} \mathbf{H}_{b_k,k,n} \mathbf{m}_{l,k,n} \right|^2}{\beta_{l,k,n}} \tag{9b}$$

$$\begin{aligned} \beta_{l,k,n} \ge \widetilde{N}_0 & + \sum_{i \in \mathcal{U} \setminus \{k\}} \sum_{j=1}^{L} |\mathbf{w}_{l,k,n}^{\mathrm{H}} \mathbf{H}_{b_i,k,n} \mathbf{m}_{j,i,n}|^2 \\ & + \sum_{i \in \mathcal{U} \setminus \mathcal{U}_{b_k}} \sum_{j=1}^{L} |\mathbf{w}_{l,k,n}^{\mathrm{H}} \tilde{\mathbf{H}}_{i,k,n} \bar{\mathbf{m}}_{j,i,n}|^2 \end{aligned} \tag{9c}$$

$$\bar{\gamma}_{l,k,n} \le \frac{\left| \bar{\mathbf{w}}_{l,k,n}^{\mathrm{H}} \mathbf{H}_{b_k,k,n}^{\mathrm{T}} \bar{\mathbf{m}}_{l,k,n} \right|^2}{\bar{\beta}_{l,k,n}} \tag{9d}$$

$$\begin{aligned} \bar{\beta}_{l,k,n} \ge \widetilde{N}_0 & + \sum_{i \in \mathcal{U} \setminus \{k\}} \sum_{j=1}^{L} |\bar{\mathbf{w}}_{l,k,n}^{\mathrm{H}} \mathbf{H}_{b_k,i,n}^{\mathrm{T}} \bar{\mathbf{m}}_{j,i,n}|^2 \\ & + \sum_{i \in \mathcal{U} \setminus \mathcal{U}_{b_k}} \sum_{j=1}^{L} |\bar{\mathbf{w}}_{l,k,n}^{\mathrm{H}} \hat{\mathbf{H}}_{b_i,b_k,n} \mathbf{m}_{j,i,n}|^2 \end{aligned} \tag{9e}$$

$$\sum_{n=1}^{N} \sum_{k \in \mathcal{U}_b} \sum_{l=1}^{L} \|\mathbf{m}_{l,k,n}\|^2 \le x_b P_{\max} \; \forall \; b \tag{9f}$$

$$\sum_{n=1}^{N} \sum_{l=1}^{L} \|\bar{\mathbf{m}}_{l,k,n}\|^2 \le \bar{x}_{b_k} P_{\max}^{\mathrm{UE}} \; \forall \; k \tag{9g}$$

$$x_b + \bar{x}_b = 1 \; \forall \; b, \quad x_b \in \{0,1\}, \quad \bar{x}_b \in \{0,1\} \tag{9h}$$

where the optimization variables are $\mathbf{m}_{l,k,n}, \mathbf{w}_{l,k,n}, \gamma_{l,k,n}, \beta_{l,k,n}, \bar{\mathbf{m}}_{l,k,n}, \bar{\mathbf{w}}_{l,k,n}, \bar{\gamma}_{l,k,n}, \bar{\beta}_{l,k,n} \; \forall \; l,k,n$ and $x_b, \bar{x}_b \; \forall \; b$. $\tilde{v}_k \triangleq \alpha_k^{1/q} v_k$ and $\tilde{u}_k \triangleq \beta_k^{1/q} v_k$ are the element of vectors $\tilde{\mathbf{v}}$ and $\tilde{\mathbf{u}}$, respectively, and α_k and β_k are the weighting factor used to alter the user priority based on the quality of service (QoS) constraints such as packet delay requirements and packet waiting time. The BS specific power constraints are considered in (21e) while the UL user power constraints are considered in (21g). The UL/DL mode selection is handled via multiplying the power constraints by binary mode selection variables, and hence, allowing non-zero DL (UL) beamformers

[1] Note that UL-DL and DL-UL interference terms in (1) and (3), and (2) and (4), respectively, include potential interference from all other-cell users. UL/DL mode selection per BS/user is handled separately via binary variables.

[2] The unit can either be packets or bits as long as the arrival and the transmission units are similar.

$$\gamma_{l,k,n} = \frac{\left|\mathbf{w}_{l,k,n}^{\mathrm{H}}\,\mathbf{H}_{b_k,k,n}\,\mathbf{m}_{l,k,n}\right|^2}{\widetilde{N}_0 + \sum\limits_{i\in\mathcal{U}\setminus\{k\}}\sum\limits_{j=1}^{L}|\mathbf{w}_{l,k,n}^{\mathrm{H}}\mathbf{H}_{b_i,k,n}\mathbf{m}_{j,i,n}|^2 + \sum\limits_{i\in\mathcal{U}\setminus\mathcal{U}_{b_k}}\sum\limits_{j=1}^{L}|\mathbf{w}_{l,k,n}^{\mathrm{H}}\tilde{\mathbf{H}}_{i,k,n}\bar{\mathbf{m}}_{j,i,n}|^2} \tag{3}$$

$$\bar{\gamma}_{l,k,n} = \frac{\left|\bar{\mathbf{w}}_{l,k,n}^{\mathrm{H}}\,\mathbf{H}_{b_k,k,n}^{\mathrm{T}}\,\bar{\mathbf{m}}_{l,k,n}\right|^2}{\widetilde{N}_0 + \sum\limits_{i\in\mathcal{U}\setminus\{k\}}\sum\limits_{j=1}^{L}|\bar{\mathbf{w}}_{l,k,n}^{\mathrm{H}}\mathbf{H}_{b_k,i,n}^{\mathrm{T}}\bar{\mathbf{m}}_{j,i,n}|^2 + \sum\limits_{i\in\mathcal{U}\setminus\mathcal{U}_{b_k}}\sum\limits_{j=1}^{L}|\bar{\mathbf{w}}_{l,k,n}^{\mathrm{H}}\hat{\mathbf{H}}_{b_i,b_k,n}\mathbf{m}_{j,i,n}|^2} \tag{4}$$

only in DL (UL) cells. Note that (9a) includes an implicit rate constraint $Q_k - \sum_{n=1}^{N}\sum_{l=1}^{L} t_{l,k,n} \geq 0$, $\bar{Q}_k - \sum_{n=1}^{N}\sum_{l=1}^{L} \bar{t}_{l,k,n} \geq 0$ on the maximum number of transmitted bits for each user as governed by the number of backlogged packets. The parameter q is used to provide tradeoff between fairness and sum queue minimization [11].

The optimization problem in (9) is non-convex, combinatorial problem with binary variables. The objective is convex and power constraints are convex. The interference constraints (9c) and (9e) are convex for fixed receivers. Similarly, the SINR constraints can be formulated as difference of convex functions as, for fixed receivers, the RHS of (9b) and (9d) are convex (quadratic over linear) functions. Note that for fixed mode selection variables, the problem is a straightforward extension of [11].

A. Approximation of the SINR constraints

The SINR expressions (9b) and (9d) can be handled as in [11]. The main steps are briefly reproduced here for convenience. Now, by fixing $\mathbf{w}_{l,k,n}, \bar{\mathbf{w}}_{l,k,n}$, (9c), (9e) can be written as a second-order cone (SOC) constraint. Thus, the difficulty is due to the nonconvexity of the constraint in (9b) and (9d). Let

$$g(u_{l,k,n}) \triangleq \frac{\left|\mathbf{w}_{l,k,n}^{\mathrm{H}}\,\mathbf{H}_{b_k,k,n}\,\mathbf{m}_{l,k,n}\right|^2}{\beta_{l,k,n}} \tag{10}$$

be the r.h.s of (9b), where $u_{l,k,n} \triangleq \{\mathbf{w}_{l,k,n}, \mathbf{m}_{l,k,n}, \beta_{l,k,n}\}$. Note that the function $g(u_{l,k,n})$ is convex for a fixed $\mathbf{w}_{l,k,n}$, since it is in fact the ratio between a quadratic form of $\mathbf{m}_{l,k,n}$ over an affine function of $\beta_{l,k,n}$ [12]. The nonconvex set defined by (9b) can be decomposed as a series of convex subsets by linearizing the convex function $g(u_{l,k,n})$ with the first order Taylor approximation around a fixed operating point $\hat{u}_{l,k,n}$ [13], also referred to as SCA in [14]. By using the reduced convex subset for (9b), the problem in (9) is solved iteratively by updating the operating point in each iteration.

Let the real and imaginary components of the complex number $\mathbf{w}_{l,k,n}^{\mathrm{H}}\,\mathbf{H}_{b_k,k,n}\,\mathbf{m}_{l,k,n}$ be represented by

$$p_{l,k,n} \triangleq \Re(\mathbf{w}_{l,k,n}^{\mathrm{H}}\,\mathbf{H}_{b_k,k,n}\,\mathbf{m}_{l,k,n}) \tag{11}$$
$$q_{l,k,n} \triangleq \Im(\mathbf{w}_{l,k,n}^{\mathrm{H}}\,\mathbf{H}_{b_k,k,n}\,\mathbf{m}_{l,k,n}) \tag{12}$$

and $g(u_{l,k,n}) = (p_{l,k,n}^2 + q_{l,k,n}^2)/\beta_{l,k,n}$. Let $\tilde{u}_{l,k,n} \triangleq \{\tilde{\mathbf{w}}_{l,k,n}, \tilde{\mathbf{m}}_{l,k,n}, \tilde{\beta}_{l,k,n}\}$ be a minimizer from the previous SCA iteration. Now, by using the first order Taylor approximation around the operating point $\tilde{u}_{l,k,n}$, we can approximate (9b) as

$$\begin{aligned}&2\frac{\tilde{p}_{l,k,n}}{\tilde{\beta}_{l,k,n}}(p_{l,k,n} - \tilde{p}_{l,k,n}) + 2\frac{\tilde{q}_{l,k,n}}{\tilde{\beta}_{l,k,n}}(q_{l,k,n} - \tilde{q}_{l,k,n}) \\ &+ \frac{\tilde{p}_{l,k,n}^2 + \tilde{q}_{l,k,n}^2}{\tilde{\beta}_{l,k,n}}\big(1 - \frac{\beta_{l,k,n} - \tilde{\beta}_{l,k,n}}{\tilde{\beta}_{l,k,n}}\big) \geq \gamma_{l,k,n}\end{aligned} \tag{13}$$

Similarly, for (9d), Let

$$g(v_{l,k,n}) \triangleq \frac{\left|\bar{\mathbf{w}}_{l,k,n}^{\mathrm{H}}\,\mathbf{H}_{b_k,k,n}^{\mathrm{T}}\,\bar{\mathbf{m}}_{l,k,n}\right|^2}{\bar{\beta}_{l,k,n}} \tag{14}$$

By using the first order Taylor approximation around the operating point $\tilde{v}_{l,k,n}$, we can approximate (9d) as

$$\begin{aligned}&2\frac{\tilde{r}_{l,k,n}}{\tilde{\beta}_{l,k,n}}(r_{l,k,n} - \tilde{r}_{l,k,n}) + 2\frac{\tilde{s}_{l,k,n}}{\tilde{\beta}_{l,k,n}}(s_{l,k,n} - \tilde{s}_{l,k,n}) \\ &+ \frac{\tilde{r}_{l,k,n}^2 + \tilde{s}_{l,k,n}^2}{\tilde{\beta}_{l,k,n}}\big(1 - \frac{\beta_{l,k,n} - \tilde{\beta}_{l,k,n}}{\tilde{\beta}_{l,k,n}}\big) \geq \bar{\gamma}_{l,k,n}\end{aligned} \tag{15}$$

where

$$r_{l,k,n} \triangleq \Re(\bar{\mathbf{w}}_{l,k,n}^{\mathrm{H}}\,\mathbf{H}_{b_k,k,n}^{\mathrm{T}}\,\bar{\mathbf{m}}_{l,k,n}) \tag{16}$$
$$s_{l,k,n} \triangleq \Im(\bar{\mathbf{w}}_{l,k,n}^{\mathrm{H}}\,\mathbf{H}_{b_k,k,n}^{\mathrm{T}}\,\bar{\mathbf{m}}_{l,k,n}) \tag{17}$$

B. Binary relaxation solution

In order to solve the combinatorial problem in (9) efficiently, the binary variables on $x_b, \bar{x}_b \in \{0,1\}$ are replaced by continuous variables $x_b, \bar{x}_b \in [0,1]$, $\forall b$. Then, the problem becomes convex (for fixed receivers, at any given linearization point of the SINR constraints). In order to ensure sparsity in the relaxed binary variables $\mathbf{x} = [x_1, \ldots, x_{N_B}, \bar{x}_1, \ldots, \bar{x}_{N_B}]$, a regularization function $f(\mathbf{x})$ is used in the objective to enforce the binary outcome for vector $\mathbf{x}$. One such function, proposed originally in [15], is given as

$$f(\mathbf{x}) = \sum_{t=1}^{N_B}(\log(x_b + \epsilon) + \log(\bar{x}_b + \epsilon)) \tag{18}$$

where ϵ is a small positive constant included in the argument to limit the dynamic range of log function. Now, by using $f(\mathbf{x})$ as a regularization function, the objective of (9) is given as

$$\text{minimize} \quad \|\tilde{\mathbf{v}}\|_q + \|\tilde{\mathbf{u}}\|_q + \psi\sum_{t=1}^{N_B}(\log(x_b + \epsilon) + \log(\bar{x}_b + \epsilon)). \tag{19}$$

where ψ is a constant. The use of log function in (19) is justified by the fact that $\log(x_b + \epsilon)$ has slope $1/\epsilon$ at the origin that increases as $\epsilon \to 0$, which is greater than ℓ_1 norm that has unit slope at the origin. Therefore, similar to ℓ_0 norm, which has infinite slope at the origin, $\log(x_b + \epsilon)$ term allows a large penalty for relatively small nonzero values of $\mathbf{x}$, thus leading to a solution, where the elements of $\mathbf{x}$ are binary valued [15].

The objective is now a sum of convex (norm) and concave (log) functions, i.e., in a difference of convex (DC) form, therefore, it can be solved by linearizing each concave function $\log(x_b + \epsilon)$ by its first order Taylor approximation around some fixed operating point

$\mathbf{x}^{(i)}$ as

$$f(\mathbf{x};\mathbf{x}^{(i)}) = \sum_{b=1}^{N_B}\Bigg(\log(x_b^{(i)}+\epsilon)+\log(\bar{x}_b^{(i)}+\epsilon) + \frac{x_b - x_b^{(i)}}{x_b^{(i)}+\epsilon} + \frac{\bar{x}_b - \bar{x}_b^{(i)}}{\bar{x}_b^{(i)}+\epsilon}\Bigg). \tag{20}$$

Now, by ignoring the constant terms in the approximate function (20), we can write an equivalent approximate problem as

$$\min. \quad \|\tilde{\mathbf{v}}\|_q + \|\tilde{\mathbf{u}}\|_q + \psi\sum_{b=1}^{N_B}\left(\frac{x_b - x_b^{(i)}}{x_b^{(i)}+\epsilon} + \frac{\bar{x}_b - \bar{x}_b^{(i)}}{\bar{x}_b^{(i)}+\epsilon}\right) \tag{21a}$$

$$\text{s. t.} \quad (13) \text{ and } (15) \tag{21b}$$

$$\beta_{l,k,n} \geq \widetilde{N}_0 + \sum_{i\in\mathcal{U}\setminus\{k\}}\sum_{j=1}^{L}|\mathbf{w}_{l,k,n}^{\mathrm{H}}\mathbf{H}_{b_i,k,n}\mathbf{m}_{j,i,n}|^2 + \sum_{i\in\mathcal{U}\setminus\mathcal{U}_{b_k}}\sum_{j=1}^{L}|\mathbf{w}_{l,k,n}^{\mathrm{H}}\tilde{\mathbf{H}}_{i,k,n}\bar{\mathbf{m}}_{j,i,n}|^2 \tag{21c}$$

$$\bar{\beta}_{l,k,n} \geq \widetilde{N}_0 + \sum_{i\in\mathcal{U}\setminus\{k\}}\sum_{j=1}^{L}|\bar{\mathbf{w}}_{l,k,n}^{\mathrm{H}}\mathbf{H}_{b_k,i,n}^{\mathrm{T}}\bar{\mathbf{m}}_{j,i,n}|^2 + \sum_{i\in\mathcal{U}\setminus\mathcal{U}_{b_k}}\sum_{j=1}^{L}|\bar{\mathbf{w}}_{l,k,n}^{\mathrm{H}}\hat{\mathbf{H}}_{b_i,b_k,n}\mathbf{m}_{j,i,n}|^2 \tag{21d}$$

$$\sum_{n=1}^{N}\sum_{k\in\mathcal{U}_b}\sum_{l=1}^{L}\|\mathbf{m}_{l,k,n}\|^2 \leq x_b P_{\max} \;\forall\, b \tag{21e}$$

$$\sum_{n=1}^{N}\sum_{l=1}^{L}\|\bar{\mathbf{m}}_{l,k,n}\|^2 \leq \bar{x}_{b_k} P_{\max}^{\mathrm{UE}} \;\forall\, k \tag{21f}$$

$$x_b + \bar{x}_b = 1 \;\forall\, b, \quad 0 \leq x_b \leq 1, \quad 0 \leq \bar{x}_b \leq 1 \tag{21g}$$

where the optimization variables are $\mathbf{m}_{l,k,n}, \mathbf{w}_{l,k,n}, \gamma_{l,k,n}, \beta_{l,k,n}, \bar{\mathbf{m}}_{l,k,n}, \bar{\mathbf{w}}_{l,k,n}, \bar{\gamma}_{l,k,n}, \bar{\beta}_{l,k,n}\ \forall\ l,k,n$ and $x_b, \bar{x}_b\ \forall\ b$. The proposed iterative algorithm involves two nested iteration loops. The outer alternating optimization (AO) loop consists of separate transmit and receive beamformer optimization, where the optimal receive beamformers given fixed TX beamformers correspond to the scaled MMSE receivers (see [11]). The inner SCA loop is used to find the transmit precoders by solving (21) iteratively until convergence for fixed receivers.[3] Upon finding a sparse solution for the relaxed vector $\mathbf{x}$, the augmented regularization term in the objective of (21) vanishes. We can directly observe that the proposed algorithm provides monotonic convergence for the objective in (21). This follows from the fact that each linear approximation is a global over-estimator of a concave function in (20), or global under-estimator of convex functions (13) and (15). The proof of monotonic progression follows from the basic successive convex approximation results provided in [16]. A more detailed convergence discussion can be found in [11, Appendix A].

IV. Numerical Examples

The numerical examples are gathered from a realistic 21-cell network with $|\mathcal{U}_b| = 2\ \forall\ b$ users per cell. Each BS is equipped with 4 antennas and the user terminals have 2 antennas. The maximum number of spatial data streams per user is limited to one for faster convergence. The simulation environment follows the guidelines given in [17]. The traffic model used in the simulations assume asymmetric traffic within the network, where a randomly chosen fraction of users, parametrized by α, have β times higher packet arrival rate $A_k = \bar{A}_k = A\ \forall\ k$. In practice, this illustrates the performance impact of having users with high or low rate demand applications, e.g, web browsing and video streaming. The number of bits in the queue are drawn from a Poisson distribution with varying packet/bit arrival rates. To further emphasize the application driven nature of the network traffic, the user specific traffic only occurs at downlink or uplink but not in both. Probability of having downlink (or uplink) traffic is 0.5 for each channel realization. For each user drop, the transceiver optimization and UL/DL mode selection is carried out over a scheduling interval consisting of two consecutive frames and the average queue backlog per user is used as the performance indicator. As a reference, we use a conventional TDD system, where one frame is fixed solely for the downlink and one for the uplink (termed as '50/50 fixed' in Fig. 1).

[3]In practice, the RX beamformers should be updated after each SCA update for faster convergence.

In Fig. 1, the queue backlog per user averaged over a large number of independent scheduling intervals is plotted as a function of arrival rate parameter A. The mean arrival rate across all low and high rate demand users can be straightforwardly calculated as $(1-\alpha)A + \alpha\beta A$. The high rate applications are simulated with two arrival rate amplifications ($\beta = 5$ and $\beta = 10$). Also, the impact of traffic asymmetric traffic is demonstrated by using high rate user fractions $\alpha = 0.2$ and $\alpha = 0.4$. Increasing the arrival rate further would results in an unstable system, where the queue backlogs would grow without limit (if a time-continuous arrival process was simulated). It is evident that asymmetric traffic causes difficulties for the conventional TDD systems, as the fixed mode assignment cannot exploit the uneven traffic patters, while significant gains in the order of 30-40% can be achieved using dynamic UL/DL mode selection. On the other hand, as the traffic patterns become more even, the performance gain from dynamic TDD diminishes.

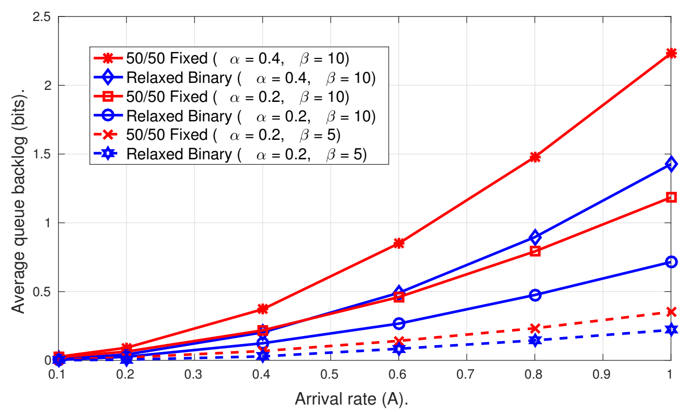

Fig. 1. Average number of queued bits per user with varying packet arrival rates.

V. Conclusion

Joint UL/DL cell mode selection, beamformer design and resource allocation among available time/frequency/space dimensions was considered for weighted queue minimization in Dynamic TDD setting with coordinated multi-cell MIMO transmission/reception. An efficient centralized solution was proposed by dividing the highly complex non-convex and combinatorial problem into solvable subproblems. Integer valued UL/DL mode selection variables were replaced by continuous variables, while enforcing a binary outcome using an appropriate regularization function in the objective. The numerical examples demonstrated significant gains from dynamic UL/DL mode selection as compared to the conventional TDD with fixed mode assignment.

REFERENCES

[1] Z. Shen, A. Khoryaev, E. Eriksson, and X. Pan, "Dynamic uplink-downlink configuration and interference management in TD-LTE," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 51–59, 2012.

[2] W. Jeong and M. Kavehrad, "Cochannel interference reduction in dynamic-TDD fixed wireless applications, using time slot allocation algorithms," *IEEE Trans. Commun.*, vol. 50, no. 10, pp. 1627–1636, 2002.

[3] P. Omiyi, H. Haas, and G. Auer, "Analysis of TDD cellular interference mitigation using busy-bursts," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2721–2731, 2007.

[4] I. Sohn, K. B. Lee, and Y. Choi, "Comparison of decentralized time slot allocation strategies for asymmetric traffic in TDD systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2990–3003, 2009.

[5] Nokia Networks, "5g radio access system design aspects," Tech. Rep., Nokia, Aug. 2015.

[6] E. Lähetkangas, K. Pajukoski, J. Vihriälä, G. Berardinelli, M. Lauridsen, E. Tiirola, and P. Mogensen, "Achieving low latency and energy consumption by 5G TDD mode optimization," in *Proc. IEEE Int. Conf. Commun.*, June 2014, pp. 1–6.

[7] Mobile and wireless communications Enablers for the Twenty-twenty Information Society (METIS), "Components of a new air interface – building blocks and performance D2.3," Tech. Rep., European Commission, Apr. 2014.

[8] 3GPP TSG RAN WG1, "Study on scenarios and requirements for next generation access technologies TR 38.913," Tech. Rep., 3rd Generation Partnership Project 3GPP, www.3gpp.org, 2016.

[9] 3GPP TSG RAN WG1, "Study on new radio access technology RP-160671," Tech. Rep., 3rd Generation Partnership Project 3GPP, www.3gpp.org, 2016.

[10] M. Neely, *Stochastic network optimization with application to communication and queueing systems*, Morgan & Claypool Publishers, 2010.

[11] G. Venkatraman, A. Tölli, M. Juntti, and L. N. Tran, "Traffic aware resource allocation schemes for multi-cell MIMO-OFDM systems," *IEEE Trans. Signal Processing*, vol. 64, no. 11, pp. 2730–2745, June 2016.

[12] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.

[13] G. R. Lanckriet and B. K. Sriperumbudur, "On the Convergence of the Concave-Convex Procedure," in *Advances in Neural Information Processing Systems*, 2009, pp. 1759–1767.

[14] B. R. Marks and G. P. Wright, "A General Inner Approximation Algorithm for Nonconvex Mathematical Programs," *Operations Research*, vol. 26, no. 4, pp. 681–683, 1978.

[15] E. J. Candes, M. B Wakin, and S. Boyd, "Enhancing Sparsity by Reweighted ℓ_1 Minimization," *Journal of Fourier analysis and applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[16] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Journal of the Operations Research Society of America*, vol. 26, no. 4, pp. 681–683, July–Aug. 1978.

[17] 3GPP, "Technical report 36.814: Further advancements for E-UTRA physical layer aspects," Tech. Rep., 3rd Generation Partnership Project 3GPP, www.3gpp.org, 2010-03.