Sequential Beamforming for Multiuser MIMO with Full-duplex Training

Xu Du, John Tadrous and Ashutosh Sabharwal

Abstract—Multiple transmitting antennas can considerably increase the downlink spectral efficiency by beamforming to multiple users at the same time. However, multiuser beamforming requires channel state information (CSI) at the transmitter, which leads to training overhead and reduces overall achievable spectral efficiency. In this paper, we propose and analyze a sequential beamforming strategy that utilizes full-duplex base station to implement downlink data transmission concurrently with CSI acquisition via in-band closed or open loop training. Our results demonstrate that full-duplex capability can improve the spectral efficiency of uni-directional traffic, by leveraging it to reduce the control overhead of CSI estimation. In moderate SNR regimes, we analytically derive tight approximations for the optimal training duration and characterize the associated respective spectral efficiency. We further characterize the enhanced multiplexing gain performance in the high SNR regime. In both regimes, the performance of the proposed full-duplex strategy is compared to the half-duplex counterpart to quantify spectral efficiency improvement. With experimental data [1] and 3D channel model [2] from 3GPP, in a 1.4 **MHz** 8×8 **system LTE system with the block length of** 500 **symbols, the proposed strategy attains a spectral efficiency improvement of 130% and 8% with closed and open loop training, respectively.**

I. INTRODUCTION

Multiuser MIMO downlink systems have the potential to increase the spectral efficiency by serving multiple users at the same time with a multiple-antenna base station. A base station with M antennas can simultaneously support up to M *half-duplex* single-antenna users at full multiplexing gain, if it has perfect channel information (CSI). Accurate channel knowledge at the transmitter is vital for achieving maximum spectral efficiency. For example, when no CSI is available at the base station, TDMA strategy is optimal [3]. Therefore, only one user can be supported with full multiplexing gain. In transmitter beamforming based systems, CSI is obtained by either closed or open loop training, which is defined as below.

- In *closed loop training* method, each user first estimates CSI by using the training pilots sent out by the base station. Then, the CSI is quantized and sent back to the base station[1].

- In *open loop training* method, base station learns downlink CSI by receiving training pilots from users through the uplink channel; channel *reciprocity* is then leveraged to estimate the downlink CSI from the uplink receptions.

For time-varying channels, overhead due to CSI acquisition leads to significant spectral efficiency loss. Thus, CSI estimation overhead reduction remains an important challenge. In this paper, we investigate the use of full-duplex capability to *reduce overhead of CSI estimation* to increase the spectral efficiency of downlink traffic.

The recently developed full-duplex radios [1], [5]–[10] allow concurrent uplink and downlink data transmission. However, the limited receiver dynamic range [11] and circuit-design [12] in small form-factor handsets for full-duplex transmission remain challenging problems. In this paper, we assume that only the base station antenna array to be an M-antenna full-duplex MIMO array [5] and all the mobile nodes to be half-duplex. The potential rate gain region of a full-duplex node is analyzed in [13]. In [14], a full-duplex base station is used to increase spectral efficiency by serving half-duplex downlink and uplink traffic simultaneously. In this paper, we propose an alternative use of full-duplex, where full-duplex capability is harnessed to increase downlink spectral efficiency by saving on the training time and thus reducing control channel overhead. In particular, our key contributions in this paper are as follows:

- We propose a sequential beamforming strategy for multiuser downlink transmissions with either closed or open loop training. Instead of waiting to receive all CSI before starting data transmission, the base station now begins transmitting to some users as it receives their CSI.

- The simultaneous transmission of feedback and data creates additional inter-node interference at the downlink receiving users. By optimizing the optimal training duration, we then analyze the spectral efficiency of the proposed sequential beamforming strategy and demonstrate its relative spectral efficiency gain over the half-duplex counterpart in closed-form. In general, due to the relatively low user training power[2], we show that inter-node interference only leads to a limited downlink rate reduction during training.

Xu Du, Ashutosh Sabharwal are with the Department of Electrical and Computer Engineering, Rice University, Houston, TX, 77005 (e-mails: {Xu.Du, ashu}@rice.edu). John Tadrous was with with the Department of Electrical and Computer Engineering, Rice University, Houston. He is now with Electrical and Computer Engineering Department, Gonzaga University, Spokane, WA, 99223 (e-mail: tadrous@gonzaga.edu) .

[1]Closed loop training with analog feedback is possible, but not considered in this paper. In [4], it is shown that, compared to open loop training, with the same amount of training symbols, closed loop training with analog feedback results in a larger error in uplink CSI training and greater interference due to precoding. In this paper, we only consider closed loop training with digital training and open loop training.

[2]In current systems, the transmission power of users is usually limited due to lower power budget compared to a base station.

- The spectral efficiency predicted by the closed-form results are further verified through simulations based on channel data from experiment [1] and 3GPP 3D channel model [2]. For example, in a typical 1.4 MHz LTE system with a block length of around 500 symbols, the proposed strategy demonstrates a spectral efficiency improvement of 130% and 8% over its half-duplex counterpart, for an 8×8 multiuser MIMO system with closed and open loop training, respectively.

The rate loss due to imperfect CSI with different types of training has been studied in [4]. In [15], the authors characterize the optimal training duration and its associated spectral efficiency for half-duplex systems. User selection [16], [17] has been proposed to reduce the number of training symbols needed by selecting users with a larger distance in channel space. Our analysis has two main differences from prior research. First, we study how to utilize the full-duplex operation to obtain gains in spectral efficiency. Second, the influence of limited training power at the mobile user is modeled and examined throughout this paper.

We first proposed to utilize the training time in systems composed of both full-duplex base station and *full-duplex* mobile in [18] and [19]. In this paper, we consider a system comprising a full-duplex base station and only *half-duplex* users.

The remainder of this paper is structured as follows. Section II describes the system model. Then the *sequential beamforming* strategy is proposed in Section III. The optimal training duration is studied in Section IV for systems with both closed and open loop training. The associated spectral efficiency is then presented in Section V with both theoretical analysis and experimental data validation. High SNR analysis is provided in Section VI to evaluate the proposed strategy. We conclude this paper by summarizing the main results in Section VII.

II. SYSTEM MODEL

We consider a symmetric multiuser MIMO downlink system consisting of an M-antenna full-duplex base station and M single-antenna non-cooperative half-duplex users. The base station aims at delivering downlink data to each user. Albeit sub-optimal, base station adopts zero-forcing (ZF) beamforming [20] for simultaneous transmission to multiple users. In ZF, the base station projects the signal intended for one user to the null space of the others. Thus, if perfect CSI is available, each user only receives the expected signal without interference.

Since CSI is obtained from finite training, it is almost always inaccurate and thus results in inter-beam interference for ZF transmissions. During the full-duplex training, the downlink data is communicated in the same band as the training signals sent by users, thus receiving users also suffer from *inter-node interference*. In this paper, we quantify the impact of inter-node interference on spectral efficiency for training-based ZF strategy. When User k sends training symbols and the base station transmits downlink data to User $1, 2, ..., k-1$, the received signal of User i is immediately captured as

$$y_i = \mathbf{h}_i^* \mathbf{V} \mathbf{s} + h_{\mathrm{U_{ik}}} x_{\mathrm{tr}_k} + n_i, \quad i = 1, \cdots, k-1. \quad (1)$$
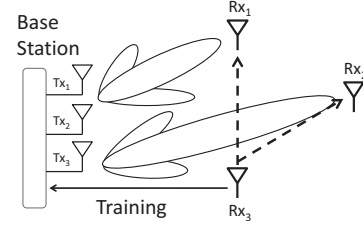

Fig. 1: A schematic of the interference in a 3×3 multiuser MIMO downlink system when User 3 sends training and others receive downlink data. The receiving users suffer inter-beam interference (side lobes) due to imperfect CSI. The receiving users also incur inter-node interference (dashed lines) resulting from User 3's training. Since users are half-duplex nodes, User 3 does not receive while it is sending the training signal.

Here $\mathbf{h}_i \in \mathcal{C}^{1 \times M}$ and $h_{\mathrm{U_{ik}}}$ stand for the channel realization from User i to the base station and User k, respectively. In this paper, the coherence time length is $T^{\mathrm{coherence}}$ symbols where the channel stays unchanged. Moreover, the block length T is the number of symbols for uplink training and downlink data transmission. We assume a Rayleigh block fading environment, i.e., each element of $\mathbf{h}_i$ and $h_{\mathrm{U_{ik}}}$ is independently complex Gaussian distributed from block to block.

The term $\mathbf{s} \in \mathcal{C}^{k-1 \times 1}$ is the actual signal intended to User $1, 2, \cdots, k-1$ and $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_{k-1}] \in \mathcal{C}^{M \times k-1}$ represents the ZF precoding matrix generated based on the quantized (estimated) CSI of users, which is presented as $\hat{\mathbf{h}}_i, i = 1, 2, .., k-1$. The precoded symbol is then $\mathbf{V} \mathbf{s}$, which is constrained to an average power constraint of P. We consider equal power allocation among symbols, i.e., $\mathbb{E} \left[|\mathbf{v}_i s_i|^2 \right] = P/M, \forall i$.

If only imperfect CSI is available, the inter-beam interference is non-zero. The signal and the inter-beam interference both are contained in term $\mathbf{h}_i^* \mathbf{V} \mathbf{s}$. Term x_{tr_k} is the uplink training symbol sent by User k. To account for the limitations of both battery and size of user devices, we consider a more strict average power constraint for users, which is described as $\mathbb{E}[|x_{\mathrm{tr}_k}|^2] \leqslant fP, \ f \in (0, 1]$. Term $h_{\mathrm{U_{ik}}} x_{\mathrm{tr}_k}$ captures the inter-node interference. We assume inter-node interference power to be proportional to the training power fP, i.e., it grows as $|h_{\mathrm{U_{ik}}} x_{\mathrm{tr}_k}|^2 \sim \alpha f P$, where $\alpha = \frac{|h_{\mathrm{U_{ik}}} x_{\mathrm{tr}_k}|^2}{|x_{\mathrm{tr}_k}|^2} > 0$ and $fP = |x_{\mathrm{tr}_k}|^2 > 0$. The signal is degraded by an independent unit variance additive white complex Gaussian noise n_i.

We assume that each User i has perfect knowledge of the downlink channel from the base station to itself $\mathbf{h}_i$ by estimating downlink pilots broadcast by base station antennas before any uplink training or downlink data transmission. However, the base station is required to obtain CSI by either closed or open loop training. The base station is assumed to be an $M-$antenna full-duplex MIMO array [5]. We assume that self-interference due to the full-duplex operation at the base station is reduced to near-noise floor by the state-of-art circulators and filter-based self-interference cancellation [5]. Since the same antennas are used for both transmission and reception, channel reciprocity also holds between uplink and downlink channels. Unlike [18], [19], the proposed sequential beamforming strategy needs only half-duplex users. Thus there is no self-interference at mobile nodes.

III. Sequential Beamforming

In this section, we propose a sequential beamforming strategy that leverages the full-duplex capability at the base station to send downlink data during CSI collection. First, we describe the sequential beamforming strategy in Section III-A. Then we characterize the influence of inter-node and inter-beam interference on downlink rate.

A. Sequential Beamforming Strategy

The proposed strategy is referred to as *sequential beamforming*. In sequential beamforming, pre-scheduled users send their channel state information in orthogonal time slots. And as the base station receives a particular user's information, it starts data transmission to that user. Thus, unlike the half-duplex system, the base station does not wait for all the users to send their channel feedback. As noted before, the proposed strategy only requires the base station to be full-duplex and all the mobiles can be half-duplex. A sequential beamforming strategy with total T^{tr} training symbols from all users is described as follows:

1) At the beginning of each block, no downlink data transmission is performed due to the lack of CSI knowledge. From Symbol 1 to Symbol $\frac{T^{\mathrm{tr}}}{M}$, User 1 sends[3] training symbols to the base station. We define symbols from Symbol $(j-1)\frac{T^{\mathrm{tr}}}{M}+1$ to Symbol $j\frac{T^{\mathrm{tr}}}{M}$ to be Cycle j where User j sends its training symbols.

2) In cycle $j+1$, the base station transmits downlink data based on the updated ZF precoding matrix and beamforms to User $1,2,..,j$ whose training symbols are collected over the previous j cycles. Users who have finished training, i.e., User $1,2,\cdots,j$, begin receiving downlink data. All receiving users decode the received signal by treating interference (both inter-beam and inter-node interference) as noise.

3) Repeat 2) till the end of T^{tr} symbols. The above full-duplex training part is referred to as *training phase.*

4) After all training is collected, only downlink data transmission takes place. This part is referred to as *half-duplex phase.* Fig. 2 provides an illustration of sequential beamforming.

We will compute the overall *spectral efficiency (SE)* for both phases as

$$\mathrm{SE}_{\mathrm{SqBf}} = \frac{T^{\mathrm{tr}}}{MT}\sum_{j=2}^{M}\frac{1}{M}\sum_{i=1}^{j-1}R(i,j) + \frac{T-T^{\mathrm{tr}}}{T}R^{\mathrm{data}}. \quad (2)$$

The first and second terms capture the spectral efficiency achieved during and after training, respectively. Rate expression $R(i,j)$ stands for the downlink rate achieved by User i during cycle j. Moreover, R^{data} is the rate achieved after training, i.e., during half-duplex phase. By only considering the second term in Eq. (2), the spectral efficiency of half-duplex counterpart is immediate as

$$\mathrm{SE}_{\mathrm{Hf}} = \frac{T-T^{\mathrm{tr}}}{T}R^{\mathrm{data}}. \quad (3)$$

[3]For a given set of users, the user index $1,2,...,M$ may be randomly assigned in every coherence block to achieve fairness among users.

Our objective is to maximize the downlink spectral efficiency. We first quantify the influence of both inter-node and inter-beam interference on $R(i,j)$ and R^{data} in Section III-B for further analysis. During training, $i-1$ users are served on downlink in Cycle $i>1$. The base station uses power budget P/M to serve each receiving user.

In this paper, the performance of the following four systems is examined: sequential beamforming strategy with closed and open loop training, half-duplex with closed and open loop training. We differentiate between sequential beamforming and half-duplex systems through the subscripts SqBf and Hf, respectively. The strategy is further detailed by the training type used by the system through another subscript Cl for closed loop and Op for open loop training. The superscript is used to denote the system status. For example, $\mathrm{SE}_{\mathrm{SqBf_{Cl}}}$ stands for the spectral efficiency of a system adopting sequential beamforming strategy with closed loop training.

B. Rate Performance with Inter-beam and Inter-node Interference

To optimize the spectral efficiency of the proposed sequential beamforming strategy, we now quantify the influence of inter-node and inter-beam interference on downlink rate.

In ZF beamforming, $\mathbf{v}_i$ is chosen to be orthogonal to other users' channel realization, i.e., $|\mathbf{v}_i\mathbf{h}_j|=0, j\neq i$. In a genie-aided system where perfect CSI is available, the base station beamforms to users without training and each user receives downlink data at rate R^{ZF} as

$$R^{\mathrm{ZF}} = \mathbb{E}\left[\log_2\left(1+\frac{P}{M}|\mathbf{h}_i^*\mathbf{v}_i|^2\right)\right]. \quad (4)$$

Rate (4) can be viewed as an upper bound for all strategies that employ ZF, since neither training overhead nor inter-beam interference is included. When User k sends the training symbols, the received SINR of User i ($i<k$) is decided by both inter-beam and inter-node interference, which can be mathematically expressed as

$$\mathrm{SINR}_i = \frac{|\mathbf{h}_i^*\mathbf{v}_i|^2\frac{P}{M}}{1+\sum_{j\neq i}\frac{P}{M}|\mathbf{h}_i^*\mathbf{v}_j|^2+|h_{\mathrm{U_{ik}}}x_{\mathrm{tr}_k}|^2}, \quad i=1,\cdots,k-1. \quad (5)$$

By adopting a Gaussian input, User i with SINR_i can achieve a longer term average rate (over both fading and training error) of

$$R_i = \mathbb{E}\left[\log_2\left(1+\mathrm{SINR}_i\right)\right], \quad i=1,\cdots,k-1.$$

We now characterize the downlink rate $R(i,j)$ with a unified lower bound that is independent of cycle number. In Cycle k of the training phase, the base station beamforms to Users $1,...,k-1$. Each receiving user suffers inter-beam interference from the signals intended to the other $k-2$ users. Our lower bound assumes that users receive additional inter-beam interference from signal intended to other $M-k$ users. Thus, in total, each receiving user suffers inter-beam interference coming from signals to $M-1$ users instead of $k-2$ users. The downlink rate of the receiving users in this scenario is denoted as R^{tr}, which is detailed as

$$R(i,j) \geqslant R^{\mathrm{tr}}\left(T^{\mathrm{tr}}\right), \quad j=2,...,M, \quad i=1,...,j. \quad (6)$$
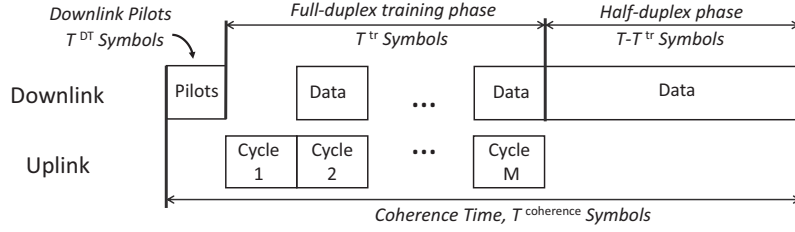
Fig. 2: A depiction of sequential beamforming: Users first obtain CSI by estimating downlink training pilots broadcast by base station antennas in the first T^{DT} symbols. Training cycles constitute T^{tr} symbols and half-duplex phase occupies the rest of $T - T^{\text{tr}}$ symbols. At the end of training cycles, the base station updates its precoding vectors and serves all users whose CSI has been collected. The whole coherence time length is $T^{\text{coherence}}$ symbols and parameter T is the total number of symbols for uplink training and downlink data transmission.

This lower bound is the same for all the receiving users in all cycles during the training phase. Therefore, the rate expression of (2) reduces to

$$\text{SE}_{\text{SqBf}} \geqslant \frac{M-1}{2M}\frac{T^{\text{tr}}}{T}R^{\text{tr}}\left(T^{\text{tr}}\right) + \left(1 - \frac{T^{\text{tr}}}{T}\right)R^{\text{data}}\left(T^{\text{tr}}\right), \tag{7}$$

here T^{tr} is the training duration of sequential beamforming strategy. Comparing to (3), we find in sequential beamforming strategy, on average, the each user utilizes $\frac{M-1}{2M}$ fraction of the training time also to receive downlink data while other users send training signal.

Since the base station is assumed to perform perfect self-interference cancellation, the influence of inter-beam interference is a function of the training method, power, and duration, which is characterized in [4]. We now present an extended lemma that quantifies the rate loss due to inter-beam and inter-node interference. Following the notations in [4], ΔR^{tr} and ΔR^{data} denote the upper bound of rate gap (compared to perfect zero-forcing) during and after training, respectively.

Lemma 1: In all cycles of training phase, the downlink data transmission rate of the receiving users, when another user is sending the training symbols are lower bounded as

$$R^{\text{tr}}\left(T^{\text{tr}}\right) \geqslant R^{\text{ZF}} - \log\left(\frac{1 + \mathcal{P}_{\text{IBI}}\left(T^{\text{tr}}\right) + \alpha fP}{1 + \frac{\alpha fP}{1 + \frac{P}{M}}}\right)$$
$$= R^{\text{ZF}} - \Delta R^{\text{tr}}\left(T^{\text{tr}}\right),$$

where $\mathcal{P}_{\text{IBI}} = P(1+fP)^{-\frac{T^{\text{tr}}}{M(M-1)}}$ and $\frac{P}{M}\frac{M-1}{1+\frac{T^{\text{tr}}}{M}fP}$ for closed and open loop training, respectively.

Proof. See Appendix A for detail. ∎

In the rate gap term ΔR^{tr}, inter-beam and inter-node interference are reflected through terms $\mathcal{P}_{\text{IBI}}$ and αfP, receptively. If more training symbols are sent, $\mathcal{P}_{\text{IBI}}$ decreases. This decrease is because that the base station has better CSI estimates, which leads to less inter-beam interference. The inter-node interference term αfP does not change during the whole training phase. It is emphasized that the lower bound present in Lemma 1 is independent of the user index and cycle index, due to the use of Eq. (6).

Many recent works have helped raise the level of self-interference cancellation [1], [5]–[10]. Hence in this work, we assume that full-duplex MIMO array with perfect self-interference cancellation is possible for the base station, which has a larger footprint than a mobile node. This assumption allows us to focus on characterizing the tradeoff between inter-beam and inter-node interference. Lemma 1 and other

analysis in this paper can be extended to system models that include the impact of limited self-interference cancellation by substituting fP, which is the effective uplink training SNR in $\mathcal{P}_{\text{IBI}}$, as $\frac{fP}{1+P_{\text{SI}}}$. Here P_{SI} is the power of residual self-interference interference. Simulation results for systems with self-interference are provided in Section V.

As $T^{\text{tr}} \rightarrow \infty$, the rate loss due to inter-beam interference vanishes and rate gap bound becomes $\log\left(\frac{1 + \alpha fP}{1 + \frac{\alpha fP}{1 + P/M}}\right)$, which stands for the influence of inter-node interference and is noted as ΔR^{INI}. Term ΔR^{INI} is as a constant rate loss caused by inter-node interference during the training phase. We will study the impact of this term in the following analysis. Even when $\alpha \rightarrow \infty$, the rate loss term is still upper bounded by $\log\left(1 + P/M\right)$, which is obviously finite. This finite rate loss suggests that positive downlink rate gain can still be achieved asymptotically under the influence of inter-node interference, which is later confirmed in Section VI.

After training, each user continues to receive data until the end of the block. Thus, only the effect of inter-beam interference exists. We can conveniently obtain the rate expression R^{data} by setting $\alpha = 0$ in Lemma 1, which characterizes inter-beam interference with the help of [4].

Proposition 1: The downlink transmission rate of User i after training is lower bounded by

$$R^{\text{data}} \geqslant R^{\text{ZF}} - \Delta R^{\text{data}} = R^{\text{ZF}} - \log\left(1 + \mathcal{P}_{\text{IBI}}\right), \tag{8}$$

where $\mathcal{P}_{\text{IBI}} = P(1+fP)^{-\frac{T^{\text{tr}}}{M(M-1)}}$ and $\frac{P}{M}\frac{M-1}{1+\frac{T^{\text{tr}}}{M}fP}$ for closed and open loop training, respectively.

Similar to the influence of inter-beam interference during training, we find that the influence of inter-beam interference also decreases as training symbols amount increases.

We assume perfect CSI at users by estimating pilots broadcast by each base station antenna. The rate loss due to imperfect CSI at users [4] is upper bounded by

$$\Delta R^{\text{CSIR}} \leqslant \log\left(1 + \frac{P/M}{1 + \frac{T_{\text{DL}}}{M}P_{\text{DL}}}\right),$$

where T_{DL}/M and P_{DL} is the number and power of downlink training symbols broadcast by base station antennas. For example, a system with $M = 8$ and $P = 15$ dB, with "pilots power boost of 3dB" from [21], when 1 training pilot is used for each base station antenna, the rate loss is upper bounded by 0.06 bps/Hz, which is smaller than 3% of the associated R^{ZF}. This rate loss is the same for half-duplex and the proposed system. Similar to [4], [15], the rate considered

in this paper is the *ergodic* rate, which can be achieved by spanning codewords across enough large number of blocks.

IV. TRAINING TIME OPTIMIZATION

In sequential beamforming strategy described in Section III-A, the base station obtains CSI from uplink training symbols. Obtaining the best spectral efficiency performance requires a balance between inter-beam and inter-node interference by optimizing training duration. In this section, we analyze the optimal training duration for sequential beamforming and traditional half-duplex strategy with both closed and open loop training. We use the metric of spectral efficiency to characterize the optimal solution for the proposed strategy. The optimization

$$T_s^{\text{tr}*} = \arg\max \text{SE}_s\left(T^{\text{tr}}\right), \qquad (9)$$

is implemented for $s = \text{SqBf}_{\text{Cl}}, \text{SqBf}_{\text{Op}}, \text{Hf}_{\text{Cl}}$ and Hf_{Op}. We use the superscript $*$ to denote optimal solution. The optimality in this paper is under the criterion of maximum spectral efficiency. We also consider T^{tr} to be continuous. The accent $\sim$ is used to represent approximation in Sections IV-A, IV-B, where closed form analytical solutions are not feasible.

A. Optimal Training Duration of Sequential Beamforming

In this subsection, we solve the optimization problem posed in (9) by applying a *Marginal Analysis* [22] technique. As shown below, the marginal analysis allows accurate closed form approximation for systems with both closed and open loop training.

Proposition 2: The optimal training duration of sequential beamforming strategy happens at the point where the spectral benefit of adding training symbols equals to loss, i.e.,

$$\frac{\partial \text{SE}_{\text{SqBf}}\left(T^{\text{tr}*}\right)}{\partial T^{\text{tr}}} = 0. \qquad (10)$$

Proof. Since the mobile nodes are half-duplex, more training implies less time for downlink data reception. The influence of inter-beam interference on the rate in Proposition 1 suggests that the increase in rate with respect to training increase is monotonically decreasing. Combining with the facts above, we conclude that the benefit in spectral efficiency from M additional training symbols is monotonically decreasing as training grows. From Lemma 1, the influence of longer inter-node interference duration is monotonically increasing. Thus, the spectral efficiency SE_{SqBf} is concave in T^{tr}. Therefore, a unique $T_{\text{SqBf}}^{\text{tr}*}$ exists to optimize the spectral efficiency. ∎

Applying Taylor's expansion to (10) and ignoring all the expansion terms yield

$$\text{SE}_{\text{SqBf}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}*}\right) \approx \text{SE}_{\text{SqBf}}\left(\widetilde{T}_{\text{SqBf}}^* + M\right). \qquad (11)$$

With the help of spectral efficiency characterization provided in Lemma 1, expanding both sides of (11) leads to

$$\begin{aligned}
&\frac{M-1}{2M}\frac{\widetilde{T}_{\text{SqBf}}^{\text{tr}}}{T}\left[R^{\text{tr}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}} + M\right) - R^{\text{tr}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}}\right)\right] \\
&+ \frac{T - \widetilde{T}_{\text{SqBf}}^{\text{tr}}}{T}\left[R^{\text{data}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}} + M\right) - R^{\text{data}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}}\right)\right] \\
&= \frac{M-1}{2T}\left[R^{\text{data}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}} + M\right) - R^{\text{tr}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}} + M\right)\right] \\
&+ \frac{M+1}{2T}R^{\text{data}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}} + M\right). \qquad (12)
\end{aligned}$$

The left side in (12) is the benefit obtained in spectral efficiency by adding M training symbols. We note this benefit as *Marginal Utility* (MU). The MU comes from the fact: more training can reduce inter-beam interference both during and after training, which corresponds to the first and second term on left side of (12), respectively.

More training symbols results in lower inter-beam interference in half-duplex phase. By using Proposition 1, it is expressed as a rate increase of

$$\begin{aligned}
&R^{\text{data}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}} + M\right) - R^{\text{data}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}}\right) \\
&= \log\left(1 + \frac{\mathcal{P}_{\text{IBI}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}}\right) - \mathcal{P}_{\text{IBI}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}} + M\right)}{1 + \mathcal{P}_{\text{IBI}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}} + M\right)}\right). \qquad (13)
\end{aligned}$$

We refer the rate improvement due to less inter-beam interference as $\delta R^{\text{data}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}}\right)$. In the same spirit, the rate increase of R^{tr} by lower inter-beam interference during training phase is

$$R^{\text{tr}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}} + M\right) - R^{\text{tr}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}}\right) \approx \delta R^{\text{data}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}}\right).$$

We find that, the rate improvement due to less inter-beam interference is almost constant during and after training. Applying the two results above, the marginal utility is then

$$\begin{aligned}
MU &= \left(1 - \frac{M+1}{2M}\frac{\widetilde{T}_{\text{SqBf}}^{\text{tr}}}{T}\right)\delta R^{\text{data}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}}\right) \\
&\approx \left(1 - \frac{1}{2}\frac{\widetilde{T}_{\text{SqBf}}^{\text{tr}}}{T}\right)\delta R^{\text{data}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}}\right), \qquad (14)
\end{aligned}$$

which suggests a rate increase of $\delta R^{\text{data}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}}\right)$ in $1 - \frac{1}{2}\frac{\widetilde{T}_{\text{SqBf}}^{\text{tr}}}{T}$ fraction of the whole block is achieved by adding M training symbols. Later we further obtain the marginal utility of half-duplex counterparts by the same process with SE_{SqBf} substituted as SE_{Hf}.

On the right side of (12) is the loss of spectral efficiency, referred to as *Marginal Cost* (MC), due to longer training and comprises two parts. The first term corresponds to the fact that additional inter-node interference is suffered in $\frac{M-1}{2}$ of the M symbols. The second term reflects that the rest $\frac{M+1}{2}$ training symbols are still not able to be utilized for downlink. With the help of Lemma 1 and Proposition 1, the rate loss due to

additional inter-node interference is

$$R^{\text{data}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}} + M\right) - R^{\text{tr}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}} + M\right)$$

$$= \log\left(\frac{1 + \frac{\alpha f P}{1 + \mathcal{P}_{\text{IBI}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}} + M\right)}}{1 + \frac{\alpha f P}{1 + \frac{P}{M}}}\right) \approx \Delta R^{\text{INI}}.$$

The downlink rate loss due to inter-node interference in the training phase is almost independent of the training duration. The downlink rate is immediate as R^{ZF}. The marginal cost is then

$$MC = \frac{M-1}{2T}\Delta R^{\text{INI}} + \frac{M+1}{2T}R^{\text{ZF}} \approx \frac{M}{2T}\left[\Delta R^{\text{INI}} + R^{\text{ZF}}\right], \tag{15}$$

which is independent of training symbol amount. The approximation made in Eq. (15) holds for large T. In the training phase, Eq. (15) suggests that, on average, each user receives downlink data during half of the training time.

The unique optimal point $T_{\text{SqBf}}^{\text{tr*}}$ happens at the point where the spectral efficiency benefit (marginal utility) and cost (marginal cost) break even, i.e., $MU = MC$. Using (14) and (15), it is mathematically captured as

$$\left(1 - \frac{M+1}{2M}\frac{\widetilde{T}_{\text{SqBf}}^{\text{tr}}}{T}\right)\delta R^{\text{data}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr}}\right)$$
$$= \frac{M-1}{2T}\Delta R^{\text{INI}} + \frac{M+1}{2T}R^{\text{ZF}}. \tag{16}$$

The optimal training duration of sequential beamforming strategy with closed and open loop training is then obtained by the inter-beam interference characterization provided in Proposition 1.

Theorem 1: The approximation of optimal training duration $T_{\text{SqBf}_{\text{Cl}}}^{\text{tr*}}$ of sequential beamforming strategy with closed loop training is

$$\widetilde{T}_{\text{SqBf}_{\text{Cl}}}^{\text{tr*}} = M\left(M-1\right)\frac{\log\left(\frac{TP}{c}\right) + \log\left(\left(1 + fP\right)^{\frac{1}{M-1}} - 1\right)}{\log\left(1 + fP\right)}, \tag{17}$$

where $c = \frac{M-1}{2}\Delta R^{\text{INI}} + \frac{M+1}{2}R^{\text{ZF}}$.
Proof. See Appendix B. ∎

Several interesting observations are made here. First, as T grows, for closed loop training based sequential beamforming strategy, the optimal training duration scales as $\log T$. We later observe similar scaling law for its half-duplex counterpart. This scaling law, to our best knowledge, has not been reported before. Second, as inter-node interference becomes stronger, less training is sent to account for the higher training *cost*. Third, the optimal training symbols amount scales as $\frac{\log P}{\log(1+fP)}$ with respect to P, which is less than $\log P$. From Theorem 4 in [23], we conclude that full multiplexing is not obtained as P grows. Fourth, the number of training symbols increases almost quadratically with respect to the number of users M, which lies well with the intuition that training symbols number scales with the number of total channel coefficients.

Fig. 3a provides both optimal training duration and its approximation of sequential beamforming strategy with closed loop training. Since optimal training duration scales as $\log T$, the fraction of training duration scales as $\frac{\log T}{T}$ and is further confirmed numerically. In the same spirit, the optimal training duration of sequential beamforming strategy with open loop training is obtained as below.

Theorem 2: The approximation of optimal training duration $T_{\text{SqBf}_{\text{Op}}}^{\text{tr*}}$ of sequential beamforming strategy with open loop training is

$$\widetilde{T}_{\text{SqBf}_{\text{Op}}}^{\text{tr*}} = \sqrt{\frac{\left(M-1\right)T}{f\frac{(M-1)\Delta R^{\text{INI}}+(M+1)R^{\text{ZF}}}{2M}}} \approx \sqrt{\frac{\left(M-1\right)T}{f\frac{\Delta R^{\text{INI}}+R^{\text{ZF}}}{2}}}.$$

Proof. See Appendix C. ∎

In Theorem 2, we observe that for large T, the optimal training duration scales as $\sqrt{T}$. The optimal fraction of time resource devoted to training then decreases as $\frac{1}{\sqrt{T}}$, which is slower than that of closed loop training systems. The scaling rate of $\sqrt{T}$ has been observed in various open loop training based systems. For example, similar scaling has been observed for both half-duplex MIMO broadcast channels with analog training [15] and point-to-point MIMO [24]. Such scaling rate has also been observed in MIMO downlink with the full-duplex base station and full-duplex node [19]. We find the same scaling law is shared by open loop training based systems. In Section IV-B, we conclude that the half-duplex counterparts also follow the respective scaling laws. Numerical results presented in Fig. 3b confirm our observation.

For sequential beamforming strategy with open loop training, as the number of transmitting antennas M increases, the optimal training duration scales as $\sqrt{M}$, unlike M^2 scaling in closed loop training based systems. The slower scaling rate in open loop training systems suggests a lower overhead cost in systems with a large number of users. Further analytical results in Section VI confirm this observation.

Similar to closed loop systems, as inter-node interference increases, larger rate loss during training is expected in open loop training systems. Thus, one should use fewer symbols for training to account for this effect. Another interesting finding is that even when no inter-node interference exists, the optimal training duration is not T. The reason is that sequential beamforming strategy is only able to recover the training overhead partially.

B. Optimal Training Duration of Half-duplex Strategy

In this subsection, we apply the marginal analysis method developed in Section IV-A to obtain approximations of optimal training duration of half-duplex systems. As a by-product of analysis in section IV-A, we find the marginal utility, which stands for the gain in spectral efficiency of adding M more training symbols, for half-duplex systems is

$$MU = \frac{T - \widetilde{T}_{\text{Hf}}^{\text{tr*}}}{T}\left[R^{\text{data}}\left(\widetilde{T}_{\text{Hf}}^{\text{tr*}} + M\right) - R^{\text{data}}\left(\widetilde{T}_{\text{Hf}}^{\text{tr*}}\right)\right]$$
$$= \frac{T - \widetilde{T}_{\text{Hf}}^{\text{tr*}}}{T}\delta R^{\text{data}}\left(\widetilde{T}_{\text{SqBf}}^{\text{tr*}}\right). \tag{18}$$

The marginal cost of half-duplex strategy is conveniently obtained by ignoring inter-beam interference after training as

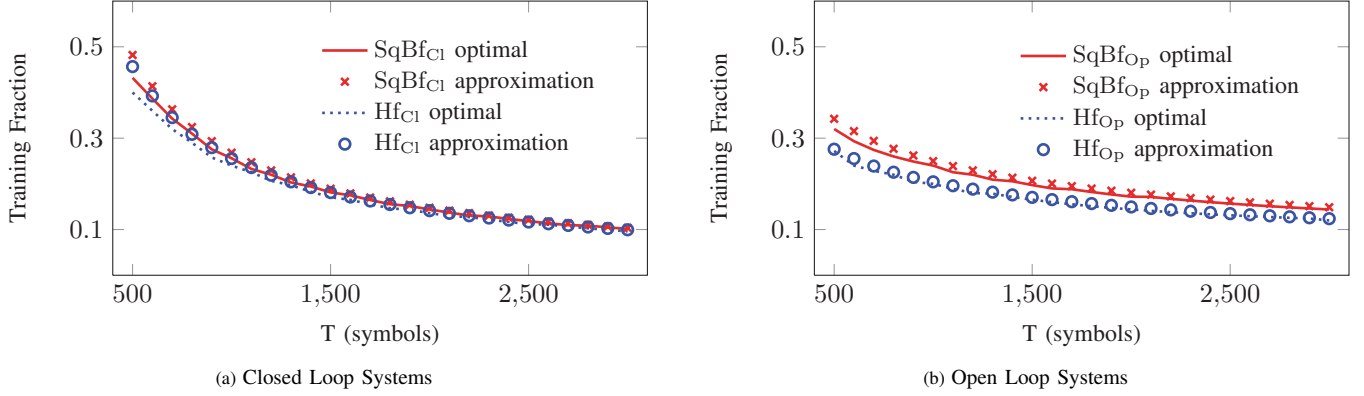$$MC = \frac{M}{T}R^{\text{ZF}}. \tag{19}$$

Fig. 3: Optimal training duration fraction of 8×8 sequential beamforming and half-duplex strategy with closed and open loop training at $P = 15$ dB with $f = 0.1$ and $\alpha = 0.3$.

The approximation is obtained by letting marginal cost and utility be equal in half-duplex systems. We further proceed by applying the rate characterization provided in Proposition 1. The result regarding closed loop training based systems is first presented with open loop result follows.

Theorem 3: The approximation of optimal training duration $T_{\mathrm{Hf_{Cl}}}^{\mathrm{tr*}}$ of half-duplex strategy with closed loop training is

$$\widetilde{T}_{\mathrm{SqBf_{Cl}}}^{\mathrm{tr*}} = M\,(M-1)\,\frac{\log\left(\frac{TP}{c}\right) + \log\left((1+fP)^{\frac{1}{M-1}} - 1\right)}{\log\,(1+fP)},$$

where $c = MR^{\mathrm{ZF}}$.

Proof. See Appendix D. ∎

We observe the optimal training duration $\widetilde{T}_{\mathrm{SqBf_{Cl}}}^{\mathrm{tr*}}$ and fraction $\frac{\widetilde{T}_{\mathrm{SqBf_{Cl}}}^{\mathrm{tr*}}}{T}$ of half-duplex counterpart to share the same scaling law as sequential beamforming strategy in Theorem 1. It should also be noted that as the number of antenna M increases, the optimal number of training symbols also increases quadratically. Comparing to Theorem 1, the only difference lies in the $\log(c)$ term in the numerator, which can be viewed as the normalized marginal cost of the strategy. This finding also suggests that the optimal training duration difference between sequential beamforming strategy and half-duplex is a constant gap which is independent of the block length. Therefore, the difference in the fraction of training time decreases as T increases.

Theorem 4: The approximation of training duration $T_{\mathrm{Hf_{Op}}}^{\mathrm{tr*}}$ that optimizes spectral efficiency of open loop training half-duplex systems is

$$\widetilde{T}_{\mathrm{Hf_{Op}}}^{\mathrm{tr*}} = \sqrt{\frac{(M-1)\,T}{fR^{\mathrm{ZF}}}}. \tag{20}$$

Proof. See Appendix E. ∎

Similar to sequential beamforming system with open loop training, the optimal training duration scales with T and M at the rate of $\sqrt{T}$ and $\sqrt{M-1}$, respectively, as block length and antennas number grows. It should also be noted that by substituting the normalized marginal cost term $\frac{\Delta R^{\mathrm{INI}} + R^{\mathrm{ZF}}}{2}$ as the half-duplex system's normalized marginal cost term R^{ZF}, we can also obtain Theorem 4. Instead of assuming each user has the same power constraint P of the base station [15], our approximation results further consider the limitation of user

power. The closed-form approximations are further applied in Section V to characterize the spectral efficiency of sequential beamforming and half-duplex strategy with optimal training duration.

V. SPECTRAL EFFICIENCY EVALUATION

In training based multiuser MIMO downlink systems, spectral efficiency is reduced due to imperfect CSI and training overhead resulting from its acquisition. To quantify the spectral efficiency loss of different systems, we compare the spectral efficiency of different systems with optimal training duration to a system where perfect CSI is available for the base station at no cost. It can be visualized as a genie provides perfect CSI to base station at the beginning of each block. Thus it serves as an upper bound for systems' performance with ZF; we label the perfect CSI system as *genie-aided system*. The spectral efficiency achieved is $\mathrm{SE}^{\mathrm{ZF}} = R^{\mathrm{ZF}}$, which is presented in (4). The rate loss due to training overhead is then

$$\Delta\mathrm{SE}_s = \mathrm{SE}^{\mathrm{ZF}} - \mathrm{SE}_s, \quad s \in \{\mathrm{SqBf_{Cl}}, \mathrm{SqBf_{Op}}, \mathrm{Hf_{Cl}}, \mathrm{Hf_{Op}}\}. \tag{21}$$

The spectral efficiency of the half-duplex counterparts are also analyzed for comparison.

Theorem 5: The spectral efficiency loss of closed loop training based sequential beamforming system with respect to genie-aided system is upper-bounded as

$$\Delta\mathrm{SE}_{\mathrm{SqBf_{Cl}}}\left(T_{\mathrm{SqBf_{Cl}}}^{\mathrm{tr*}}\right)$$
$$\leqslant \left(\frac{M-1}{2M}\Delta R^{\mathrm{INI}} + \frac{M+1}{2M}R^{\mathrm{ZF}}\right)\frac{M\,(M-1)}{\log\,(1+fP)}\frac{\log T}{T}$$
$$+ o\left(\frac{\log T}{T}\right). \tag{22}$$

Proof. See Appendix F. ∎

Here $o\left(\frac{\log T}{T}\right)$ is a term that vanishes as T increases, i.e., $\lim_{T \to \infty}\frac{o(\log T/T)}{\log T/T} = 0$. By employing higher training power f, or by using longer block length T, the spectral efficiency overhead decreases. However, in a more realistic scenario where user power and block length are inherently limited, the spectral efficiency loss cannot be neglected. Based on expression (22), some observations are made for sequential beamforming strategy with closed loop training. i) The spectral efficiency loss scales quadratically as M increases, which

indicates sequential beamforming strategy with closed loop training is not a good choice for systems with a large number of antennas. ii) The spectral efficiency loss decreases rapidly as $\frac{\log T}{T}$ as T increases. Fig. 4 presents the spectral efficiency policy for different strategy versus T. We observe that as T grows, spectral efficiency loss drops rapidly for systems with closed loop training, which agrees with our analysis. iii) As inter-node interference level decreases, smaller term ΔR^{INI} suggests less spectral efficiency loss which is confirmed in Fig. 4.

Theorem 6: The spectral efficiency loss of open loop training based sequential beamforming system with respect to genie-aided system is upper-bounded as

$$\Delta \mathrm{SE}_{\mathrm{SqBf_{O_p}}}\left(T^{\mathrm{tr}*}_{\mathrm{SqBf_{O_p}}}\right)$$
$$\leqslant 2\sqrt{\frac{(M-1)\left[\frac{M-1}{2M}\Delta R^{\mathrm{INI}} + \frac{M+1}{2M}R^{\mathrm{ZF}}\right]}{fT}} + o\left(\frac{1}{\sqrt{T}}\right). \tag{23}$$

Proof. See Appendix G. ∎

In (6), the term $o(\frac{1}{\sqrt{T}})$ shows that the additional spectral efficiency loss term vanishes in systems with a large T. Interestingly, we observe a different scaling law with respect to both block length and antenna number. For sequential beamforming with open loop training, the spectral efficiency loss grows only at the rate of $\sqrt{M-1}$, which is slower than $M(M-1)$ in Theorem 5. Thus, for systems with a large number of users, sequential beamforming with open loop training is advisable.

On the other hand, the spectral efficiency loss decreases as $\frac{1}{\sqrt{T}}$, which is confirmed from Fig. 4. It should be further noted that the decreasing rate (w.r.t. T) is slower than that of systems with closed loop training ($\log T/T$). From Fig. 4, an increase spectral efficiency is achieved by sequential beamforming strategy at low inter-node interference level.

Fig. 4 plots closed and open loop training based systems with power controlled sequential beamforming, sequential beamforming, and half-duplex strategy. We observe a further spectral efficiency increase by allowing power adaptation during training. Having established performance bounds for the spectral efficiency loss of the sequential beamforming policy, we now investigate the performance of the half-duplex counterpart to compute the gains of the proposed sequential beamforming strategy.

Theorem 7: The spectral efficiency loss of closed loop training based half-duplex systems with respect to genie-aided system is upper bounded as

$$\Delta \mathrm{SE}_{\mathrm{Hf_{Cl}}}\left(T^{*}_{\mathrm{Hf_{Cl}}}\right) \leqslant R^{\mathrm{ZF}} \frac{M(M-1)}{\log(1+fP)} \frac{\log T}{T} + o\left(\frac{\log T}{T}\right). \tag{24}$$

Proof. See Appendix H. ∎

Here we observe the same scaling of spectral efficiency loss with respect to the number of antennas M and block length T as in sequential beamforming strategy with closed loop training. Actually, we can obtain Theorem 7 by replacing the normalized marginal cost term $\frac{M-1}{2M}\Delta R^{\mathrm{INI}} + \frac{M+1}{2M}R^{\mathrm{ZF}}$ in Theorem 5 with R^{ZF}. The main reason is the similarity

between the marginal utility term in sequential beamforming and half-duplex system.

Theorem 8: The spectral efficiency loss of open loop training based half-duplex systems with respect to genie-aided system is upper bounded as

$$\Delta \mathrm{SE}_{\mathrm{Hf_{O_p}}}\left(T^{*}_{\mathrm{Hf_{O_p}}}\right) \leqslant 2\sqrt{\frac{(M-1)R_{\mathrm{ZF}}}{fT}}. \tag{25}$$

Proof. See Appendix I. ∎

The spectral efficiency loss scaling with both block length T and the number of antennas M are identical to that of sequential beamforming system with open loop training. Theorem 8 can be viewed as changing the normalized marginal cost of sequential beamforming strategy into its half-duplex counterpart. Comparing Theorem 7 and Theorem 8 to their sequential beamforming counterparts, spectral efficiency loss is substantially reduced by adopting sequential beamforming strategy. Sequential beamforming strategy improves spectral efficiency performance significantly.

We now further validate the sequential beamforming strategy for an 8×8 system with experimental data from [1] and 3GPP 3D channel model [2]. The number of base station antennas is chosen to be 8, which is the maximum number currently supported by LTE. In the experiment, the authors of [1] measure the channel realization between an 8×9 two-dimensional antenna array and 12 randomly located users. The measurement is conducted in both indoor and outdoor environment. In 3GPP channel generation, the users are randomly uniformly placed in Urban Macrocell with equal probability to be indoor and outdoor. The simulations for systems containing self-interference is modeled as follows. The base station has a transmit power budget of 0 dBm and a noise floor of -90 dB. Self-interference is managed through a combination of transmitter-receiver isolation of 40 dB, analog domain cancellation of 30 dB, and digital domain cancellation of 30 dB.

The spectral efficiency of both sequential beamforming and half-duplex counterparts are evaluated through Monte Carlo simulations with $10,000$ iterations for both the proposed strategy and half-duplex counterpart. In each iteration, 8 random users and 8 antennas in a random horizontal antenna array are simulated. For closed loop training, feedback bits are equally divided into real and imaginary parts with 3 bits for integer part and rest for the fractional part. In the simulation, we assume that the downlink transmit power in the training phase is adapted, which is not allowed in the previous theoretical analysis due to the analytical intractability in optimal power allocation when imperfect CSI exists. Thus, in cycle $i > 1$, when Users $1, .., i-1$ receive data on downlink, each receiving user signal will be precoded with power constraint $\frac{P}{i-1}$.

Fig. 5 confirms the spectral efficiency improvement achieved by sequential beamforming. The spectral efficiency improvement achieved in Fig. 4 is also shown for reference. Similar to results in Fig. 4, sequential beamforming demonstrates a significant spectral efficiency improvement.

For example, in a typical LTE system, there are around 500 to 2100 symbols in each slot depending on the available
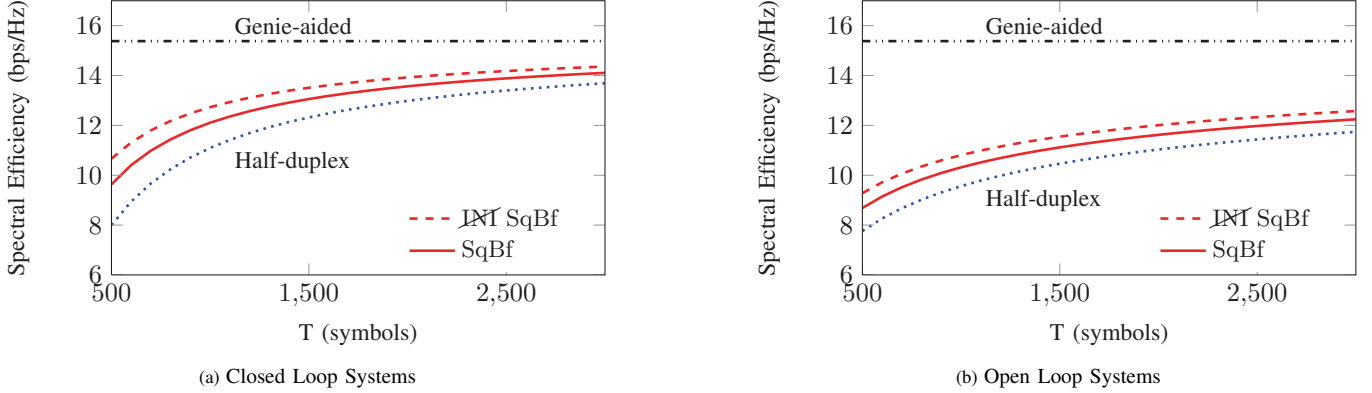
(a) Closed Loop Systems

(b) Open Loop Systems

Fig. 4: Spectral efficiency of an 8×8 system with optimal training duration with and without inter-node interference and half-duplex counterparts with $P = 15$ dB for $f = 0.1$.
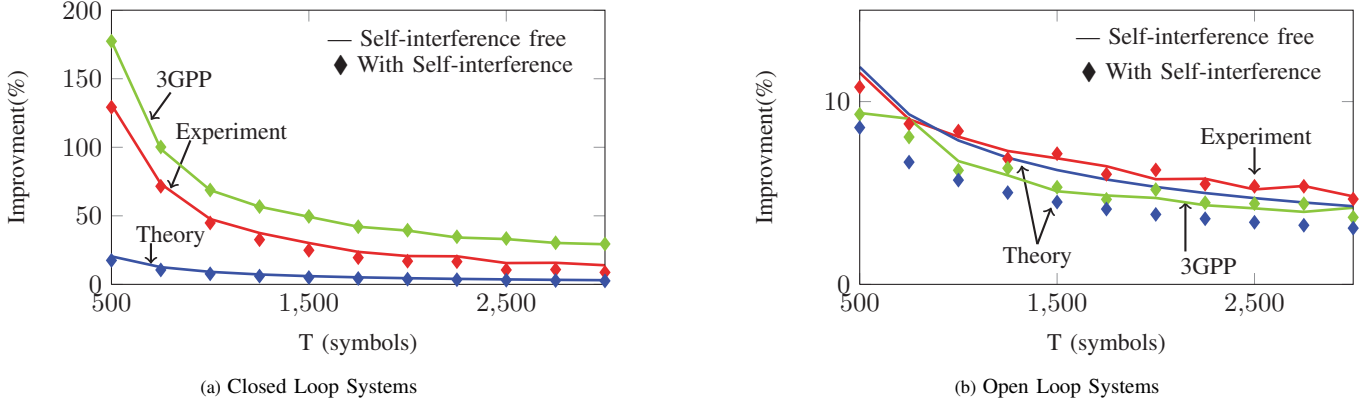

(a) Closed Loop Systems

(b) Open Loop Systems

Fig. 5: Percent spectral efficiency improvement of an 8×8 system with sequential beamforming strategy for systems with and without self-interference at $P = 15$ dB with $f = 0.1$ and $\alpha = 0.3$. Here experiment and 3GPP refer to the simulation results obtained with experimental data [1] and 3GPP 3D channel model [2], respectively. Theory refers to results in Theorem 5-8.

bandwidth (1.4 MHz to 5 MHz). When the block length equals 500 symbols, proposed sequential beamforming strategy attains an over 130% and 12% spectral improvement under the influence of inter-node interference for closed and open loop training systems, respectively. As T grows, the performance of half-duplex counterparts grows. Thus the improvement in sequential efficiency decreases. From Fig. 4, we conclude that a notable spectral efficiency improvement is still observed even for systems with long block length ($T = 3000$). Lower inter-node level does show a better spectral efficiency improvement in Fig. 5.

Remark 1: For closed loop systems, sequential beamforming demonstrates a higher spectral efficiency compared to the results in Fig. 4a, where the base station serves each downlink users at a fixed power P/M in the training phase. This improvement suggests that proper power adaptation can increase the performance of sequential beamforming dramatically. On the other hand, we find power adaptation does not influence the spectral efficiency improvement of open loop system.

In this section, significant spectral efficiency improvement by adopting sequential beamforming is observed. In Section VI, we compare the spectral efficiency asymptotically where equal power allocation and ZF can achieve the full multiplexing gain.

VI. HIGH SNR ANALYSIS

In Section IV and SectionV, with optimized training duration, sequential beamforming strategy exhibits significant spectral efficiency improvement in the finite SNR regime. We continue our investigation of sequential beamforming strategy in the high SNR regime where equal power allocation and ZF can achieve the full multiplexing gain. Notation $\doteq$ is used to denote exponential equality, i.e.,

$$g\left(P\right) \doteq P^{\zeta} \Leftrightarrow \lim_{P \to \infty} \frac{\log g\left(P\right)}{\log P} = \zeta.$$

Since $fP \doteq P$, we now assume the power constraint for training is P^{ζ} to account for the limitation of training power. We use a multiplexing gain metric r, which can be mathematically captured as

$$\lim_{P \to \infty} \frac{\text{SE}_S\left(\zeta, T^{\text{tr}}\right)}{\log P} \doteq r_s, \quad s \in \{\text{SqBf}_{\text{Cl}}, \text{SqBf}_{\text{Op}}, \text{Hf}_{\text{Cl}}, \text{Hf}_{\text{Op}}\}.$$
(26)

Our objective is to maximize the spectral efficiency asymptotically under certain training power constraint, which is mathematically captured as, for $s = \text{SqBf}_{\text{Cl}}, \text{SqBf}_{\text{Op}}, \text{Hf}_{\text{Cl}}$ and Hf_{Op},

$$\max_{T^{\text{tr}}} \quad r_s\left(\zeta, T^{\text{tr}}\right).$$
(27)

We first present the results regarding sequential beamforming system with closed loop training. The results for sequential beamforming strategy with open loop training then follows.

In the asymptotic characterization of sequential beamforming strategy, for mathematical concision, we consider the fraction of full-duplex transmission term $\frac{M-1}{2M}$ in (7) to be $\frac{1}{2}$. This approximation is valid for systems with large numbers of antennas.

A. Sequential Beamforming with Closed Loop Training

In this subsection, we consider the relationship between multiplexing gain r and training power constraint ζ. Similar to the approach in the finite SNR regime; we first present a lemma capturing the influence of inter-beam and inter-node interference in the high SNR regime, then the spectral efficiency is characterized. We define $\theta = \left(M \left(M - 1 \right) \right)/T$, which is useful in analysis.

Lemma 2: In closed loop systems, the downlink data transmission rate during training phase, under the influence of inter-beam and inter-node interference, is

$$\lim_{P \to \infty} \frac{R^{\mathrm{tr}} \left(T^{\mathrm{tr}} \right)}{\log P} = \max \left(\min \left(\frac{\zeta}{\theta} \frac{T^{\mathrm{tr}}}{T}, 1 - \zeta \right), 0 \right). \quad (28)$$

Proof. The proof is obtained by substituting the training power in Appendix A fP with P^{ζ}. ∎

Interestingly, we observe the impact of inter-node interference in the high SNR regime to be divided into two scenarios. If only coarse CSI is available, the influence of inter-beam interference dominates the rate performance during training, i.e., there is no impact of inter-node interference on performance. Otherwise, the influence of inter-node interference dominates the rate performance during training. Now we present the maximal multiplexing gain as a function of training power constraint ζ for different closed loop training systems.

Applying Lemma 2 to characterize (7), we have

$$\lim_{P \to \infty} \frac{\mathrm{SE}_{\mathrm{SqBf_{Cl}}}}{\log P} = \frac{1}{2} \frac{T^{\mathrm{tr}}}{T} \max \left(\min \left(\frac{T^{\mathrm{tr}}}{T} \frac{\zeta}{\theta}, 1 - \zeta \right), 0 \right)$$
$$+ \left(1 - \frac{T^{\mathrm{tr}}}{T} \right) \min \left(\frac{T^{\mathrm{tr}}}{T} \frac{\zeta}{\theta}, 1 \right). \quad (29)$$

The results regarding sequential beamforming system without inter-node interference are first presented as an upper bound for the performance of proposed strategy. Then the results regarding the half-duplex systems are presented for comparison. Finally, the performance of sequential beamforming system with inter-node interference is presented.

Theorem 9 (Inter-node interference free sequential beamforming strategy): The maximal multiplexing gain of sequential beamforming strategy with closed loop training, without inter-node interference, under training power constraint ζ is

$$r_{\mathrm{SqBf_{Cl}INI}}^*(\zeta) = \begin{cases} \frac{1}{2} \frac{\zeta}{\theta}, & \zeta < \theta \\ 1 - \frac{1}{2} \frac{\theta}{\zeta}, & \zeta \geqslant \theta \end{cases}.$$

Proof. The multiplexing gain of sequential beamforming without inter-node interference is

$$r_{\mathrm{SqBf_{Cl}INI}}(\zeta, T^{\mathrm{tr}}) = \left(1 - \frac{1}{2} \frac{T^{\mathrm{tr}}}{T} \right) \min(\frac{\zeta}{\theta} \frac{T^{\mathrm{tr}}}{T}, 1).$$

By maximizing the multiplexing gain in the cases of $\frac{\zeta}{\theta} \frac{T^{\mathrm{tr}}}{T} \geqslant 1$ and $\frac{\zeta}{\theta} \frac{T^{\mathrm{tr}}}{T} < 1$ by choosing the optimal training duration, the theorem is directly obtained. ∎

The multiplexing gain is composed of two regimes. When ζ is small, spectral efficiency increases linearly as training power increases. In this regime, the growth of rate performance during and after training is the primary reason. As more training power is allowed, users send training symbols until no spectral efficiency loss is observed due to inter-beam interference. The spectral efficiency improvement attributes to using less time to send the same amount of training information. Thus, spectral efficiency performance increases less as training power grows. Now the asymptotic performance of half-duplex counterpart is presented for comparison.

Theorem 10 (Half-duplex system): The maximal multiplexing gain of closed loop training half-duplex system under training power constraint ζ is

$$r_{\mathrm{Hf_{Cl}}}^*(\zeta) = \begin{cases} \frac{1}{4} \frac{\zeta}{\theta}, & \zeta < 2\theta \\ 1 - \frac{\theta}{\zeta}, & \zeta \geqslant 2\theta \end{cases}.$$

Proof. Similar to inter-node interference free sequential beamforming strategy, omitting the extra spectral obtained during full-duplex training (VI-A), we first express the multiplexing gain of half-duplex system as

$$r_{\mathrm{Hf_{Cl}}}(\zeta, T^{\mathrm{tr}}) = \left(1 - \frac{T^{\mathrm{tr}}}{T} \right) \min(\frac{\zeta}{\theta} \frac{T^{\mathrm{tr}}}{T}, 1).$$

Directly optimizing training duration in cases of $\frac{\zeta}{\theta} \frac{T^{\mathrm{tr}}}{T} \geqslant 1$ and $\frac{\zeta}{\theta} \frac{T^{\mathrm{tr}}}{T} < 1$ leads to the proof. ∎

It should be noted that similar to sequential beamforming strategy with closed loop training, its half-duplex counterpart's spectral efficiency consists of two regimes. Compared to Theorem 9, a significant multiplexing gain improvement is observed. Thus, the proposed sequential beamforming strategy doubles the spectral efficiency of a unidirectional downlink communication asymptotically when $\zeta < \theta$. Finally, we look at the influence of inter-node interference on the asymptotic spectral efficiency of sequential beamforming strategy.

Theorem 11 (Sequential beamforming strategy with inter-node interference): The maximal multiplexing gain of closed loop training sequential beamforming strategy under training power constraint ζ is

$$r_{\mathrm{SqBf_{Cl}}}^*(\zeta) = \begin{cases} r_{\mathrm{SqBf_{Cl}INI}}^*(\zeta), \zeta \leqslant \frac{3\theta}{2+3\theta} \\ r_{\mathrm{SqBf_{Cl}INI}}^*(\zeta), \frac{3\theta}{2+3\theta} < \zeta < \min \left(1, \max \left(\frac{3\theta}{2+3\theta}, \frac{3\theta}{2-\theta} \right) \right) \\ r_{\mathrm{Hf_{Cl}}}^*(\zeta), \min \left(1, \max \left(\frac{3\theta}{2+3\theta}, \frac{3\theta}{2-\theta} \right) \right) \leqslant \zeta \end{cases},$$

where

$$r_{\mathrm{SqBf_{Cl}INI}}^*(\zeta) = \begin{cases} \frac{((2-\theta)\zeta+\theta)^2}{16\zeta\theta}, & \zeta < \frac{3\theta}{2-\theta} \\ 1 - \frac{\theta}{2} - \frac{\theta}{2\zeta}, & \frac{3\theta}{2-\theta} \leqslant \zeta \end{cases}.$$

Proof. Detailed proof can be found in Appendix J. ∎

The influence of inter-node interference on the spectral efficiency, interestingly, can be divided into three regimes. For systems targeting small multiplexing gain, only small amount of training power is needed. In this regime, inter-beam interference dominates the downlink performance during

full-duplex training and no inter-node interference penalty is observed. However, if the higher multiplexing gain is targeted, the inter-node interference dominates the downlink performance during full-duplex training. In this case, inter-node interference will reduce the potential benefit obtained from sequential beamforming strategy. Finally, if a relatively high training power ($\zeta > 1$) is used to achieve high spectral efficiency, the high inter-node interference level leads to no benefit from the downlink transmission during the training phase. This observation is confirmed by simulations shown in Fig. 6.

Remark 2: As the number of antennas increases (with respect to block length), θ increases. Interestingly, we observe that higher θ actually increases the regime where sequential beamforming strategy does not suffer from inter-node interference. As $\theta \to \infty$, sequential beamforming strategy suffers no inter-node interference as long as training power constraint is smaller than 1.

B. Sequential Beamforming with Open Loop Training

Following the same approach in Section VI-A, we now investigate the spectral efficiency for different open loop training based systems. The influence of inter-beam interference in high SNR regime is first characterized, which is followed by the multiplexing gain analysis.

Lemma 3: For open loop training based systems, the rate performance under inter-beam interference after training is

$$\lim_{P \to \infty} \frac{R^{\mathrm{data}}}{\log P} = \lim_{P \to \infty} \frac{R^{\mathrm{ZF}} - \log \left[1 + \frac{P}{M} \frac{M-1}{1 + \frac{T^{\mathrm{tr}}}{M} P^\zeta} \right]}{\log P}$$
$$= \max \left(1 - \zeta, 0 \right).$$

Proof. Substituting the training power to P^ζ in Proposition 1 leads to the theorem. ∎

The rate performance achieved after training, surprisingly, is only decided by the training power constraint ζ. More training symbols do not help to reduce inter-beam interference after training. Thus, the optimal training duration goes to zero in high SNR regime. Therefore, the maximal multiplexing gain performance is obtained as follows.

Theorem 12: For open loop training based systems, the multiplexing gain of both sequential beamforming and half-duplex strategy is only decided by training power constraint as

$$r^*_{\mathrm{SqBf_{Op}}} = r^*_{\mathrm{Hf_{Op}}} = \zeta. \tag{30}$$

This theorem is valid for both half-duplex and sequential beamforming strategy with open loop training. Sequential beamforming strategy does not provide extra spectral benefit at high SNR. It should be emphasized that in low-to-moderate SNR regime, from the analysis in Section V, sequential beamforming does obtain significant spectral efficiency gain. This difference is not observed for closed loop training systems, where significant spectral efficiency improvement is observed in all SNR regime.

Before comparing the spectral efficiency performance of closed and open loop system asymptotically, we first validate the influence of the assumption that each user has a perfect knowledge of its channel on both closed and open loop training. This information is crucial for the decoding at the user side. It has been shown [4] that, asymptotically, one training pilot from each base station antenna is both necessary and good enough for the influence of imperfect CSI on downlink rate to vanish for both closed and open loop training.

Based on multiplexing gain characterization above, we compare the spectral efficiency of systems with different types of training. For systems with longer block length and few users, in general, closed loop training outperforms open loop training. The major reason is that the closed loop training significantly reduces inter-beam interference by learning from more training symbols. Despite the longer training duration, closed loop training is still more advantageous. However, if there are many antennas and block length is short, then it is better to use open loop training, whose training duration is asymptotically short.

VII. Conclusion

With accurate CSI, Multiuser MIMO downlink has the potential to increase the spectral efficiency tremendously. In systems with many users, CSI acquisition leads to unavoidable training overhead. This paper aims at reducing the overhead of multiuser MIMO downlink systems by utilizing full-duplex radios.

With characterization of inter-node interference due to full-duplex training, we optimize the training duration of sequential beamforming strategy with both closed and open loop training. The proposed sequential beamforming strategy demonstrates significant spectral efficiency improvement compared to its half-duplex counterpart. Sequential beamforming strategy can also be applied in frequency-division duplex systems where uplink and downlink are orthogonal by nature. The orthogonality will also prevent the generation of inter-node interference.

The closed and open loop training methods exhibit distinct spectral efficiency performance. Asymptotically, the number of users, block length and training power jointly decide which type of training should be adopted. It has been observed in [25] that closed loop training is more favorable than open loop training to reduce estimation error, while common sense suggests that open loop training is preferable for large systems. Our results quantify the decision region of training method and bridge these two observations.

We close this paper by noting some relevant issues. The training symbols are symmetrically allocated among users. Since the downlink receiving time decreases from User 1 to User M, extra spectral efficiency can be obtained by assigning training symbols decreasingly from User 1 to User M. Moreover, we focus on providing the spectral efficiency characterization of a particular users subset to the scheduler. Thus, the MAC layer designs, such as deafness effect, are not considered. Past studies have revealed that picking users whose channel are more orthogonal reduces the inter-beam interference [16], [17]. Finally, the considered base station is a full-duplex MIMO array where each antenna is used for both transmission and reception [5]. Another line of
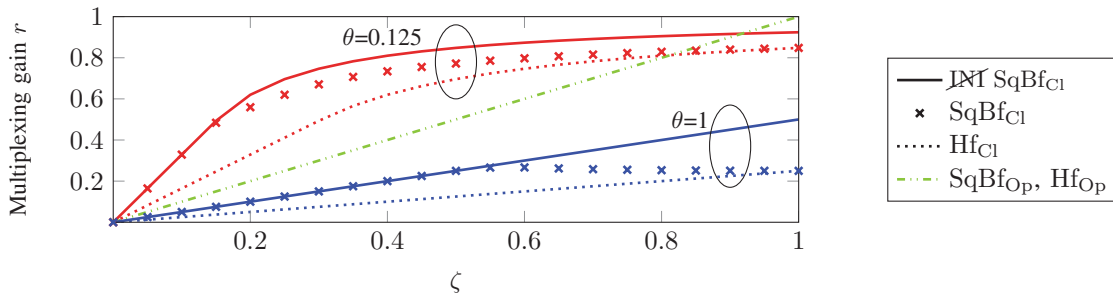
Fig. 6: Multiplexing gain r as a function of training power constraint ζ.

recent development of full-duplex designs [1], [6]–[9] adopt separating transmission and reception antennas. For such array, the joint design and optimization of data service, CSI training, and full-duplex operation is an important future direction.

APPENDIX A
PROOF OF LEMMA 1

Since perfect CSI is assumed available at each user, the rate loss can be upper bounded by

$$
\begin{aligned}
\Delta R^{\mathrm{tr}} \leqslant &\; \mathbb{E}\left[\log\left(1 + |\mathbf{h}_i \mathbf{v}_i|^2 \frac{P}{M}\right)\right] \\
&- \mathbb{E}\left[\log\left(1 + |\mathbf{h}_i \mathbf{v}_i|^2 \frac{P}{M} + |\mathbf{h}_{ik} x_{\mathrm{tr}_k}|^2\right)\right] \\
&+ \mathbb{E}\left[\log\left(1 + \sum_{j \neq i} \frac{P}{M}|\mathbf{h}_i \mathbf{v}_j|^2 + |\mathbf{h}_{ik} x_{\mathrm{tr}_k}|^2\right)\right].
\end{aligned}
\tag{31}
$$

The first step is to follow the same recipe in Appendix II step (a) of [4], and then ignoring the positive term $\sum_{j \neq i} \frac{P}{M}|\mathbf{h}_i \mathbf{v}_j|^2$ leads to the result above. The sum $\sum_{j \neq i} \frac{P}{M}|\mathbf{h}_i \mathbf{v}_j|^2$ and the term $|\mathbf{h}_{ik} x_{\mathrm{tr}_k}|^2$ quantify the interference due to imperfect precoding and full-duplex training, i.e., inter-beam and inter-node interference, respectively. We refer them as $\mathcal{P}_{\mathrm{IBI}}$ and $\mathcal{P}_{\mathrm{INI}}$. The inter-node interference is assumed to scale proportional to the training power of users, i.e., $\mathcal{P}_{\mathrm{INI}} = \alpha f P$. Noting the concavity of logarithm function, similar to step (33) in Remark 4.2 in [4], we apply Jensen's inequality to obtain Lemma 1.

APPENDIX B
PROOF OF THEOREM 1

The rate increase term $\delta R^{\mathrm{data}}\left(T^{\mathrm{tr}}\right)$ of (IV-A) is characterized by applying Proposition 1 as

$$
\delta R^{\mathrm{data}}\left(T^{\mathrm{tr}}\right) \approx P\left(1 + fP\right)^{-\frac{T^{\mathrm{tr}}}{M(M-1)}}\left[\log\left(1 + fP\right)^{\frac{1}{M-1}} - 1\right].
$$

Here the last step is directly obtained by using Taylor expansion. Substituting into (IV-A), we have

$$
\begin{aligned}
&\left(1 - \frac{M+1}{2M}\frac{T^{\mathrm{tr}}}{T}\right) P\left(1 + fP\right)^{-\frac{T^{\mathrm{tr}}}{M(M-1)}}\left[(1 + fP)^{\frac{1}{M-1}} - 1\right] \\
&= \frac{M-1}{2T}\Delta R^{\mathrm{INI}} + \frac{M+1}{2T}R^{\mathrm{ZF}}.
\end{aligned}
$$

Noticing that it is an transcendental equation which is challenging to solve. Then omitting the $\frac{M+1}{2M}\frac{T^{\mathrm{tr}}}{T}$ term leads us to the theorem. This approximation is valid for large T.

APPENDIX C
PROOF OF THEOREM 2

For open loop training based systems, the rate improvement due to more training symbols can be obtained by using Proposition 1 as

$$
\delta R^{\mathrm{data}}\left(T^{\mathrm{tr}}\right) \approx \frac{(M-1)M}{f\left(\widetilde{T}^{\mathrm{tr}}_{\mathrm{SqBf}}\right)^2},
$$

where the last step is the direct result of Maclaurin series. Combining with (IV-A) leads to

$$
\left(1 - \frac{M+1}{2M}\frac{\widetilde{T}^{\mathrm{tr}}_{\mathrm{SqBf}}}{T}\right) \frac{(M-1)M}{f\left(\widetilde{T}^{\mathrm{tr}}_{\mathrm{SqBf}}\right)^2} = \frac{M-1}{2T}\Delta R^{\mathrm{INI}} + \frac{M+1}{2T}R^{\mathrm{ZF}},
$$

whose solution leads to the theorem.

APPENDIX D
PROOF OF THEOREM 3

The rate increase term $\delta R^{\mathrm{data}}\left(T^{\mathrm{tr}}\right)$ is immediately approximated by using results from Appendix B. Applying this rate characterization term to evaluate (IV-B) gives

$$
\left(1 - \frac{T^{\mathrm{tr}}}{T}\right) \frac{(1 + fP)^{\frac{1}{M-1}} - 1}{P\left(1 + fP\right)^{-\frac{T^{\mathrm{tr}}}{M(M-1)}}} = \frac{M}{T}R^{\mathrm{ZF}},
$$

which is a transcendental equation. Following the same step in Appendix B, we omit the $\frac{T^{\mathrm{tr}}}{T}$ to obtain the theorem.

APPENDIX E
PROOF OF THEOREM 4

Using the rate increase characterization term $\delta R^{\mathrm{data}}\left(T^{\mathrm{tr}}\right)$ in Appendix C and further applying (IV-B), (19) gives

$$
\left(1 - \frac{T^{\mathrm{tr}}}{T}\right) \frac{(M-1)M}{f\left(\widetilde{T}^{\mathrm{tr}}_{\mathrm{SqBf}}\right)^2} = \frac{M}{T}R^{\mathrm{ZF}},
$$

whose solution is the theorem.

APPENDIX F
PROOF OF THEOREM 5

Evaluating the achieved spectral efficiency of sequential beamforming strategy with approximated optimal training duration $\widetilde{T}^{\mathrm{tr}*}_{\mathrm{SqBf_{Cl}}}$ obtained in Theorem 1 gives upper bound

$$
\begin{aligned}
\Delta \mathrm{SE}_{\mathrm{SqBf_{Cl}}}\left(T^{\mathrm{tr}*}_{\mathrm{SqBf_{Cl}}}\right) \leqslant &\; R^{\mathrm{ZF}} + \frac{M-1}{2M}\frac{\widetilde{T}^{\mathrm{tr}*}_{\mathrm{SqBf_{Cl}}}}{T}\Delta R^{\mathrm{INI}} \\
&- \left(1 - \frac{M+1}{2M}\frac{\widetilde{T}^{\mathrm{tr}*}_{\mathrm{SqBf_{Cl}}}}{T}\right)\left[R^{\mathrm{ZF}} - \log\left(1 + P\left(1 + fP\right)^{-\frac{\widetilde{T}^{\mathrm{tr}*}_{\mathrm{SqBf_{Cl}}}}{M(M-1)}}\right)\right].
\end{aligned}
$$

Omitting the negative term

$$-\frac{M+1}{2M}\frac{\widetilde{T}_{\mathrm{SqBf_{Cl}}}^{\mathrm{tr}*}}{T}\log\left(1+P\left(1+fP\right)^{-\frac{\widetilde{T}_{\mathrm{SqBf_{Cl}}}^{\mathrm{tr}*}}{M(M-1)}}\right)$$

and sorting the small term with respect to $\frac{1}{\sqrt{T}}$ lead to the theorem.

APPENDIX G
PROOF OF THEOREM 6

Following similar analysis as that of Appendix F, we substitute approximation of optimal training duration from Theorem 2 into (7) to characterize the spectral efficiency loss as

$$\Delta\mathrm{SE}_{\mathrm{SqBf_{Op}}}\left(T_{\mathrm{SqBf_{Op}}}^{\mathrm{tr}*}\right)\leqslant R^{\mathrm{ZF}}+\frac{M-1}{2M}\frac{\widetilde{T}_{\mathrm{SqBf_{Op}}}^{\mathrm{tr}*}}{T}\Delta R^{\mathrm{INI}}$$
$$-\left(1-\frac{M+1}{2M}\frac{\widetilde{T}_{\mathrm{SqBf_{Op}}}^{\mathrm{tr}*}}{T}\right)\left[R^{\mathrm{ZF}}-\log\left(1+\frac{(M-1)\frac{P}{M}}{1+f\widetilde{T}_{\mathrm{SqBf_{Op}}}^{*}P/M}\right)\right].$$

Then dropping the negative term $-\frac{M+1}{2M}\frac{\widetilde{T}_{\mathrm{SqBf_{Op}}}^{\mathrm{tr}}}{T}\log\left(1+\frac{(M-1)\frac{P}{M}}{1+f\widetilde{T}_{\mathrm{SqBf_{Op}}}^{*}P/M}\right)$ and sorting small term with respect to $\frac{\log T}{T}$ lead to the theorem.

APPENDIX H
PROOF OF THEOREM 7

Similar to systems adopting sequential beamforming strategy, the spectral efficiency gap of the half-duplex counterparts with respect to the genie-aided scenario can be immediately upper bounded by evaluating the sub-optimal scheme $\mathrm{SE}_{\mathrm{Hf_{Cl}}}\left(\widetilde{T}_{\mathrm{Hf_{Cl}}}^{*}\right)$:

$$\Delta\mathrm{SE}_{\mathrm{Hf_{Cl}}}\left(T_{\mathrm{Hf_{Cl}}}^{\mathrm{tr}*}\right)\overset{(a)}{\leqslant}\frac{T_{\mathrm{Hf_{Cl}}}^{\mathrm{tr}*}}{T}R^{\mathrm{ZF}}+\log\left(1+P\left(1+fP\right)^{-\frac{\widetilde{T}_{\mathrm{Hf_{Cl}}}^{\mathrm{tr}*}}{M(M-1)}}\right)$$
$$=R^{\mathrm{ZF}}\frac{M\left(M-1\right)}{\log\left(1+fP\right)}\frac{\log T}{T}+o\left(\frac{\log T}{T}\right).$$

Inequality (a) is the result of dropping negative term $-\frac{\widetilde{T}_{\mathrm{Hf_{Cl}}}^{*}}{T}\log\left(1+P\left(1+fP\right)^{-\frac{\widetilde{T}_{\mathrm{Hf_{Cl}}}^{\mathrm{tr}*}}{M(M-1)}}\right)$. Applying training time approximation in Theorem 3 gives the final step.

APPENDIX I
PROOF OF THEOREM 8

Inspired by [15], spectral efficiency gap with respect to genie-aided situation can be immediately upper bounded by evaluating $\mathrm{SE}_{\mathrm{Hf_{Op}}}\left(\widetilde{T}_{\mathrm{Hf_{Op}}}^{*}\right)$

$$\Delta\mathrm{SE}_{\mathrm{Hf_{Op}}}\left(T_{\mathrm{Hf_{Op}}}^{*}\right)\leqslant\mathrm{SE}^{\mathrm{ZF}}-\mathrm{SE}_{\mathrm{Hf}}\left(\widetilde{T}_{\mathrm{Hf_{Op}}}^{*}\right)$$
$$\overset{(a)}{\leqslant}\frac{\widetilde{T}_{\mathrm{Hf}}^{*}}{T}R^{\mathrm{ZF}}+\log\left(1+\frac{(M-1)\frac{P}{M}}{1+f\widetilde{T}_{\mathrm{Hf_{Op}}}^{*}P/M}\right)$$
$$\leqslant\sqrt{\frac{(M-1)R^{\mathrm{ZF}}}{fT}}+\sqrt{\frac{R^{\mathrm{ZF}}(M-1)}{fT}}$$
$$=2\sqrt{\frac{(M-1)R^{\mathrm{ZF}}}{fT}}.$$

Inequality (a) is obtained by dropping negative term $-\frac{\widetilde{T}_{\mathrm{Hf_{Op}}}^{*}}{T}\log\left(1+\frac{(M-1)\frac{P}{M}}{1+f\widetilde{T}_{\mathrm{Hf}}^{*}P/M}\right)$. The next step is the result of Maclaurin expansion of the logarithm term, which is tight for large T.

APPENDIX J
PROOF OF THEOREM 11

Studying (VI-A) in different regimes of operation gives the following.
1) When $\zeta\geqslant 1$, the multiplexing gain is the same as the multiplex gain of half-duplex systems. Thus, $r_{\mathrm{SqBf_{Cl}}}^{*}(\zeta)=r^{*}\left(\zeta\right)_{\mathrm{Hf_{Cl}}}$.
2) When $\zeta<1$
   * $\frac{T^{\mathrm{tr}}}{T}\leqslant\theta\frac{1-\zeta}{\zeta}$:
   $$r_{\mathrm{SqBf_{Cl}}}\left(\zeta,T^{tr}\right)=\frac{T^{\mathrm{tr}}}{T}\frac{\zeta}{\theta}-\frac{1}{2}\left(\frac{T^{\mathrm{tr}}}{T}\right)^{2}\frac{\zeta}{\theta}$$
   $$=r_{\mathrm{SqBf_{Cl}}\mathrm{INI}}\left(\zeta,T^{tr}\right)$$
   * $\theta\frac{1-\zeta}{\zeta}<\frac{T^{\mathrm{tr}}}{T}<\frac{\theta}{\zeta}$:
   $$r_{\mathrm{SqBf_{Cl}}}\left(\zeta,T^{tr}\right)=\frac{1}{2}\frac{T^{\mathrm{tr}}}{T}\left(1-\zeta\right)+\left(1-\frac{T^{\mathrm{tr}}}{T}\right)\frac{T^{\mathrm{tr}}}{T}\frac{\zeta}{\theta}$$
   * $\frac{\theta}{\zeta}\leqslant\frac{T^{\mathrm{tr}}}{T}$:
   $$r_{\mathrm{SqBf_{Cl}}}\left(\zeta,T^{tr}\right)=\frac{1}{2}\frac{T^{\mathrm{tr}}}{T}\left(1-\zeta\right)+\left(1-\frac{T^{\mathrm{tr}}}{T}\right)$$

By carefully evaluating the derivative in different regimes and applying the optimized training duration into equation (VI-A) lead to the theorem.

ACKNOWLEDGMENT

REFERENCES

[1] Evan Everett, Clayton Shepard, Lin Zhong, and Ashutosh Sabharwal, "Measurement-driven evaluation of all-digital many-antenna full-duplex communication," *arXiv preprint arXiv:1508.03765*, 2015.
[2] 3GPP, "Study on 3D channel model for LTE," TS 36.873, 3rd Generation Partnership Project (3GPP), 2015.
[3] Syed Ali Jafar and Andrea J Goldsmith, "Isotropic fading vector broadcast channels: The scalar upper bound and loss in degrees of freedom," *Information Theory, IEEE Transactions on*, vol. 51, no. 3, pp. 848–857, 2005.
[4] Giuseppe Caire, Nihar Jindal, Mari Kobayashi, and Niranjay Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *Information Theory, IEEE Transactions on*, vol. 56, no. 6, pp. 2845–2866, 2010.
[5] Dinesh Bharadia and Sachin Katti, "Full duplex MIMO radios," in *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, 2014, pp. 359–372.
[6] Melissa Duarte, Chris Dick, and Ashutosh Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *Wireless Communications, IEEE Transactions on*, vol. 11, no. 12, pp. 4296–4307, 2012.
[7] Mayank Jain, Jung Il Choi, Taemin Kim, Dinesh Bharadia, Siddharth Seth, Kannan Srinivasan, Philip Levis, Sachin Katti, and Prasun Sinha, "Practical, real-time, full duplex wireless," in *Proceedings of the 17th annual international conference on Mobile computing and networking*. ACM, 2011, pp. 301–312.

[8] Ehsan Aryafar, Mohammad Amir Khojastepour, Karthikeyan Sundaresan, Sampath Rangarajan, and Mung Chiang, "MIDU: enabling MIMO full duplex," in *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, pp. 257–268.

[9] Bei Yin, Michael Wu, Christoph Studer, Joseph R Cavallaro, and Jorma Lilleberg, "Full-duplex in large-scale wireless systems," in *2013 Asilomar Conference on Signals, Systems and Computers*. IEEE, 2013, pp. 1623–1627.

[10] Elsayed Ahmed, Ahmed M Eltawil, and Ashutosh Sabharwal, "Self-interference cancellation with nonlinear distortion suppression for full-duplex systems," in *2013 Asilomar Conference on Signals, Systems and Computers*. IEEE, 2013, pp. 1199–1203.

[11] Brian P Day, Adam R Margetts, Daniel W Bliss, and Philip Schniter, "Full-duplex bidirectional MIMO: Achievable rates under limited dynamic range," *Signal Processing, IEEE Transactions on*, vol. 60, no. 7, pp. 3702–3713, 2012.

[12] Ashutosh Sabharwal, Philip Schniter, Dongning Guo, Daniel W Bliss, Sampath Rangarajan, and Risto Wichman, "In-band full-duplex wireless: Challenges and opportunities," *Selected Areas in Communications, IEEE Journal on*, vol. 32, no. 9, pp. 1637–1652, 2014.

[13] Elsayed Ahmed, Ahmed M Eltawil, and Ashutosh Sabharwal, "Rate gain region and design tradeoffs for full-duplex wireless communications," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3556–3565, 2013.

[14] Jingwen Bai and Ashutosh Sabharwal, "Distributed full-duplex via wireless side-channels: Bounds and protocols," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 8, pp. 4162–4173, 2013.

[15] Mari Kobayashi, Nihar Jindal, and Giuseppe Caire, "Training and feedback optimization for multiuser MIMO downlink," *Communications, IEEE Transactions on*, vol. 59, no. 8, pp. 2228–2240, 2011.

[16] Taesang Yoo, Nihar Jindal, and Andrea Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 7, pp. 1478–1491, 2007.

[17] Goran Dimic and Nicholas D Sidiropoulos, "On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm," *Signal Processing, IEEE Transactions on*, vol. 53, no. 10, pp. 3857–3868, 2005.

[18] Xu Du, John Tadrous, Chris Dick, and Ashutosh Sabharwal, "MIMO broadcast channel with continuous feedback using full-duplex radios," in *Proceedings of IEEE Asilomar Conference on Signals, Systems and Computers*, Nov 2014.

[19] Xu Du, John Tadrous, Chris Dick, and Ashutosh Sabharwal, "MU-MIMO beamforming with full-duplex open-loop training," in *Proceedings of IEEE 16th Workshop on Signal Processing Advances in Wireless Communications*. IEEE, June 2015.

[20] Quentin H Spencer, A Lee Swindlehurst, and Martin Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *Signal Processing, IEEE Transactions on*, vol. 52, no. 2, pp. 461–471, 2004.

[21] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," TS 36.213, 3rd Generation Partnership Project (3GPP), Jan. 2016.

[22] William J Baumol and Prentice Hall, "Economic theory and operations analysis," 1977.

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

[24] Babak Hassibi and Bertrand M Hochwald, "How much training is needed in multiple-antenna wireless links?," *Information Theory, IEEE Transactions on*, vol. 49, no. 4, pp. 951–963, 2003.

[25] Thomas L Marzetta and Bertrand M Hochwald, "Fast transfer of channel state information in wireless systems," *Signal Processing, IEEE Transactions on*, vol. 54, no. 4, pp. 1268–1278, 2006.