Full Dimension MIMO for Frequency Division Duplex under Signaling and Feedback Constraints

Martin Kurras, Lars Thiele and Thomas Haustein
Fraunhofer Institute for Telecommunications
Heinrich Hertz Institute
Einsteinufer 37, 10587 Berlin, Germany
Email: martin.kurras@hhi.fraunhofer.de

Wang Lei, Chen Yan
Central Research Institute
Huawei Technologies Co., Ltd.
No. 2222 Xin Jinqiao Road, Shanghai, China, 201206

***Abstract*—It is a open research problem of great interest how to fully utilize the benefits of massive MIMO original designed for Time Division Duplex (TDD) also in Frequency-Division Duplexing (FDD). For efficient multi-user downlink transmission Channel State Information (CSI) has to be obtained at the receiver side and fed back to the transmitter. Due to signaling/pilot overhead scaling with the number of antennas and the feedback of the CSI the beamforming gain and spatial multiplexing gains of massive MIMO are limited.**

In this paper we present a novel scheme called User Non-aware Precoding and Effective Channel Estimation (UNP & ECE) for full dimension MIMO in FDD under signaling and feedback constraints. The idea is to construct a large codebook and distribute subsets of it on time-frequency resources as precoded pilots. The same precoders are also used for downlink transmission thus no additional demodulation reference signals are required. This limits the number of signaling overhead and a high quantization of the channel can be utilized. To limit the feedback we show that it is sufficient to report only the strongest stream index of a precoder subset for only a subset of Time Frequency (TF) resources. By sending just the index of the best stream the devices don't require knowledge of the applied precoders. This enables optimization of the codebook construction, the splitting in subsets, the distribution on the TF resources at the base station without awareness at the users.

I. INTRODUCTION

Spectrum for mobile communication is a finite resource. With respect to the growing demand in throughput it is of high interest to further increase the spectral efficiency (per area) in already available spectrum for cellular mobile communications systems, which is below 6 GHz [1]. Higher order spatial multiplexing enabled by *massive* Multiple-Input Multiple-Output (MIMO) is considered as a solution. Furthermore, *massive* MIMO will most likely be an integral component of flexible next generation 5G systems [2]. Original designed for Time Division Duplex (TDD) [3] it is still an open research problem how to operate massive MIMO equally efficient in Frequency-Division Duplexing (FDD) mode due to signaling overhead and feedback constraints, see [4]. This topic has gained more and more relevance since the promising advantages of *massive* MIMO already been verified by hardware demonstrations assuming TDD. However, most of the available spectrum in the sub 6 GHz is assigned to FDD by regulation [1]. One solution is the exploitation of spatial channel correlation as in [5] or another hybrid beamforming [6]. There is also a study item in 3rd generation partnership project (3GPP) standardization called full dimension (FD) MIMO which explicitly targets massive MIMO in FDD [7]. Therein the construction of codebooks used for downlink precoding as in Long Term Evolution (LTE) is straightforward extended for two dimensional (2D) antenna arrays by the Kronecker product of two codobooks for linear antenna arrays. A drawback already formulated in [4] is that by increasing the number of pilots (or Common Reference Signals (CRSs)) the overhead scales linear and limits overall spectral efficiency by limiting the number simultaneously served devices.

In this paper we tackle the problem of the increasing signaling overhead by restricting the number of maximum active streams but utilizing the advantage of a large codebook (channel quantization) to increase the spectral efficiency per beam. Thus in the range of a moderate number of users (up to 500) the performance in terms of sum spectral efficiency between TDD and FDD is expected to be similar.

The idea is called User Non-aware Precoding (UNP) by using precoded CRS and the same precoder for data transmissions and therefore gets rid of the overhead from Demodulation Reference Signal (DMRS). The overhead of the high number of CRS from a large codebook is limited by distribution of limited subsets of the codebook over Time Frequency (TF) resources. This also limits the feedback per device which has to be send to the Base Station (BS) for downlink transmission. We show that it is sufficient to report only the strongest stream index (similar to Precoding Matrix Indicator in LTE) and Channel Quality Indicator (CQI) value to achieve most of the performance. Furthermore, our proposed scheme also operates very efficient in an open loop mode. Also an important aspect is that multiple antennas on the user side for interference cancellation are inherently utilized in the feedback and therefore any number of antennas at the device side is supported.

II. SYSTEM MODEL

We consider a system with K devices and M BSs , where $\mathcal{M}$ is the set of BSs indices $\{1, \ldots, M\}$ and $\mathcal{K}$ the set of device indices $\{1, \ldots, K\}$. Each device is equipped with N_r receive antennas and each BS with N_t transmit antennas. With

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

this the receive signal $\mathbf{y}_k$ at Mobile Station (MS) $k \in \mathcal{K}$ is given by

$$\mathbf{y}_k = \sum_{m=1}^{m\in\mathcal{M}} \mathbf{H}_{k,m}\sqrt{\mathbf{P}_m}\mathbf{B}_m\mathbf{x}_m + \mathbf{n}_k, \tag{1}$$

where $\mathbf{H}_{k,m}$ is the $N_r \times N_t$ channel matrix from BS m to device k, $\mathbf{P}_m$ is the $N_t \times N_t$ diagonal power allocation matrix with the power constraint $P_{\text{max}} \geq \text{trace}(\mathbf{P})$, $\mathbf{B}_m$ is the $N_t \times T$ precoding matrix applied by BS m, $\mathbf{x}_m$ is the $T \times 1$ vector of transmit symbols and $\mathbf{n}_k$ denotes the Additive White Gaussian Noise (AWGN) vector with covariance $\mathbb{E}[\mathbf{n}\mathbf{n}^H] = \mathbf{I}\sigma_\nu^2$. $\mathbb{E}[.]$ is the expectation value and σ_ν^2 comprises the noise power comprising the receiver noise figure and thermal noise. Each column of the precoding matrix $\mathbf{B}_m = [\mathbf{b}_1, .., \mathbf{b}_T]$ is a spatial multiplexed stream or layer and assigned to a certain device and $\mathcal{T}$ is the set of stream indices $\{1, ..., T\}$ with the constraint $T \leq N_t$. Assuming that stream $t \in \mathcal{T}$ from BS l is allocated to device k the receive signal from (1) can be divided into three parts according to

$$\begin{aligned}\mathbf{y}_{k,t} = {} & \underbrace{\mathbf{H}_{k,l}\sqrt{p_t}\mathbf{b}_{t,l}}_{\tilde{\mathbf{h}}_{k,t}} x_{t,l} + \underbrace{\sum_{j\neq t}^{j\in\mathcal{T}} \mathbf{H}_{k,l}\sqrt{p_j}\mathbf{b}_{j,l}x_{j,l}}_{\boldsymbol{\vartheta}_{k,t}} \\ & + \underbrace{\sum_{m\neq l}^{m\in\mathcal{M}} \mathbf{H}_{k,m}\sqrt{\mathbf{P}_m}\mathbf{B}_m\mathbf{x}_m + \mathbf{n}_k}_{\mathbf{z}_k}, \end{aligned} \tag{2}$$

where the effective channel is denoted by $\tilde{\mathbf{h}}_{k,t}$, the intra-sector interference caused by streams $j \neq t$ is aggregated in $\boldsymbol{\vartheta}_{k,t}$ and streams from other BSs are denoted as inter-sector interference in $\mathbf{z}_k$. At the device we assume a linear receive beamformer based on the estimated channel denoted by $\mathbf{w}_{k,t}$ as a $1 \times N_r$ vector. With this the effective Signal to Interference and Noise Ratio (SINR) is obtained by:

$$\text{SINR}_{k,t} = \frac{\mathbf{w}_{k,t}\tilde{\mathbf{h}}_{k,t}\tilde{\mathbf{h}}_{k,t}^H\mathbf{w}_{k,t}}{\mathbf{w}_{k,t}\mathbf{Z}_{k,t}\mathbf{w}_{k,t}^H}, \tag{3}$$

where $\mathbf{Z}_{k,t}$ is the interference covariance matrix plus noise obtained by $\mathbf{Z}_{k,t} = \mathbb{E}[\boldsymbol{\vartheta}\boldsymbol{\vartheta}_{k,j}^H + \mathbf{z}_k\mathbf{z}_k^H + \mathbf{n}_k\mathbf{n}_k^H]$ and $(\cdot)^H$ is the operator for transpose conjugate complex of a vector or matrix.

III. DISTRIBUTED FULL DIMENSION PRECODER

We consider a $[N_x, N_y]$ antenna array consisting of the same elements, where N_x and N_y are number of the antennas in vertical and horizontal dimension respectively and $N_t = N_xN_y$. Similar to the 2D codebook construction in [7] separate codebooks for horizontal and vertical beamforming are obtained by the Kronecker product, see [8]. The Kronecker product assumes inherently independent correlation properties in horizontal and vertical direction, which is a valid assumption [9]. However the construction of these codebooks is a research topic itself and out of the scope of this paper (see [9] and references therein) and Discrete Fourier Transform (DFT) matrices are used justified by the following arguments. We know that DFT based codebooks are robust and well suited for spatially correlated channels [10] and the construction of any number of antennas N_t is straightforward. A DFT matrix is also a structured, which means that neighbor columns correspond to neighbor beams and each column-vector is orthogonal to all other, so they form an orthogonal basis. Furthermore, the same norm per column ensures the preferred equal gain per stream. At last there are very efficient algorithms and hardware available for implementation supporting the use of DFT matrices. It was also shown in [11] that DFT matrices are optimal under certain assumptions.

The codebook is represented by a large matrix $\mathbf{B}_{\text{GoB}}$ of dimension $[N_t \times S]$ where S is a design parameter $S = S_{\text{ver}}S_{\text{hor}}$ depending on the dimension of the antenna array and quantization of the channel in azimuth domain by S_{hor} or vertical domain by S_{hor}. Without limiting the generality we assume $S = N_t$ and split $\mathbf{B}_{\text{GoB}}$ into N_q subsets $\mathbf{B}_q$ of size $[N_t \times N_s]$ with $N_s \leq N_t$, where N_s is the number of streams per subsets and $N_q = \lceil S/N_s \rceil$. Each subset $\mathbf{B}_q$ with $q = 1 \dots N_q$ is assigned to a TF resource. The subscript GoB is an acronym for Grid of Beams. The idea of the distributed codebook or precoder is illustrated in Fig. 1 for $N_t = S = 256 = N_xN_y = 8 \cdot 32$, and $N_s = 16$. Note, that the matrix of a subset allocated to a certain TF resource in Fig. 1 represents beams with different azimuth and elevation angles. Each active beam (green squares) is a column vector of size $[N_t \times 1]$ of the precoder subset $\mathbf{B}_q$.

IV. USER NON-AWARE PRECODING AND FEEDBACK

A. *Effective Channel Estimation*

With the distribution of the precoder subsets on different TF resources according to Sec. III each user estimates an effective channel per resource given by $\tilde{\mathbf{H}}_q = \mathbf{H}\mathbf{B}_q$ of size $[N_r \times N_s]$ from the precoded pilots. Compared to the original channel size of $[N_r \times N_t]$ the signaling overhead is reduced by the factor N_t/N_s.

Remark 1. *The distribution of the precoders can be consecutive similar to subbands in LTE, distributed as illustrated in Fig. 1 or a mix of both. An optimization on the distribution of the precoders with respect to certain channel conditions is for further study and not in the scope of this paper.*

The idea is that each user finds at least on some resource blocks a beam with high SINR. It is clear from multi-path and non line of sight propagation that inter-beam interference is the limiting factor and the distribution of the beams per subset has to be optimized to maximize the SINR per beam. In Fig. 1 a particular distribution is shown where the objective was to maximize the distance between two neighboring beams. This optimization also depends on the coverage in the azimuth and elevation domain of the beams and is affected by the geometry and placement of the deployed antenna, e.g. above roof top or on a wall.

Due to the fact that users directly estimate the effective channel based on the precoded reference signals, and the same

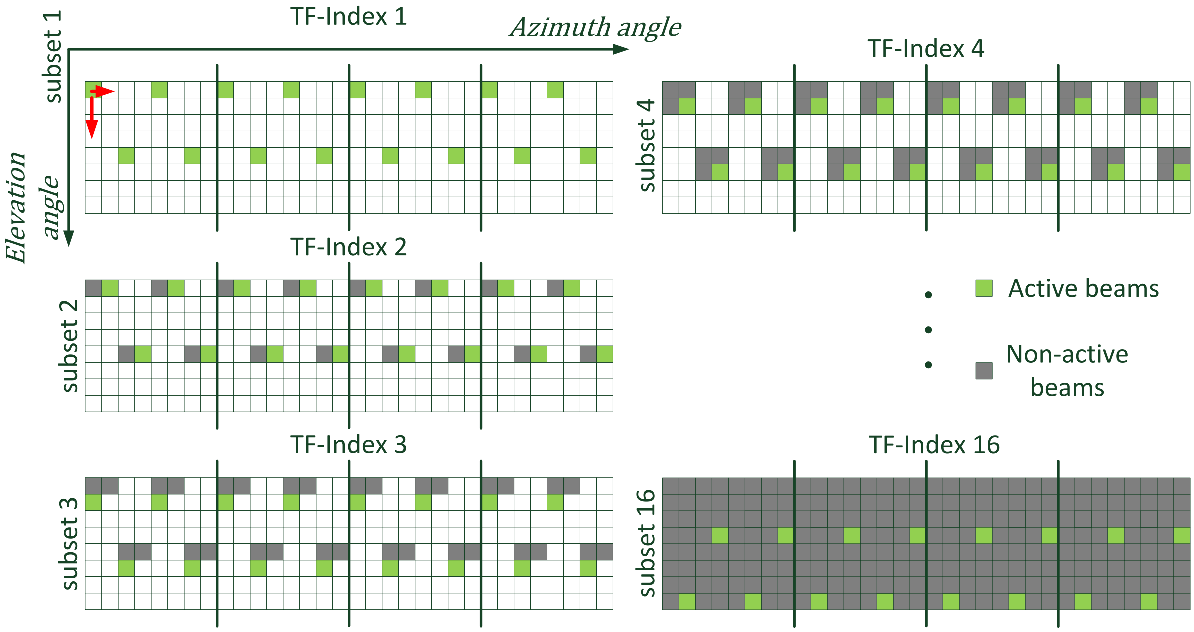

Figure 1. Example of a $[8 \times 32]$ full dimension (FD) codebook distributed over 16 Time Frequency (TF) resources by 16 disjunct subsets with 16 simultanosly active beams each.

precocer is later used for data-transmission, the precoding becomes completely transparent to the user. This means by reporting the maximum beam index related to certain TF resource the user doesn't have to know the specific $\mathbf{B}_q$ used by the BS. Therefore, we call this scheme User Non-aware Precoding and Effective Channel Estimation, short UNP & ECE.

Remark 2. *The UNP & ECE scheme enables BSs to optimize selection and distribution of the precoders according to certain key Performance Indicators (KPIs) without the need to inform the users in the system. As a consequence each BS can construct and optimize its own codebook independently according to deployment, environment, or device/load distribution.*

B. Feedback

Besides the signaling and pilot overhead which is balanced by the parameter N_s discussed in Sec. IV-A the second major limitation for massive MIMO in FDD systems is the amount of feedback required. Each user has to send the obtained Channel State Information (CSI) back to the BS necessary for efficient multi-user MIMO downlink transmission.

In contrast the given UNP & ECE transmission scheme the following information is required per device at the BS for multi-user scheduling:

1) CQI (including inter beam interference)
2) the corresponding stream id and
3) the corresponding TF resource (or Resource Block (RB)).

A huge advantage of UNP & ECE is that by the predefined grid-of-beams per resource the inter-beam interference is known in advance and can be considered in the CQI reported to the BS. Furthermore, we know from previous work that a significant portion of the performance in multi-user downlink transmission can be achieved with the strongest stream id per user available for scheduling [12]. Therefore, the number of reported CQIs values can be $\leq N_s$ and is evaluated in Sec. V. The stream id or index is similar to the Precoding Matrix Indicator (PMI) in LTE terminology. Furthermore, there is a high probability that in the given non line of sight and multi-path environment and the additional applied precoder per subset the maximum CQI varies from one TF resource to the next. Thus we foresee that only a subset of resources in time/frequency corresponding to the highest CQIs has to be reported (similar to the spatial domain) in order to achieve most of the performance compared to the "unlimited" feedback case.

For the calculation of the feedback F in bit per user and per Transmission Time Interval (TTI) the following equations are used

- Full frequency and limited spatial feedback:

$$F = N_{\text{RB}} N_{\tilde{S}} (B_{\text{CQI}} + B_S) \tag{4}$$

- Limited frequency and limited spatial feedback:

$$F = N_{\tilde{\text{RB}}} N_{\tilde{S}} (B_{\text{CQI}} + B_S), \tag{5}$$

where N_{RB} is the number of RBs (in frequency) used for transmission, $N_{\tilde{S}}$ is the number of streams per RB reported back to the BS with the constraint $N_{\tilde{S}} \leq N_S$, B_{CQI} is the number of bits for CQI levels, (here the number of bits for quantization of the Shannon rate), B_S is the number of bits required for a stream index such that $2^{B_S} \geq N_S$, and $N_{\tilde{\text{RB}}} \leq N_{\text{RB}}$ is the number RBs for which the stream index and CQI value is reported (sorted by the receive beam power). Note, that there is a switching point when reducing $N_{\tilde{\text{RB}}}$ where it is more efficient to use (5) instead of (4) in terms of feedback rate which is considered for feedback calculations in Sec. V.

V. NUMERICAL RESULTS

For evaluation of the proposed scheme we conducted system level simulations where the SINR per beam and per device is calculated according to Eq. (3). Note, that no large-system approximations are considered and each link of each device-BS antenna pair is taken into account enabled by a high performance computation cluster [13]. The source code of the MATLAB based Quasi Deterministic Radio Channel Generator (QUADRIGA) channel model is public available can be downloaded for free [14]. If not stated otherwise the simulations assumptions listed in Tab. I are considered.

First we investigate how much feedback can be reduced by limiting the number of reported streams and RBs in Fig. 2. This confirms the results from [12] also for full dimension MIMO, that most of the performance can be achieved by reporting only the strongest stream index per device. Here 96 % of the upper performance with $N_S = 16$ are achieved by $N_{\tilde{S}} = 1$ at $N_{\tilde{\text{RB}}} = 100$. Reducing the number of feedback RBs $N_{\tilde{\text{RB}}}$ results in a monotonic decrease in the sum Spectral Efficiency (SE), e.g. at $N_{\tilde{\text{RB}}} = 64$ the performance loss is less than 3 % compared to $N_{\tilde{\text{RB}}} = 100$.

Second we look into the feedback rate per user and TTI given in Fig. 3 obtained with Eq. (4) and (5). The feedback rate is given over the median sum SE for $N_{\tilde{S}} = [1, 2]$ and the number next to each marker is the number of reported RBs $N_{\tilde{\text{RB}}}$. From this we observe for a given feedback rate it is more efficient in terms of SE to report only a single stream index for more RBs instead of more streams for less RBs.

Remark 3. *These observations strongly depend on the selection of the scheduler. It would be of interest to investigate the spatial-frequency diversity trade-off in terms of feedback rate and SE also for other schedulers, e.g. Round Robin or proportional fair.*

Next the performance of the UNP & ECE scheme for full dimension MIMO in FDD is compared with the TDD mode assuming perfect CSI at the transmitter (CSIT) in Fig. 4. The precoder in TDD mode is constructed by Singular Value Decomposition (SVD)-Zero-Forcing (ZF) (see [15]). We also include the case of optimal beam selection in terms of sum SE instead of fixed precoder subsets assuming that all streams on all RBs are fed back. The increase in sum SE from device diversity is saturating for all schemes, at 60 for SVD-ZF and at 210 for the others. Note, that for fair comparison a maximum of 16 devices is spatially multiplexed per TF resource. With optimal beam selection approx. 63 % of SVD-ZF is achieved. With our proposed UNP & ECE scheme under limited feedback and signaling, and subset distribution according to Fig. 1 the loss compared to optimal beam selection is approx. 4 %.

Table I
SIMULATION ASSUMPTIONS.

Parameter	**Value**
Channel model	QUADRIGA [16]
Scenario	Urban macro 100% NLOS
Center frequency f_c	2.68 GHz
Simulation type	Monte-Carlo
Number of samples	500
ISD	500 m
Bandwidth	18 MHz
Number of RBs in frequency	100
BS height	32 m
Antenna array $[N_x, N_y]$	$[32, 8]$
Antenna element spacing	$\lambda/2$
Antenna element typ	Patch antenna
Half power beamwidth [azimuth, elevation]	$[65, 65]^\circ$
Antenna element gain	9.4 dBi
Transmit power	43 dBm
Scheduler for spatial multiplexing	maximum throughput
Precoder B_{GoB}	2D DFT according to Section III
$[N_s, N_q, S]$ (see Sec. III)	[16, 16, 256,]
Precoder distribution	According to Fig. 1
Number of devices	512 if not given otherwise
Device antenna array	2 cross polarized dipoles
Device height	2 m
Channel estimation	Perfect
CQI quantization	Shannon with 8 bits from -5 to 25 dB SINR

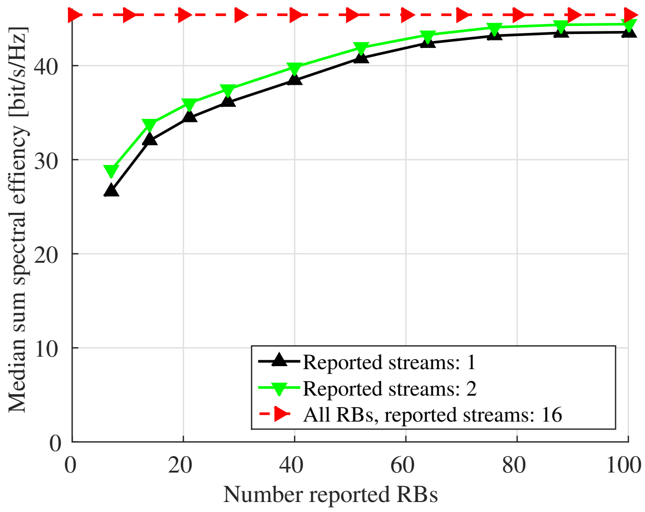

Figure 2. Median sum spectral efficiency over the number of reported RBs in frequency domain for [1, 2, 16] reported streams per RB.

VI. CONCLUSION

In this paper we addressed the questions how to operate *massive* MIMO in a FDD system considering signaling and feedback constraints. With the proposed UNP & ECE scheme a large precoder is split into subsets and distributed over TF resources. This results in device transparent precoding which means that devices estimate directly the effective channel from the precoded reference signals. Thus it is sufficient to report only the stream index with the maximum receive

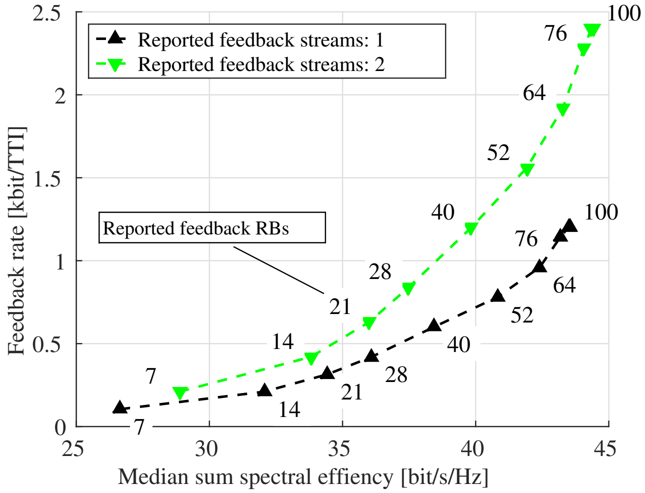

Figure 3. Feedback rate in kbit per TTI over the median sum spectral efficiency for 1 and 2 reported streams decreasing the number of reported RBs. 100 RBs and 8 bit per feedback value are assumed.

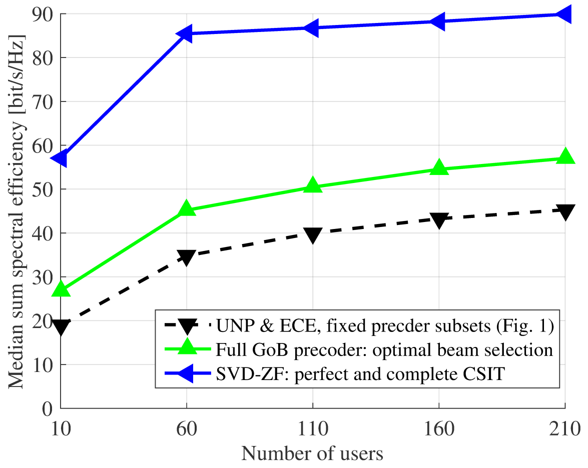

Figure 4. Comparison of sum SE between "optimal" precoding in TDD with perfect CSIT, optimal beam selection in FDD from a limited codebook, and the proposed UNP & ECE.

power, without required knowledge about the used precoder. This enables system wide per BS optimization of codebooks according to the deployment, environment and propagation scenario or other KPIs.

Thus huge performance gains are possible compared to LTE without further increasing the signaling overhead. Also a limited feedback channel was considered and the trade-off between frequency and spatial diversity in terms of feedback and sum SE is discussed. We can conclude that with the strongest stream index per precoder-subset reported by the devices results in negligible performance loss. Thus it is always better to report more TF resources instead of more streams to achieve higher sum SE in the given setup. The overall performance degradation compared to a TDD system with perfect CSIT is not too large, and will become smaller if further impairments for TDD are considered. One advantage of FDD is that the number of devices to do channel estimation is not limited. The UNP & ECE scheme also utilizes inherently multiple receive antennas at the device thus enabling advanced receive beamforming techniques. In TDD channel estimation for multiple device antennas would increase the time required for estimation.

For future work, it is of great interest to find the optimal distribution of the beams per precoder subsets, distribution of the subsets on TF resources, and to study the impact of other schedulers than maximum throughput.

REFERENCES

[1] METIS, "Intermediate description of the spectrum needs and usage principles," Mobile and wireless communications Enablers for the Twenty-twenty Information Society (METIS), Tech. Rep. D5.1, Aug 2013.

[2] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 186–195, February 2014.

[3] T. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 11, pp. 3590 –3600, Nov 2010.

[4] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: 10 Myths and One Grand Question," *CoRR*, vol. abs/1503.06854, pp. 1–10, 2015. [Online]. Available: http://arxiv.org/abs/1503.06854

[5] Z. Jiang, A. Molisch, G. Caire, and Z. Niu, "Achievable Rates of FDD Massive MIMO Systems with Spatial Channel Correlation," *Wireless Communications, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[6] M. Kurras, L. Thiele, and G. Caire, "Interference Mitigation and Multiuser Multiplexing with Beam-Steering Antennas," in *WSA 2015; 19th International ITG Workshop on Smart Antennas; Proceedings of*, March 2015, pp. 1–5.

[7] 3GPP, "Study on Elevation Beamforming/Full-Dimension (FD) MIMO for LTE," 3rd Generation Partnership Project, Tech. Rep. 36897, Apr 2015.

[8] Y. Yuan, Y. Wang, W. Zhang, and F. Peng, "Separate Horizontal and Vertical Codebook Based 3D MIMO Beamforming Scheme in LTE-A Networks," in *Vehicular Technology Conference (VTC Fall), 2013 IEEE 78th*, Sept 2013, pp. 1–5.

[9] D. Ying, F. Vook, T. Thomas, D. Love, and A. Ghosh, "Kronecker product correlation model and limited feedback codebook design in a 3D channel model," in *Communications (ICC), 2014 IEEE International Conference on*, June 2014, pp. 5865–5870.

[10] D. Love and R. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *Information Theory, IEEE Transactions on*, vol. 51, no. 8, pp. 2967–2976, Aug 2005.

[11] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint Spatial Division and Multiplexing: The Large-Scale Array Regime," *Information Theory, IEEE Transactions on*, vol. 59, no. 10, pp. 6441–6463, 2013.

[12] L. Thiele, M. Kurras, M. Olbrich, and T. Haustein, "Boosting 4G Networks with Spatial Interference Management Under Feedback Constraints : A Noncooperative Downlink Transmission Scheme," *Vehicular Technology Magazine, IEEE*, vol. 8, no. 1, pp. 40–48, Mar. 2013.

[13] L. Thiele, T. Wirth, M. Olbrich, T. Schierl, T. Haustein, and V. Frascolla, "High performance cluster computing as a tool for 4g wireless system development." *Intel Technology Journal*, vol. 18, no. 3, pp. 1–20, 2014.

[14] Fraunhofer HHI, *Quasi Deterministic Radio Channel Generator (QUADRIGA)*, Online, Fraunhofer, Oct 2012. [Online]. Available: http://www.quadriga-channel-model.de

[15] L. Thiele, M. Kurras, K. Börner, and T. Haustein, "A system-level study on multi-user MIMO Transmission for dense FDD networks," in *Signals, Systems and Computers, 2013 Asilomar Conference on*, Nov 2013, pp. 1885–1889.

[16] S. Jaeckel, L. Raschkowski, K. Börner, and L. Thiele, "QuaDRiGa: A 3-D Multi-Cell Channel Model With Time Evolution for Enabling Virtual Field Trials," *Antennas and Propagation, IEEE Transactions on*, vol. 62, no. 6, pp. 3242–3256, June 2014.