# Generalized Semi-Orthogonal Multiple-Access for Massive MIMO

Majid Nasiri Khormuji
Huawei Technologies Sweden AB, Stockholm
Emails: majid.nk@huawei.com

***Abstract*—We propose a novel framework to enable concurrent transmission of multiple users to an access node equipped with a massive multiple-input multiple-output (mMIMO), which encompasses and extends the conventional time-division duplex (TDD) protocol and the recently proposed semi-orthogonal multiple-access (SOMA). The new solution is referred to generalized semi-orthogonal multiple-access (GSOMA). To enable GSOMA, the users are grouped such that the users within each group are assigned mutually orthogonal pilot sequences. However the pilot sequences can be reused in different groups and the users in different group are coordinated such that their pilot sequences do not interfere by a semi-orthogonal mechanism enabled by a resource-offset sequence. We describe the general framework and analyze a case of GSOMA and show that a properly designed GSOMA can outperform the conventional TDD and SOMA.**

## I. Introduction to mMIMO

Massive multiple-input multiple-output (mMIMO) communication is considered as one of promising physical-layer solutions to enable multi-user communication for future communications systems [1]–[4] in response to ever-increased demand for high data rate applications as well as to a more homogenous quality-of-service across the service area. Time-division duplex (TDD) protocol in which the uplink pilot symbols are used to estimate both uplink and downlink is proposed to cope with the pilot overhead for mMIMO arrays. In TDD protocol [1]–[4], the transmission time over each coherence time is divided into four non-overlapping phases: Channel training to learn the channel between the users and the base station, each user transmits some known pilot symbols; Uplink Data: the uplink data of all users are transmitted over the same time–frequency resources in a *non-orthogonal* fashion such that the base station receives a superposition of all transmitted symbols; Processing Time: the time which is needed to perform the channel estimation and precode the users data for the downlink transmission; and Downlink Data: finally the downlink data of all users are precoded using the estimated channels and transmitted over the same time–frequency resources in a *non-orthogonal* fashion. The duration of uplink and downlink transmissions may vary and can be adjusted based on the amount of the users' data and the traffic demands.

In this paper, we investigate the communication of $K$ single-antenna users to a common receiver with a massive antenna array over an uplink shared channel. To estimate the channel between two antenna ports, the transmitting node sends pilot symbols which are known at the receive node (the time–frequency location and the associated value are generally preset). The pilot symbols from each user should have the periodicity of $N = T_c B_c$ symbols in order to track the channel variation over the time and frequency where $T_c$ and $B_c$ denote the coherence time and bandwidth, respectively. It is desirable to coordinate as many users as possible for uplink transmission such that the receiver can perform spatial-division multiple-access (SDMA); i.e. the receiver is able to obtain interference-free signals associated to each user via spatial filtering (e.g. projection of the received signals in the space that is spanned by the associated channels from the users to the antenna array). In mMIMO, the receiver can asymptotically separate the uplink data of different users without inter-user interference if the spatial channels are for example independent [1]. Hence ignoring the processing time, the maximum number of the coordinated users, under successful decoding, can be obtained by maximizing the total number of transmitted data symbols that are separable at the receiver, which is given by $K(B_c T_c - K)$, where the maximum is taken over the active number of users, $K$, for a given coherence interval spanning $N$ symbols. Under assumption that all data symbols contain equal information; i.e. all users employ the same modulation order, this concludes that the coherence interval should be equally divided between the channel training and data transmission as considered in [1]. So the optimal number of active users operating over the same time–frequency resources should be set to $\lfloor \frac{1}{2} N \rfloor$ and the total number of data symbols that can be transmitted hence becomes $\lfloor \frac{1}{2} N \rfloor \cdot \lceil \frac{1}{2} N \rceil$ symbols.

In [5], it is shown that the above number of users and the associated aggregate rate can be improved by the so-called a semi-orthogonal feature. In this paper, we further extend this concept and present a generalized one which can bridge both SOMA and the conventional TDD [1] and provide an enhanced aggregate throughput.

The remainder of the paper is organized as follows. Section II briefly reviews semi-orthogonal multiple-access (SOMA) described in [5]. Section III generalizes SOMA scheme to enhance its potential performance. Section IV discusses two realizations of GSOMA and presents an iterative receiver. This section also analyzes the sum-rate for two modes of GSOMA using matched filtering (MF) and zero-forcing (ZF) receiver. Section V presents numerical results and Section VI finally concludes the paper.

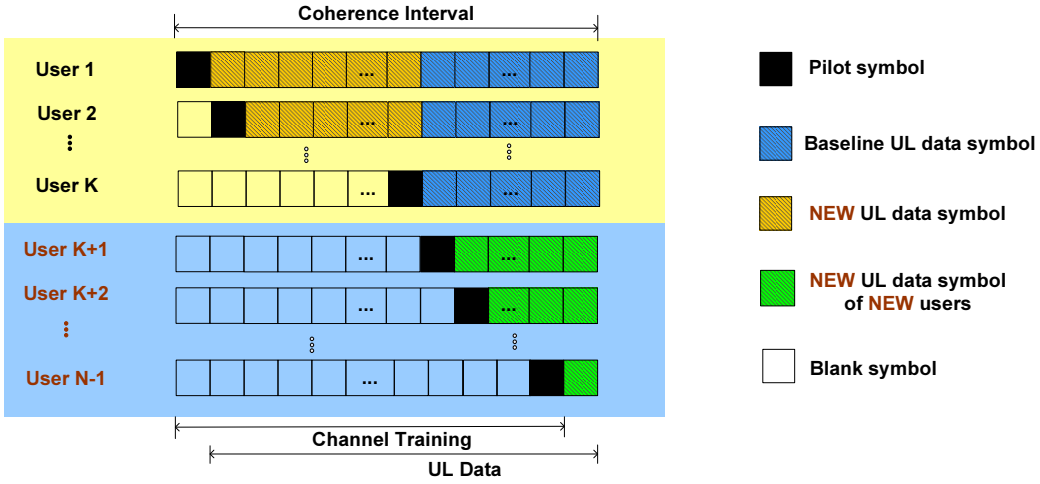


Fig. 1. Illustration of Semi-Orthogonal Multiple-Access (SOMA) transmission.

## II. SOMA

We next briefly discuss the semi-orthogonal multiple-access in [5] which is designed to address the uplink capacity shortage problem in mMIMO. SOMA coordinates up to $K \leq N-1$ users, where $N = T_c B_c$ is the number of resources in time and frequency over which the channel is approximately constant, i.e. the coherence interval. Fig. 1 depicts an example of how the uplink transmission is configured in which $K = N-1$ users are scheduled. Over each coherence block, the user $1 \leq j \leq N-1$, transmits one pilot symbol and $N-j$ data symbols such that the pilot symbols of users are transmitted over orthogonal time–frequency resources, and the data symbols of user $j$, for all $1 \leq j \leq N-1$, consume all the time–frequency resources of users $j+1 \leq k \leq N-1$, which are used for both pilot and data transmission. The received signals at the access node in a coherence window is illustrated in Fig. 1. In a given time-slot, some of the users are silent and appears *orthogonal* while some other users transmit *non-orthogonally*. For example, in time slot one, only user one transmits its pilot and all other users appear orthogonal and in the second time slot the first user transmits its data symbol and second user transmits its pilot symbol and so forth; cf. *semi-orthogonal* feature. SOMA when used with an access node with a very high number of antennas can schedule $K$ users where user $k$ can transmit $N-k$ asymptotically error-free symbols. Therefore, for very large antenna arrays, SOMA solution nearly doubles the throughput as compared to the baseline TDD with optimal number of users. The receiver is constructed using a sequential channel estimation and data detection enabled by the designed semi-orthogonal feature embedded in the transmitted signals.

## III. Generalized SOMA

We next present the generalized SOMA (GSOMA). Fig. 2 depicts multi-user GSOMA transmission in which the users are grouped into $J$ groups where each group contains $k_j$ users for $j \in \{1, 2, \ldots, J\}$ and $K = \sum_{j=1}^{J} k_j$ is the total number of the users. The user $i$ in group $j$ uses the pilot sequences $s_i$ for all $j \in \{1, 2, \ldots, J\}$. That is the pilot sequences are reused. The pilot sequences within each group are orthogonal such that it allows interference-free channel estimation for the users in each group. The maximum number of pilot sequences therefore is $\max_j k_j$. The conventional reuse of pilot sequences where the pilot symbols interference with one another results to "pilot contaminations" which severely degrades the performance of the users. However with this new solution, it is allowed to reuse the pilot sequences in a controlled fashion. A pilot reuse is performed *semi-orthogonally* to boost the spectral efficiency of the system. However, the interference is controlled by a transmission of a *resource-offset* (e.g. timing-offset or frequency-offset) such that the received packets at the access node have the following structure

- The pilot signals of different groups are received non-orthogonally (for example non-overlapping time slots with timing-offset)
- The pilot sequences of the group $j \in \{1, 2, \ldots, J\}$, are only interfered by data symbols of users in the groups 1 to $j-1$. The pilot sequences of the first group are received interference-free. That is the other users appear silent at the receiver side.

With GSOMA, the user may use partial blanking which has the same granularity as the length of pilot region to control to the inter-group interference. GSOMA includes both SOMA scheme described in [5] and the conventional TDD solution proposed by Marzetta in [1] as special cases. When each group contains one user and no blanking is used, GSOMA reduces to SOMA. When there is only one group with maximum number of users, then GSOMA reduces to the conventional TDD wherein only orthogonal pilot sequences are used. Therefore, a properly designed GSOMA can combine the advantages of both SOMA and conventional TDD. The advantage of GSOMA with respect to the conventional TDD is that it schedules more groups, which enhances the aggregate rate. Since

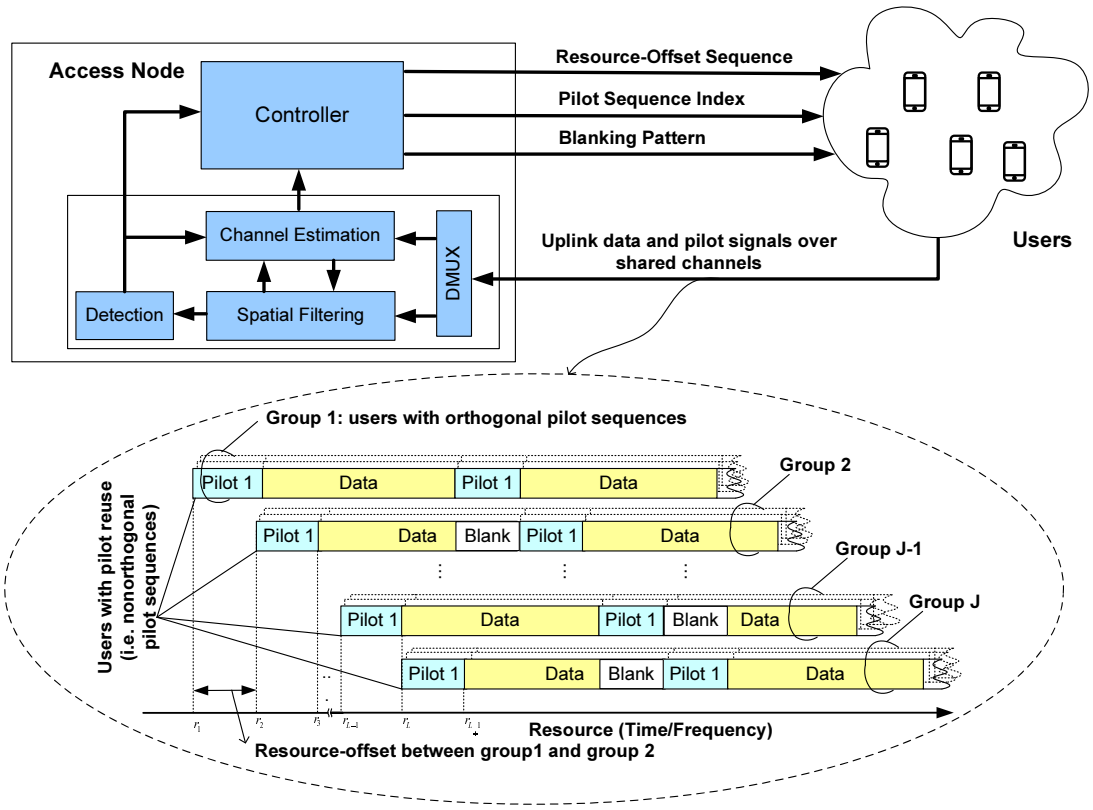


Fig. 2. Generalized Semi-Orthogonal Multiple-Access (GSOMA) transmission.

within each group the pilot sequences are mutually orthogonal as that in the conventional TDD, it allows performing joint channel estimation without interference for all users in each group and employing joint spatial filtering such as zero-forcing per group to suppress the inter-user interference among the users in each group. This also increases the aggregate rate.

## IV. Two-Group GSOMA

In this section, we tailor the design of GSOMA to the cases with two groups and discuss two modes of transmission.

### A. Mode 1: without blanking

Fig. 3 shows the transmission protocol and the receiver for the case with two groups where there is no blanking. The users are grouped into two groups such that each group is designed according the conventional TDD wherein the half of the coherence time, i.e. $\frac{1}{2}T_c$, is used for pilot transmission and the other half is consumed for the data transmission. To enable the channel estimation without interference, a time-offset equal to $\frac{1}{2}T_c$ is used. The receiver first estimates the channel of the users in the first group using the knowledge of the orthogonal pilot sequences and then performs joint spatial filtering to decode the data of the first group. The decoded data are fed back to the channel estimator of the second group to perform interference cancellation prior to the channel estimation. After the interfere cancellation, the channels of the second group are estimated using the known orthogonal pilot sequences used for the users in the second group. The decoded data of the second group is used to cancel the interference over the pilot symbols of first group. This sequential channel estimation and decoding is iteratively continued until all data are successfully decoded.

We next present the single-cell uplink throughput for the protocol in Fig. 4 with $L$ consecutive sub-frames when the receiver has $n_t$ antennas and the channels are i.i.d. Rayleigh fading with unit variance. We first consider matched-filtering (MF) and then zero-forcing (ZF). The proofs, which are omitted due the space limitation, can be obtained by similar approaches as those in [5].

*Proposition 1:* The transmission protocol in Fig. 3 where each group contains $K$ users and the receiver employs MF achieves the sum-rate

$$R_{\Sigma,1}^{\mathrm{MF}} = \frac{K}{2L+1}\sum_{l=1}^{L}(R_{1,l}+R_{2,l}) \tag{1}$$

where

$$R_{1,l} = \log\left(1+\frac{(n_t-1)(1-N_{e_{1,l}})P_{d_{1,l}}}{N_0+N_{e_{1,l}}P_{d_{1,l}}+P_{p_{2,l}}+(K-1)P_{d_{1,l}}}\right)$$

$$R_{2,l} = \log\left(1+\frac{(n_t-1)(1-N_{e_{2,l}})P_{d_{2,l}}}{N_0+N_{e_{2,l}}P_{d_{2,l}}+P_{p_{1,l+1}}+(K-1)P_{d_{2,l}}}\right)$$

$$N_{e_{1,l}} = \frac{N_0+KN_{e_{2,l-1}}P_{d_{2,l-1}}}{N_0+P_{p_{1,l}}+KN_{e_{2,l-1}}P_{d_{2,l-1}}}$$

$$N_{e_{2,l}} = \frac{N_0+KN_{e_{1,l}}P_{d_{1,l}}}{N_0+P_{p_{2,l}}+KN_{e_{1,l}}P_{d_{2,l}}}$$

and $P_{p_{j,l}}$, $P_{d_{j,l}}$ denote the average power consumed for the pilot and data symbols of the users in group $j \in \{1,2\}$ and sub-frame $l \in \{1,2,\ldots,L\}$, and $N_{e_{j,l}}$ and $N_0$ respectively de-

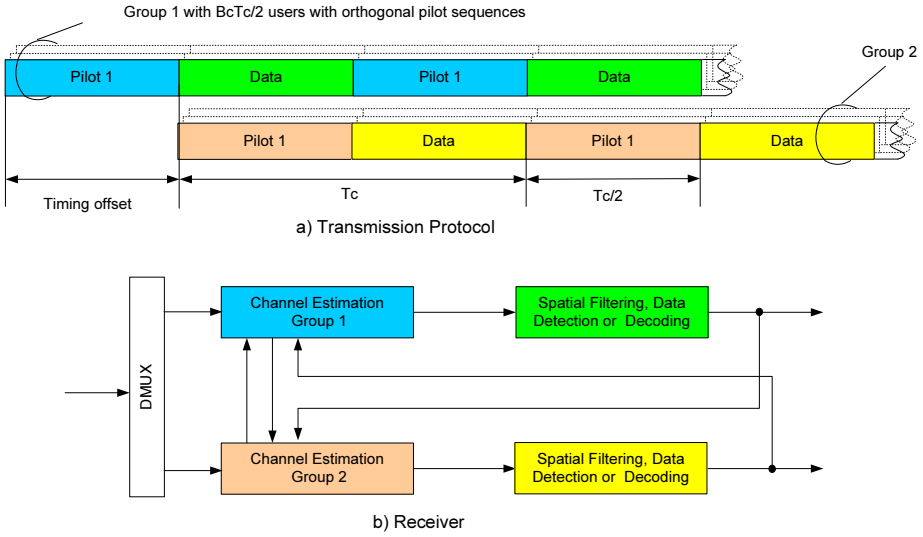


Fig. 3. Two-Group GSOMA without blanking.

note variance of AWGN at the receiver and channel estimation error, and $N_{e_{1,0}} = P_{d_{1,0}} = P_{p_{1,L+1}} = 0$.

*Proposition 2:* The transmission protocol in Fig. 3 where each group contains $K$ users and the receiver employs ZF achieves the sum-rate

$$R^{\text{ZF}}_{\sum,1} = \frac{K}{2L+1}\sum_{l=1}^{L}\log\left(1+\frac{(n_t-K)(1-N_{e_{1,l}})P_{d_{1,l}}}{N_0+KN_{e_{1,l}}P_{d_{1,l}}+P_{p_{2,l}}}\right) + \log\left(1+\frac{(n_t-K)(1-N_{e_{2,l}})P_{d_{2,l}}}{N_0+KN_{e_{2,l}}P_{d_{2,l}}+P_{p_{1,l+1}}}\right) \quad (2)$$

where $N_{e_{j,l}}$ is defined in Prop. 1.

### B. Mode 2: with blanking

For the case the array contains not so many antenna elements with respect to the scheduled number of users in each group, it is beneficial to partially blank some part of sub-frames to enhance the channel estimation and to improve consequently the performance of spatial filtering which in turn improves the spectral efficiency of the system. The blanking pattern, should be chosen based on the inter-user interference when for higher interference a blanking pattern with a higher density is selected and similarly a sparser blanking pattern should be selected for lower interference to improve the spectral efficiency of the system. Fig. 4 shows the transmission protocol and the receiver for the case with two groups where the blanking is used. The users are grouped into two groups and transmission is arranged such that the pilot symbols of first group do not experience any interference. This is useful for the case that one to enhance the channel estimation for the users in the first group, this also enhances the interference cancelation for channel estimation of the second group. Additionally, this arrangement is also useful for the case where first and second groups compromise far and near users, respectively.

*Proposition 3:* The transmission protocol in Fig. 4 where each group contains $K$ users and the receiver employs MF achieves the sum-rate

$$R^{\text{MF}}_{\sum,2} = \frac{1}{3}K\log\left(1+\frac{(n_t-1)(1-N_{e_1})P_{d_1}}{N_0+N_{e_1}P_{d_1}+P_{p_2}+(K-1)P_{d_1}}\right) + \frac{1}{3}K\log\left(1+\frac{(n_t-1)(1-N_{e_2})P_{d_2}}{N_0+N_{e_2}P_{d_2}+(K-1)P_{d_2}}\right) \quad (3)$$

where $P_{p_j}$, $P_{d_j}$ denote the average power consumed for the pilot and data symbols of the users in group $j \in \{1,2\}$, and $N_0$ denotes variance of AWGN at the receiver. The quantity $N_{e_j} = N_{e_{j,1}}$ denotes the variance of channel estimation error for users in group $j \in \{1,2\}$.

*Proposition 4:* The transmission protocol in Fig. 4 where each group contains $K$ users and the receiver employs ZF achieves the sum-rate

$$R^{\text{ZF}}_{\sum,2} = \frac{1}{3}K\log\left(1+\frac{(n_t-K)(1-N_{e_1})P_{d_1}}{N_0+KN_{e_1}P_{d_1}+P_{p_2}}\right) + \frac{1}{3}K\log\left(1+\frac{(n_t-K)(1-N_{e_2})P_{d_2}}{N_0+KN_{e_2}P_{d_2}}\right) \quad (4)$$

where $N_{e_j} = N_{e_{j,1}}$ for $j \in \{1,2\}$ is given in Prop. 3.

From the sum-rates in Propositions 1–4, one can see that the imperfections due to inter-user interference, the channel estimation error and AWGN are linearly reduced by the factor of $\frac{1}{n_t-1}$ and $\frac{1}{n_t-K}$ for MF and ZF, respectively. We additionally see that ZF removes the inter-user interference in each group due the data transmission which is one of the advantages of GSOMA with respect to SOMA.

## V. Numerical Evaluations

For comparison, we consider time-shared TDD as a baseline when the resources are shared between two groups of users when each group is designed according to the conventional TDD. This, using ZF, gives the sum-rate

$$R^{\text{ZF}}_{\sum,\text{TS}} = \frac{1}{4}K\log\left(1+\frac{(n_t-K)(1-N_{e_1})P_{d_1}}{N_0+KN_{e_1}P_{d_1}}\right) + \frac{1}{4}K\log\left(1+\frac{(n_t-K)(1-N_{e_2})P_{d_2}}{N_0+KN_{e_2}P_{d_2}}\right) \quad (5)$$

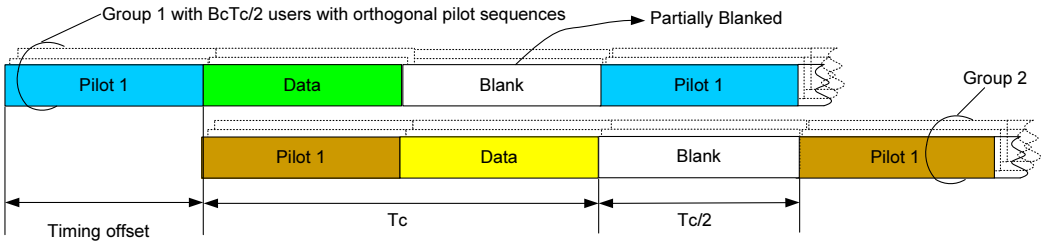


Fig. 4. Two-Group GSOMA with blanking.

where $N_{e_j} = \frac{N_0}{N_0+P_{p_j}}$.

We next discuss two numerical examples of the sum-rate. Fig. 5 and 6 show the sum-rate of the schemes as a function of number of antennas for $N = B_c T_c = 100$, where the number of users in each group is $K = \frac{1}{2}N = 50$. We set $P_{d_{1,l}} = P_{d_1}$, $P_{d_{2,l}} = P_{d_2}$ (i.e. uniform power allocation across the sub-frames) and $N_0 = 0$ dB and the power of the associated pilots are set 10 dB higher than the data for all users to ensure a good channel estimation. In each figure, four schemes are considered: time-shared TDD with ZF, SOMA with MF, GSOMA with transmission protocol in Fig. 3 (i.e. Mode 1) with ZF and $L = 100$, GSOMA with transmission protocol in Fig. 4 with ZF (i.e. Mode 2). In Fig. 5 we set $P_{d_1} = 0$, $P_{d_2} = -20$ dB and in Fig. 6 we set $P_{d_1} = 10$, $P_{d_2} = -5$ dB. In both cases, GSOMA provides an enhanced aggregate rate as compared to the time-shared TDD. The gain is more pronounced for the case that the groups are higher difference in the received signal strength for which Mode 1 performs better. Fig. 6, Mode 2 performs better than Mode 1 due to the fact that the channel estimation for the high-power users in group 1 is less degraded as compared to that in Mode 1.

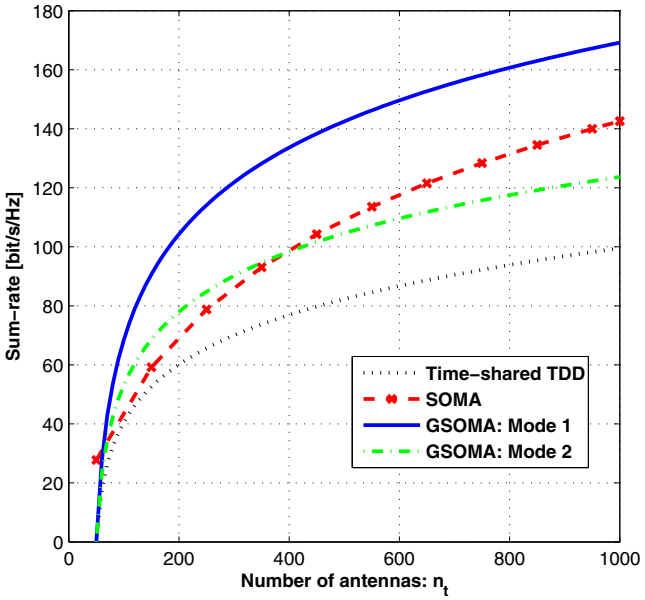


Fig. 5. The sum-rate of the schemes for $P_{d_1} = 0$, $P_{d_2} = -20$ dB.

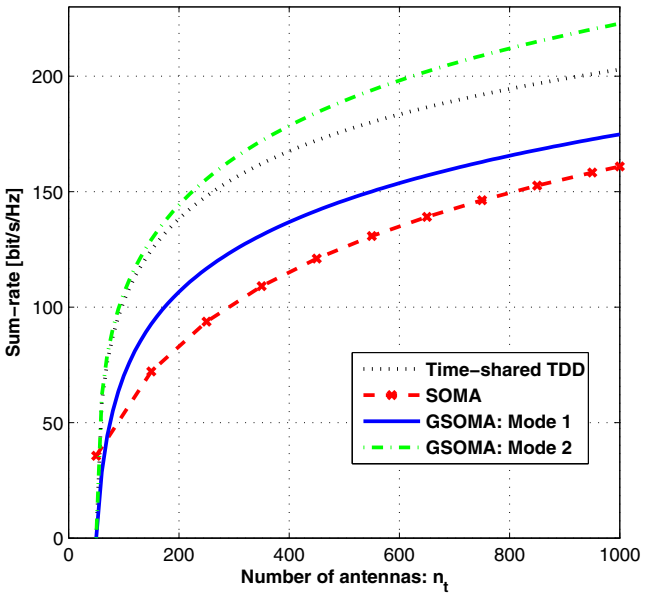


Fig. 6. The sum-rate of the schemes for $P_{d_1} = 10$, $P_{d_2} = -5$ dB.

## VI. Conclusions

We presented a new multiple-access solution and analyzed its aggregate rate. The new solution is constructed using a semi-orthogonal feature for a group of users wherein each group employs the conventional TDD. The numerical evaluation showed that the new proposal can provide a higher aggregate rate as compared to the conventional TDD solution.

## References


[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[2] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, pp. 186–195, 2014.

[3] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *Signal Processing Magazine, IEEE*, vol. 30, no. 1, pp. 40–60, 2013.

[4] J. Hoydis, S. Ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, 2013.

[5] M. N. Khormuji and B. M. Popović, "Semi-orthogonal multiple-access for massive MIMO," *Submitted to IEEE Trans on Wireless Communications*, 2014.