# Pattern Division for the Massive MIMO Networks with the Two-stage Precoding

Jianpeng Ma, Hongyan Li, Shun Zhang, Nan Zhao, Arumugam Nallanathan

***Abstract*—In massive multiple-input multiple-output (MIMO) networks with two-stage precoding, the user clusters with serious angle-spreading-range (ASR) overlapping should be divided into different patterns and scheduled in orthogonal sub-channels to achieve optimal performance. In this letter, we propose one graph theory based pattern division (GT-PD) scheme to deal with the ASR overlapping with a limited number of sub-channels. First, we depict the ASR overlapping as an undirected weighted graph, where the weight of each edge indicates the strength of the ASR overlapping between two connected clusters. Then, we separately denote each user cluster and patten as a vertex and a color, and transform the pattern division into a graph coloring problem with limited colors. In addition, the GT-PD scheme is developed with the help of the Dsatur algorithm. Finally, numerical results are provided to corroborate the efficiency of the proposed scheme.**

***Index Terms*—Graph theory, massive MIMO, pattern division, two-stage precoding.**

## I. Introduction

Due to its significant improvement in spectral efficiency, massive multiple-input multiple-output (MIMO) has been widely considered as one promising technology for the 5G cellular system [1]. The channel state information (CSI) at the base station (BS) side plays an important role for the downlink precoding and uplink detection in the massive MIMO system. For the time-division duplex (TDD) systems, the CSI can be obtained through the uplink training with the uplink-downlink reciprocity. However, in the frequency-division duplex (FDD) system, the CSI at the BS side should be acquired by the downlink training and the information feedback, which will lead to unacceptable overhead [2]. To overcome this bottleneck, Adhikary *et al.* proposed an idea of two-stage precoding [3], which can divide the massive MIMO channel into several independent equivalent channels with reduced dimensions.

Several two-stage precoding schemes have been developed [4]–[6]. In [4], an iterative algorithm was proposed to design the prebeamforming matrix by maximizing the signal-to-leakage-plus-noise ratio (SLNR). In [5], Chen and Lau developed a low-complexity online iterative algorithm to track the prebeamforming matrix. In [6], Sun *et al.* proposed a beam division multiplex access transmission scheme.

In [3]–[6], it was assumed that the angle-spreading-ranges (ASRs) of scattering rays for different user clusters do not overlap. Hence, orthogonal transmission can be achieved with the two-stage precoding. However, the users are randomly distributed in practice, and the ASRs for different user clusters are overlapped with high probability, as shown in Fig. 1. An intuitive approach to deal with this problem is to schedule the user clusters with ASR overlapping into orthogonal sub-channels. But this approach will occupy too much sub-channels and is not efficient. To handle this problem, we should schedule user clusters into a limited number of patterns, where the user clusters without (or with slight) ASR overlapping are allocated into the same pattern to be served on the same sub-channel, while those with serious ASR overlapping will be divided into different patterns and work on the orthogonal sub-channels. Theoretically, the optimal pattern division problem can be formulated as a combinational optimization aiming at minimizing the ASR overlapping among users clusters within each pattern.

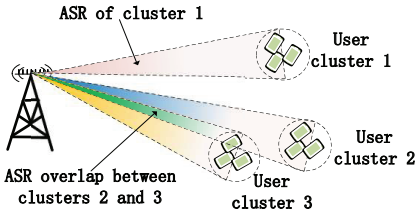


Fig. 1. Illustration of ASR overlapping.

In this letter, we propose a graph theory based pattern division (GT-PD) scheme to implement the optimal pattern scheduling. We first depict the ASR overlapping relationship among all clusters as an undirected weighted graph, which is called ASR overlapping graph (ASR-OG) here. The weight of each edge in the ASR-OG represents the strength of the ASR overlapping between two connected clusters. Then, we separately denote each user cluster and patten as a vertex and a color, and transform the pattern division problem into a graph coloring problem with limited colors. Finally, the proposed GT-PD scheme is developed with the Dsatur algorithm [7].

*Notations*: We use lowercase (uppercase) boldface to denote vector (matrix). $(\cdot)^T$ and $(\cdot)^H$ represent the transpose and Hermitian transpose, respectively. $\mathbf{I}_N$ represents a $N \times N$ identity matrix. $\mathbf{n} \sim \mathcal{CN}(0, \mathbf{I}_N)$ means that $\mathbf{n}$ is complex circularly-symmetric Gaussian distributed with zero mean and covariance $\mathbf{I}_N$. $|\mathcal{A}|$ is the number of elements in set $\mathcal{A}$.

## II. System Model and Preliminaries

We consider a single-cell FDD massive MIMO downlink system, where the BS employs a uniform linear array (ULA) with $N \gg 1$ antennas to serve $K$ single-antenna users. It is assumed that the users can be partitioned into $C$ clusters, where the users in the same cluster are almost co-located. Denote the number of the users in the cluster $c$ as $K_c$. Then, we have $\sum_{c=1}^{C} K_c = K$. Moreover, orthogonal frequency division multiplexing (OFDM) is adopted to divide one specific wideband channel into $L$ flat fading sub-channels in the frequency domain. Correspondingly, we utilize the $1 \times N$ vector $(\mathbf{h}_{c,k}^l)^T$ to represent the channel vector in the $l$-th sub-channel along the link from the BS to the user $k$ in the cluster $c$. The notation $\mathbf{H}_c^l = [\mathbf{h}_{c,1}^l, \mathbf{h}_{c,2}^l, \cdots, \mathbf{h}_{c,K_c}^l]^T$ is also defined for further use.

Similar to [3]–[6], the one-ring channel model is adopted. Define the channel covariance matrix at the BS for the user $k$ in the cluster $c$ as the $N\times N$ matrix $\mathbf{R}_{c,k}^{l}=\mathbb{E}\{\mathbf{h}_{c,k}^{l}(\mathbf{h}_{c,k}^{l})^{H}\}$. The users in the same cluster are co-located and share the same one-ring model parameters. Furthermore, we know that $\mathbf{R}_{c,k}^{l}$ is determined by the one-ring model parameters and independent of the sub-channels [6]. Hence, we can obtain $\mathbf{R}_{c}=\mathbf{R}_{c,k}^{l}$, where $l=1,2,\cdots L, k=1,2,\cdots K_c, c=1,2,\ldots,C$. With eigen-decomposition, we have

$$\mathbf{R}_{c}=\mathbf{E}_{c}\mathbf{\Lambda}_{c}(\mathbf{E}_{c})^{H}, \tag{1}$$

where the $r_c\times r_c$ diagonal matrix $\mathbf{\Lambda}_c$ is the nonzero eigenvalue matrix and $\mathbf{E}_c$ contains the eigenvectors corresponding to the nonzero eigenvalues. Interestingly, in the massive MIMO system, $\mathbf{E}_c$ can be constructed by $r_c$ columns of the $N\times N$ unitary discrete Fourier transform (DFT) matrix $\mathbf{F}_N$ as [8]

$$\mathbf{E}_{c}=\left[\mathbf{f}_{n}:n\in\mathcal{Q}_{c}\right], N\to\infty \tag{2}$$

where $\mathbf{f}_n$ is the $n$-th column of $\mathbf{F}_N$, and the index set $\mathcal{Q}_c$ is defined as

$$\mathcal{Q}_{c}=\left\{n:n\in\left[N\frac{\tau}{\lambda}\sin(\theta_{c}+\Delta_{c}),N\frac{\tau}{\lambda}\sin(\theta_{c}-\Delta_{c})\right]\right\}. \tag{3}$$

where $\theta_c$ and $\Delta_c$ are the central azimuth angle and the angle spread (AS) for the scattering ring of cluster $c$, respectively. Since the antennas of the BS are usually fixed at one higher tower, $\Delta_c$ is very small, and $\mathbf{R}_c$ processes low-rank property, which means that $r_c=|\mathcal{Q}_c|$ is much less than $N$.

Resorting to the Karhunen-Loeve representation, we can rewrite the channel vector $\mathbf{h}_{c,k}^{l}$ as [3]

$$\mathbf{h}_{c,k}^{l}=\mathbf{E}_{c}\{\mathbf{\Lambda}_{c}\}^{\frac{1}{2}}\mathbf{w}_{c,k}^{l}, \tag{4}$$

where $\mathbf{w}_{c,k}$ is the $r_c\times 1$ noise vector, whose entries are i.i.d. complex Gaussian distributed with zero mean and unit variance.

We divide the $C$ user clusters into $L$ patterns. Then, the user clusters with the same pattern $l$ are simultaneously served in the $l$-th sub-channel, while those with different patterns are scheduled into orthogonal sub-channels, as shown in Fig. 2(c). Define the set $\mathcal{C}_l$ to gather the user clusters with the pattern $l$. Then the received signal of the users in the cluster $c\in\mathcal{C}_l$ can be expressed as

$$\mathbf{y}_{c\in\mathcal{C}_{l}}=\mathbf{H}_{c}^{l}\mathbf{P}_{c}^{l}\mathbf{d}_{c}+\sum_{c'\in\mathcal{C}_{l},c'\neq c}\mathbf{H}_{c}^{l}\mathbf{P}_{c'}^{l}\mathbf{d}_{c'}+\mathbf{n}_{c}, \tag{5}$$

where $\mathbf{P}_c^l$ is the $N\times S_c$ precoding matrix for the cluster $c$; $\mathbf{d}_c$ is the $S_c\times 1$ data vector to the cluster $c$; and the $K_c\times 1$ vector $\mathbf{n}_g\sim\mathcal{CN}(\mathbf{0},\mathbf{I}_{K_c})$ is the additive complex Gaussian noise. Clearly, the first part on the right hand side of (5) is the desired signal, and the second part is the inter-cluster interferences.

For the two-stage precoding, the precoding process can be divided into two steps as $\mathbf{P}_c^l=\mathbf{B}_c^l\mathbf{V}_c^l$, where the $N\times M_c$ prebeamformer $\mathbf{B}_c^l$ only depends on the channel covariance matrices, and is used to eliminate the inter-cluster interferences; the $M_c\times S_c$ inner precoder $\mathbf{V}_c^l$ is determined by the $K_c\times M_c$ equivalent effective channel $\overline{\mathbf{H}}_c^l=\mathbf{H}_c^l\mathbf{B}_c^l$ and deals with the intra-cluster interferences; $M_c$ is the dimension of $\overline{\mathbf{H}}_c^l$ seen by the inner precoder and satisfies $S_c\le M_c\le r_c$.

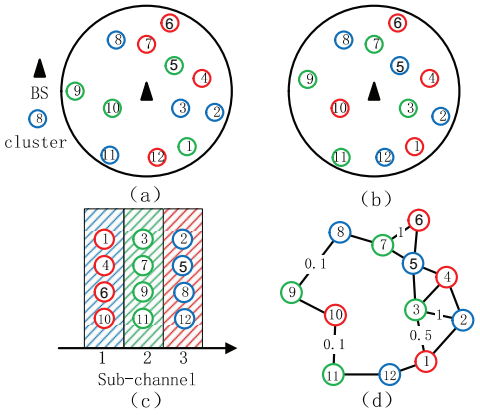


Fig. 2. A pattern division example, where $C=12$ user clusters are divided into $L=3$ patterns and each color represents one pattern. (a) (b) Two pattern division strategies. (c) Sub-channels scheduling for patterns. (d) ASR-OG.

The common idea of the prebeamformer design is to choose $\mathbf{B}_c^l$ to satisfy the following constraint,

$$\mathbf{E}_{c'}^{H}\mathbf{B}_{c}^{l}=0,\forall c'\neq c, c\in\mathcal{C}_{l}\text{ and }c'\in\mathcal{C}_{l}. \tag{6}$$

With the two-stage precoding, the inter-cluster interference is eliminated, and the sum-rate of the cluster $c$ can be written as

$$\mathbb{R}_{c}=\log\det\left\{\mathbf{I}+\overline{\mathbf{H}}_{c}^{l}\mathbf{V}_{c}^{l}\mathbf{d}_{c}(\overline{\mathbf{H}}_{c}^{l}\mathbf{V}_{c}^{l}\mathbf{d}_{c})^{H}\right\}. \tag{7}$$

## III. Proposed GT-PD Scheme

### A. Problem Formulation

In Fig. 2(a) and Fig. 2(b), we present two pattern division strategies as examples. It can be checked that the strategy in Fig. 2(a) suffers more serious ASR overlapping than that in Fig. 2(b). For the system with $C$ clusters and $L$ patterns, the total number of the pattern division strategies is $L^C$. Here, we aim to find one strategy to maximize the sum-rate of all the clusters, which can be formulated as

$$\begin{aligned}\text{(P1)}\quad &\max_{\mathcal{C}_1,\mathcal{C}_l,\cdots,\mathcal{C}_L}\sum_{l=1}^{L}\sum_{c\in\mathcal{C}_l}\mathbb{R}_c,\\ &\text{s.t.}\quad \cup_{l=1}^{L}\mathcal{C}_l=\{1,2,\cdots C\}.\end{aligned} \tag{8}$$

However, solving this combinatorial optimization problem is not straightforward. Firstly, the searching complexity, $L^C$, is unacceptable for large $C$ and $L$. Secondly, it would consume too much channel resources to estimate $\{\overline{\mathbf{H}}_c^l\}$ for each possible patter division strategy. To deal with the above challenges, we propose a GT-PD scheme to achieve a near-optimal patter division strategy with low complexity and low signaling overhead.

### B. Graph Construction

With the property of the DFT matrix, $\mathbf{B}_c^l$ for the user cluster $c\in\mathcal{C}_l$ can be written as

$$\mathbf{B}_{c}^{l}=\left[\mathbf{f}_{n}:n\in\left(\mathcal{Q}_{c}-\cup_{c'\in\mathcal{C}_{l},c'\neq c}\mathcal{Q}_{c'}\right)\right], \tag{9}$$

where $\mathcal{A}-\mathcal{B}$ contains all the elements in the set $\mathcal{A}$ but not in set $\mathcal{B}$, i.e., $\mathcal{A}-\mathcal{B}=\{x:x\in\mathcal{A}\text{ and }x\notin\mathcal{B}\}$. Hence, we can obtain the dimension of the equivalent effective channel $M_c$ as

$$M_{c}=r_{c}-\left|\mathcal{Q}_{c}\cap(\cup_{c'\in\mathcal{C}_{l},c'\neq c}\mathcal{Q}_{c'})\right|. \tag{10}$$

For two specific user clusters $c$ and $c'$, it can be derived that

$$M_c + M_{c'} = \psi_{<c,c'>} + \xi_{<c,c'>}, \tag{11}$$

where $\psi_{<c,c'>} = r_c + r_{c'} - 2|\mathcal{Q}_c \cap \mathcal{Q}_{c'}|$, $\xi_{<c,c'>} = |\mathcal{Q}_c \cap (\cup_{i\in\mathcal{C}_l, i\neq c, i\neq c'} \mathcal{Q}_i)| + |\mathcal{Q}_{c'} \cap (\cup_{j\in\mathcal{C}_l, j\neq c, j\neq c'} \mathcal{Q}_j)|$. From (10) and (11), we can obtain the following observations: if the ASRs for the clusters $c$ and $c'$ do not overlap, i.e., $\mathcal{Q}_c \cap \mathcal{Q}_{c'} = \emptyset$, the two-stage precoding achieves the orthogonal transmission. However, if the ASRs overlap, both $M_c$ and $M_{c'}$ would decrease, which would reduce the system sum-rate. The impact of ASR overlapping between clusters $c$ and $c'$ on the sum-rate is closely related to the parameter $\psi_{<c,c'>}$.

Hence, we define $d_{<c,c'>}$ as a metric to measure the ASR overlapping between the clusters $c$ and $c'$, i.e.,

$$d_{<c,c'>} = \frac{1}{\psi_{<c,c'>}} = \frac{1}{r_c + r_{c'} - 2|\mathcal{Q}_c \cap \mathcal{Q}_{c'}|}. \tag{12}$$

We can see that the larger $d_{<c,c'>}$ indicates more serious ASR overlapping between cluster $c$ and $c'$. Due to the property $d_{<c,c'>} = d_{<c',c>}$, the ASR overlapping among all clusters can be described by an undirected weighted graph as

$$\mathcal{G} = \{\underbrace{\{1, 2, \cdots, C\}}_{\mathcal{V}}, \underbrace{\{d_{<c,c'>} : c \neq c'\}}_{\mathcal{E}}\}, \tag{13}$$

where the vertexes set $\mathcal{V}$ represents the clusters, and the edges set $\mathcal{E}$ denotes the ASR overlapping strength among all clusters. Here, we refer to $\mathcal{G}$ as ASR-OG, and presents one ASR-OG in Fig. 2(d), where the edges with neglectable $d_{<c,c'>}$ are omitted and only a fraction of the edge weights are marked for simplicity.

### C. Graph Coloring

In the ASR-OG, the larger $d_{<c,c'>}$ will cause more serious sum-rate decrease, if the clusters $c$ and $c'$ are assigned to the same pattern. Hence, two clusters with large $d_{<c,c'>}$ should be scheduled to different patterns as far as possible. Correspondingly, the pattern division can be transformed into a graph coloring problem, where one specific color denotes a pattern and the coloring of a vertex represents assigning a pattern to a cluster. The graph coloring problem is proper, if any two connected vertexes corresponds to different colors [10]. Such a condition is called the coloring constraint here. In fact, many efficient algorithms have been proposed to solve the proper graph coloring problem [7]. However, in the ASR-OG, due to the limited number of the patterns, the coloring constraint may not be met. Thus, two clusters with small $d_{<c,c'>}$ may be assigned to the same pattern. To address the above challenges, the GT-PD algorithm is proposed with the help of the Dsatur method [7]. We present the pseudocodes in **Algorithm 1**, and will describe the procedures of the proposed algorithm in detail as follows.

*1) Initialization steps (1-5):* Firstly, the cluster sets for each pattern is set as empty in the step 2. In the steps 3 and 4, the two clusters $c_1$, $c_2$ with the largest $d_{<c_1,c_2>}$ are selected and assigned to patterns $\mathcal{C}_1$, $\mathcal{C}_2$, respectively. The set of the unsigned user clusters $\Omega$ is initialized through deleting the clusters $c_1$, $c_2$ from $\mathcal{V}$ in step 5.

**Algorithm 1** The Proposed GT-PD Algorithm

**Input:** The total number of patterns: $L$;
The constructed ASROG: $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$.
**Output:** The pattern division solution: $\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_L$.
1: **Initialization:**
2: $\mathcal{C}_l = \emptyset,\ \forall l = 1, 2, \cdots, L$.
3: $< c_1, c_2 >= \arg\max_{c_1 \neq c_2} \{d_{<c_1,c_2>}\}$.
4: $\mathcal{C}_1 = \{c_1\},\ \mathcal{C}_2 = \{c_2\}$.
5: $\Omega = \mathcal{V} - \{c_1, c_2\}$.
6: **repeat**
7: $\lambda_{c,l} = \sum_{c'\in\mathcal{C}_l} d_{<c,c'>},\ \forall c \in \Omega, l = 1, 2, \cdots, L$.
8: $\lambda_c^{\max} = \max_{l\in\{1,\ldots,L\}} \{\lambda_c^l\},\ \lambda_c^{\min} = \min_{l\in\{1,\ldots,L\}} \{\lambda_c^l\}$.
9: $\delta_c = \lambda_c^{\max} - \lambda_c^{\min},\ c^* = \max_{c\in\Omega} \{\delta_c\}$.
10: $l^* = \arg\min_{l=\{1,\ldots,L\}} \{\lambda_{c^*}^l\}$.
11: $\mathcal{C}_{l^*} = \mathcal{C}_{l^*} \cup \{c^*\},\ \Omega = \Omega - \{c^*\}$.
12: **until** $\Omega = \emptyset$
13: **return** $\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_L$.

*2) Cluster selection steps (7-9):* To select clusters in the order that maximizes the sum-rate, an unassigned cluster with highest degree is selected and assigned in each iteration. In the step 7, we define $\lambda_c^l$ as the sum-weight of the corresponding edges, which connect the unassigned cluster $c$ and the ones in $\mathcal{C}_l$. With respect to one specific cluster $c$, we obtain the maximal (minimal) $\lambda_c^l$ as $\lambda_c^{\max}$ ($\lambda_c^{\min}$) in the step 8. Then, the variable $\delta_c = \lambda_c^{\max} - \lambda_c^{\min}$ is defined as the degree of cluster $c$. In fact, $\delta_c$ has the following meanings: the smaller $\delta_c$ indicates the pattern selection for the cluster $c$ makes less sense, while the larger $\delta_c$ implies the proper pattern selection for the cluster $c$ achieves more sum-rate gain. Hence, the cluster $c^*$ with the largest $\delta_c$ is greedily selected in the step 9.

*3) Cluster assignment steps (10-11):* We assign the cluster $c^*$ to the pattern $l^*$, where the cluster $c^*$ has the weakest ASR overlapping with other clusters in $\mathcal{C}_{l^*}$. Then, the cluster $c^*$ is removed from the set $\Omega$.

*4) Repetition:* Repeat steps 7-11 until all clusters are assigned to the corresponding pattern, i.e., $\Omega = \emptyset$.

The proposed GT-PD scheme only requires $\{\mathbf{R}_c\}$ instead of the instant CSI $\{\mathbf{H}_c\}$. Since $\{\mathbf{R}_c\}$ is determined by the ASR characteristics and slowly-varying, it can be tracked from the uplink training with the low overhead [9]. As a result, the proposed GT-PD method will have a low signaling overhead. Furthermore, the computational complexity of the proposed GT-PD algorithm is $O(LC^3)$, which is much smaller than $O(L^C)$ for the exhaustive search. Taking a system with $C = 20$ and $L = 5$ as example, we can obtain that $\frac{LC^3}{L^C} = 4.194 \times 10^{-10}$.

## IV. Numerical Results

In this section, we demonstrate the performance of the proposed GT-PD scheme through numerical simulations. The number of the antennas at the BS is $N = 128$, the BS antenna spacing equals the half wavelength and the carrier frequency is 2GHz. The variance of the noise is set as 1, and the signal-to-noise ratio (SNR) is defined as SNR $= P_t$, where $P_t = \mathbf{d}_c^H \mathbf{d}_c$

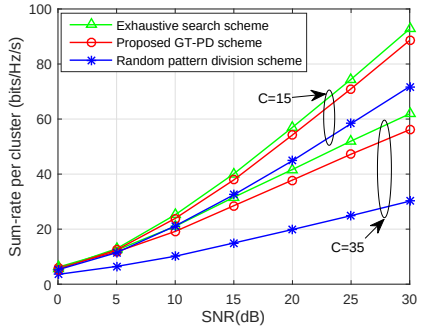


Fig. 3. Sum-rate per cluster versus SNR with $C = 15, 35$.

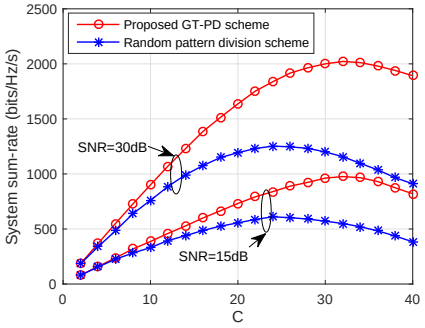


Fig. 4. The system sum-rate versus $C$ with SNR$= 15, 30$ dB.

is the transmitting power for each cluster. The channel vector for each user is generated with (4), and the number of the patterns is $L = 5$. Furthermore, we adopt the regularized zero-forcing method to calculate $\{\mathbf{V}_c\}$ [11].

Fig. 3 presents the sum-rate per cluster versus the SNR. We compare our proposed GT-PD scheme with both the exhaustive search method and the random pattern division scheme. It can be seen from Fig. 3 that the sum-rate achieved by the proposed GT-PD scheme is much higher than that of the random pattern division scheme and is very close to that of exhaustive search.

Fig. 4 shows the curves of the total sum-rate for the whole clusters versus the clusters number $C$. From Fig. 4, we can obtain the following observations. The total sum-rates of both the proposed GT-PD scheme and the random pattern division can be improved by increasing $C$ within its small-value range. However, in $C$'s large-value range, the total sum-rates for both schemes will fall off. Nonetheless, the total sum-rate of the proposed GT-PD scheme drops much more slowly than that of the random pattern division method. Furthermore, the GT-PD scheme achieves its maximal total sum-rate at $C = 32$, while the random pattern division scheme obtains the maximal sum-rate at $C = 24$. The observations verify that the proposed scheme can efficiently deal with the ASR overlapping, especially when $C$ is larger and ASR overlapping is serious.

Fig. 5 compares the proposed GT-PD scheme with the scheme that scheduling any clusters with ASR overlapping into orthogonal sub-channels (fully separating scheme). We define the sub-channel utilization rate, i.e., $(\sum_{c=1}^{C} \mathbb{R}_c)/L$, as a metric. It can be seen that in $C$'s big value range, the proposed GT-PD scheme achieves higher sub-channel

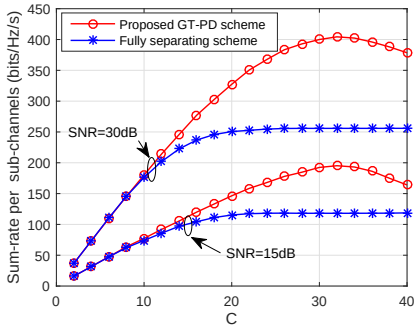


Fig. 5. The sub-channel utilization rate versus $C$ with SNR$= 15, 30$ dB.

utilization rate than the fully separating scheme. The reason is that the proposed GT-PD scheme uses a limited number of sub-channels to deal with ASR overlapping as far as possible, while the number of sub-channels used by fully separating scheme increases with $C$.

## V. Conclusion

In this letter, a graph theory based pattern division scheme with low complexity and signaling overhead has been proposed. Firstly, the ASR overlapping among the user clusters was described by one undirected weighted graph. Then, the pattern division was transformed into a graph coloring problem and solved through greedily assigning two connected clusters with a large weight to different patterns. Finally, simulation results demonstrated that the performance of the GT-PD scheme is very close to that of the exhaustive search.

## References

[1] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[2] S. Noh, M. D. Zoltowski, and D. J. Love, "Training sequence design for feedback assisted hybrid beamforming in massive MIMO systems," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 187–200, Jan. 2016.

[3] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing the large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[4] D. Kim, G. Lee, and Y. Sung, "Two-stage beamformer design for massive MIMO downlink by trace quotient formulation," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2200–2211, Jun. 2015.

[5] J. Chen and V. K. N. Lau, "Two-tier precoding for FDD multi-cell massive MIMO time-varying interference networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1230–1238, Jun. 2014.

[6] C. Sun, X. Gao, S. Jin, M. Matthaiou, Z. Ding, and C. Xiao, "Beam division multiple access transmission for massive MIMO communications," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2170–2184, Jun. 2015.

[7] V. Doshi, D. Shah, M. Médard, and M. Effros, "Functional compression through graph coloring," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3901–3917, Aug. 2010.

[8] H. Xie, B. Wang, F. Gao and S. Jin, "A Full-Space Spectrum-Sharing Strategy for Massive MIMO Cognitive Radio Systems," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 10, pp. 2537-2549, Oct. 2016.

[9] H. Xie, F. Gao and S. Jin, "An Overview of Low-Rank Channel Estimation for Massive MIMO Systems," *IEEE Access*, vol. 4, no., pp. 7313-7321, Nov. 2016.

[10] R. Y. Chang, Z. Tao, J. Zhang, and C. C. Kuo, "A graph approach to dynamic fractional frequency reuse (FFR) in multi-cell ofdma networks," in *Proc. IEEE ICC 2009*, Dresden, Germany, pp. 1–6, Jun. 2009.

[11] J. Mao, J. Gao, Y. Liu and G. Xie, "Simplified Semi-Orthogonal User Selection for MU-MIMO Systems with ZFBF," *IEEE Wireless Commun. Lett.*, vol. 1, no. 1, pp. 42-45, Feb. 2012.