User Scheduling in Wireless Networks for Deterministic Service: An Efficient Genetic Algorithm Method

You Zhao, Yuchao Zhu, *Graduate Student Member, IEEE*, and Shaowei Wang, *Senior Member, IEEE*

***Abstract*—To meet the stringent latency and reliability requirements of deterministic service in 5G and beyond networks, we investigate the user selection scheme jointly considering multiple transmission time intervals in a radio access network system with multiple access technology. The objective is to maximize the number of users that satisfy both data volume and latency requirements. Due to the NP-hardness of the formulated optimization task, we propose an efficient genetic algorithm to find promising user subsets at low complexity. Numerical results demonstrate that our proposal outperforms the compared greedy algorithm and round-robin algorithm, which can provide a guideline for the design of systems with deterministic requirements in practice.**

***Index Terms*—Deterministic service, genetic algorithm, radio access network.**

I. Introduction

To MEET the demands of the highly diverse and data-intensive domains, future networks need to efficiently utilize wireless resources in the spatial, temporal, and frequency domains. Several key radio access technologies, such as massive multiple-input multiple-output (MIMO) technology, have been developed and play a pivotal role in modern wireless communication systems to enhance spectral efficiency. Since a multi-antenna base station (BS) will simultaneously serve multiple users with the same time-frequency resources [1], user scheduling becomes crucial for interference mitigation. In this process, the BS selects a subset of users for concurrent transmission so as to optimize the overall system performance.

Over the past few decades, various user scheduling schemes have been proposed in terms of diverse metrics to achieve promising solutions with manageable complexity. Maximizing the throughput to enhance spectral efficiency is widely recognized as a crucial objective. In [2], the problem is relaxed into a semi-definite programming task, where the authors employ a randomized rounding-based greedy user selection algorithm to approximate the maximum throughput solution. In [3], multiuser wireless systems utilizing zero-forcing dirty-paper coding adopt a greedy scheduling scheme to achieve the maximum sum rate. Fairness is also a commonly considered criterion in scheduling, with round-robin and its variants highly valued for prioritizing short-term fairness [4], [5], [6]. Greedy scheduling and opportunistic round-robin scheduling is compared in [7], and the efficiency of opportunistic round-robin in overcoming the limitations of greedy scheduling is demonstrated. Furthermore, in [8], high-performance communication sets are generated through an iterative weighted sum-rate maximization procedure, followed by solving an integer programming problem to ensure fairness. The scheme proposed in [9] supports diverse users with equitable service satisfaction and flexible access by maximizing the average user utility while minimizing its distribution variance.

Recently, 5G and beyond networks are required to support a range of applications, including multimedia services with high data rates, vehicular communications with ultra-reliable and low-latency requirements, and industrial Internet with massive machine connections, where user experience is greatly emphasized [10]. Specifically, real-time applications, such as augmented reality and telesurgery, require ultra-high data volume at the gigabit level and ultra-low latency at the millisecond level [11], [12]. Such emerging applications place critical emphasis on deterministic service, the imperative requirement of which is to transmit multiple data streams within prescribed time limitations.

Traditional user selection schemes, which perform scheduling at each individual transmission time interval (TTI) based on the instantaneous channel state information (CSI) due to the time-varying nature of wireless channels, cannot meet the data volume and latency requirements in 5G and beyond networks. Hence, there is a need to develop a user scheduling scheme that takes into account multiple TTIs simultaneously to ensure deterministic communications. In most practical scenarios, the channel exhibits temporal correlation due to the nearly stationary scattering environment [13]. Channel prediction schemes have been proposed to obtain CSI of future time slots [14], which enable us to make scheduling decisions for multiple upcoming TTIs at a time.

Motivated by the aforementioned points, we focus on the multi-TTI joint user selection to provide deterministic service in a wireless network system that utilizes MIMO technology, where a data stream transmitted completely within a given latency is considered valid. The optimization task is to maximize the number of users receiving valid service. To address the challenge of the enormous search space caused by jointly considering multiple TTIs, we employ evolutionary computation principles and develop a genetic algorithm (GA) based scheme to explore promising solutions, where user selection decisions for multiple TTIs are encoded as a long

Manuscript received 8 October 2023; revised 30 November 2023; accepted 4 December 2023. Date of publication 13 December 2023; date of current version 6 March 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 61931023. The associate editor coordinating the review of this article and approving it for publication was T. Wolf. *(Corresponding author: Shaowei Wang.)*

The authors are with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: youzhao@smail.nju.edu.cn; dz20230030@smail.nju.edu.cn; wangsw@nju.edu.cn).

Digital Object Identifier 10.1109/LNET.2023.3342424

binary code string. Numerical results demonstrate that our proposed GA-based scheme can offer valid service to more users compared to other typical schemes.

The remainder of this letter is organized as follows. Section II introduces the system model and the problem formulation. Section III presents the proposed GA-based scheme in detail. Numerical results are provided in Section IV, followed by concluding remarks in Section V.

II. System Model and Problem Formulation

We consider a downlink transmission network with one BS equipped with M antennas, serving K single-antenna users. In this downlink transmission network, overlapping multiple user signals on the same resource block is allowed. Let $\Omega = \{1, \ldots, K\}$ represent the index set of all K users, from which a subset $J = \{j_1, \ldots, j_\kappa\}$ is selected for concurrent transmission at a specific TTI.

A spatially uncorrelated flat Rayleigh fading channel is assumed between each user and the BS, i.e., the channel gain vector follows a complex Gaussian distribution $\mathcal{CN}(0, \boldsymbol{I})$. The signal received by the selected user i at the tth TTI can be written as

$$y_{i,t} = \boldsymbol{h}_{i,t}\boldsymbol{x}_{J,t} + n_{i,t}, \quad i \in J, \tag{1}$$

where $\boldsymbol{x}_{J,t} \in \mathbb{C}^{M\times 1}$ is the transmitted signal from the BS antennas. $\boldsymbol{h}_{i,t} \in \mathbb{C}^{1\times M}$ is the channel gain vector of user i at the tth TTI. $n_{i,t}$ represents the additive white Gaussian noise, which is normalized to have unit variance.

The precoding matrix for the selected users is defined as $\boldsymbol{W}_{J,t} = [\boldsymbol{w}_{j_1,t}, \ldots, \boldsymbol{w}_{j_\kappa,t}] \in \mathbb{C}^{M\times\kappa}$. Let $\boldsymbol{s}_{J,t} \in \mathbb{C}^{\kappa\times 1}$ denote the transmitted source information before precoding, with its power normalized. In this letter, we adopt equal power allocation, i.e., the allocated power for each selected user is denoted as $\rho = \frac{P}{\kappa}$, where P is the maximum transmission power. Thus the transmitted signal designed for the user set J at the tth TTI is given by

$$\boldsymbol{x}_{J,t} = \sum_{i=j_1}^{j_\kappa} \sqrt{\rho}\boldsymbol{w}_{i,t}s_{i,t}. \tag{2}$$

Then, the received signal of the selected user i can be expressed as follows:

$$y_{i,t} = \sqrt{\rho}\boldsymbol{h}_{i,t}\boldsymbol{w}_{i,t}s_{i,t} + \sum_{j\neq i}\sqrt{\rho}\boldsymbol{h}_{i,t}\boldsymbol{w}_{j,t}s_{j,t} + n_{i,t}, \tag{3}$$

where the first term indicates the desired signal of the selected user i, while the second term is the interference among users. Zero forcing precoding technique is adopted to eliminate such interference for simplicity [15]. We collect the channel gain vector of selected users into a matrix $\boldsymbol{H}_{J,t} = [\boldsymbol{h}_{j_1,t}^T, \ldots, \boldsymbol{h}_{j_\kappa,t}^T] \in \mathbb{C}^{M\times\kappa}$ for notation convenience, and the precoding matrix of the selected users can be written as

$$\boldsymbol{W}_{J,t} = \sqrt{\alpha}\boldsymbol{H}_{J,t}^*\left(\boldsymbol{H}_{J,t}^T\boldsymbol{H}_{J,t}^*\right)^{-1}, \tag{4}$$

where α is a scaling factor to normalize signal power such that $tr(\boldsymbol{W}_{J,t}\boldsymbol{W}_{J,t}^H) = 1$.

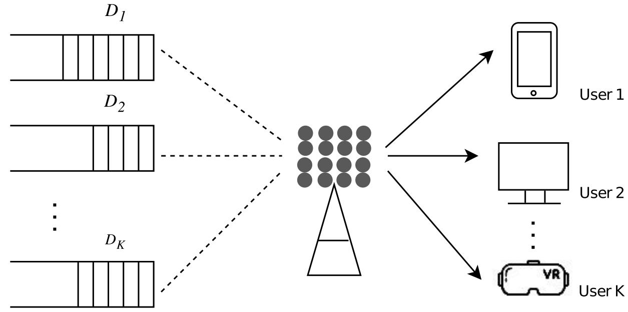

Fig. 1. A downlink system with requirements of deterministic service.

Since the interference is nullified through zero forcing precoding, the signal to interference plus noise ratio of the kth user at the tth TTI can be expressed as follows:

$$SINR_{k,t} = \begin{cases} \frac{\rho|\boldsymbol{h}_{k,t}\boldsymbol{w}_{k,t}|^2}{|n_{k,t}|^2} & , k \in J, \\ 0 & , k \notin J. \end{cases} \tag{5}$$

The corresponding transmission rate $r_{k,t}$ can be calculated by

$$r_{k,t} = B\log_2\left(1 + SINR_{k,t}\right), \tag{6}$$

where B represents the channel bandwidth, and we set $B = 1$.

As shown in Fig. 1, we consider a transmission scenario requiring deterministic service, aiming to provide valid service as many as possible. Let D_k and τ_k represent the data volume and the latency constraint of user $k \in \Omega$, respectively. The scheduling time window is determined by the maximum latency constraint, which is denoted by $T = \max_k \tau_k$.

Let λ_k represent a binary variable that signifies whether user k receives valid service or not. Note that if the data volume is not completely transmitted within the given latency, the previous transmission is deemed invalid. Thus, λ_k can be written as

$$\lambda_k = \begin{cases} 1, & R_k \geq D_k, \\ 0, & R_k < D_k, \end{cases} \tag{7}$$

where $R_k = \sum_{t=1}^{\tau_k} r_{k,t}t_0$ denotes the throughput of the user k within its latency constraint. t_0 represents a unit of time, equivalent to one TTI.

Our objective is to maximize the number of users receiving valid service throughout the scheduling time window. Let $c_{k,t}$ denote a binary variable, which indicates that user k is selected at tth TTI to transmit its data if $c_{k,t} = 1$, and 0 otherwise. Then, the optimization problem can be formulated as

$$\begin{aligned} \max_{c_{k,t}} \quad & \sum_{k=1}^{K}\lambda_k \\ s.t. \quad & C_1: c_{k,t} \in \{0,1\}, \quad \forall k \in \{1,\ldots,K\}, t \in \{1,\ldots,T\}, \\ & C_2: \sum_{k=1}^{K} c_{k,t} \leq M, \quad \forall t \in \{1,\ldots,T\}. \end{aligned} \tag{8}$$

III. Genetic Algorithm for Multistage User Selection

The optimization task is a type of combinatorial optimization problem with a solution space of $(\sum_{i=1}^{M}\binom{K}{i})^T$ that is astronomically large. Due to the NP-hardness of the task, we adopt a GA-based method to derive the near-optimal

Algorithm 1 GA-Based User Selection Algorithm

Input: N, $maxIteration$, $p_{crossover}$, $p_{mutation}$
Output: The best chromosome $\boldsymbol{C}_I$
Generate a population $\mathcal{P} = \{\boldsymbol{C}_1, \boldsymbol{C}_2, \ldots, \boldsymbol{C}_N\}$ randomly;
$iteration \leftarrow 1$, $maxfitness \leftarrow 0$;
while $iteration \leq maxIteration$ and $maxfitness < K$ **do**
  for $i = 1{:}N$ **do**
    for $t = 1{:}T$ **do**
      if $\sum_{k=1}^{K} c_{k,t} > M$ **then**
        Repair chromosome $\boldsymbol{C}_i$ by random toggling;
      end if
    end for
    Calculate fitness f_i of each chromosome $\boldsymbol{C}_i$ by (9);
  end for
  $maxfitness = \max_{i \in \{1,2,\ldots,N\}} f_i$, $I = \arg\max_{i \in \{1,2,\ldots,N\}} f_i$;
  for $i = 1{:}N$ **do**
    Calculate the selection probability p_i by (10);
  end for
  Create a new population $\mathcal{P}' = \emptyset$;
  while $|\mathcal{P}'| < N$ **do**
    Select two parents from $\mathcal{P}$ with probability $p_i, i \in \{1, 2, ..., N\}$;
    Perform crossover with probability $p_{crossover}$;
    Perform mutation with probability $p_{mutation}$;
    Insert the two offspring to $\mathcal{P}'$;
  end while
  Use $\boldsymbol{C}_I$ to replace a randomly selected chromosome in $\mathcal{P}'$;
  $\mathcal{P} \leftarrow \mathcal{P}'$;
  $iteration \leftarrow iteration + 1$;
end while
return $\boldsymbol{C}_I$

user scheduling decisions with low complexity. Specifically, we encode the user selection decisions for multiple TTIs as a long binary string, resulting in a discrete neighborhood that facilitates the exploration of GA.

GA is an optimization algorithm based on the principles of natural selection and evolution. Potential solutions are encoded in a set called population, of data structures known as chromosomes. The chromosomes undergo the process of crossbreeding, mutation, and evolution over multiple iterations, or generations, moving towards the near-optimal solution. The details of our proposed GA-based algorithm are presented in Algorithm 1, which can be summarized as the following steps.

- Initialization. As shown in Fig. 2, a chromosome $\boldsymbol{C} \in \{0,1\}^{K \times T}$ is designed to characterize the situation of user selection over T TTIs, where $c_{k,t}$ is represented by the $(K(t-1)+k)$th bit of $\boldsymbol{C}$. We create an initial population consisting of N chromosomes to represent feasible solutions, where every bit of the binary string is initialized randomly. Considering constraint C_2 in (8) which requires that the number of selected users should not exceed the antenna number, a repair process is required to avoid the appearance of invalid chromosomes. We check the count of 1 within each TTI of every chromosome and randomly toggle 1 if the count is larger than M.

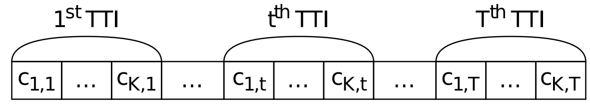

Fig. 2. Chromosome representation of user selection decisions for K users over T TTIs.

- Evaluation. Every chromosome within the population undergoes evaluation based on the value of fitness in each generation. Since our objective is to maximize the number of users receiving valid service, the fitness value of one given chromosome $\boldsymbol{C}_i$, $i \in \{1, \ldots, N\}$ is set as
$$f_i = \sum_{k=1}^{K} \lambda_k, \tag{9}$$
where λ_k is calculated by (3). Generally speaking, the higher the fitness value is, the better the quality of the chromosome will be. The fitness is upper bounded by K, indicating that all user requirements are fulfilled.
- Breeding. Selection, crossover and mutation are the GA operators to breed the offspring. The selection operation is to select two chromosomes as parents and let them pass their characteristics to the next generation. Here, we adopt fitness proportionate selection. The probability p_i of chromosome $\boldsymbol{C}_i$, $i \in \{1, \ldots, N\}$ being selected can be calculated by
$$p_i = \frac{f_i}{\sum_{j=1}^{N} f_j}. \tag{10}$$
Then, the two selected chromosomes are combined to create new offspring by the crossover operation. Specifically, a crossover position is randomly selected within the chromosome, and then the two selected parents swap all bits after that position to generate new offspring with a probability of $p_{crossover}$. Mutation is introduced to avoid trapping in a local optimal, where the bits in the bit string of newly generated offspring can be flipped with a probability of $p_{mutation}$. Elite policy is also employed, where the best chromosome in the last generation is reproduced directly to the next generation, thereby guaranteeing at least that the offspring will perform as well as their parents. To be precise, N chromosomes are obtained after the breeding operators, among which a chromosome is randomly selected and replaced by the best chromosome in the last generation.

IV. Numerical Results

We consider a downlink wireless system where the BS is equipped with $M = 8$ antennas. Unless otherwise specified, D_k is randomly chosen from an interval from 25 to 55. τ_k is randomly chosen from an interval from 15 to 20. The range of user quantities set in the experiment is from 10 to 50 [16].

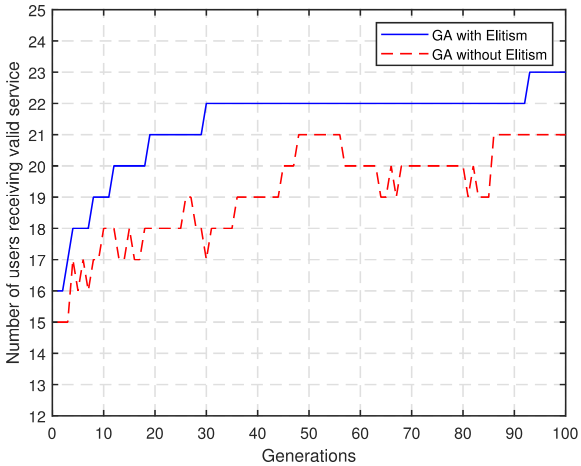

Fig. 3. Number of users receiving valid service per generation.

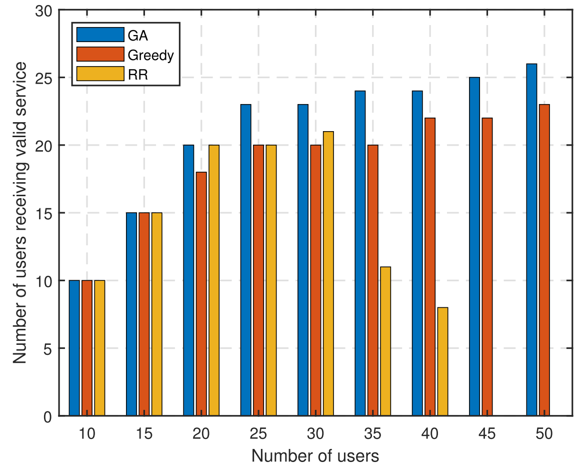

Fig. 4. Users that receive valid service as a function of total users.

For our proposed GA-based method, $p_{crossover}$ and $p_{mutation}$ are set as 0.6 and 10^{-3}, respectively. The population size N and the maximum number of iterations are set as 1000 and 100, respectively.

First, we investigate the convergence of the GA and the influence of the elite policy. Fig. 3 exhibits the number of users receiving valid service during each iteration of GA with and without elitism when $K = 30$. We can observe that the fitness increases along with the generations and finally reaches a steady state. Also, the performance of GA with elitism is better than that of GA without elitism. Generally speaking, we should balance the convergence speed and the solution accuracy in practice, which can be adjusted by the parameters in the algorithm. The computational complexity of the proposed algorithm is $\mathcal{O}(M^3TNN_I)$ flops (please refer to the Appendix). M and T are network parameters. N and the iteration count N_I can be adjusted. Larger population sizes and iteration counts imply higher computational costs but may yield better results. Hence, we can adjust the population size and maximum iteration count to achieve a trade-off between the computation cost and the result accuracy.

Then, we compare our proposal with algorithms based on two typical and popular ideas, i.e., greedy-based and round-robin-based algorithms. These algorithms are commonly employed in wireless networks and often serve as baseline algorithms for comparison against other algorithms.

- Greedy: The subset of users is selected to maximize the sum rate for the current TTI. In each TTI, the process begins by adding the user with the highest rate to the selected set. Subsequently, users are incrementally added based on their contribution to the cumulative data rate when combined with the previously selected users. This process continues until either the data rate ceases to increase or the number of users reaches the antenna count.
- RR: Before scheduling begins, multiple subsets of users, referring to as scheduled user groups, are pre-selected from the entire set of K users. Throughout the scheduling process, these user groups are sequentially chosen in a round-robin manner for each TTI.

Fig. 4 shows the number of users receiving valid service as a function of the total number of users. When the number of users is relatively small, e.g., $K \leq 15$, the system capacity is significantly greater than the user load, thus all algorithms can meet the requirements of all users. We can observe that the gain of our proposal is remarkable as the number of users increases. Specifically, when $K \geq 20$, the GA method satisfies the requirements of at least 9.1% more users as compared to the Greedy method. For the RR, extra performance gain from the increasing number of users cannot be obtained since the round-robin strategy neglects the volume and latency constraints, leading to a failure to validly serve even one user when the number of users is relatively large.

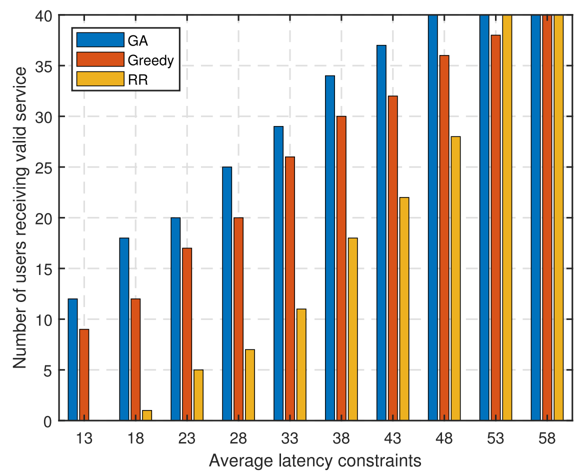

Fig. 5. Users that receive valid service as a function of latency constraints.

Fig. 5 shows the number of users receiving valid service under different latency constraints. The number of total users is 40. The horizontal axis represents average latency constraint, denoted by $\bar{\tau}$. τ_k is randomly chosen from an interval from $\bar{\tau}-2$ to $\bar{\tau}+2$. D_k is randomly chosen from an interval from 40 to 80. When faced with stringent latency constraints, e.g., $\bar{\tau} \leq 50$, the GA method significantly outperforms the other two algorithms, achieving at least 11.1% and 42.9% performance improvement over the Greedy and the RR, respectively. Additionally, with the relaxation of latency constraints, our proposal satisfies the requirements of all users more quickly and efficiently as compared with the Greedy and RR.

V. Conclusion

In this letter, we investigated a user selection problem in a downlink wireless system requiring deterministic service, where the data must be completely transmitted under the latency constraints of users. Aiming at maximizing the number of users receiving valid service, we proposed an efficient GA-based method to optimize the selected subset of users with joint consideration of multiple TTIs. More specifically, user selection within the time window is jointly coded and searched by GA using a fitness function based on the number of users receiving valid service. Numerical results demonstrate that the proposed algorithm can successfully serve more users as compared with the greedy and round-robin methods.

Appendix

Computational Complexity Analysis

For a complex-valued matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$, we outline the computational complexity of various matrix operations utilized in our proposed scheduling algorithm.

- The multiplication of an $m \times n$ matrix by an $n \times p$ matrix necessitates mnp complex additions and mnp complex multiplications, totaling $8mnp$ flops.
- The Frobenius norm $||\boldsymbol{A}||_F^2$ takes $2mn$ real multiplications and $2mn$ real additions, resulting in a total flop count of $4mn$.
- The flop count for SVD of real-valued $m \times n$ matrices is $4m^2n + 8mn^2 + 9n^3$, where $m \geq n$.

The user selection process in the GA is mainly centered on bit manipulation operations within chromosomes, which are generally negligible in computational complexity. The primary complexity stems from the evaluation of the fitness function represented by each chromosome, which is related to the transmission rate of the chosen users as defined in (6). The following analysis will omit some steps with a complexity of $\mathcal{O}(1)$. The complexity of calculating the precoding matrix of the selected users' channel gain lies in (4). The computation of the pseudo-inverse offers various methods. Assuming it is computed using SVD decomposition, the flop count would be $4M^2\kappa + 8M\kappa^2 + 9\kappa^3$. To get the α, we calculate $tr(W_{J,t}W_{J,t}^H)$. The flop count is $4M\kappa$. The flop count of (5) is $12M$ and it is $12M\kappa$ for computing κ users. As there are T time slots, one fitness calculation requires $\mathcal{O}((4M^2\kappa + 8M\kappa^2 + 9\kappa^3 + 4M\kappa + 12M\kappa)T) \approx \mathcal{O}(M^3T)$ flops. The GA calculates this metric $N \times N_I$ times, where N_I is the maximum number of iterations. Thus, the entire scheduling process is $\mathcal{O}(M^3TNN_I)$ flops.

References

[1] W. Yu, T. Wang, and S. Wang, "Multi-label learning based antenna selection in massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 7255–7260, Jul. 2021.

[2] Z. Wang, Y. Cao, D. Zhang, X. Hua, P. Gao, and T. Jiang, "User selection for MIMO downlink with digital and hybrid maximum ratio transmission," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 11101–11105, Oct. 2021.

[3] J. Jiang, R. M. Buehrer, and W. H. Tranter, "Greedy scheduling performance for a zero-forcing dirty-paper coded system," *IEEE Trans. Commun.*, vol. 54, no. 5, pp. 789–793, May 2006.

[4] O.-S. Shin and K. Bok, "Antenna-assisted round robin scheduling for MIMO cellular systems," *IEEE Commun. Lett.*, vol. 7, no. 3, pp. 109–111, Mar. 2003.

[5] B. Li, A. Eryilmaz, and R. Srikant, "Emulating round-robin in wireless networks," in *Proc. 18th ACM Int. Symp. MobiHoc*, 2017, pp. 1–10.

[6] H. H. Yang, Y. Wang, and T. Q. S. Quek, "Delay analysis of random scheduling and round robin in small cell networks," *IEEE Wireless Commun. Lett.*, vol. 7, no. 6, pp. 978–981, Dec. 2018.

[7] P. Patil and C. Borse, "Fair resource allocation to MIMO wireless system using opportunistic round robin scheduling algorithm," in *Proc. IEEE ICPC*, 2015, pp. 1–3.

[8] M. Ge and D. M. Blough, "High throughput and fair scheduling for multi-AP multiuser MIMO in dense wireless networks," *IEEE/ACM Trans. Netw.*, vol. 26, no. 5, pp. 2414–2427, Oct. 2018.

[9] W. Han, X. Wang, and J. Mei, "EDI-driven multi-dimensional resource allocation for inclusive 6G communications," *IEEE Netw. Lett.*, vol. 5, no. 2, pp. 110–114, Jun. 2023.

[10] A. Nasrallah et al., "Ultra-low latency (ULL) networks: The IEEE TSN and IETF DetNet standards and related 5G ULL research," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 88–145, 1st Quart., 2019.

[11] S. Dang, O. Amin, B. Shihada, and M. S. Alouini "What should 6G be?" *Nat. Electron.*, vol. 3, no. 1, pp. 20–29, Jan. 2020.

[12] S. Xu, M. Perez, K. Yang, C. Perrenot, J. Felblinger, and J. Hubert, "Determination of the latency effects on surgical performance and the acceptable latency levels in telesurgery using the dv-trainer simulator," *Surg. Endosc.*, vol. 28, pp. 2569–2576, Sep. 2014.

[13] J. Yuan, H. Q. Ngo, and M. Matthaiou, "Machine learning-based channel prediction in massive MIMO with channel aging," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 2960–2973, May 2020.

[14] W. Jiang, M. Strufe, and H. D. Schotten, "Long-range MIMO channel prediction using recurrent neural networks," in *Proc. 17th IEEE CCNC*, Las Vegas, NV, USA, 2020, pp. 1–6.

[15] N. Fatema, G. Hua, Y. Xiang, D. Peng, and I. Natgunanathan, "Massive MIMO linear precoding: A survey," *IEEE Syst. J.*, vol. 12, no. 4, pp. 3920–3931, Dec. 2018.

[16] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath, and B. L. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3658–3663, Sep. 2006.