Data-Driven Spectrum Partition for Multiplexing URLLC and eMBB

Haoran Peng, *Member, IEEE*, Li-Chun Wang, *Fellow, IEEE*, and Zhuofu Jian

Abstract—**Multiplexing ultra-reliable low-latency communications (URLLC) and enhanced mobile broadband (eMBB) are critical in the next generation mobile network. URLLC requires ultra-high reliability and extremely low latency, whereas eMBB demands high data rates. Thus, the coexistence system of URLLC and eMBB faces the challenge of sharing the spectrum efficiently and effectively. In this study, we comprehensively investigate the state of the art spectrum partition methods in combined URLLC and eMBB services. We formulate a joint optimization problem for maximizing the eMBB throughput and guaranteeing the URLLC performance. For the eMBB and URLLC multiplexing system, a full separative spectrum partition scheme based on data-driven genetic algorithm-based spectrum partition (DDGSP) is proposed. Our simulation results demonstrate that the proposed DDGSP can make the URLLC and eMBB coexistence system outperform the state-of-the-art methods in terms of the error rate and computational efficiency.**

Index Terms—**5G new radio, spectrum partition, wireless scheduling, URLLC, eMBB.**

Fig. 1. Coexistence schemes.

I. INTRODUCTION

A. Motivation and Problem Statement

5G NEW radio (NR) wireless systems are expected to efficiently support enhanced mobile broadband (eMBB) and ultra-reliable low-latency communication (URLLC) services [1]. The services of URLLC aims to achieve an extremely low delay (0.25–0.3 msec/packet) while guaranteeing high reliability with 99.999% packet success transmission probability [2]. Furthermore, eMBB is required to provide high throughput (gigabit per second) data rates with millisecond-level latency [3], [4]. More importantly, 5G NR wireless systems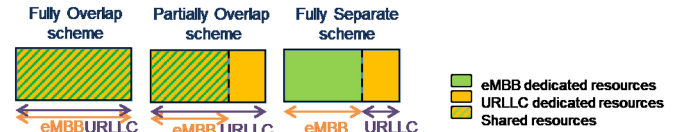 require eMBB and URLLC to be dynamically multiplexed on the same channel to improve the spectrum efficiency [5], [6]. Hence, to satisfy the communication quality of service (QoS), URLLC and eMBB adopt different random access technologies.

Furthermore, URLLC adopts grant-free random access to meet strict latency requirements by skipping the grant acquisition phase [7]. User equipment (UE) transmits data using grant-free transmission without sending a scheduling request (SR) and receiving a resource allocation (RA). This arrival-and-go manner can significantly reduce latency [8]. On the other hand, grant-based transmissions provide reliable access and high peak data rate for eMBB services, but barely reach the latency requirement [3]. Specifically, eMBB UEs continue transmitting large-sized packets and thus require high throughput [9]. The packet size of URLLC transmissions is much smaller than that of eMBB transmissions. Due to reliability and time-critical transmission constraints, URLLC UEs have a higher priority than eMBB UEs [9], [10], [11]. How to design an efficient resource allocation mechanism to maximize spectrum usage is a challenge for the coexisting eMBB and URLLC system.

To address this issue, various multiplexing mechanisms, such as fully separate (FS), fully overlap (FO), and partially overlap (PO), are proposed for sharing spectrum between URLLC and eMBB users [2]. Fig. 1 shows that the overlapped spectrum sharing schemes can outperform the separate-based method in terms of spectrum efficiency. However, PO and FO schemes must overcome interference issues because packet collisions deteriorate the performance of high-quality communication services. The FS scheme achieved better communication quality than the overlap-based spectrum partition due to the lack of packet collisions and improve the spectrum efficiency of the multiplexed eMBB and URLLC systems [12]. However, the optimal spectra-sharing strategy for multiplexed URLLC and eMBB is a challenging research issue because it is a multi-objective optimization (MOO) problem subject to complex constraints [13].

Three kinds of solutions have been proposed to resolve the MOO problem, including mathematical optimization, genetic

algorithm (GA), and reinforcement learning (RL) [14], [15], [16], [17], [18]. Mathematical optimization, like linear regression and the least-squares method (LSM), provides effective solutions to convex optimization problems with low computational complexity [16], [19]. However, the considered Moo spectrum sharing problems are nonconvex and intractable because of the coupling of multiple variables and nonconvex constraints. To address this issue, the RL-based method has been suggested [20]. Unfortunately, RL-based optimization is limited by memory capacity and computing resources [14]. Inspired by natural selection, GA was proposed to resolve nonconvex optimization using an iterative search process with biological-like operations [15]. The GA-based method has been widely used to solve the communication resource allocation problem thanks to low computational complexity and the ease of design [21], [22].

B. State-of-the-Art Methods

1) Mathematical Optimization: The authors of [17] adopted the weighted LSM to smoothen the binary matrix of photographs and optimize the quality. A generalized iteratively reweighted LSM was proposed to effectively solve the joint low-rank and sparse minimization problems [23]. Despite several benefits in [17] and [23], the simple mathematics-based method can only be used to solve the convex-structured optimization problem. Further, generalized bender decomposition (GBD) was proposed to decompose the nonconvex mixed-integer nonlinear programming problem (MINLP) into primal and master problems, and then resolve the decomposed two sub-problems iteratively until their solutions converge [24]. The authors of [25] used multi-generation cuts for the master problem to improve GBD's computational efficiency. ML techniques were proposed to further reduce the computational complexity of multi-cut GBD by dropping the cuts irrelevant to convergence [26]. The use of GBD-based optimization must satisfy the primal problem's convexity requirement after decomposition [27]. Hence, it is impractical to adopt GBD to solve the spectrum sharing optimization problem.

2) Reinforcement Learning-Based Optimization: A robust deep RL-based energy harvesting method was proposed in [20] to maximize the energy efficiency of unmanned aerial vehicle (UAV)-assisted communications subject to the QoS requirements. In [28], RL was used to allocate communication resources to improve the users' quality of experience in the Internet-of-Things networks. Another paradigm of RL-based resource allocation is to efficiently improve the throughput and reduce the interference of dynamic ultra-dense small cells [29]. Cooperative-aware multi-agent RL (MARL) was used to optimize the spectrum efficiency in vehicular networks by encouraging each agent to make decisions independently [30]. The authors of [31] adopted the Nash equilibrium-based MARL to optimize spectrum usage by determining the power-level and the sub-band without information exchange between UAVs.

3) Genetic Algorithm-Based Optimization: The effectiveness and efficiency of GA-based optimization have been demonstrated in various practical applications since it can optimize the resource allocation performance with acceptable computational complexity [15], [18], [32]. The authors of [32]

developed a GA-based model to investigate the optimal sink locations on the trajectory for sensor clusters and increase the lifetime of wireless sensor networks. A multi-objective GA was proposed to improve the average CPU usage and reduce the energy consumption of cloud data centers by dynamically predicting resources usage in the next time slot [33]. The work of [34] used multi-objective GA to optimize the complex multi-task planning for UAVs and achieved the desired performance.

4) Limitations: Mathematically based optimization algorithms, such as linear regression, LSM, and GBD, have different limitations in solving MOO problems [19], [24], [26], [35]. Linear regression solves optimization problems with a single objective, reducing its effectiveness in the MOO problem of spectrum sharing between URLLC and eMBB [19]. GBD-based methods require convexity and linear separateness for the decomposed primal and master subproblems, respectively, suitable only for specific problems [24], [26]. The LSM-based approach loses effectiveness in large-scale optimization problems and faces the challenge of high error rates [35]. However, LSM can smooth the throughput curves of URLLC and eMBB services on the same channel, which can help analyze the relationship between URLLC and eMBB to improve the spectrum efficiency [16]. Deep RL has demonstrated its ability to determine a satisfactory solution in dynamic environments [14]. Thus, the deep RL-based approaches have significant potential since they can optimize the spectrum resource allocation problem in a constantly changing communications environment. However, deep RL has high computational complexity and its performance can be easily affected by the unavoidable inaccurate estimation of the action-value function [20]. GA-based optimization method, which takes advantage of both the mathematical method and the learning algorithm, has been widely used to solve the communication resource allocation problems because of its global search capability and low computational complexity [32]. Further, the GA-based optimization algorithm performs well even when the environment changes slightly [15].

Based on the aforementioned review, this study adopts GA to optimize the spectrum allocation between URLLC and eMBB traffic, whereas the mathematical and RL-based methods are the benchmark for evaluating the proposed GA-based spectrum-sharing method.

C. Contributions

This study maximizes the spectrum usage efficiency subject to the critical QoS requirements when URLLC and eMBB transmissions share the same channel. Practically, we investigate the overlap and separate-based multiplexing mechanisms for the URLLC and eMBB coexistence system that can reduce packet collisions and provide high-quality communication. Further, this study formulates the spectrum resource allocation problem as the MOO problem because the targets of the URLLC and eMBB services are different. To address the formulated problem, we explore state-of-the-art optimization methods, such as the mathematical, GA, and RL methods, and provide a comprehensive comparison of these methods. The main contributions of this study are as follows:
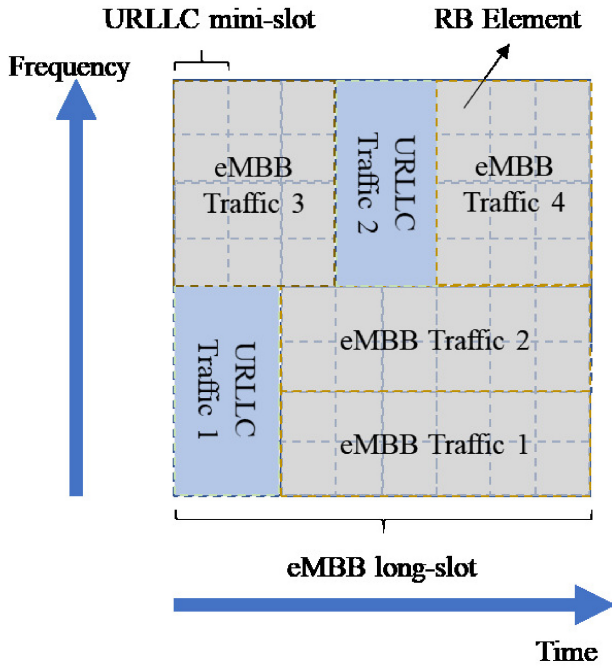
Fig. 2.   The multiplexing mechanism of URLLC and eMBB.

1) We present the multiplexing URLLC and eMBB model based on the FS mechanism and then formulate the corresponding spectrum-sharing problem as the MOO problem.
2) We propose a data-driven GA-based spectrum partition (DDGSP) to maximize eMBB throughput while satisfying URLLC latency and reliability requirements. The proposed DDGSP can perform well even under slight changes in the communication environment.
3) We verify the effectiveness and efficiency of the proposed DDGSP through rigorous numeral simulations. Moreover, the detailed comparison of the proposed DDGSP and state-of-the-art methods demonstrates the superiority of the proposed spectrum partition in terms of trade-off efficiency and practicality.

D. Paper Organization

The rest of this article is organized as follows. The system model and problem formulation are given in Sections II and III, respectively. The design of the benchmarks is detailed in Section IV. Section V presents the proposed spectrum partition mechanism. The numerical results and conclusions are presented in Sections VI and VII, respectively.

Notations: Table I summarizes a partial of the important notations in this paper.

II. SYSTEM MODEL

A. System Assumption

This study considers that multiple URLLC and eMBB UEs perform transmission in a grant-free and grant-based manner, respectively. Further, the arrival of the UE data packet for URCCL and eMBB traffic follows the Poisson process with data rates ϵ_u and ϵ_e, respectively. Denote N_u and N_e as the UE number of the URLLC and that of the eMBB, respectively.

TABLE I
GLOSSARY OF NOTATIONS

Symbol	Description
N_u	The number of URLLC UEs
N_e	The number of eMBB UEs
ϵ_u	The arrival data rate of URLLC traffic
ϵ_e	The arrival data rate of eMBB traffic
N_{RB}	The total number of RBs
$N_{RB}^{u_i}$	The number of RBs assigned to the ith URLLC receiver
$N_{RB}^{e_j}$	The number of RBs assigned to the jth eMBB receiver
M	The number of URLLC mini-slots in each eMBB long slot
N_p	The number of total packets provided by the URLLC trasmitter in each long slot
$N_{ACK_1}^{u_i}$	The number of received packets of the first transmission for the i-th URLLC receiver
$N_{ACK_2}^{u_i}$	The number of received packets of the retransmission for the i-th URLLC receiver
t	The index of time (long) slot
τ	The number of total time (long) slots
T	The duration of a time (long) slot t
$\mathcal{P}_u(t)$	The total reliability of URLLC in each long slot t
f_b	The bandwidth of the RB b
$l \in \{1, 2\}$	The index of the first transmission or retransmission
$K_b^l(t)$	The number of mini-slots for the l-th transmission during the time slot, t
$E_b(t)$	The downlink transmission power on the RB b
$h_b(t)$	Time-varying Rayleigh fading channel gain on the RB b
GFO	The entire time of a single grant-free transmission
P_u	The data size of a packet for the URLLC UE
P_e	The data size of a packet for the eMBB UE
$\mathcal{L}_{ACK_l}^{u_i}$	The latency for each transmission l of the URLLC receiver i
$N_{ACK}^{e_j}$	The number of ACK packets received by the eMBB receiver, j, in each time slot t
$\mathcal{R}_{e_j}(t)$	The effective throughput of the eMBB UE, j, in each time slot t
$\mathcal{P}_{target}$	The required URLLC reliability threshold
p_{s_k}	The probability of a strategy k being selected
$\mathcal{V}_{fitness}^k$	The fitness value of a strategy k
N_{RB}^k	The total number of RBs assigned to URLLC traffic in the strategy k
p_m	The probability of the mutation in DDGSP
r_t	The reward of Q-learning spectrum partition in each time slot t

Referring to [12], we consider that the grant-free opportunity (GFO) presents the entire time of a single grant-free transmission, in which the URLLC UE receives the acknowledgment (ACK) from the BS until its timer expires. The transmitter retransmits the same ACK data after the first transmission to minimize the negative effect of packet collision.

However, retransmission is allowed only once due to a latency constraint [36]. Based on 3GPP standards [1], URLLC transmissions are scheduled with a short transmission time interval size of a mini-slot (0.125 msec), whereas eMBB transmissions are scheduled with 1 ms. Figure 2 shows that the URLLC packet preempts the eMBB transmission and spans multiple frequency bands [5].

B. URLLC Transmission Model

The resource blocks (RBs) assigned to URLLC transmissions should not be accessed by eMBB transmissions to avoid packet collisions between eMBB and URLLC. Let N_{RB} represent the total number of RBs. Denote $N_{RB}^{u_i}$ and $N_{RB}^{e_j}$ as the number of RBs assigned to the i-th URLLC receiver and the j-th eMBB receiver in each time (long) slot t, respectively. Consider M URLLC mini-slots in each eMBB long slot for this assumption. The BS does not have precise parameters for URLLC packets due to the single link dependency of grant-free transmission [8]. As shown in Figure 2, each URLLC traffic provides two transmissions for a URLLC UE, i.e., a first transmission and retransmission, which occupies two mini-slots. During each sampling period, the URLLC transmitter provides receivers with information on the number of total packets, N_p. The total reliability of URLLC, $\mathcal{P}_u$, in each long-slot (1 ms), t, can be calculated as

$$\mathcal{P}_u(t) = \sum_{i=1}^{N_u} \frac{N_{ACK_1}^{u_i} + N_{ACK_2}^{u_i}}{N_p}, \quad (1)$$

where $N_{ACK_1}^{u_i}$ and $N_{ACK_2}^{u_i}$ denote the number of URLLC ACK packets of the first transmission and retransmission for the i-th URLLC receiver, respectively.

Assume that the RBs assigned for the i-th URLLC UE are consistent throughout the first transmission and retransmission phase. Without loss of generality, it is assumed that each transmission occupies one mini-slot. The number of ACK packets received by the i-th URLLC receiver at the l-th transmission during the time-slot t can be obtained by

$$N_{ACK_l}^{u_i} = \sum_{b=1}^{N_{RB}^{u_i}} \frac{f_b K_b^l(t)}{M \times P_u} \log_2\left(1 + \frac{E_b(t)h_b(t)}{\sigma^2}\right) \times T, \quad (2)$$

where f_b is the bandwidth of the RB b, and $K_b^l(t)$ is the number of mini-slots for the l-th ($l \in \{1, 2\}$) transmission of the i-th URLLC UE during the time slot, t. $E_b(t)$ and $h_b(t)$ represent the downlink transmission power and time-varying Rayleigh fading channel gain on the RB b, respectively. T is the duration of the time slot, t, equal to the long slot of eMBB. P_u denotes the data size of a packet for the URLLC UE, i. Therefore, the latency for each transmission l of the URLLC receiver i can be denoted as:

$$\mathcal{L}_{ACK_l}^{u_i} = \frac{T}{M \times N_{ACK_l}^{u_i}}$$

$$= \frac{1}{\sum_{b=1}^{N_{RB}^{u_i}} \frac{f_b K_b^l(t)}{P_u} \log_2\left(1 + \frac{E_b(t)h_b(t)}{\sigma^2}\right)}, \quad \forall l \in \{1, 2\}.$$
$$(3)$$

C. eMBB Transmission Model

The eMBB traffic uses the grant-based transmission, in which the transmitter knows the receiver. As collision-free access, the grant-based transmission contains three links, which include SR, RA, and ACK transmissions. The effective throughput, $\mathcal{R}_{e_j}(t)$, of the eMBB UE, j, can be calculated throughout the Shannon capacity and is given as follows:

$$\mathcal{R}_{e_j}(t) = f_b \sum_{b=1}^{N_{RB}^{e_j}} \left(1 - \frac{K_b^l(t)}{M}\right) \log_2\left(1 + \frac{E_b(t)h_b(t)}{\sigma^2}\right). \quad (4)$$

Therefore, the number, $N_{ACK}^{e_j}$, of ACK packets received by the eMBB receiver, j, in the time slot t can be expressed as follows:

$$N_{ACK}^{e_j} = \frac{\mathcal{R}_{e_j}(t) \times T}{P_{e_j}}, \quad (5)$$

where P_{e_j} is the size of the data packets for each eMBB receiver j. The total capacity of eMBB transmissions during the time slot, t, can be obtained as follows:

$$\mathcal{R}(t) = \sum_{j=1}^{N_e} \mathcal{R}_{e_j}(t)$$

$$= \frac{1}{T} \sum_{j=1}^{N_e} N_{ACK}^{e_j} * P_{e_j}. \quad (6)$$

III. PROBLEM FORMULATION

The objective of this study is to reduce RB waste under the premise of satisfying URLLC reliability requirements and eMBB throughput. Following [12], we give URLLC traffic a higher priority than eMBB traffic. Therefore, this study maximizes eMBB throughput while satisfying the constraints of the URLLC reliability requirement. The problem can be formulated as follows:

$$\max_{N_{RB}^{u_i}} \mathcal{R}(t) = \sum_{j=1}^{N_e} \mathcal{R}_{e_j}(t),$$

$$s.t. \quad C1 : \mathcal{P}_u(t) \geq 99.999\%$$

$$C2 : \mathcal{L}_{ACK_l}^{u_i} \leq 0.25\text{msec/packet}, \forall i \in [1, N_u], l \in \{1, 2\}$$

$$C3 : \sum_{i=1}^{N_u} N_{RB}^{u_i} + \sum_{j=1}^{N_e} N_{RB}^{e_j} = N_{RB}, \quad (7)$$

where $C1$ and $C2$ represent URLLC reliability and delay requirements, respectively. $C3$ is the upper limited number of RBs for all URLLC and eMBB UEs. The optimization problem is nonconvex due to the discrete feasible region and the coupling of multiple variables, implying that it is difficult to effectively solve this problem using standard convex optimization methods. This study develops a novel DDGSP to optimize the spectrum allocation between URLLC and eMBB traffic, while mathematical and RL-based methods are the benchmarks for evaluating the proposed approach.

IV. DESIGN OF THE BENCHMARKS

In the literature review, we examined several optimization methods such as linear regression, LSM, GBD, RL, and GA. Linear regression performs a regression task that requires a linear relationship of the input and output variables. However, because of multiple complexity constraints, the MOO problem formulated in (7) is not linear. The principle of GBD is to decompose the MINLP problem into a primal and master problem [24]. The primal problem of GBD is convex and corresponds to the original MINLP problem when the integer variables are fixed. The master problem must be linear given the continuous variables [26]. Unfortunately, all the variables in (7) are integer variables that cannot be efficiently solved by GBD. Therefore, we design LSM- and RL-based spectrum sharing approaches as benchmarks to evaluate the proposed DDGSP. Notably, URLLC reliability and eMBB throughput are observed to be directly and inversely proportional to the total number of RBs assigned to URLLC, $N_{RB}^U = \sum_{i=1}^{N_u} N_{RB}^{u_i}$, respectively.

A. Least-Squares Method-Based Spectrum Sharing

The relationships between URLLC reliability and eMBB throughput are nonlinear due to various channels, and cannot be obtained directly. LSM is a common algorithm for solving the regression problem by approximating the minimum sum of the squares of the residuals between the observed and predicted points. Therefore, this study adopts LSM to fit the curve of the relationship between URLLC reliability, eMBB throughput, and the total number of RBs assigned to URLLC, i.e., $\sum_{i=1}^{N_u} N_{RB}^{u_i}$. The nonlinear relationship between URLLC reliability can be denoted as $(\mathcal{P}_u(t), N_{RB}^U)$. We approximate the nonlinear relationship with the linear curve $\mathcal{P}_u(t) = \beta_1 \times N_{RB}^U + \beta_2$, based on LSM [37]. Then, we can approach the N_{RB}^U threshold that satisfies the requirement of $\mathcal{P}_u(t) \geq 99.999\%$. The value of N_{RB}^U must accurately reach the point $\mathcal{P}_u(t) = 99.999\%$ to avoid resource waste and maximize the performance of eMBB.

The discrete values of N_{RB}^U result in a constant fluctuation in URLLC reliability, which can be observed using an exhaustive algorithm. Following [17], we further smoothen the LSM curve by reconstructing N_{RB}^U into $N_{RB}^U = N_{RB}^{U'} + \delta$, where δ is small and rapidly varying and $N_{RB}^{U'}$ is close to N_{RB}^U. The weighted LS energy function for the smoothed curve can be expressed as follows:

$$\min \left\{ \left\| N_{RB}^{U'} - N_{RB}^U(i) \right\|^2 + \lambda \sum_{i=1}^m \left(N_{RB}^U(i+1) - N_{RB}^U(i) \right)^2 \right\}, \quad (8)$$

where m is the number of sampled data and $N_{RB}^U(i)$ is the number of RBs assigned to URLLC transmission in the i-th data. λ is a smoothing parameter. The LS problem can be resolved by the following:

$$\left(I + \lambda D^T D \right) N_{RB}^{U'} = N_{RB}^U, \quad (9)$$

where D is a matrix $n \times (n+1)$, which is defined as follows:

$$D = \begin{bmatrix} -1 & 1 & 0 & & 0 & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 0 & -1 & & 0 & 0 & 0 \\ & \vdots & & \ddots & & \vdots & \\ 0 & 0 & 0 & & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & 0 & & 0 & -1 & 1 \end{bmatrix} \quad (10)$$

The reconstructive data can be obtained using the above equation.

B. Q-Learning-Based Spectrum Sharing

This study considers the RL expressed as $\langle S, A, P_T, R, \gamma \rangle$. $A = \{a_1, a_2, \ldots, a_t\}$ and $S = \{s_1, s_2, \ldots, s_t\}$ are finite sets of actions and states, respectively. By applying the action $a_t \in A$ of the time step t, which is the time slot t in this study, in the current state, $s_t \in S$, the system will return an instant reward $r_t \in R : S \times A \times S \to \mathbb{R}$ to the agent. Then, the communication environment changes from the current state, s_t, to a new state, $s_{t+1} \in S$, of the next time step, $t + 1$. The transition function based on a probability distribution over the set of possible transitions is denoted as $P_T : S \times A \times S \to [0, 1]$. Typically, the reward function, $\mathcal{R}$, and the transition function, $\mathcal{P}$, comprise the model, $\pi_* : S \to A$, to maximize the long-term reward calculated by the following:

$$\max_{\pi_*} J(\pi_*) := \mathbb{E} \left[\sum_t \gamma^t r_t(s_t, \pi_*(s_t)) \right], \quad (11)$$

where $\gamma \in [0, 1]$ is the discounting factor that determines the importance of future rewards based on the current state.

We consider a Q-learning algorithm with low computational complexity and rapid convergence since the time-critical URLLC service requires fast decisions. The state and observation space, action space, and reward design are illustrated in the following.

1) Action and Observation Space: For simplicity, we define the action space as increasing or decreasing the number $\Delta_t \in \{-\Delta, +\Delta\}$ of RBs assigned to URLLC transmission. The action of the agent at the time step t can be expressed as:

$$a_t = \pm \Delta. \quad (12)$$

The Fibonacci sequence is used to generate the value of action Δ according to [38]. It follows that

$$a_t = \pm \left(C_{n-1}^0 + C_{n-2}^1 + \cdots + C_{n-1-m}^m \right), n \in [2, 7]. \quad (13)$$

The state of the communication environment directly affects the policy's decision and the reward of the current action. The number of RBs assigned to URLLC directly affects the reliability under the assumption of the system model, implying that it is also closely related to the reward of this action. The more the RBs assigned to the URLLC transmission, the greater the reliability of URLLC. Therefore, we define the number of RBs currently assigned to the URLLC transmissions as the observation at the time step, t, which is given by:

$$s_t = N_{RB}^U. \quad (14)$$

Therefore, the transited observation of the next time step s_{t+1} can be expressed as follows:

$$s_{t+1} = s_t + a_t = N_{RB}^U(t) \pm \Delta_t. \tag{15}$$

2) Reward Design: Before the agent performs the action of the next step, the system assesses the quality of this action and performs the corresponding state transition. The positive reward represents the objective of the proposed framework, which is to maximize the total eMBB throughput of the coexistence system while guaranteeing the required URLLC reliability. Further, eMBB throughput and URLLC reliability are inversely and positively correlated with N_{RB}^U, respectively. Therefore, eMBB throughput reaches the maximum value when URLLC reliability meets the demand threshold simultaneously. We define the reward function as follows:

$$r_t = \theta - \alpha \times \left(\mathcal{P}_{target} - \mathcal{P}_u(t)\right)^2, \tag{16}$$

where $\mathcal{P}_{target} = 99.999\%$ and $\mathcal{P}_u(t)$ represents the current URLLC reliability based on the action performed at the time step t. θ is a given constant, and $\alpha > 0$ represents the effect of the difference between the reliability on the overall score. The designed reward function shows that the closer the current URLLC reliability is to the target reliability, the higher the reward obtained. Furthermore, the maximum number of RBs assigned to eMBB is achieved when the URLLC requirements are satisfied. The throughput of eMBB decreases as the reliability of URLLC increases. Therefore, the reward function (16) aims to reduce the deviation between the required URLLC reliability threshold and the current URLLC reliability.

3) Training Process: Figure 3 illustrates the training process of the proposed Q-learning-based spectrum partition. After applying the current state and action pairs (s_t, a_t) in the environment, the environment returns the corresponding reward r_t of the time step t and transfers the state to the next step s_{t+1}. The goal of Q-learning is to determine the optimal strategy, $\pi_* : s \rightarrow \Delta_t$, in which the agent is to update its Q table based on the reward and state obtained. The Bellman function for the agent to update its policy can be expressed as follows:

$$Q(s_t, a_t) = (1-\omega)Q(s_t, a_t) + \omega \max_{a_{t+1}}(r_t + \gamma Q(s_{t+1}, a_{t+1})), \tag{17}$$

where ω is the learning rate. At each time step, the optimal action is selected based on the current strategy of the environment state and the reward. After several iterations, the Q-table finally converges to the optimal strategy.

Algorithm 1 shows the Q-learning-based spectrum sharing for the URLLC and eMBB coexistence system. Furthermore, the coordinate of the designed Q-learning table is the state representing the number of RBs assigned to URLLC UEs, N_{RB}^U, and the abscissa is the action. The value of $s_t + \Delta_t$ cannot exceed $[0, N_{RB}]$ since there is a limited number of total RBs. Otherwise, the reward is set to -100 if the value of $s_t + \Delta_t$ exceeds $[0, N_{RB}]$.

V. DATA-DRIVEN GENETIC ALGORITHM-BASED SPECTRUM PARTITION

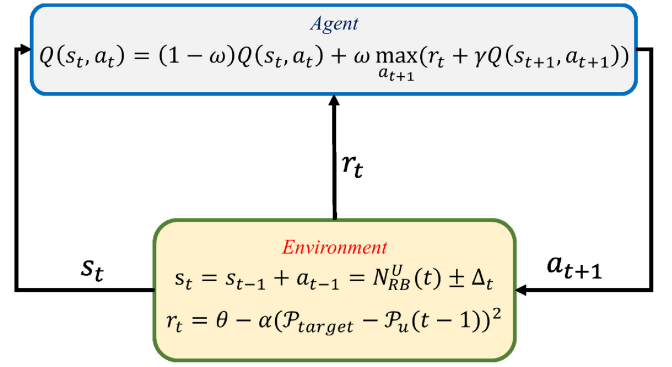In this section, we propose a novel DDGSP to efficiently allocate spectrum resources for the URLLC and eMBB

Fig. 3. The learning process of the Q learning-based spectrum partition.

Algorithm 1: Q-Learning-Based Spectrum Sharing

1 **Initial**: All state and action pairs, $Q(s_t, a_t)$, are initialized to zero;
2 **Initial**: A given number of total steps, τ;
3 **for** $t = 0$ *to* τ **do**
4   Select an optimal action, a_t, based on the current Q-value strategy, $Q(s, \cdot)$, for the current environment state s_t;
5   Applying the current state and action pairs (s_t, a_t) in the envirnment;
6   Obtain the corresponding reward, r_t, of the time step, t, and transfers the state to the next step s_{t+1};
7   Using the Bellman Equation (17) to update the policy.
8 **end**

---

multiplexing system. Inspired by Darwin's theory of evolution, GA transforms specific problems into basic chromosome-like information sequences. It solves a problem with data that have a similar chromosome-like structure by encoding the solution, retaining key information, and then reorganizing the data to obtain the corresponding results. GA can continuously optimize the problem and determine the most suitable results using selection techniques such as crossovers, and mutations. The proposed DDGSP, which is a combination of the original GA and the data collection process, is introduced in this section. The DDGSP population is the number of RBs N_{RB}^u assigned to URLLC UEs.

A. Evaluation

The evaluation function in the proposed DDGSP assesses each individual in the population to obtain its corresponding fitness value. The throughput of eMBB and the reliability of URLLC are inversely and positively correlated with the number of RBs assigned to URLLC UEs, respectively. The higher the URLLC reliability, the lower the eMBB. This is because the total RBs resources are finite in the coexistence system. Based on the aforementioned discussions, it can be concluded that when URLLC reliability and latency just meet the demand threshold, the throughput of eMBBs simultaneously reaches a relative maximum. The fitness value is derived from the
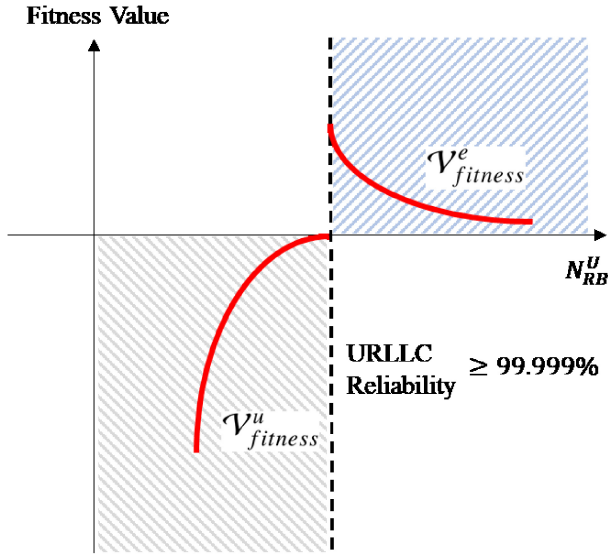
Fig. 4. Fitness Function.

optimization problem's objective to evaluate the resource allocation strategy. Therefore, the balance of the URLLC QoS requirements and eMBB throughput reaches the optimal solution when the fitness value is maximized. If the URLLC latency requirement cannot be met, the fitness value for the current strategy can be denoted as:

$$\mathcal{V}_{fitness}^{u} = -1, \text{ if } \exists \mathcal{L}_{ACK_l}^{u_i} > 0.25\text{msec/packet.} \quad (18)$$

If the URLLC latency is guaranteed but its reliability is not satisfied ($< 99.999\%$), the fitness value can be expressed as follows:

$$\bar{\mathcal{V}}_{fitness}^{u} = -\left(\mathcal{P}_{target} - \mathcal{P}_u(t)\right)^2, \mathcal{P}_u(t) < 99.999\%, \quad (19)$$

where $\mathcal{P}_{target}$ represents the URLLC reliability threshold (99.999%). The fitness value will be calculated from (18) if both the reliability and latency of URLLC do not satisfy the requirement. $\mathcal{V}_{fitness}$ approaches to zero when the reliability of URLLC $\mathcal{P}_u(t)$ closes to the target reliability of 99.999% and the latency requirement is guaranteed. Further, the fitness value is re-generated to maximize the eMBB throughput when URLLC reliability and latency are guaranteed and is written as

$$\mathcal{V}_{fitness}^{e} = \frac{1}{T} \sum_{j=1}^{N_e} N_{ACK}^{e_j} * P_{e_j}. \quad (20)$$

Referring to (18), (19), and (20), the fitness value for the current strategy k can be expressed as

$$\mathcal{V}_{fitness}^{k} = \begin{cases} \mathcal{V}_{fitness}^{u}, \text{ if } \exists \mathcal{L}_{ACK_l}^{u_i} > 0.25\text{msec/packet} \\ \bar{\mathcal{V}}_{fitness}^{u}, \text{ if } \mathcal{P}_u(t) < 99.999\% \\ \mathcal{V}_{fitness}^{e}, \text{ otherwise.} \end{cases} \quad (21)$$

Figure 4 shows that one can obtain the fitness function by classification when the URLLC latency requirement is satisfied. In the figure, the abscissa represents the number of RBs assigned to URLLC, and the ordinate represents the fitness value corresponding to the current number of RBs. The left part is the fitness value when URLLC reliability does not satisfy the requirements. The right part is the fitness value to maximize the eMBB throughput. Fitness value of different individuals can assess whether the allocation strategy is appropriate for the current network. Therefore, the fitness value is designed to be negative when URLLC reliability and latency requirements cannot be met. Furthermore, the positive fitness value is positively correlated with the eMBB throughput when all URLLC requirements are satisfied. The optimization objective of this study is equivalent to finding the strategy to maximize the fitness value. However, to exhuasively find the maximized fitness value is time consuming. To this end, we proposed a DDGSP algorithm to reach the maximum fitness value.

B. Selection

The input of the GA operation consists of each individual and its corresponding fitness value. The allocation strategy with fitness values exceeding a given threshold will be selected as the next-generation strategy according to Darwin's theory of survival of the fittest. The GA operation creates an intermediate population by the selection step before generating the next population from the current population. Following [39], the proposed DDGSP uses the roulette-wheel random method to implement the selection operation.

Let each spectrum resource allocation strategy be an individual. With n individuals, the fitness value of the k-th individual is defined as $\mathcal{V}_{fitness}^{k}$. The total number of RBs assigned to URLLC traffic in this strategy is indicated as N_{RB}^{k}. The corresponding area of each individual in the roulette, representing the probability of being selected, is denoted as

$$p_{s_k} = \frac{\mathcal{V}_{fitness}^{k}}{\sum_n \mathcal{V}_{fitness}^{k}}. \quad (22)$$

The strategy that maximizes the eMBB throughput under the URLLC requirements has the highest probability of being selected according to (22), which is equivalent to the objective of the problem described in (7). After obtaining the selected probability p_{s_k} of each resource allocation strategy, the individual is sent to the roulette wheel for stochastic sampling. Thus, the first batch of the intermediate population is generated. For each selection, the higher the individual's fitness value, the greater the probability of it being selected. Therefore, the individual with a high fitness value must be selected multiple times after several independent repeated selections. In this way, the DDGSP can naturally simulate the survival rules of the fittest and choose the optimal resource allocation strategy quickly.

C. Crossover and Mutation

In the intermediate generation, different individuals will be randomly selected as parents to create new offspring by mutation and crossover. The next generation is formed by the intermediate generation and new offspring. The offspring are formed by implementing crossover and mutation operations. Initially, we define the set of all allocation strategies as the

population. We do not use binary string encoding for each individual, rather, we use the allocation strategy as the processing object.

The crossover and the mutation operations are described below. We randomly select two individuals from the intermediate population as father and mother and define them as N_{RB}^f and N_{RB}^m, respectively. Their corresponding fitness values are $\mathcal{V}_{fitness}^f$ and $\mathcal{V}_{fitness}^m$, and the offspring are calculated based on the following formula:

$$N_{RB}^{OS} = \frac{\mathcal{V}_{fitness}^m}{\mathcal{V}_{fitness}^f + \mathcal{V}_{fitness}^m} N_{RB}^m + \frac{\mathcal{V}_{fitness}^f}{\mathcal{V}_{fitness}^f + \mathcal{V}_{fitness}^m} N_{RB}^f \qquad (23)$$

According to (23), the higher the parent's fitness value, the higher the probability that it will affect the offspring. This is because the offspring can approach to the optimal strategy without falling into the local optimum. This process illustrates how the crossover operation produces new offspring.

Naturally, mutations are uncertain and have an extremely low probability of occurrence. Thus, not every generation has mutants. The probability of mutation is determined by the given probability $p_m \ll 1$. In this study, we use a uniformly distributed function to generate a random number $\phi \in (0, 1)$. If $\phi \geq p_m$, the mutation operation is applied to the current generation; otherwise, there are no mutants in this generation.

For the mutation operation, we randomly generate a number $\mu \in (0, 2)$ and use (24) to mutate the current individual k to generate a new offspring

$$N_{RB}^{om} = \mu \times N_{RB}^k \qquad (24)$$

We can get different new offspring using the aforementioned cross-mutation. These new offspring and the intermediate population generated in the selection replace the current generation as the next generation population. Mutation and crossover operations can prevent the DDGSP from falling into the local optimum.

The aforementioned steps form one iteration of the proposed DDGSP. The pseudocode of the proposed DDGSP is summarized in Algorithm 2. The allocation strategy is continually optimized until it obtains the appropriate spectrum resource allocation. Specifically, this study defines $\S(\mathcal{G})$ as the highest fitness value in each iteration $\mathcal{G}$. Therefore, the stopping condition of the proposed DDGSP can be defined as $\S(\mathcal{G}) == \S(\mathcal{G} + 1)$, indicating that the highest fitness in the new generation no longer increases.

D. Convergence Analysis

The proposed DDGSP performs the roulette wheel selection to select individuals to generate offspring. Individuals with high fitness have a higher probability of being selected to produce offspring according to (22). Due to multiple selection operations being performed in each iteration, the population size of the offspring is several times greater than that of their parents. The high-fitness individual's offspring will increase with iterations, similar to Darwin's theory of evolution. The

---

Algorithm 2: The Proposed Data-Driven Genetic Algorithm-Based Spectrum Partition

1 **Input**: The index of iterations $\mathcal{G} = 0$, the index of selection, $\varpi = 0$, in each iteration, the highest fitness value, $\S(\mathcal{G}) = -1$, in the $\mathcal{G}$-th iteration;

2 **Input**: The ratio, χ, of the individual being selected for *crossover* in each iteration, the number of selections, ζ, in each iteration;

3 **Input**: The given probability of mutation $p_m \ll 1$;

4 **Initial**: Initial the dataset, $\mathcal{D}$, of URLLC and eMBB traffics in the multiplexing system;

5 **Initial**: Randomly generate a population of n individuals;

6 **Evaluate**: Calculate the fitness value of each individual k using (21) on $\mathcal{D}$;

7 **while** $\mathcal{G} = 0$ *or* $\S(\mathcal{G})! = \S(\mathcal{G} - 1)$ **do**

8   **Selection:**

9   Calculate the selection probability, p_{s_k}, of each individual k using (22);

10   **for** $\varpi < \zeta$ **do**

11   Use roulette wheel selection method and the selection probability, p_{s_k}, to select $\chi \times n$ individuals as the set of *crossover*;

12          $\varpi = \varpi + 1$;

13   **end**

14       $\varpi = 0$;

15       $(1 - \chi) \times n$ individuals with the greatest fitness value are directly put into the set of the next-generation individuals;

16   **End**

17   **Crossover:**

18   Pair the *crossover* individuals up;

19   Produce offspring by (23) to generate the next-generation individuals;

20   **End**

21   **Mutation:**

22   Generate a random number $\phi \in (0, 1)$ follows the uniformly distributed function;

23   **if** $\phi \geq p_m$ **then**

24   Randomly generate a number $\mu \in (0, 2)$;

25   Randomly select $\phi \times n$ individuals from the generation $\mathcal{G} + 1$ for mutation;

26   Use (24) to mutate the selected individuals generate new offspring;

27   **end**

28   **End**

29   **Evaluation:**

30   Calculate the fitness value of each individual k in the geneartion $\mathcal{G} + 1$ using (21) on $\mathcal{D}$;

31       $\mathcal{G} = \mathcal{G} + 1$;

32   **End**

33 **end**

34 **Return**: The individual with the highest fitness value;

---

convergence of GA has been proved using Markov chain theory [40]. The convergence performance of GA depends on the probabilities of crossover and mutation [41]. However,

TABLE II
SIMULATION PARAMETERS

Environment Parameters	Default Value
N_{RB}	200
N_u	40
N_e	40
ϵ_u	3 packets/mini-slot
ϵ_e	10 packets/mini-slot
f_b	15 KHz
P_u	32 bytes
P_e	160 bytes
GFO	0.25 msec
Reward Function Parameters	
$\mathcal{P}_{target}$	99.999%
θ	100

the classical GA-based method can converge to a suboptimal solution [40]. This happens because the individual with the globally optimal solution may be eliminated by selection. To this end, the proposed DDGSP selects $(1-\chi) \times n$ individuals with the highest fitness value directly into the next generation. Therefore, once a globally optimal solution is found, it must survive in the population and produce offspring. The DDGSP must converge to a globally optimal solution [40], [42].

VI. SIMULATION

In this section, the proposed DDGSP-based method is evaluated through rigorous numerical simulations. In addition, we provide comprehensive comparisons between the proposed DDGSP and the state-of-the-art techniques in terms of computational complexity and spectrum sharing performance. Following [43], we consider the propagation scenario in an urban area. The BS sends small packets to URLLC UEs using the grant-free transmission mechanism, whereas large packets are sent to eMBB UEs using the grant-based transmission mechanism. The communication environment in this study is constantly changed because the packets are transmitted randomly. Table II lists the simulation parameters.

A. System Performance of Least Square Method

Figure 5 shows the simulation results of the LSM-based spectrum partition. In the figure, the URLLC reliability and EMBB throughput curves have been smoothened. The objective of the LSM-based spectrum partition is to determine the point closest to 99.999%. The figure shows that the number of RBs assigned to URLLC is 24 when the reliability of URLLC just reaches 99.999% and the eMBB throughput achieves the corresponding maximum value. We conducted 1,000 repetitions of the URLLC and eMBB multiplexing experiments to verify the obtained spectrum partition strategy. In addition, we define the evaluation error rate as follows:

$$\mathrm{K}_{error} = \frac{\mathcal{N}_{fail}}{\mathcal{N}_{total}}, \quad (25)$$

where $\mathcal{N}_{fail}$ is the number of experiments in which URLLC reliability did not meet demand and $\mathcal{N}_{total}$ is the total number
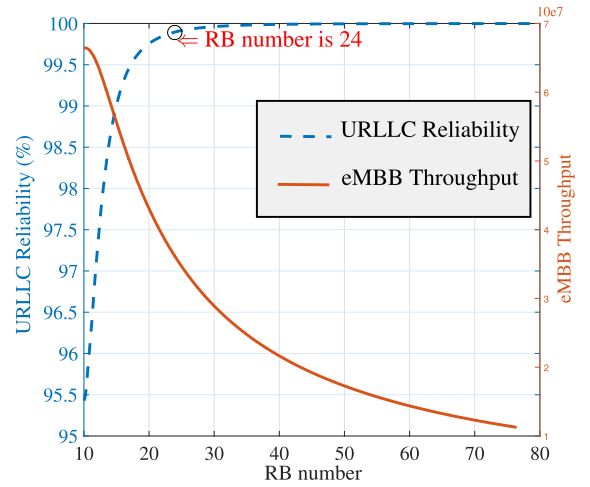

Fig. 5. Simulation results of the LSM-based spectrum partition.
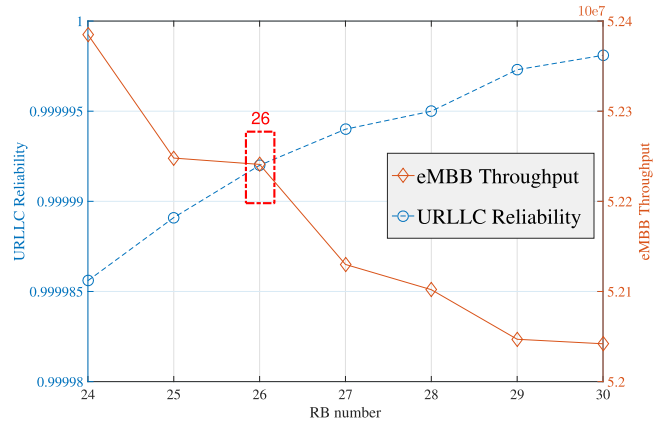

Fig. 6. Simulation results of the proposed DDGSP.

of experiments. The error rate, K_{error}, of the LSM-based approach is 48.2% after 1,000 experiments, which cannot guarantee system performance. This is due to the changes in the communication environment in different experiments.

B. System Performance of the Q-Learning Algorithm

Table III describes the optimization performance of the Q-learning-based spectrum sharing approach for the URLLC and eMBB multiplexing system. The maximum reward is obtained when the state and action are 28 and –3, respectively. Therefore, the optimal N_{RB}^u for the current strategy is $(28 - 3 =)25$. We used the obtained strategy to perform repeated independent experiences to further evaluate the performance of the Q-learning-based spectrum partition. The error rate of the Q-learning-based approach is 16.2% after 1,000 experiences, which significantly outperforms the LSM-based spectrum partition.

C. System Performance of the Proposed DDGSP

Figure 6 illustrates the optimization performance of the proposed DDGSP for multiplexing of URLLC and eMBB transmissions. The abscissa is the number of RBs N_{RB}^u assigned to URLLC UEs. The figure shows that the URLLC reliability requirement, 99.999%, is addressed when the optimal value

TABLE III
RESULTS OF THE Q-LEARNING METHOD

Action / State	−13	−8	−5	−3	−2	−1	+1	+2	+3	+5	+8	+13
1	−100	−100	−100	−100	−100							
...												
28				MAX								
...												
200							−100	−100	−100	−100	−100	−100

Fig. 7. The convergence performance of the proposed DDGSP.

TABLE IV
COMPARISON OF DIFFERENT METHODS

Method	Result	Error Rate	Complexity
DDGSP	26	4.1%	$O(n^2)$
LSM	24	48.2%	$O(n)$
Q-Learning	25	16.2%	$O(n^3)$
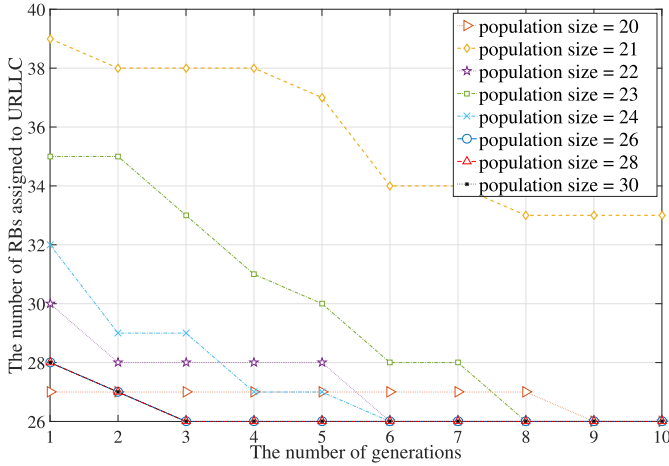
of N_{RB}^u is 26. The corresponding maximum eMBB throughput was also achieved in the same strategy. We conducted 1,000 independent experiments to verify the obtained spectrum partition strategy of the proposed DDGSP. The error rate of the proposed DDGSP is 4.1% based on Eq. (25), which is much lower than the LSM-based spectrum partition approach. Performance fluctuations and error rates are unavoidable since the communication environment changes randomly. However, the proposed DDGSP-based algorithm reaches the best and most stable performance among the state-of-the-art approaches in the dynamic communication environment.

The size of the initial population directly affects the performance and convergence speed of GA-based methods [15]. The larger initial population necessitates more computational resources, but the convergence process is accelerated [44]. This is because the larger population has the higher probability of finding the optimal solution [45]. However, there is no solution to find the optimal population size that can converge quickly and reduce computational costs. Therefore, we conducted several experiments with different initial populations to determine the optimal value of the initial population. Due to the total RBs $N_{RB} = 200$, the initial population size and the running population size are empirically set from 20 to 30 for the test, which is 10%–15% of the total solutions [44], [45], [46]. Figure 7 compares the convergence performance with population values ranging from 20 to 30. All curves in the figure converge to 26 during 10 generations, except the line of experience with a population value of 21. Further, the convergence speeds of different population

sizes are different. Figure 7 shows that the convergence speed increases as the population size increases. The learning process converges in two generations when the population size is greater than 26. This is because the probability that the initial individuals contain the near-optimal solution is very high when the population size is greater than 26 in this experimental scenario. For this study, 26 is the optimal initial population to compensate for the computational complexity and convergence performance.

Table IV compares the proposed DDGSP with the conventional techniques in terms of optimizing the spectrum partition strategy of the URLLC and eMBB multiplexing system. This table shows that the LSM-based approach has the lowest computational complexity, but suffers the highest error rate (48.2%). The Q-learning-based approach has the highest computational complexity with an error rate of 16.2%. The proposed DDGSP achieves the lowest error rate (4.1%) with accepted computational complexity of $O(n^2)$.

In summary, the proposed DDGSP achieves the best performance among the three spectrum partition methods because DDGSP reaches the trade-off between the error rate and the computational complexity of the URLLC and eMBB multiplexing system. The LSM-based method has the lowest computational complexity, but with an extremely high error rate. Compared to the Q-learning-based spectrum sharing, the proposed DDGSP has a lower computational complexity and a lower error rate. Therefore, the proposed DDGSP outperforms LSM- and Q-learning-based methods in the considered data-driven URLLC and eMBB multiplexing system. Furthermore, the convergence speed of the proposed DDGSP increases as the population size increases in the data-driven architecture. Clearly, the larger the population size, the higher the probability of including individuals with a globally optimal solution. As such, the computational complexity of the proposed DDGSP also increases as the population size increases. Hence, finding the optimal population size that achieves fast convergence at acceptable computational complexity is worthy of future research.

VII. CONCLUSION

This study investigates the spectrum partition mechanisms for a URLLC and eMBB multiplexing system. The time-critical URLLC service necessitates extremely low latency (0.25–0.3 msec/packet) while guaranteeing 99.999% reliability. In contrast, eMBB services require the maximization of their throughput for high-speed transmissions. We formulate the objective of spectrum sharing in URLLC and eMBB multiplexing as a multi-objective optimization (MOO) problem and propose a data-driven genetic algorithm-based spectrum partition (DDGSP) to solve it. The proposed DDGSP outperforms conventional techniques, least squares method (LSM), and Q-learning, in terms of optimization performance and computational complexity. Our simulation results demonstrated the effectiveness and efficiency of the proposed DDGSP, which significantly reduced the system error rate from 48.2% to 4.1%. It is worthwhile further investigating a technique to provide resilience for the DDGSP-based spectrum partition against variability in the communication environment to improve the system performance for the coexistence of URLLC and eMBB. Additionally, finding the optimal population size to achieve fast convergence at acceptable computational complexity is also worthy of future research.

REFERENCES

[1] "3GPP release 15." Apr. 26, 2019. [Online]. Available: https://www.3gpp.org/release-15

[2] A. Anand, G. de Veciana, and S. Shakkottai, "Joint scheduling of URLLC and eMBB traffic in 5G wireless networks," *IEEE/ACM Trans. Netw.*, vol. 28, no. 2, pp. 477–490, Apr. 2020.

[3] Y. Huang, Y. T. Hou, and W. Lou, "A deep-learning-based link adaptation design for eMBB/URLLC multiplexing in 5G NR," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2021, pp. 1–10.

[4] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view," *IEEE Access*, vol. 6, pp. 55765–55779, 2018.

[5] M. Alsenwi, N. H. Tran, M. Bennis, S. R. Pandey, A. K. Bairagi, and C. S. Hong, "Intelligent resource slicing for eMBB and URLLC coexistence in 5G and beyond: A deep reinforcement learning based approach," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4585–4600, Jul. 2021.

[6] H. Peng, P.-C. Chen, P.-H. Chen, Y.-S. Yang, C.-C. Hsia, and L.-C.Wang, "6G toward metaverse: Technologies, applications, and challenges," in *Proc. IEEE VTS Asia–Pacific Wireless Commun. Symp. (APWCS)*, Aug. 2022, pp. 6–10.

[7] N. H. Mahmood, R. Abreu, R. Böhnke, M. Schubert, G. Berardinelli, and T. H. Jacobsen, "Uplink grant-free access solutions for URLLC services in 5G new radio," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Oulu, Finland, Aug. 2019, pp. 607–612.

[8] T. Jacobsen et al., "System level analysis of uplink grant-free transmission for URLLC," in *Proc. IEEE Global Commun. Conf. Workshops (GC Wkshps)*, Dec. 2017, pp. 1–6.

[9] N. U. Ginige, K. B. S. Manosha, N. Rajatheva, and M. Latva-aho, "Admission control in 5G networks for the coexistence of eMBB-URLLC users," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–6.

[10] M. Almekhlafi, M. A. Arfaoui, M. Elhattab, C. Assi, and A. Ghrayeb, "Joint resource allocation and phase shift optimization for RIS-aided eMBB/URLLC traffic multiplexing," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 1304–1319, Feb. 2022.

[11] M. Alsenwi, N. H. Tran, M. Bennis, A. K. Bairagi, and C. S. Hong, "eMBB-URLLC resource slicing: A risk-sensitive approach," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 740–743, Apr. 2019.

[12] S.-F. Cheng, L.-C. Wang, C.-H. Hwang, J.-Y. Chen, and L.-Y. Cheng, "On-device cognitive spectrum allocation for coexisting URLLC and eMBB users in 5G systems," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 1, pp. 171–183, Mar. 2021.

[13] R. T. Marler and J. S. Arora, "Survey of multi-objective optimization methods for engineering," *Struct. Multidiscip. Optim.*, vol. 26, no. 6, pp. 369–395, Apr. 2004.

[14] N. Mazyavkina, S. Sviridov, S. Ivanov, and E. Burnaev, "Reinforcement learning for combinatorial optimization: A survey," *Comput. Oper. Res.*, vol. 134, Oct. 2021, Art. no. 105400.

[15] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: Past, present, and future," *Multimed. Tools. Appl.*, vol. 80, no. 5, pp. 8091–8126, Feb. 2021.

[16] G. Grisetti, T. Guadagnino, I. Aloise, M. Colosi, B. D. Corte, and D. Schlegel, "Least squares optimization: From theory to practice," *Robotics*, vol. 9, no. 3, p. 51, Jul. 2020.

[17] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, and M. N. Do, "Fast global image smoothing based on weighted least squares," *IEEE Trans. Image Process.*, vol. 23, pp. 5638–5653, 2014.

[18] S. H. da Mata and P. R. Guardieiro, "A genetic algorithm based approach for resource allocation in LTE uplink," in *Proc. Int. Telecommun. Symp. (ITS)*, Sao Paulo, Brazil, Aug. 2014, pp. 1–5.

[19] E. Carrizosa, C. Molero-Río, and D. R. Morales, "Mathematical optimization in classification and regression trees," *Top*, vol. 29, no. 1, pp. 5–33, Apr. 2021.

[20] H. Peng, L.-C. Wang, G. Y. Li, and A.-H. Tsai, "Long-lasting UAV-aided RIS communications based on SWIPT," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Austin, TX, USA, Apr. 2022, pp. 1–6.

[21] T. Yang, Y. Hu, X. Yuan, and R. Mathar, "Genetic algorithm based UAV trajectory design in wireless power transfer systems," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2019, pp. 1–6.

[22] C. Guerrero, I. Lera, and C. Juiz, "Genetic algorithm for multi-objective optimization of container allocation in cloud architecture," *J. Grid Comput.*, vol. 16, no. 1, pp. 113–135, Nov. 2018.

[23] C. Lu, Z. Lin, and S. Yan, "Smoothed low rank and sparse matrix recovery by iteratively reweighted least squares minimization," *IEEE Trans. Image Process.*, vol. 24, pp. 646–654, 2015.

[24] A. M. Geoffrion, "Generalized benders decomposition," *J. Optim. Theory Appl.*, vol. 10, no. 4, pp. 237–260, Oct. 1972.

[25] L. Su, L. Tang, and I. E. Grossmann, "Computational strategies for improved MINLP algorithms," *Comput. Chem. Eng.*, vol. 75, no. 1, pp. 40–48, Apr. 2015.

[26] M. Lee, N. Ma, G. Yu, and H. Dai, "Accelerating generalized benders decomposition for wireless resource allocation," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1233–1247, Feb. 2021.

[27] H. Peng, C.-Y. Ho, Y.-T. Lin, and L.-C. Wang, "Energy-efficient symbiotic radio using generalized benders decomposition," in *Proc. IEEE 96th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2022, pp. 1–5.

[28] K. Gai and M. Qiu, "Optimal resource allocation using reinforcement learning for IoT content-centric services," *Appl. Soft Comput.*, vol. 70, no. 1, pp. 12–21, Sep. 2018.

[29] L. Xiao et al., "Reinforcement learning-based downlink interference control for ultra-dense small cells," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 423–434, Jan. 2020.

[30] L. Liang, H. Ye, and G. Y. Li, "Spectrum sharing in vehicular networks based on multi-agent reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2282–2292, Oct. 2019.

[31] J. Cui, Y. Liu, and A. Nallanathan, "Multi-agent reinforcement learning-based resource allocation for UAV networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 729–743, Feb. 2020.

[32] M. K. Singh, S. I. Amin, and A. Choudhary, "Genetic algorithm based sink mobility for energy efficient data routing in wireless sensor networks," *Int. J. Electron. Commun.*, vol. 131, Mar. 2021, Art. no. 153605.

[33] F.-H. Tseng, X. Wang, L.-D. Chou, H.-C. Chao, and V. C. Leung, "Dynamic resource prediction and allocation for cloud data center using the multiobjective genetic algorithm," *IEEE Syst. J.*, vol. 12, no. 2, pp. 1688–1699, Jun. 2018.

[34] C. Ramirez-Atencia, G. Bello-Orgaz, M. D. R. Moreno, and D. Camacho, "Solving complex multi-UAV mission planning problems using multi-objective genetic algorithms," *Soft Comput.*, vol. 21, no. 17, pp. 4883–4900, Sep. 2017.

[35] A. Dieuleveut, N. Flammarion, and F. Bach, "Harder, better, faster, stronger convergence rates for least-squares regression," *J. Mach. Learn. Res.*, vol. 18, no. 101, pp. 1–51, Oct. 2017.

[36] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai, "A survey on low latency towards 5G: RAN, core network and caching solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3098–3130, 4th Quart. 2018.

[37] S. H. Brown, "Multiple linear regression analysis: A matrix approach with MATLAB," *Alabama J. Math.*, vol. 34, no. 1, pp. 1–3, 2009.

[38] A. A. Chaves and L. H. N. Lorena, "An adaptive and near parameter-free BRKGA using Q-learning method," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Kraków, Poland, Jun. 2021, pp. 2331–2338.

[39] S. Mirjalili, "Genetic algorithm," in *Evolutionary Algorithms and Neural Networks*. Cham, Switzerland: Springer, 2019, pp. 43–55.

[40] G. Rudolph, "Convergence analysis of canonical genetic algorithms," *IEEE Trans. Neural Netw.*, vol. 5, no. 1, pp. 96–101, Jan. 1994.

[41] M. Srinivas and L. Patnaik, "Adaptive probabilities of crossover and mutation in genetic algorithms," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 24, no. 4, pp. 656–667, Apr. 1994.

[42] R. F. Hartl and R. Belew, "A global convergence proof for a class of genetic algorithms," Inst. Manage., Univ. Vienna, Vienna, Austria, Rep., Jan. 1990. [Online]. Available: https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.330.1662

[43] J. Medbo et al., "Radio propagation modeling for 5G mobile and wireless communications," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 144–151, Jun. 2016.

[44] T. Chen, K. Tang, G. Chen, and X. Yao, "A large population size can be unhelpful in evolutionary algorithms," *Theor. Comput. Sci.*, vol. 436, no. 1, pp. 54–70, Jun. 2012.

[45] P. A. Diaz-Gomez and D. F. Hougen, "Initial population for genetic algorithms: A metric approach," in *Proc. Int. Conf. Genetic Evol. Methods*, Las Vegas, NV, USA, Jun. 2007, pp. 43–49.

[46] S. E. De León-Aldaco, H. Calleja, and J. Aguayo Alquicira, "Metaheuristic optimization methods applied to power converters: A review," *IEEE Trans. Power Electron.*, vol. 30, no. 12, pp. 6791–6803, Dec. 2015.

Li-Chun Wang (Fellow, IEEE) received the Ph.D. degree from the Georgia Institute of Technology, Atlanta, in 1996. From 1996 to 2000, he was a Senior Technical Staff Member with AT&T Laboratories. Since August 2000, he has been with the College of Electrical and Computer Engineering, National Yang Ming Chiao Tung University, Taiwan. He is currently the Chair Professor and is jointly appointed by the College of Computer Science and College of AI. He holds 26 U.S. patents and has published over 300 journal and conference papers, and co-edited the book, "*Key Technologies for 5G Wireless Systems*," (Cambridge University Press 2017). He was recognized as Top 2% Scientists Worldwide in a study from Stanford University. His recent research interests are in the areas of cross-layer optimization for wireless systems, AI-enabled radio resource management for heterogeneous mobile networks, and big data analysis for industrial Internet of Things.

Dr. Wang won the two Distinguished Research Awards from National Science and Technology Council in 2012 and 2017, the IEEE Communications Society Asia–Pacific Board Best Award in 2015, the Y. Z. Hsu Scientific Paper Award in 2013, and the IEEE Jack Neubauer Best Paper Award in 1997. He was elected to the IEEE Fellow in 2011 for his contributions to cellular architecture and radio resource management in wireless networks.

Haoran Peng (Member, IEEE) received the B.Eng. degree in software engineering from the University of Electronic Science and Technology of China in 2015, and the Ph.D. degree in electrical and computer engineering from National Yang Ming Chiao Tung University in 2022. From 2015 to 2018, he was a Full-Time Software Engineer. From June 2021 to August 2021, he was a Visiting student Research Collaborator with the Global Cybersecurity Institute, Golisano College of Computing and Information Sciences, Rochester Institute of Technology. His main research interests are in the field of optimization and machine learning for wireless communications. He was awarded the IEEE VTS Student Travel Grant in Vehicular Technology Conference Fall (VTC2022-Fall). He served or is serving as a Reviewer for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, and IEEE WIRELESS COMMUNICATIONS and the 2022 IEEE International Conference on Communications Workshop. He served as a TPC Member for the 2022 IEEE 96th VTC2022-Fall.

Zhuofu Jian received the B.S. degree in electrical engineering and automation from Southwest Jiaotong University in 2018, and the M.S. degree in electrical and computer engineering from National Yang Ming Chiao Tung University in 2020.