Cell-Free XL-MIMO Meets Multi-Agent Reinforcement Learning: Architectures, Challenges, and Future Directions

Zhilong Liu, Jiayi Zhang, Ziheng Liu, Hongyang Du, Zhe Wang, Dusit Niyato, Mohsen Guizani, and Bo Ai

Abstract

Cell-free massive multiple-input multiple-output (mMIMO) and extremely large-scale MIMO (XL-MIMO) are regarded as promising innovations for the forthcoming generation of wireless communication systems. Their significant advantages in augmenting the number of degrees of freedom have garnered considerable interest. In this article, we first review the essential opportunities and challenges induced by XL-MIMO systems. We then propose the enhanced paradigm of cell-free XL-MIMO, which incorporates multi-agent reinforcement learning (MARL) to provide a distributed strategy for tackling the problems of high-dimensional signal processing and costly energy consumption. Based on the unique near-field characteristics in XL-MIMO systems, we propose two categories of the low-complexity algorithm design, that is, antenna selection and power control, to adapt to different cell-free XL-MIMO scenarios and meet the increasing data rate requirements. For inspiration, several critical future research directions pertaining to green cell-free XL-MIMO systems are presented.

Introduction

The next generation of wireless communication systems, that is, the sixth-generation (6G), is expected to deliver unprecedented levels of performance, particularly in digital twins, integrated sensing and communication, and extended reality scenarios. The commercialization of massive multiple-input multiple-output (mMIMO) technology has played a significant role in the development of wireless networks. However, conventional MIMO techniques face limitations in meeting the complex requirements of 6G use cases. In light of this challenge, emerging technologies such as cell-free mMIMO and extremely large-scale MIMO (XL-MIMO) are being proposed to overcome the capacity constraints of conventional MIMO. These advanced technologies are critical to fulfill the massive connectivity and all-round multidimensional access to the space, air, ground, and sea, which will enable the Internet of Everything.

As a high-profile technology, the novel cell-free mMIMO holds great promise in meeting the growing demand for increasing network throughput and low-latency transmission. By deploying a large number of geographically distributed access points (APs) connected to a central processing unit (CPU), cell-free mMIMO can effectively address the inter-cell interference that exists in the intrinsic implementation of "cell-centric" network [1, 2]. Similarly, the promising XL-MIMO technology inherits the prior cellular network with the world-shaking change of base stations (BSs) to adapt the communication variations from far-field to near-field since the deployment of massive antennas [3, 4]. Moreover, the XL-MIMO can also provide a much stronger beamforming gain as well as harvest abundant degrees of freedom (DoFs) to compensate for the severe path loss in the millimeter-wave and terahertz band communications.

In cell-free mMIMO systems, the data processing procedures can be performed locally using the large-scale fading decoding (LSFD) method [1]. This approach is highly effective in relieving the computational load on the CPUs. From the perspective of electromagnetic (EM) fields, the addition of antennas in XL-MIMO is a superficial phenomenon. In fact, the significant changes occur in the analysis methods, where the spherical wavefront-based analysis framework replaces the planar wavefront-based one [5, 6]. In parallel, the interdisciplinary Electromagnetic Information Theory (EIT) is undergoing a global roll-out research. By integrating cell-free mMIMO and XL-MIMO, namely cell-free XL-MIMO, this prototype will be a forward-looking architecture that can accommodate full scenarios and hot-spot venues to extend the range of near-field communication (NFC), as shown in Fig. 1.

To reduce the overall system computational complexity and energy consumption, low-complexity baseband signal processing algorithms are in demand. Multi-agent reinforcement learning (MARL) has been widely used for decision-making in large-scale network scenarios [9–11], for example, unmanned aerial vehicles (UAVs), swarm intelligence, and traffic scheduling. We hasten to say that the algorithms are proficient to improve spectral efficiency (SE), enhance interference management in XL-MIMO systems, and increase coverage, improve user fairness, and achieve distributed resource allocation in cell-free mMIMO systems. In multi-agent systems, interactions between intel-

Zhilong Liu, Jiayi Zhang (corresponding author), Ziheng Liu, Zhe Wang, and Bo Ai are with Beijing Jiaotong University, China;
Hongyang Du and Dusit Niyato are with the Nanyang Technological University, Singapore;
Mohsen Guizani is with Mohamed Bin Zayed University of Artificial Intelligence, UAE.

Digital Object Identifier: 10.1109/MWC.007.2300176

1536-1284/24/$25.00 © 2024 IEEE

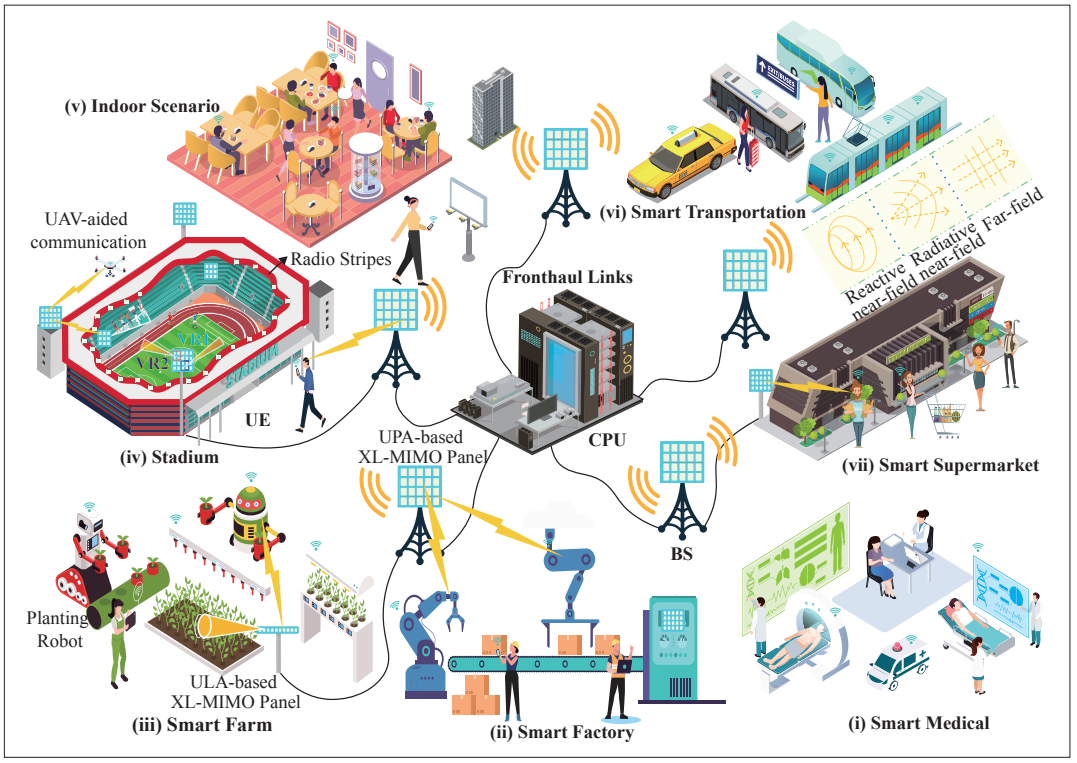

FIGURE 1. System architecture and application scenarios of cell-free XL-MIMO systems around NFC. The BSs are equipped with XL-MIMO panels, and the user equipments are equipped with different numbers of antennas, from single to hundreds. BSs and users are distributed in the service area. The BSs are connected via fronthaul links to a CPU with a high computational power. The communication regions are divided into reactive near-field, radiative near-field, and far-field [4]. In XL-MIMO systems, the communication focuses on the radiative near-field. The boundary between radiative near-field and far-field is decided by the Rayleigh distance [7, 8], and visibility regions (VRs) induced by the non-stationary channel are illustrated since sheltering from different buildings and obstacles. In 2019, Ericsson proposed radio stripes, the prototype of cell-free mMIMO systems, which is an ideal deployment solution for outdoor and indoor areas such as shopping malls, stadiums, smart factories, and other scenarios [1].

ligent agents and environments drive the achievement of goals. In particular, RL algorithms become almost indispensable tools for exploring complex dynamic scenarios, which can effectively reduce the overall power consumption. Notable advances include the development of low-complexity RL-based power control algorithms that can be scaled to XL-MIMO systems and the exploration of hybrid analog-digital precoding schemes that can considerably enhance the energy efficiency (EE).

Motivated by the aforementioned works, we investigate the cell-free XL-MIMO systems with MARL techniques. The main contributions of this work are summarized as follows:

- We introduce new NFC characteristics, basic system scheme, and application scenarios of cell-free XL-MIMO systems. More important, we comprehensively introduce the crucial challenges of power consumption, computational complexity, and user mobility.
- We investigate three technical frameworks, for example, fully decentralized, fully centralized, and centralized training and decentralized execution (CTDE), algorithm categories, and applications of MARL methods in existing literature, as shown in Fig. 2.
- To strive for the undiscovered performance, we focus on two critical methods, that is, antenna selection (AS) and power control, to reduce the power consumption and improve SE with MARL methods. Numerical results are given to illustrate the ability to improve SE and EE.

Finally, the article concludes by discussing open problems toward uncovering the potential of cell-free XL-MIMO systems.

Opportunities and Challenges of XL-MIMO Communication Systems

In this section, we focus on the newly discovered EM wave transmission characteristics in the NFC domain. Through the unique near-field properties uncovered by the XL-MIMO, such as the spherical wave model (SWM), spatial non-stationary effect, and effective DoF (EDoF), they can be well designed to enhance the communication performance. In addition, the power consumption, computational complexity, and mobility problems present us with new challenges.

New Opportunities

Spherical Wave Model: SWM is a mathematical tool used to describe the behavior of EM waves in three-dimensional space [4]. An accurate SWM is essential for excavating the capacity upper bound of XL-MIMO systems as it facilitates the efficient processing and manipulation of EM waves, thereby improving signal quality and enhancing data throughput. In previous research, channel models mainly focused on the basic assumption of Rayleigh or Rician fading channels. However, once the communication distance is shorter than the Rayleigh distance, for example, for an XL-MIMO panel with a diagonal of 10 m at 3 GHz, the boundary is up

In particular, RL algorithms become almost indispensable tools for exploring complex dynamic scenarios, which can effectively reduce the overall power consumption.

Paper	Training	Execution	Research Topic	Scenarios	Algorithm Feature
[7]	Decentralized	Decentralized	Minimize energy consumption	UAV	• A fully-decentralized MARL mechanism was proposed, jointly optimizing resource management and UAV mobility
[8]	Centralized	Decentralized	Power allocation	Cell-free mmWave MIMO	• The core idea of MADDPG is centralized training with decentralized execution for saving the computational overhead
[9]	Centralized	Decentralized	Pilot assignment	Cell-free mMIMO	• A two-level hierarchical dynamic multi-agent structure was proposed to optimize the system pilot assignment with better average SE performance and convergence behavior
[10]	Centralized	Decentralized	Antenna selection	UC mMIMO	• Select antennas for users with A3C algorithm in the selection stage based on user locations • Adjust selected antennas with A3C algorithm in the adjustment stage according to the nearest distance criterion

※ The ***centralized training and centralized execution (CTCE) paradigm*** is not widely used in MIMO systems since the difficulty in coordination and heavy communication overhead.

Technical Frameworks ⇓ ⇑ **Algorithm Categories**

Methods	Core	Classical Algorithms	Benefits	Limitations
Value Decomposition	The expected long-term cumulative reward is decomposed into a set of value functions that capture the contribution of each agent to the joint action value	Independent Q-Learning; Deep Recurrent Q-Network [9]; Q-function Transfer	• Centralized Training and Decentralized Execution (CTDE) • Improves the scalability	• Overlook the dynamic non-stationary environment, causing learned value functions to be outdated or inaccurate • Assumption of independent agents and undisturbed actions is no longer held in real-world scenarios, e.g., XL-MIMO systems
Actor-Critic	***Actor*** is responsible for selecting actions based on the current state of environment; ***Critic*** evaluates the quality of actions taken by ***Actor***	Asynchronous Advantage Actor-Critic (A3C) [10]; Multi-agent Deep Deterministic Policy Gradient (MADDPG) [7] [8]	• Adapt non-stationary environment • Better coordination between agents during execution • ***Critic*** provides feedback to ***Actor*** based on the expected reward over time rather than immediate reward	• Sensitive to choose hyperparameters • Easy to fall into a suboptimal state
Experience Replay	Agents store experiences from their past interactions with environment and train agents' policies by randomly sampling from the replay buffer.	Deep Policy Inference Q-Network; Parallel Subspace Trust Region Policy Optimization	• Improved stability • Reduce the total interactions with the environment	• High memory requirements • Low optimal policy search efficiency

FIGURE 2. Summary of mainstream MARL technical frameworks, including Decentralized Training and Decentralized Execution (DTDE) and Centralized Training and Decentralized Execution (CTDE), algorithm categories, and applications in the existing literatures [7-10].

to 2 km, the communication domain focuses on the near-field rather than the far-field. Therefore, the existing channel models used to analyze the conventional MIMO systems are not suitable for XL-MIMO systems as the NFC dominates [7].

Furthermore, the integration of massive antennas can make it difficult to obtain accurate channel state information (CSI) in XL-MIMO systems. Regarding the near-field effects, the channel should be properly modelled to ensure accuracy in the near-field under the spherical wavefront assumption. Based on EIT, the exploration of SWM has revolutionized for enabling high-speed data transmission, broadening coverage and improving user experience [12].

Spatial Non-Stationary Effect: In XL-MIMO systems, the spatial non-stationary effect arises because only partial antennas in BSs can receive spherical EM waves from specific UEs propagated by different scatters. This can lead to fluctuations in the channel gain, phase, and delay over time [12]. Similarly, each UE can only observe a subset of the antenna array, which is called the visibility region (VR), as shown in Fig. 1. As a result, the channel capacity and quality vary significantly, and the traditional channel estimation (CE) and equalization techniques may not be effective at mitigating the effects of non-stationary channels. Thus, effectively exploiting this peculiarity would be a tutorial for green communication systems, as a cost-effective way to reduce the computational complexity for crowded scenarios.

Effective Degree of Freedom: An important parameter characterizing the performance of XL-MIMO systems is the EDoF, referring to the number of significant electromagnetic modes. It represents the potential capacity of a MIMO system to spatially multiplex multiple data streams. The EDoF considers the effects of various factors, for example, channel correlations, signal-to-noise (SNR) ratios, and interference. However, increasing the number of antennas may not always improve the EDoF, as it may increase channel correlation, interference between different data streams, and energy consumption, all of which degrade the system performance [4]. Therefore, to achieve a high EDoF in practical XL-MIMO systems, appropriate antenna numbers and configurations should be chosen based on specific wireless channels and system requirements, for example, the maximum EDoF is around 1600 for a 2.25 m × 2.25 m panel size at 0.1 m wavelength to satisfy the hot-spot scenarios.

New Challenges

Power Consumption: Although the XL-MIMO technology can effectively improve the speed and reliability of the signal transmission, implementing enormous sub-processing units in the XL-MIMO transceiver can result in a high hardware cost and power consumption. Reducing the power consumption in the XL-MIMO is essential to ensure the energy-efficient, cost-effective, and sustainable while maintaining its high performance capabilities [13]. Currently, existing methods capable of solving the above challenges mainly focus on traditional approaches, for example, heuristic fractional power control laws and deep learning-based power control methods. It is necessary

to balance the system performance and power consumption factors considering the characteristics in practical NFC scenarios.

Computation Complexity: Complex computations significantly increase latency in wireless communication systems with limited computing power. Distributed signal processing methods are the trends to deal with the problem of high-dimensional computation. By shifting the processing responsibilities to the local processing unit (LPU), XL-MIMO can reduce latency and lessen the need for specialized processors or additional memory. In general, the antenna selection technique involves selecting the appropriate subset of antennas from the antenna array, which can help minimize the computational burden of signal processing and reduce the power consumption, as well as improve SNR [14].

User Mobility: In XL-MIMO systems, user mobility leads to time-varying channel conditions. As users move within the coverage area, channel characteristics, such as path loss, fading, and interference, change dynamically. Moreover, the movement of UEs can switch the propagation mode between the near-field and far-field, and thus, the channel estimation and codebook design in the hybrid-field should be re-examined. Additionally, user mobility necessitates dynamic adaptation of transmission strategies, handover management, user scheduling, and mobility prediction. By considering these factors, XL-MIMO systems can be effectively optimized to maintain reliable connectivity.

With these aspects, the XL-MIMO can be seen as an extended version of the conventional MIMO, which involves more than just increasing the number of antennas deployed from 64 antennas to thousands of antennas. From an environmentally-friendly perspective, by optimizing the transceiver power and adopting appropriate distributed processing algorithms, XL-MIMO systems can achieve a superior performance while minimizing the energy consumption and reducing the carbon footprint of wireless communication systems.

System Architecture of Multi-Agent Cell-Free XL-MIMO

With the increasing computational dimension and time-varying configurations and parameters, the traditional optimization methods do not work well with XL-MIMO systems. We have to seek a better solution to resolve it. In what follows, by integrating the advantages of the cell-free mMIMO and MARL methods, we propose a novel cell-free XL-MIMO system with the MARL optimization scheme to further improve the performance of cell-free XL-MIMO systems.

Multi-Agent Reinforcement Learning

MARL, a subfield of artificial intelligence, has been widely used in real-world scenarios focusing on the interaction with the environment and multiple agents. Extending from a single-agent domain to a multi-agent environment, this method arises from the need to develop intelligent systems that can interact with other intelligent agents in complex and dynamic environments. This approach integrates the methodologies of RL, game theory, and multi-agent systems, empowering agents to learn effective interaction strategies with both their environment and other agents. This learning framework is designed to optimize the achievement of specific objectives through adaptive and strategic behaviors [9]. The main idea behind MARL is to model the behavior of a group of agents that can cooperate, compete, and even negotiate. More intuitively, different training schemes, for example, fully decentralized, fully centralized, and centralized training and decentralized execution (CTDE), are considered as promising paradigms to adapt to different environments. Furthermore, the existing MARL algorithms can be divided into three categories.

Value Decomposition: Value decomposition (VD) based algorithms are usually based on value functions, that is, *Deep Recurrent Q-Network*, to decompose value functions into local value functions for agents, so as to deal with the interaction between multiple agents. This type of algorithm usually combines the actions and states of multiple agents as global states, and then uses single-agent algorithms such as *Q-learning* to learn local value functions.

Actor-Critic: Actor-Critic (AC) based algorithms combine *value functions* and *strategy functions* with two networks, *Actor* and *Critic*, where the *Actor network* learns the strategy and the *Critic network* evaluates the value of the action and updates the actor's policy. Examples of AC-based methods include Asynchronous Advantage Actor-Critic (A3C) [15] and Multi-agent Deep Deterministic Policy Gradient (MADDPG) [11]. To illustrate, MADDPG follows the CTDE paradigm, where the additional information has been gathered in *Critic networks* to facilitate the training process while *Actor networks* take actions based on their own local observations.

Experience Replay: Experience replay (ER) based algorithms typically use experience replay caches to store past experiences and randomly sample them for training. This approach speeds up training by making more efficient use of data, and is usually applied experiential playback to single-agent algorithms, that is, *Deep Policy Inference Q-Network*. However, in multi-agent scenarios, the implementation of experience reply is more complicated, and the interaction between the multi-agent needs to be considered.

As shown in Fig. 2, these three categories have been widely used in communication scenarios for resource allocation. While the centralized learning method is advantageous for global assessment with unified decision-making, distributed learning using the MARL methods is more feasible for local processing, which is beneficial for real-time processing.

In multi-agent environments, agents' actions affect the state of the environment, and each agent must learn a policy that not only maximizes its rewards but also takes into account the actions of other agents. The MADDPG algorithm extends the popular DDPG algorithm by introducing a centralized *Critic network* that can observe the joint actions of all agents and provide feedback to each agent's policy network, as shown in Fig. 3. In turn, the *Actor network* learns to optimize their policies, taking into account the feedback from the Critic network and the observations of other agents.

In the signal processing phase of the XL-MIMO, high-dimensional matrix operations, and time-sensitive actions are critical to achieve the optimal system performance. Therefore, traditional data processing schemes fail to meet the requirements of cell-free XL-MIMO systems. As such, we have to

MARL, a subfield of artificial intelligence, has been widely used in real-world scenarios focusing on the interaction with the environment and multiple agents. Extending from a single-agent domain to a multi-agent environment, this method arises from the need to develop intelligent systems that can interact with other intelligent agents in complex and dynamic environments.

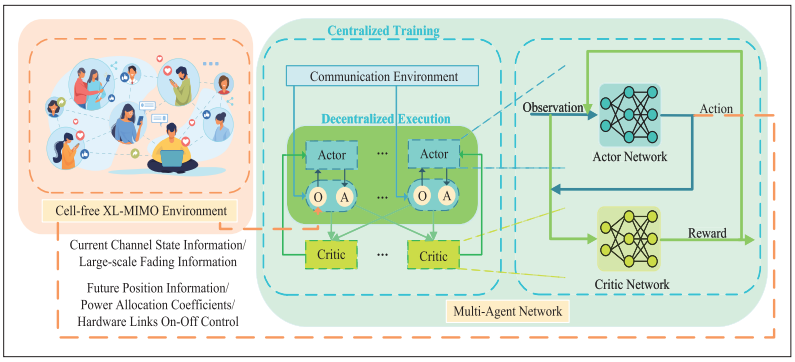

FIGURE 3. The basic scheme of CTDE-based MADDPG algorithm interacting with cell-free XL-MIMO systems.

concentrate on local processing or distributed signal processing to reduce the load on the fronthaul links. For example, we can apply the MARL methods to approach the SE or EE maximum by defining a Markov decision process that includes states, actions, and rewards [9]. The agents interact with the environment in the current state and move to the next state. Then, the next state is sent to the agent, which decides to take an action against the environment. The environment then sends the next state and reward to the agent.

System Architecture of Multi-Agent Cell-Free XL-MIMO

In conventional massive MIMO systems, centralized processing methods lack the ability to parallelize operations. Furthermore, scaling up the dimensions of the array proves to be an arduous feat owing to the significant amount of interconnections and overwhelming burden placed on the central node. Therefore, various decentralized techniques have been proposed. Among them is the cell-free architecture, which aims at eliminating cell boundaries and focusing on user-centric communication [1], providing more flexible transmission/reception of UEs. To adapt to the requirements of distributed architectures, we propose a modified embodiment of distributed XL-MIMO that exploits the advantages of cell-free mMIMO systems while considering multi-agent systems simultaneously.

As shown in Fig. 1, a distributed-processing XL-MIMO system architecture drawing on the merits of cell-free mMIMO is illustrated. The so-called LSFD method can be used to detect the signals using maximum ratio combining or minimum mean squared error combining [1, 13]. For each BS equipped with XL-MIMO panels, it completes the signal processing as well as the channel estimation with all the CSI. All processed signals are then transmitted to the CPU via fronthaul links. In cell-free XL-MIMO systems, there are multiple antennas at the transmitter and receiver sides, and a large number of UEs communicating simultaneously. The communication and resource allocation between these antennas and users can be optimized using MARL, a technique that allows agents to learn how to behave in an environment by interacting with it and receiving feedback in the form of rewards.

Using MARL, agents, that is, UEs, BSs, and even antennas, can learn to allocate physical layer resources and optimize the transmission strategy further. They interact with the system environment with their CSI and location for acquiring the future decision until the SE or EE maximum is reached. Besides, the MARL-based approach can adapt to dynamic changes in the environment, such as UE mobility and time-varying channels.

Directions and Solutions of Multi-Agent Cell-Free XL-MIMO Systems

Multi-antenna technology has been widely recognized as an effective means of improving SE with diversity gain and multiplexing gain. However, to achieve more performance gains, the computational complexity of cell-free XL-MIMO systems increases rapidly with the number of antennas and grievous interference causes signal quality degradation.

Having introduced the new opportunities, in this section, we provide a new look to solve the urgent challenges with MARL methods, for example, AS and power control.

Challenge 1: Antenna Selection

MARL-Empowering Antenna Selection: In cell-free XL-MIMO systems, it is necessary to explore effective AS techniques to reduce the number of antennas used in work patterns, enhance performance, and minimize complexity, especially in energy-constrained environments [14]. Not all antennas serve uplink or downlink UEs simultaneously, making it possible to reduce the number of radio-frequency (RF) links and signal processing units to lower hardware cost and power consumption. As the number of antennas tends to be enormous, the circuit cost and computational complexity of conventional methods based on fully-digital receive arrays will increase dramatically.

AS provides a low hardware-complexity mentality for exploiting the spatial-diversity benefits of a multiple antenna technology with solely partial antennas activated to serve different UEs and can be considered at both transmitters and receivers in cell-free XL-MIMO systems. The basic idea of AS is to choose the optimal subset of antennas from the available antennas in the whole antenna array, based on some selection criteria [14], as shown in Fig. 4. In a cell-free XL-MIMO, AS can be achieved either statically or dynamically. In static AS, a fixed set of antennas is selected that remains unchanged during transmission, while in dynamic AS, the optimal set of antennas is determined based on the channel conditions at each transmission.

In Fig. 5, we draw the boxplot of the sum SE and average EE under different AS strategies. Each UE antenna is regarded as an individual agent to select different BS antenna for achieving the maximum SE. For a fair comparison, we assume the case without AS as a benchmark. The traditional optimal channel selection based on LSF coefficients decreases the system performance by sacrificing DoFs. However, with the introduction of "multi-agent," each antenna can dynamically adjust the selected antennas. It is noteworthy that the MADDPG algorithm can effectively improve the SE of poor quality UEs and nearly achieve a 26 percent EE improvement compared with the benchmark.

Future Research Directions:

Hardware Design: To overcome the computing bottlenecks, one promising solution is to partition the uniform planar array (UPA) or uniform linear array (ULA)-based XL-MIMO into *subarrays-disjoint units* with partial-connected structure and individual processing units. Instead of connect-

ing all the antennas, only a subset of antennas is interconnected, allowing antennas to be connected in a flexible and scalable manner.

Subarray Selection: Apart from AS, subarray selection is worth investigating with fixed or adjustable format depending on whether they correspond to separate hardware entities or software-defined logical connections between different antenna elements, as shown in Fig. 4. The use of subarrays enables more efficient and distributed processing, enabling the system to handle larger and more complex data sets without compromising on performance and accuracy.

Non-Stationary Perspective: One approach to achieving AS in non-stationary channels is to use multiple antennas in combination with channel estimation and equalization techniques, such as space-time coding and beamforming. These techniques can help mitigate the effects of non-stationary channels by using multiple antennas to create a more robust signal and adapting the transmit and receive strategies to the changing channel conditions.

Challenge 2: Power Control Design

Existing Power Control Method: Apart from the AS, designing an effective power allocation algorithm is another open challenge for reducing power consumption in cell-free XL-MIMO systems. With limited communication resources, the dynamic power allocation is worth optimizing based on the real-time channel information. The existing power control methods solving the inter-user interference are focused on the following optimization objectives: max-min, max-product, and max-sum. Traditional power control methods, such as linear optimization techniques, have limitations in large-scale MIMO systems due to the increased complexity and static configuration. Though the non-convex problem can be easily solved using supervised learning-based methods or centralized mechanisms, the prior optimal output data is challenging to obtain in large-scale networks.

With the benefits of massive antennas, the cell-free XL-MIMO poses new challenges for power optimization. Due to near-field propagation and spherical wavefronts, signal strengths vary across the extremely large array in XL-MIMO systems. Furthermore, some antennas may contribute minimally to overall system performance, a result of non-stationarities and visibility regions (VRs). These lead to the activation of power-intensive RF links for these antennas becoming burdensome and significantly reducing the total EE of systems. In this case, existing algorithms are not always able to harvest the global optimal solution, especially when dealing with high-dimensional matrix operations. To overcome these limitations, MARL algorithms are promising to deal with power control in cell-free XL-MIMO systems.

Proposed MARL-Based Power Control Method: RL algorithms enable real-time optimization of power control decisions based on the current state of the system, including channel conditions and signal quality. The basic idea behind using MARL algorithms for power control in large-scale MIMO systems is to model each BS or antenna as an individual agent and to optimize the joint behavior of all agents using RL techniques. This allows for a more flexible and data-driven power control solution compared to traditional methods.

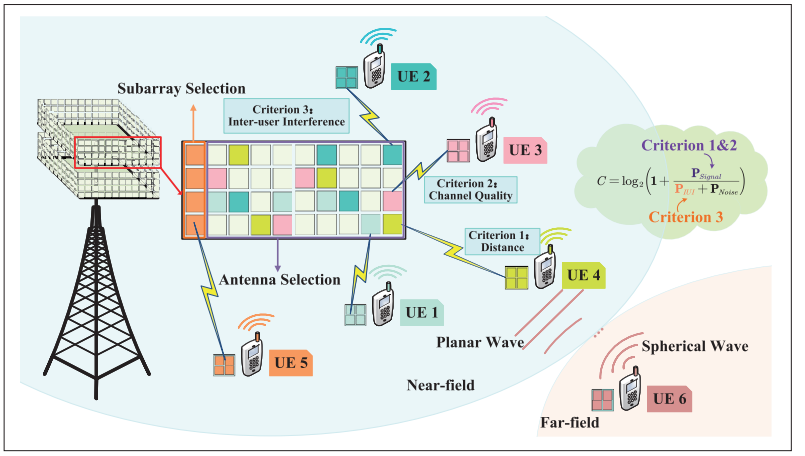

FIGURE 4. The antenna selection of the BS that is equipped with an XL-MIMO panel. A BS simultaneously serves four UEs, and different antennas serve different UEs without reuse. Note that different AS strategies depend on various criteria, for example, maximizing the received signal power (criterion 1 and criterion 2), minimizing the inter-user interference (criterion 3) and so on.

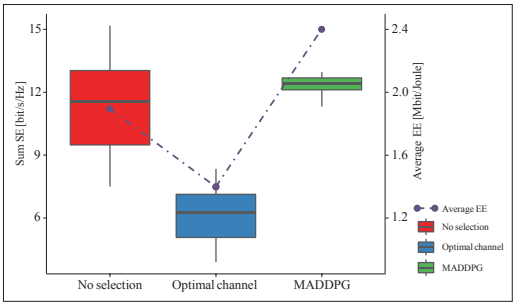

FIGURE 5. Boxplot of sum uplink SE and average EE of a cell-free XL-MIMO system under three circumstances: no selection (with all antennas activated), optimal channel selection based on the large-scale fading coefficients, and MADDPG-based selection. In the simulation, we consider the LSFD architecture [13] with single BS and multiple UEs. All BS and UEs are equipped with XL-MIMO panels. The BS with N_r = 625 antennas serves K = 6 UEs with N_s = 9 antennas simultaneously within a square of size 100 m × 100 m. Additionally, the data transmission power p = 200 mW.

To achieve power control in large-scale MIMO systems using MARL algorithms, the following steps can be taken.

Select Individual Agent: Each antenna, BS, or UE can be modeled as an independent agent, with its unique state, action, and reward, depending on the uplink or downlink transmission. The state of the agent should represent the current channel conditions and interference, while the action should represent the transmit power of the antenna.

Define Reward Function: The reward function should reflect the performance objective of the power control algorithm, such as maximizing SE or minimizing interference.

Train MARL Algorithm: MARL algorithms should be trained using the defined reward function and the modelled agents. The training process involves multiple iterations of the agents taking actions, observing the environment, and updating their policies based on the reward received.

Implement Power Control Algorithm: Once the training process is complete, the power control algorithm can be implemented in large-scale MIMO systems. The algorithm will use the learned

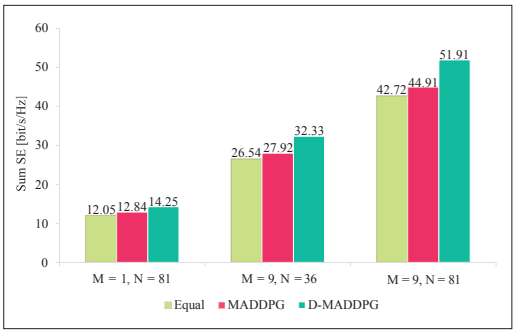

FIGURE 6. Sum uplink SE of cell-free XL-MIMO systems with different power control algorithms: the equal power method, MADDPG strategy, and D-MADDPG under different BS number M and XL-MIMO antennas number N. In the D-MADDPG architecture, we model the power control problem as the MARL framework with two layers. In the first layer, each agent corresponds to a BS in the cell-free XL-MIMO system, and the objective is to optimize its transmit power level to maximize system performance while taking into account interference from other agents. Then, with the constraint of BS power obtained in the first layer, the second layer is responsible for the allocation of each antenna of XL-MIMO panels. The simulation parameters are the same with Fig. 5.

policies of the agents to determine the optimal transmit power of each antenna.

Evaluate Performance: The performance of power control algorithms should be evaluated in a realistic simulation or test environment to ensure their effectiveness in cell-free XL-MIMO systems.

The application of MARL algorithms for power control in large-scale MIMO systems is a growing area of research, and recent studies have demonstrated the potential of these algorithms for improving the performance and efficiency of MIMO systems [13]. Based on existing MARL methods, we successfully apply the MADDPG algorithm to solve power control problems for better performance. In addition, we introduce a double-layer power control architecture called D-MADDPG that is based on LSF coefficients between antennas. This architecture differs from the conventional single-layer architecture, which considers all antennas subjected to an agent as a whole, and it demonstrates a notable advantage in increasing the sum SE, as shown in Fig. 6.

Future Research Directions:

Precoding Design: The hybrid precoding is promising to relieve the pressure of excessive power consumption by decomposing the high-dimensional full-digital precoder into the realization of an analog beamformer and digital precoder. Through effective precoding design, the RF links and power costs can be significantly reduced. Additionally, advanced precoding designs can mitigate the beam split effect that severely degrades the achievable rate degradation.

Partial Interaction Design: Designing a distributed MARL algorithm with partial-interaction architecture is a promising way to lessen the quantity of network training and information interaction. Partial interaction allows agents to select appropriate agents for flexible interaction based on distance, service relationship, and other factors, rather than interacting with all agents in cell-free XL-MIMO systems, which is more practical for scalable networks.

Jointly Optimization Design: The jointly optimized design of AS and power control is promising to enhance the robustness of the system, eliminating the need for separate optimization. And the power allocation can be re-examined with appropriate antennas selected from XL-MIMO systems.

Based on the above discussion, the design of AS and power control utilizing real-time interactions with the MADDPG method achieves a higher performance gain in the near-field. Accordingly, such effective MARL methods can be extended to other resource allocation schemes.

Future Research Directions

Hybrid-Field Channel Estimation

To obtain accurate CSI, the CE is a key challenge because the near-field angle-domain channel is not sparse. Faced with huge data streams, lightweight CE methods with reduced computational complexity, fast convergence, and exhaustive channel feature capture are essential to adapt to the near-field characteristics and non-stationary channels. Furthermore, the accurate models based on the spherical wavefronts even the hybrid spherical- and planar-wavefronts, capturing more channel details are essential to reduce the bit error rate due to the user mobility.

Hybrid-Field Beamforming

First, for the near-field beam training, the array response vector of near-field channels is not only related to the angle but also the distance, resulting in a high-dimension codebook set. Thus, a polar-transform codebook should be utilized instead of a discrete Fourier transform codebook to capture the information of the channel paths. Second, the near-field beam split effect occurs when the transmitting antennas are placed close to each other and the distance between the antennas is comparable to the wavelength of the signal being transmitted. In such cases, the transmitted signal may split into multiple near-field beams that interfere with each other. Third, it is a hybrid-field joint design optimization for solving the switch from a far-field beamsteering to a near-field beamfocusing, and vice versa.

RIS-Aided Cell-Free XL-MIMO

With the ability to dynamically reconfigure the electromagnetic environment, reconfigurable intelligent surface (RIS) can improve channel quality and overcome the limitations of the propagation environment. In the future, the evolution of RIS will perhaps develop toward extremely large-scale RIS (XL-RIS) for the future 6G wireless communications, which makes beam training complicated and data throughput exploded. Additionally, since the RIS is deployed in the near-field of XL-MIMO, the RIS codebook should be well-designed considering the NFC characteristics.

Green Communications

To achieve green communications, next-generation communication systems propose sustainable, energy-efficient, and energy-aware requirements. Low-resolution devices, for example, analog-to-digital converters (ADCs), are the trend to cope with the great expense of cell-free XL-MIMO systems. On the one hand, hardware impairments still confuse signal processing, especially when the dimension is gigantic. Accordingly, the fruitful compensation algorithm design is necessary to approach the opti-

mum. On the other hand, the simultaneous wireless information and power transfer technology should focus on elaborate near-field beamforming design to achieve a higher performance.

Conclusion

In this article, the fundamental opportunities in the near-field communication of cell-free XL-MIMO systems and open challenges have been discussed in terms of SWM, spatial non-stationary effect, EDoF, power consumption, and computational complexity, respectively. In particular, we investigated the existing MARL categories and proposed the basic scheme of promising cell-free XL-MIMO systems using MARL methods. Then, we investigated two existing challenges namely AS and power control. Accordingly, we successfully applied MADDPG-based algorithms to solve them. Finally, we pointed out the critical and promising future research directions, which are hybrid-field CE, hybrid-field beamforming, RIS-aided cell-free XL-MIMO architecture, and green communications.

Acknowledgments

This research is supported in part by the Fundamental Research Funds for the Central Universities under Grant No. 2023YJS001, in part by National Key R&D Program of China under Grant 2020YFB1807201, in part by National Natural Science Foundation of China under Grant 62221001, in part by Natural Science Foundation of Jiangsu Province, Major Project under Grant BK20212002, in part by the Fundamental Research Funds for the Central Universities under Grant 2022JBQY004, in part by ZTE Industry-University-Institute Cooperation Funds under Grant No. HC-CN-20221202003, in part by the National Research Foundation, Singapore, and Infocomm Media Development Authority under its Future Communications Research & Development Programme, DSO National Laboratories under the AI Singapore Programme (AISG Award No: AISG2-RP-2020-019), and MOE Tier 1 (RG87/22).

References

[1] J. Zhang *et al.*, "Prospective Multiple Antenna Technologies for Beyond 5G," *IEEE JSAC*, vol. 38, no. 8, Aug. 2020, pp. 1637–60.
[2] M. Matthaiou *et al.*, "The Road to 6G: Ten Physical Layer Challenges for Communications Engineers," *IEEE Commun. Mag.*, vol. 59, no. 1, Jan. 2021, pp. 64–69.
[3] H. Iimori et al., "Joint Activity and Channel Estimation for Extra-Large MIMO Systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, 2022, pp. 7253–70.
[4] Z. Wang *et al.*, "Extremely Large-Scale MIMO: Fundamentals, Challenges, Solutions, and Future Directions," *IEEE Wireless Commun.*, early access, 2023, pp. 1–9.
[5] H. Zhang *et al.*, "6G Wireless Communications: From Far-Field Beam Steering to Near-Field Beam Focusing," *IEEE Commun. Mag.*, vol. 61, no. 4, Apr. 2023, pp. 72–77.
[6] C. Huang *et al.*, "Holographic MIMO Surfaces for 6G Wireless Networks: Opportunities, Challenges, and Trends," *IEEE Wireless Commun.*, vol. 27, no. 5, July 2020, pp. 118–25.
[7] Y. Liu *et al.*, "Near-Field Communications: What Will Be Different?" arXiv:2303.04003, 2023.
[8] M. Cui *et al.*, "Near-Field MIMO Communications for 6G: Fundamentals, Challenges, Potentials, and Future Directions," *IEEE Commun. Mag.*, vol. 61, no. 1, 2023, pp. 40–46.
[9] S. Hwang *et al.*, "Decentralized Computation Offloading With Cooperative UAVs: Multi-Agent Deep Reinforcement Learning Perspective," *IEEE Wireless Commun.*, vol. 29, no. 4, Aug. 2022, pp. 24–31.
[10] M. Rahmani *et al.*, "Multi-Agent Reinforcement Learning-Based Pilot Assignment for Cell-Free Massive MIMO Systems," *IEEE Access*, vol. 10, Nov. 2022, pp. 120,492–502.
[11] Q. Fan *et al.*, "MADDPG-Based Power Allocation Algorithm for Networkassisted Full-Duplex Cell-Free Mmwave Massive MIMO Systems With DAC Quantization," *Proc. 2022 WCSP*, 2022, pp. 556–61.
[12] H. Lu and Y. Zeng, "Communicating with Extremely Large-Scale Array/ Surface: Unified Modeling and Performance Analysis," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, June 2022, pp. 4039–53.
[13] Z. Liu *et al.*, "Uplink Power Control for Extremely Large-Scale MIMO With Multi-Agent Reinforcement Learning and Fuzzy Logic," *Proc. IEEE Conf. Computer Commun. Workshops*, Hoboken, NJ, USA, 2023, pp. 1–6.
[14] J. C. Marinello *et al.*, "Antenna Selection for Improving Energy Efficiency in XL-MIMO Systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, Nov. 2020, pp. 13,305–18.
[15] X. Chai *et al.*, "Reinforcement Learning Based Antenna Selection in Usercentric Massive MIMO," *Proc. IEEE VTC2020-Spring*, 2020, pp. 1–6.

Biographies

Zhilong Liu received the B.S. degree from the School of Information and Control Engineering, Qingdao University of Technology, Qingdao, China, in 2022. He is currently pursuing the Ph.D. degree with Beijing Jiaotong University, Beijing, China. His research interests include massive MIMO systems, signal processing, reinforcement learning, and performance analysis of wireless systems.

Jiayi Zhang [SM'20] is a Professor with the School of Electronic and Information Engineering, Beijing Jiaotong University. His research interests include cell-free massive MIMO, XL-MIMO, and RIS. He was an Associate Editor for IEEE Trans. Communications and IEEE Trans. Wireless Communications.

Ziheng Liu received the B.S. degree from the School of Information and Control Engineering, Qingdao University of Technology, Qingdao, China, in 2023. He is currently pursuing the Ph.D. degree with Beijing Jiaotong University, Beijing, China. His research interests include massive MIMO systems, signal processing, and reinforcement learning.

Hongyang Du received the B.S. degree from Beijing Jiaotong University, Beijing, China, in 2021. He is working toward his Ph.D. degree with the School of Computer Science and Engineering, Energy Research Institute at NTU, Nanyang Technological University, Singapore, under the Interdisciplinary Graduate Program. His research interests include semantic communications, generative artificial intelligence, and communication theory.

Zhe Wang received the B.S. degree from the College of Electronic Information, Qingdao University, Qingdao, China, in 2020. He is currently pursuing the Ph.D. degree with Beijing Jiaotong University, Beijing, China. His research interests include massive MIMO systems, signal processing, and performance analysis of wireless systems.

Dusit Niyato [F'17] is a Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He received B.Eng. from King Mongkuts Institute of Technology Ladkrabang (KMITL), Thailand, in 1999 and Ph.D. in Electrical and Computer Engineering from the University of Manitoba, Canada, in 2008. His research interests are in the areas of sustainability, edge intelligence, decentralized machine learning, and incentive mechanism design.

Mohsen Guizani [F'09] is a Professor of Machine Learning and the Associate Provost at Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI), Abu Dhabi, UAE. His research interests include applied machine learning, artificial intelligence, Internet of Things, smart city, and cybersecurity. He was listed as a Clarivate Analytics Highly Cited Researcher in Computer Science in 2019, 2020, 2021, and 2022.

Bo Ai [F'22] is a professor with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University. His interests include high-power amplifier linearization techniques, radio propagation and channel modeling, global systems for mobile communications for railway systems, and LTE for railway systems.

In the future, the evolution of RIS will perhaps develop toward extremely large-scale RIS for the future 6G wireless communications, which makes beam training complicated and data throughput exploded.