Future Vision on Artificial Intelligence Assisted Green Energy Efficiency Network

CHEN Jiajun[1,2], GAO Yin[1,2], LIU Zhuang[2], LI Dapeng[1,2]

(1. State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518055, China;
2. Wireless Product R&D Institute, ZTE Corporation, Shanghai 201203, China)

DOI: 10.12142/ZTECOM.202302006

https://kns.cnki.net/kcms/detail/34.1294.TN.20230426.1116.002.html, published online April 26, 2023

Manuscript received: 2023-02-01

Abstract: To meet the key performance requirement of the 5G network and the demand of the growing number of mobile subscribers, millions of base stations are being constructed. 5G New Radio is designed to enable denser network deployments, which raises significant concerns about network energy consumption. Machine learning (ML), as a kind of artificial intelligence (AI) technologies, can enhance network optimization performance and energy efficiency. In this paper, we propose AI/ML-assisted energy-saving strategies to achieve optimal performance in terms of cell shutdown duration and energy efficiency. To realize network intelligence, we put forward the concept of intrinsic AI, which integrates AI into every aspect of wireless communication networks.

Keywords: machine learning; energy efficiency; traffic distribution; load prediction; intrinsic AI

Citation (Format 1): CHEN J J, GAO Y, LIU Z, et al. Future vision on artificial intelligence assisted green energy efficiency network [J]. ZTE Communications, 2023, 21(2): 34–39. DOI: 10.12142/ZTECOM.202302006
Citation (Format 2): J. J. Chen, Y. Gao, Z. Liu, et al., "Future vision on artificial intelligence assisted green energy efficiency network," *ZTE Communications*, vol. 21, no. 2, pp. 34–39, Jun. 2023. doi: 10.12142/ZTECOM.202302006.

1 Introduction

Wireless communication has witnessed rapid development, especially in terms of higher data rates, considerably smarter devices, and diverse applications. Moreover, compared with the 4G technology, 5G uses high-frequency bands, which makes the nodes denser. To achieve optimal performance in the radio access network (RAN) and meet the demand of increasing mobile subscribers, millions of base stations (BSs) are constructed. The number of BSs in developing regions has increased by over 2 million from 2007 to 2015, and data transmission rates have increased tenfold every five years[1]. However, the expected surge in traffic load requires 5G New Radio to enable denser network deployment and network densification, which results in higher energy consumption. Most of the energy is consumed by BSs in the typical RAN. However, with the deployment of more base stations with massive multiple-input multiple-output (MIMO), energy efficiency in NR becomes more urgent and challenging.

One of the energy-saving schemes that have received extensive attention from academia and the industry is cell activation/deactivation based on load prediction[2–3]. In the third Generation Partnership Project (3GPP), energy-saving standard cases have been specified in Releases 15 and 16, such as the intra-radio access technology (RAT) case with the central unit-distributed unit (CU-DU) split, the intra-system inter-RAT case, and multiple radio access technology-dual connectivity (MR-DC)[4]. An approach has also been recently developed to optimize wireless communications and introduced into self-organizing networks (SON) to allow for smarter operation and maintenance of operators' daily tasks[5]. The inclusion of AI-based tools enables a more proactive approach to exploiting the vast number of data available and incorporating additional dimensions, such as end-user experience and behavior characterization[6–8]. The cell providing capacity booster can be switched off autonomously according to its cell traffic load status. Ref. [9] leverages AI/ML methods to predict load and achieve energy efficiency performance through dynamic threshold configuration.

In this paper, we introduce and provide related works on AI/ML based energy efficiency, simulation and evaluation in real environments, and future vision on AI/ML based wireless networks. The main contributions of this work can be summarized as follows:

1) The benefits of AI/ML enabled wireless networks and the deployment of RAN intelligence are provided.

2) Compared with no-energy saving schemes and traditional energy saving strategies, the proposed AI/ML based energy saving scheme achieves great performance on power consumption and energy efficiency.

3) Based on further consideration of future wireless communication networks, we propose to integrate AI into every aspect of wireless communication systems to depict a vision of the intrinsic AI through intelligent data perception, intelligent modeling, distributed architecture, and intelligent monitoring.

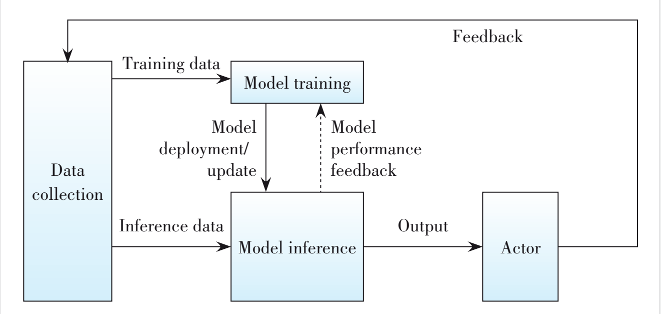
The rest of this paper is organized as follows. Section 2 introduces AI/ML assisted wireless networks and describes their benefits. The definition of typical energy-saving features and the AI/ML based energy efficiency strategy is provided in Section 3, followed by the simulation and results in Section 4. Future vision on AI/ML based wireless communication networks is provided in Section 5. Section 6 shows the conclusions.

2 Machine Learning Assisted Wireless Networks

As an important research direction of AI technologies, machine learning takes advantage of the depth of the neural network's non-linear processing capability, which successfully solves a series of previously intractable problems. In image recognition, speech processing, and natural language processing, AI shows greater performance than humans and has better ability than traditional algorithms[10]. It therefore has been successfully applied in a variety of technologies, services and applications, including telecommunications. Many optimization issues in the wireless network, such as network energy saving, mobility optimization, and load balancing, can be resolved through powerful tool-machine learning, analyzing the data pattern and historical information to predict the trends or generate optimization decisions. Fig. 1 illustrates the functional framework for RAN intelligence[11].

1) Data collection is the function that provides input data to model training and model inference functions. Training data represent the data needed as input for the AI/ML model training function, while inference data are those needed as input for the AI/ML model inference function.

2) Model training is the function that performs the AI/ML model training, validation, and testing, which may generate model performance metrics as part of the model testing procedure. The model training function is also responsible for data preparation (e. g., data pre-processing and cleaning, formatting, and transformation).



▲Figure 1. Framework of artificial intelligence/machine learning (AI/ML) enabled wireless networks

3) Model inference is the function that provides AI/ML model inference output (e. g., predictions or decisions). The model inference function may provide model performance feedback to the model training function when applicable. The model inference function is also responsible for data preparation (e. g., data pre-processing and cleaning, formatting, and transforming). The inference output of the AI/ML model is produced by the model inference function.

4) Actor is the function that receives the output from the model inference function and triggers or performs corresponding actions. The actor may trigger actions directed to other entities or at itself.

5) Model deployment/update is used to initially deploy a trained, validated, and tested AI/ML model to the model inference function or to deliver an updated model to the model inference function.

6) Model performance feedback is used for monitoring the performance of the AI/ML model. After model inference is executed, the model performance is generated and returned to the model training function to evaluate whether the model performance is good or not. If the performance is not good, the model training function can trigger model retraining and reselection.

7) Feedback is the information needed to derive training data and inference data, or to monitor the performance of the AI/ML model and its impact on the network through updating key performance indicators (KPIs) and performance counters.

The 5G network and even future networks require the introduction of AI/ML to achieve automated and intelligent operations. For network energy saving, ML algorithms may predict the energy efficiency and load state of the next period, which can be used to make better decisions on cell activation/deactivation for energy saving. Based on the predicted load, the system may dynamically configure the energy-saving strategy. For mobility optimization, many radio resource management (RRM) actions related to mobility (e.g., selecting handover target cells) can benefit from the predicted user equipment (UE) location/trajectory. For load balancing, based on a collection of various measurements and feedback from UE, network nodes, historical data, etc., AI/ML model-based solutions and predicted load could improve load balancing performance, in order to provide a higher quality user experience and to improve the system capacity.

3 Strategies for Energy Efficiency

3.1 Energy Efficiency Features

Typical energy efficiency strategies used for wireless networks include symbol shutdown, channel shutdown, carrier shutdown and deep sleep, which can be categorized into a symbol level, a physical channel level, and a machine level. Following are the definitions of each energy efficiency strategy.

1) Symbol shutdown: A base station detects that some downlink symbols have no data to send, and thus it turns off the PA and other analogue components, thereby reducing the power consumption of the base station, which basically has no effect on the user's latency. Moreover, by adjusting the number of synchronization signal block (SSB) beams when cells have no traffic or light traffic in a specified period of time, the proportions of symbols for which symbol power saving can take effect can be increased.

2) Channel shutdown: It refers to the technology of multichannel base stations such as 64/32 channels by muting some RF channels of the base station with low traffic, thereby reducing the power consumption of the base station. But for services with higher throughput requirements, when the user channel environment deteriorates, it is necessary to consider that the coverage cannot be reduced.

3) Carrier shutdown: When the service volume of the entire BS is low during off-peak hours at night ,the BS energy consumption can be reduced by retaining only the coverage-layer cells and shutting down the capacity-layer cells. If the service load is lower than a specified threshold, the capacity layers are dynamically shut down. When the load of the carrier providing basic coverage is higher than a specified threshold, the base station dynamically turns on the carriers that have been shut down for service provisioning.

4) Deep sleep: The power requirements of the radio base station vary with the cell traffic load. As the service load of the cell increases, the power amplifier gradually becomes the most energy-consuming component of the base station. Additionally, in the scenario of no traffic load, the power demand of the wireless base station mainly comes from the digital intermediate frequency module. It is worth noting that in the absence of traffic load, and between control signaling transmissions, the BS part consumes energy even when transmissions are not required. Thus, a strategy arises to reduce unnecessary radio BS energy consumption by gradually deactivating components when they remain unused for transmission.

Energy efficiency strategies can be adopted based on the various wireless network environment, and different energy efficiency strategies produce different energy-saving results. In addition, AI/ML technologies can be used to help choose which energy-saving strategy for a certain scenario. In this paper, the channel shutdown and symbol shutdown are mainly used in simulation and evaluation in Section 4.

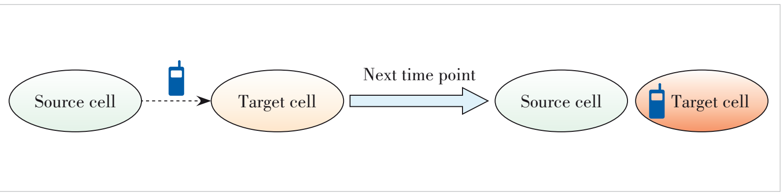
3.2 AI/ML Assisted Energy Efficiency Strategy

Cell activation/deactivation is an energy saving scheme in the spatial domain that exploits traffic offloading in a layered structure to reduce the energy consumption of the whole RAN. When the expected traffic volume is lower than a fixed threshold, the cells may be switched off, and the served UE may be offloaded to a new target cell. Efficient energy consumption can also be achieved by other means such as reduction of load, coverage modification, or other RAN configuration adjustments. The optimal energy saving decision depends on many factors including the load situation at different RAN nodes, RAN nodes capabilities, KPI/quality of service (QoS) requirements, number of active user devices, UE mobility, cell utilization, etc. AI/ML techniques could be utilized to optimize the energy saving decisions by leveraging the data collected in the RAN network. AI/ML algorithms may predict the energy efficiency and load state of the next period, which can be used to make better decisions on cell activation/deactivation for energy saving. Based on the predicted load, the system can dynamically configure the energy-saving strategy (such as the switch-off timing and granularity and offloading actions) to keep a balance between system performance and energy efficiency and to reduce energy consumption.

Moreover, using statistics of past and current cell traffic and mobility management events, radio resource and mobility management strategies can be optimized to devise appropriate green strategies to minimize network energy consumption, while avoiding the degradation of network performance in terms of coverage quality, user rate, and handover failures. Since both the load of the serving cell and those of the neighboring ones play a role in the carrier shutdown procedure, optimized traffic distribution calls for centralized cell load predictions or exchange of predictions across cells. Only when the predicted load of the source and target cells is low, the source cell may be deactivated and the target cell is handed over to its UE to avoid QoS degradation. An optimized traffic distribution should also account for hard-to-predict traffic fluctuations and future cell loads as well as the change in signal quality of the UE in the neighborhood of a shutdown carrier, which in turn may also significantly affect cell loads in the area. A new target cell may be handed over to UE, but the load of this target cell may rapidly increase soon after, making the target cell fail to meet the required QoS, and require an immediate reactivation of the recently shutdown carrier, shown in Fig. 2.

4 Evaluation and Performance

The energy-saving strategy for the 5G system adopted in this paper comprises three components: all-day symbol shutdown, all-day AI-based channel shutdown, and AI-based deep sleep from 0 a.m. to 6 a.m. The test area is located in the Panyu District of Guangzhou, China, and involves 54 active an-



▲Figure 2. Potential scenarios cause deterioration of energy efficiency tenna units (AAU). The energy-saving strategy is triggered by the predicted load threshold, as described in Section 3. Three-time ranges are set to compare the performance of no energy-saving, traditional energy-saving, and AI/ML-based energy-saving. T0 represents the time without energy-saving strategies, T2 represents the time when the AI/ML technology is used, and T1 represents the time with traditional energy-saving strategies. The simulation's configuration is illustrated in Table 1.

Table 2 displays statistical information on the duration of energy-saving strategies using AI/ML techniques compared with traditional methods. The data indicates a significant increase in the duration of deep sleep and channel shutdown, which has resulted in even greater energy savings. Specifically, the use of AI/ML techniques has extended the duration of shutdown for these strategies, resulting in a reduction of power consumption of 452.18 W and a 2.48% improvement in energy efficiency, compared with traditional methods. This information highlights the potential benefits of using AI/ML techniques for energy-saving purposes, particularly in extending the duration of energy-saving strategies, bringing significant energy savings. Fig. 3 shows the average power supply

▼**Table 1. Configuration information of evaluation**

Time	Range	Type	Energy Saving Strategy
T0	2022-06-09 ~ 2022-06-15	None	W/O channel shutdown W/O symbol shutdown W/O deep sleep
T1	2022-05-27 ~ 2022-06-22	Tradition	Channel shutdown Symbol shutdown Deep sleep
T2	2022-06-25 ~ 2022-07-01	AI/ML assisted	AI/ML channel shutdown AI/ML symbol shutdown Deep sleep

AI: artificial intelligence ML: machine learning

▼**Table 2. Time statistics of the duration of shutdown**

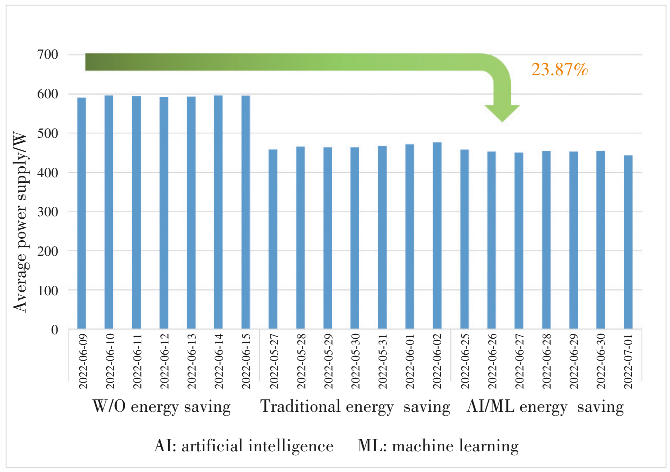
Phase	Deep Sleep/h	Channel Shutdown/h	Symbol Shutdown/h	Power Consumption/W	Improvement (Compared with T0)	Improvement (Compared with T1)
T0-W/O ES	0	0	0	593.97	-	-
T1-Traditional ES	2.61	0.58	9.63	466.92	21.39%	2.48%
T2-AI/ML ES	3.48	5.94	9.03	452.18	23.87%	-

AI: artificial intelligence ES: energy saving ML: machine learning

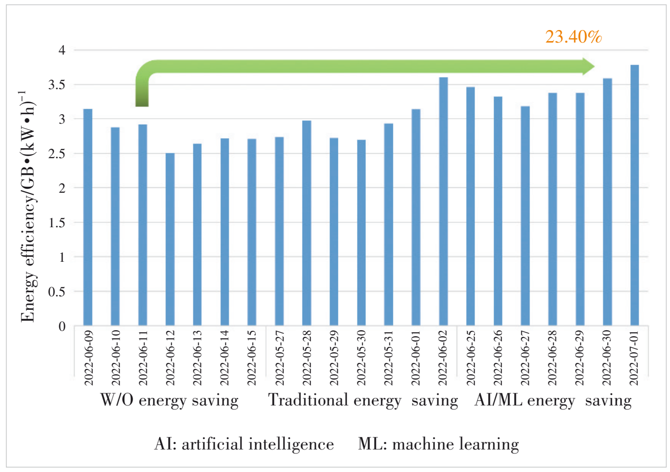
▼**Table 3. Time statistics of the duration of energy efficiency**

Phase	5G Energy Efficiency /GB•(kW•h)$^{-1}$	Improvement (Compared with T0)
T0-W/O ES	2.78	-
T1-Tranditional ES	2.97	6.55%
T2-AI/ML ES	3.44	23.40%

AI: artificial intelligence ES: energy saving ML: machine learning



▲**Figure 3. Average power supply**



▲**Figure 4. Energy efficiency improvement**

and a 23.87% reduction in power consumption with the use of AI/ML techniques. Fig. 4 shows the energy efficiency improvement of 23.40% achieved by using AI/ML techniques.

Overall, by utilizing AI/ML techniques to determine energy-saving strategies, energy efficiency can be significantly improved while simultaneously reducing energy consumption, leading to an increase in the energy-saving duration of base stations.

5 Future Vision on AI/ML Assisted Wireless Networks

The current 5G communication system is designed as a service-based architecture, providing a modular framework for meeting stringent latency and reliability requirements. Barely introducing AI technologies in wireless networks to solve a certain network optimization problem does not enable networks to be intelligent. Continuously changing radio environment requires retraining and updating the fixed ML models, resulting in repetitive work and hindering the intelligence of the wireless system. Future communication systems are not only considered to apply AI to enhance a certain function, such as energy saving, mobility management, etc., but also designed to integrate AI into every aspect of wireless communication systems to depict a vision of the intrinsic AI.

The key requirements of the 5G network are stringent latency and reliability in user scenarios, e.g, ultra-reliable low-latency communications (URLLC), enhanced mobile broadband (eMBB), and massive machine type of communication (mMTC), while future wireless networks (6G) seamlessly integrates with AI, communication networks, and edge computing. To support AI function, 5G networks preset functional modules to monitor and enhance the performance of service-based architecture (SBA), but in the future wireless network, it involves self-sensing, self-learning, self-decision and self-evolution to support self-capability and realize AI/ML integration with communication networks. The transition in the AI/ML structure from 5G networks to future wireless networks is shown in Fig. 5.

Therefore, to bring autonomous learning, autonomous decision-making, self-optimization, and self-evolution, AI-native radio networks will be an intelligent loop including intelligent data perception, intelligent modeling, distributed architecture, and intelligent monitoring.

1) Intelligent data perception

A large quantity of data transportation will bring burdens to the current interface. On the other hand, data sensed from the radio environment sometimes do not have the corresponding labels. Now with the generative adversarial networks (GANs), it will avoid transferring a large number of data between various nodes in the network and protect the data privacy. GAN can generate the required data to simulate real data and improve the performance of models to a certain extent.

2) Intelligent modeling

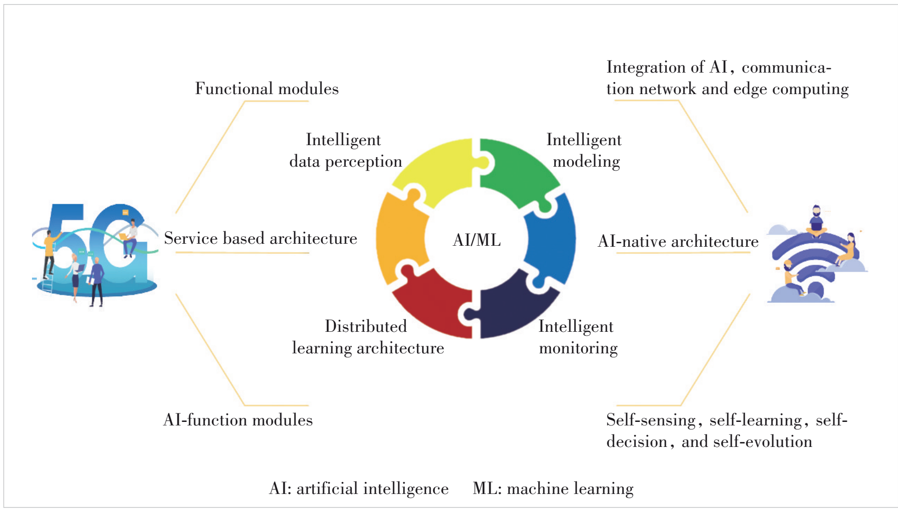
The implementation of machine learning usually requires a lot of manual intervention, such as data pre-processing, feature selection, model selection, hyper-parameter, and adjustment. Each ML model or algorithm has a specific structure and usually comes with a set of strategies or rules for model construction. The purpose of intelligent modeling is to reduce manual intervention so that the radio network can automatically generate and train the AI/ML model to achieve autonomous learning capabilities. Currently, auto-ML technologies can automatically select a machine learning model on given data and tasks, and automatically select an optimization algorithm, so that the model has the characteristics of high performance and low computational complexity of the task.

3) Distributed architecture

The centralized AI server collects the data of each network element and each node in the network, which will bring problems such as time delay and a large number of data transmission. The distributed AI server architecture can effectively solve this problem. The AI units are distributed on each network node to jointly perform the calculation tasks of the same set of AI models. Distributed AI network architecture, where each network node can be used as a part of AI training/execution and a large number of related devices can jointly build a common model based on locally collected data sets, will be the trend of intelligent AI network architecture in the future. It will reduce data transmission load and data privacy leaks on the radio interface, improve model performance, and alleviate delay problems. Distributed AI will play an important role in the subsequent evolution of the network architecture.

4) Intelligent monitoring

Intelligent monitoring is the introduction of human control into the decision-making process of the network to improve the decision-making ability of AI algorithms and help the machine better understand user preferences and make more user-preferred decisions. For example, when the AI model itself cannot make the correct decision or the cost of making the wrong decision is high, the AI algorithm can decide with human intelligence. With reinforcement learning, the agent obtains reward through interaction with the environment or feedback from users, learns the characteristics of the external environment, and improves decision-making strategies to adapt to the external environment changes.



▲Figure 5. Transition on AI/ML structure from 5G to future wireless network

6 Conclusions

In this paper, we introduce the benefits of AI/ML enabled wireless networks and provide the deployment of RAN intelligence. Compared with no energy saving schemes and traditional energy saving strategies, our proposed AI/ML based energy saving schemes achieve great performance on power consumption and energy efficiency. Moreover, we put forward further consideration on future wireless communication networks, which integrate AI into every aspect of wireless communication systems to depict a vision of the intrinsic AI through intelligent data perception, intelligent modeling, distributed architecture, and intelligent monitoring.

Reference

[1] ELKAZZI R, KHALIL A. Energy-saving solution for future cellular systems [C]//The 4th International Conference on Renewable Energies for Developing Countries (REDEC). IEEE, 2019: 1 – 6. DOI: 10.1109/REDEC.2018.8597671

[2] TONG E, WANG Y, DING F, et al. A practical eNB off/on based energy saving scheme for real LTE networks [C]//The 17th International Conference on Advanced Communication Technology (ICACT). IEEE, 2015: 12 – 17

[3] HAN T, ANSARI N. On greening cellular networks via multicell cooperation [J]. IEEE wireless communications, 2013, 20(1): 82 – 89. DOI: 10.1109/MWC.2013.6472203

[4] 3GPP. Eutran overall description: TS 36.300 [S]. 2019

[5] 3GPP. Study on ran-centric data collection and utilization for LTE and NR: TS 37.816 [S]. 2019

[6] FU Y, WANG S, WANG C X, et al. Artificial intelligence to manage network traffic of 5G wireless networks [J]. IEEE network, 2018, 32(6): 58 – 64. DOI: 10.1109/MNET.2018.1800115

[7] LI R P, ZHAO Z F, ZHOU X, et al. Intelligent 5G: when cellular networks meet artificial intelligence [J]. IEEE wireless communications, 2017, 24(5): 175 – 183. DOI: 10.1109/MWC.2017.1600304WC

[8] P'EREZ-ROMER J O, SALLENT O, FERR´US R, et al. Knowledge-based 5G radio access network planning and optimization [C]//The International Symposium on Wireless Communication Systems (ISWCS). IEEE, 2016: 359 – 365

[9] GAO Y, CHEN J J, LIU Z, et al. Machine learning based energy saving scheme in wireless access networks [C]//Proceedings of 2020 International Wireless Communications and Mobile Computing (IWCMC). IEEE, 2020: 1573 – 1578. DOI: 10.1109/IWCMC48107.2020.9148536

[10] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84 – 90. DOI: 10.1145/3065386

[11] 3GPP. Study on enhancement for data collection for NR and EN-DC: TR 37.817 [S]. 2022

Biographies

Chen Jiajun (chen.jiajun1@zte.com.cn) received his master's degree in communication engineer from Shanghai University, China in 2019. He is currently a 5G research engineer at the R&D center of ZTE Corporation and State Key Laboratory of Mobile Network and Mobile Multimedia Technology, China. His research interests include 5G wireless communications and artificial intelligence.

GAO Yin received her master's degree in circuit and system from Xidian University, China in 2005. Since then she has been engaged in 4G/5G technologies and is currently working as a wireless expert and project manager at the research center of ZTE Corporation and the State Key Laboratory of Mobile Network and Mobile Multimedia Technology, China. She has (co)-authored hundreds of proposals for 3GPP meeting and journal papers in wireless communications and has filed more than 200 patents. She has been served as RAN3 Vice Chair of 3GPP RAN3 since 2017 and was elected as Chair of 3GPP RAN3 in May 2021.

LIU Zhuang received his master's degree in computer science from Xidian University, China in 2003. He is currently a senior 5G research engineer at the R&D center of ZTE Corporation and the State Key Laboratory of Mobile Network and Mobile Multimedia, China. His research interests include 5G wireless communications and signal processing. He has filed more than 100 patents and submitted several hundreds of 3GPP contributions.

LI Dapeng received his MS degree in computer science from University of Electronic Science and Technology of China in 2003. He is currently a senior researcher in ZTE corporation and mainly focuses on research and implementation of the wireless access network system.