Enabling Energy Efficiency in 5G Network

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

(1. R&D Center of ZTE Corporation, Shanghai 201203, China;
2. State Key Laboratory of Mobile Network and Mobile Multimedia, Shenzhen 518057, China)

Abstract: The mobile Internet and Internet of Things are considered the main driving forces of 5G, as they require an ultra-dense deployment of small base stations to meet the increasing traffic demands. 5G new radio (NR) access is designed to enable denser network deployments, while leading to a significant concern about the network energy consumption. Energy consumption is a main part of network operational expense (OPEX), and base stations work as the main energy consumption equipment in the radio access network (RAN). In order to achieve RAN energy efficiency (EE), switching off cells is a strategy to reduce the energy consumption of networks during off-peak conditions. This paper introduces NR cell switching on/off schemes in 3GPP to achieve energy efficiency in 5G RAN, including intra-system energy saving (ES) scheme and inter-system ES scheme. Additionally, NR architectural features including central unit/distributed unit (CU/DU) split and dual connectivity (DC) are also considered in NR energy saving. How to apply artificial intelligence (AI) into 5G networks is a new topic in 3GPP, and we also propose a machine learning (ML) based scheme to save energy by switching off the cell selected relying on the load prediction. According to the experiment results in the real wireless environment, the ML based ES scheme can reduce more power consumption than the conventional ES scheme without load prediction.

Keywords: cell switch off; energy efficiency; energy saving; 5G; machine learning

DOI: 10.12142/ZTECOM.202101004

https://kns.cnki.net/kcms/detail/34.1294.TN.20210220.1440.004.html, published online February 20, 2021

Manuscript received: 2020-12-10

Citation (IEEE Format): Z. Liu, Y. Gao, D. P. Li, et al., "Enabling energy efficiency in 5G network," *ZTE Communications*, vol. 19, no. 1, pp. 20–29, Mar. 2021. doi: 10.12142/ZTECOM.202101004.

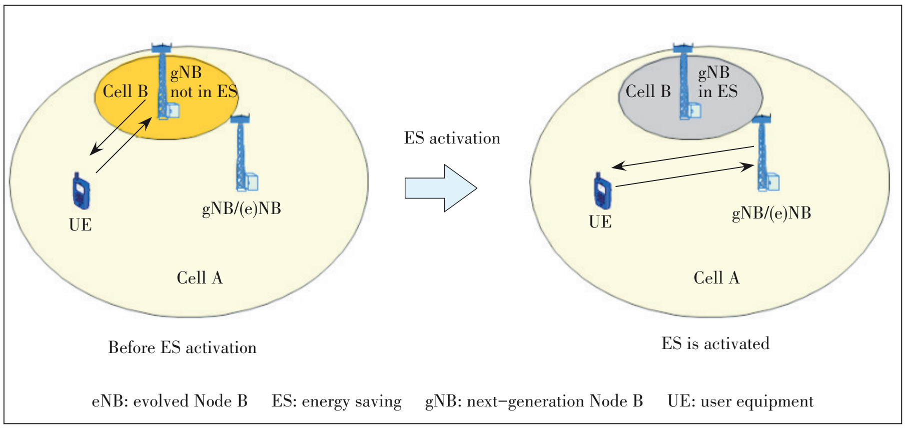
1 Introduction

To cope with expected drastic data traffic growth, 5G new radio (NR) is designed to enable denser network deployments, while the densification of networks has implied higher energy expenditure. In a typical radio access network (RAN), most energy is consumed by base stations. However, with the foreseen NR deployment of more base stations with massive multiple-input multiple-output (MIMO), energy efficiency (EE) in NR becomes even more urgent and challenging.

Energy consumption (EC) is a main part of operational expense (OPEX). The telecommunication operators are seeking for a better way to expand market shares while energy consumption in networks can be decreased to lower their OPEX. Energy efficiency in NR networks is also a significant research topic in 3GPP. Switching off cells is a widely used strategy to reduce the energy consumption of networks during off-peak conditions. Thus network elements with low power consumption become more and more important and the shutdown of unused capacity cells is also valuable. The important aspect of RAN energy efficiency is, during the network running, how to ensure cell switching-off without affecting the customer satisfac e.g., calls dropped; quality of service (QoS) degraded. A typical energy saving (ES) scenario is that capacity booster cells are deployed under the umbrella of cells providing basic coverage and that the capacity booster cells can be switched off to enter into the dormant mode when its capacity is no longer needed and to be reactivated on a need basis. This paper introduces the 3GPP schemes for switching on/off NR cells to achieve energy efficiency in 5G networks, including the 5G intra-system energy saving scheme and 4G/5G inter-system energy saving scheme involving different core networks (CN), e. g., evolved packet core (EPC) and 5G core (5GC) networks. We also propose a machine learning (ML) based scheme to save energy by switching off the cell selected relying on load prediction. According to the experiment results in the real wireless environment, the ML based ES scheme can reduce more power consumption than the conventional ES scheme without load prediction.

2 Cell Switch on/off for Energy Saving

An NR cell, which acts as a capacity booster, may be switched off and enter into the ES dormant state if there is radio coverage by another cell. **Fig. 1** shows an example of the next-generation Node B (gNB) capacity booster cell fully overlaid by a coverage providing cell. The gNB is a node providing the NR user plane and control plane with protocol terminations towards user equipment (UE), and connected to the 5GC via the next-generation (NG) interface. In the figure, Cell A is deployed to provide continuous coverage of the area, while Cell B provides more capacity only for special sub-areas, such as hot spots. The ES activation procedure of Cell B may be triggered in case that light traffic in Cell B is detected. Then the cell will be switched off and enter into the ES dormant state; if there are some users in service in Cell B, the cell will be switched off only after the handover actions to offload its traffic to Cell A is completed. The ES activation of Cell B may be triggered, that is, the cell is switched on again, when the traffic of the ES area (measured by Cell A) resumes to a high level[1].



▲**Figure 1. Capacity booster cell overlaid by coverage providing cell**

In real network deployment, ES can be divided into centralized ES and distributed ES. For the distributed ES, the NR capacity booster cell may decide to switch off when it detects that its traffic load is below a certain threshold, and its coverage can be provided by the coverage providing cell. The coverage providing cell decides to reactivate the NR capacity booster cell when it detects additional capacity is needed. For the centralized ES, a centralized entity, such as the operation and maintenance (O&M) entity, collects the traffic load performance measurements from the NR capacity booster cell and coverage providing cells, and may request a NR capacity booster cell to switch off when its traffic is below certain threshold.

In general, NR energy saving solutions include the 5G intra-system energy saving scheme and 4G/5G inter-system energy saving scheme, involving different core networks (CN), such as EPC and 5GC. Additionally, NR architectural features including central unit/distributed unit (CU/DU) split and dual connectivity (DC) are also considered in NR energy saving. We will discuss these NR ES scenarios in the following sections.

3 5G Intra-System Energy Saving

3.1 Scenarios

In a 5G network, a next-generation RAN (NG-RAN) node is either a gNB or a next-generation evolved Node B (ng-eNB), providing Long Term Evolution (LTE) services or Evolved Terrestrial Radio Access Network (E-UTRAN) services towards the UE. The gNB provides services of the NR user plane and control plane; the ng-eNB provides the E-UTRA user plane and control plane with protocol terminations towards the UE, and connected to the 5GC via the NG interface. The gNBs and ng-eNBs are inter-connected with each other by means of the Xn interface, while they are connected to the 5GC by means of the NG interfaces. The scenarios for NR intra-system energy saving are summarized in **Table 1**.

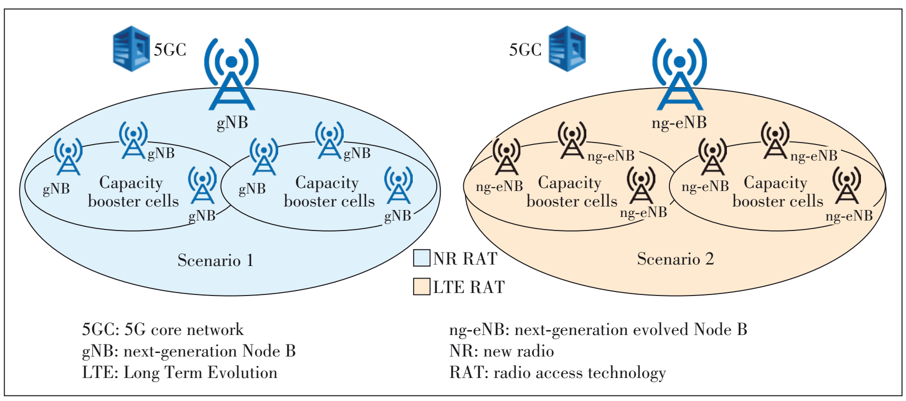
In 5G intra-radio access technology (Intra-RAT) ES cases (Scenario 1; Scenario 2), some gNB (or ng-eNB) cells are deployed to provide basic coverage, while the other gNB (or ng-eNB) cells boost the capacity (**Fig. 2**). Therefore, the coverage provider and the capacity provider are using the same RAT, e. g., 5G NR or LTE, to provide NR services or LTE/E-UTRAN services to UE. The NG-RAN cell providing the capacity booster can decide to switch off autonomously; the switch-off decision may also be taken by the O&M entity that will inform the neighbor NG-RAN cell about its deactivation action over the Xn Application Protocol (XnAP). On the other hand, the NG-RAN node providing the basic coverage can request to reactivate the switched-off booster cell over XnAP.

In 5G inter-radio access technology (Inter-RAT) ES cases (Scenario 3; Scenario 4), some gNB (or ng-eNB) cells are deployed to provide basic coverage, while the other ng-eNB (or gNB) cells boost the capacity (**Fig. 3**). Obviously, the booster cells and coverage cells are in different RAT networks. That is, the booster cells are in NR RAT while the coverage cells in LTE RAT, or the booster cells are in LTE RAT while the coverage cells in NR RAT. The Xn signaling support for inter-RAT ES is the same as that for intra-RAT ES.

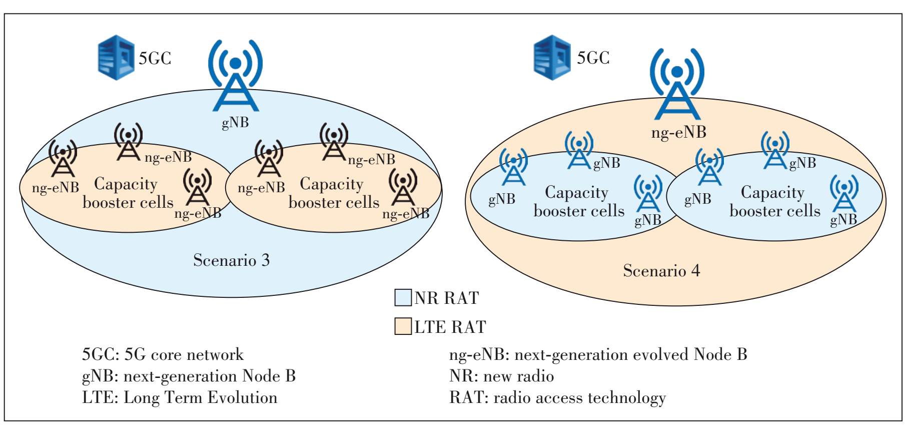
▼**Table 1. 5G intra-system energy saving scenarios (only connected with 5GC)**

5G Intra-System ES Scenario	Coverage Provider	Capacity Booster Provider	Description
1	gNB connected with 5GC	gNB connected with 5GC	intra-RAT ES
2	ng-eNB connected with 5GC	ng-eNB connected with 5GC	
3	gNB connected with 5GC	ng-eNB connected with 5GC	inter-RAT ES
4	ng-eNB connected with 5GC	gNB connected with 5GC	

5GC: 5G core network
eNB: evolved Node B
ES: energy saving
gNB: next-generation Node B
ng-eNB: next-generation evolved Node B
RAT: radio access technology



▲**Figure 2. 5G Intra-RAT energy saving**



▲**Figure 3. 5G Inter-RAT energy saving**

3.2 Signaling Support

The Xn signaling support for both the intra-RAT ES and inter-RAT ES scenarios is same. As shown in **Fig. 4**, NG-RAN Node 1 that owns a capacity booster cell and can autonomously decide whether to switch off this cell based on cell load information, while the switch-off decision may also be taken by the O&M entity. All neighbor NG-RAN nodes are informed by the NG-RAN Node 1 owning the concerned cell about the switch-off actions over the Xn interface, by means of the NG-RAN node configuration update message.

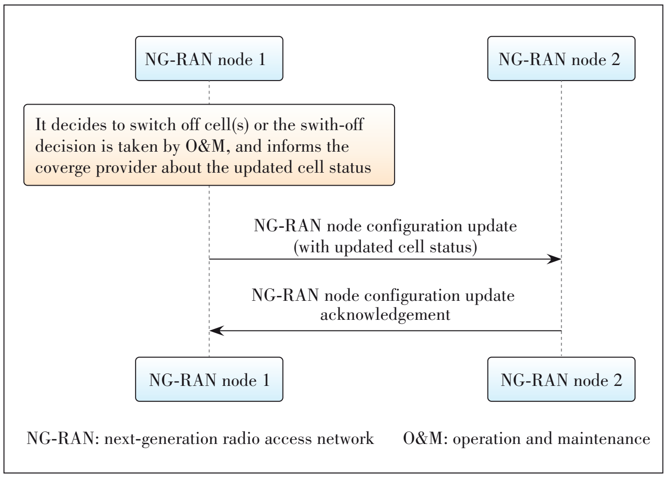
The purpose of the cell activation procedure is to enable an NG-RAN node to request the neighboring NG-RAN node to switch on one or more cells that are previously reported as inactive due to energy saving reasons. As shown in **Fig. 5**, if the basic coverage is ensured by NG-RAN node cells, the NG-RAN node owning non-capacity boosting cells may request a reactivation over the Xn interface via the cell activation procedure if needed. Upon receipt of a cell activation request message, the booster NG-RAN node activates the cells indicated in the message and these cells are also indicated in the cell activation response message when the request is fulfilled.

3.3 Intra-System Energy Saving for Multi-Radio Dual Connectivity

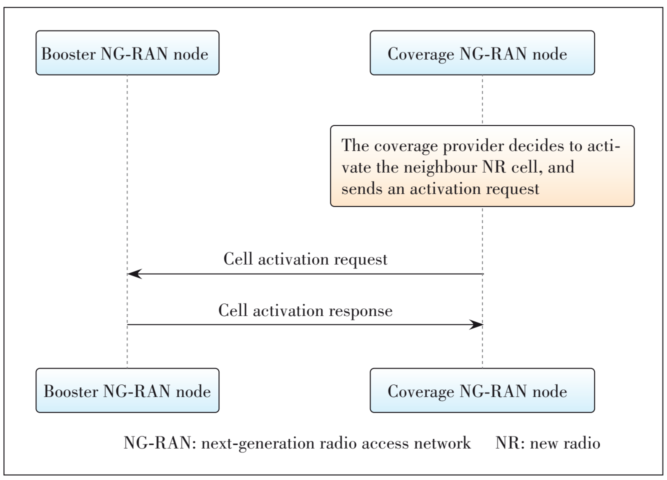
In order to implement multi-radio (MR) dual connectivity (DC), UE may be configured to utilize the resources provided by two different nodes, one providing NR access and the other one providing either E-UTRA or NR access. One node acts as the master node (MN) and the other as the secondary node (SN). The MN and SN are connected via an Xn or X2 interface and at least the MN is connected to a core network. In 3GPP Release 15, the MR DC energy saving is already supported, which is intra-system ES.

When the MN is connected to the 5GC, the DC cases include NG-RAN E-UTRA-NR dual connectivity (NGEN-DC), NE-NR-E-UTRA dual connectivity (DC) and NR-NR dual connectivity (NR-DC):

- NGEN-DC: one ng-eNB is connected with the 5GC and acts as an MN, while one gNB acts as an SN;
- NE-DC: one gNB is connected with the 5GC and acts as an MN, and one ng-eNB



▲Figure 4. NG-RAN node informs the neighbor NG-RAN node about cell status over Xn



▲Figure 5. Coverage NG-RAN node requests to activate booster cells over Xn

acts as an SN;

• NR-DC: one gNB is connected with the 5GC and acts as an MN, and another gNB acts as an SN.

The energy saving scheme for the above cases is similar to 5G intra-system ES (Section 3.1), where the SN can act as a capacity booster provider and MN provides continuous coverage of the area. The SN can autonomously decide to switch off cell(s) based on cell load information or the switch-off decision is taken by O&M; it then informs the MN about the cell deactivation action over XnAP. The MN can request to reactivate the switched-off booster cell at the SN over XnAP. The Xn ES Signaling support for MR DC with 5GC is the same with 5G Intra-system ES Signaling described in Section 3.2.

In the case that the MN is connected to the EPC, that is E-UTRA-NR dual connectivity (EN-DC), one eNB acts as an MN and one en-gNB acts as an SN. The eNB is connected to the EPC via the S1 interface and to the en-gNB via the X2 interface. The EN-DC scenario is shown in **Fig. 6**.

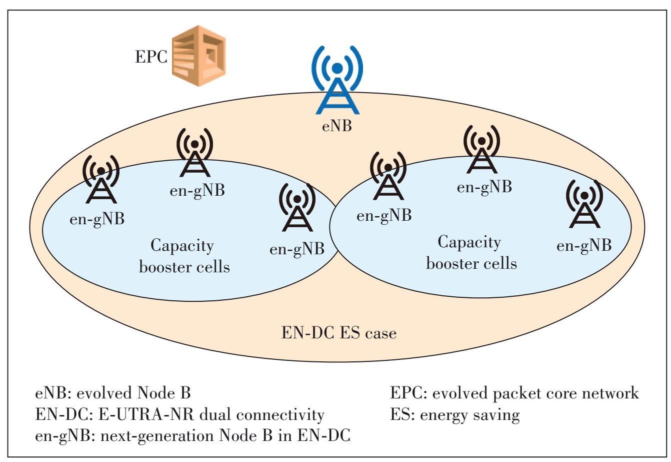
The EN-DC configuration update procedure and EN-DC cell activation procedure are used to support EN DC for intra-system energy saving over the X2 interface. The EN-DC configuration update procedure can be used to exchange updated cell statuses of eNB and en-gNB over the X2 interface. The EN-DC cell activation procedure enables an eNB to request the neighboring en-gNB to switch on one or more cells that are previously reported as inactive due to energy saving reasons. Upon receipt of this message, the en-gNB should activate the cell/s indicated both in the cell activation request message and in the EN-DC cell activation response message sent after the activation request is fulfilled. **Fig. 7** shows the detailed signaling flows.

4 Inter-System Energy Saving of 4G and 5G Systems

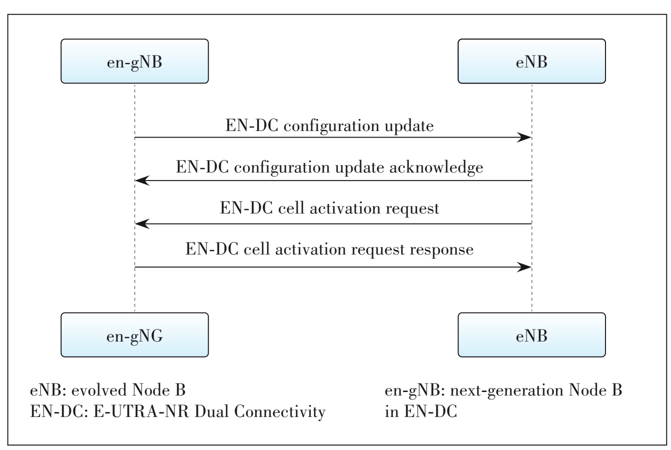
4.1 Scenarios

As shown in **Fig. 8**, 3G users continued to move to 4G from 2009 to Q3, 2018. Although 4G users kept growing steadily, the number of 2G and 3G users could not be ignored for a long time yet[2].

Similar to what happened in the 4G era shown in Fig. 8, it



▲Figure 6. EN-DC energy saving



▲Figure 7. EN-DC energy saving signaling over X2

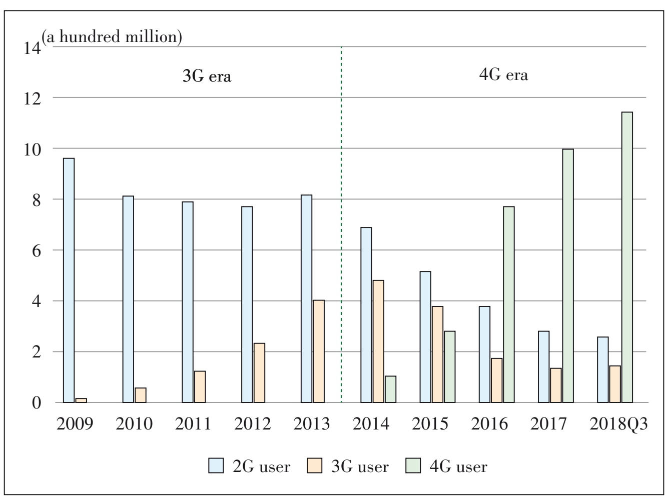
can be predicted that 4G users of operators will gradually decrease when 5G networks are deployed, but 4G networks will coexist with 5G networks for a long time. Therefore, inter-system energy saving solutions to 4G and 5G coexisting scenarios (**Table 2**) should be considered.

For the inter-system ES cases (Scenario 1 and Scenario 2 in Table 2), the NG-RAN node (a gNB, providing NR services; or an ng-eNB, providing E-UTRAN services) owns a capacity booster cell and can autonomously switch off this cell to the dormant state. The switch-off decision is typically based on cell load information, but may also be taken by the O&M entity. The NG-RAN node indicates the switch-off action to the eNB over the NG and S1 interfaces. The NG-RAN node could also indicate the switch-on action to the eNB over the NG and S1 interfaces. The eNB providing basic coverage may request an NG-RAN node's cell reactivation based on its own cell load information or neighbor cell load information, and the switch-on decision may also be taken by O&M. The eNB requests an NG-RAN node's cell reactivation and receives the NG-RAN node's cell reactivation reply from the NG-RAN node over the S1 and NG interfaces. The scenarios in Table 2 are shown in **Fig. 9**, where the E-UTRAN cell associated eNB and the NR-RAN cell associated gNB are connected to the EPC and the 5GC.

4.2 Signaling Support

3GPP Release15 (R15) defines signaling for cell activation/deactivation over X2 and Xn interfaces for intra-system energy saving. However, the signaling defined by R15 fails to support inter-system energy saving scenarios (**Fig. 10**[3]) without a direct interface between the eNB and gNB/ng-eNB. The coverage eNB cannot directly send a request to reactivate the switched-off NR booster cell.

Therefore, NG and S1 interfaces are enhanced in 3GPP Release 16 (R16) to support the inter-system scenarios. Specifically, when the NR capacity booster cell is switched off, the LTE eNB for basic coverage should be informed by the gNB via the NG/S1 message; when the LTE eNB is going to acti-

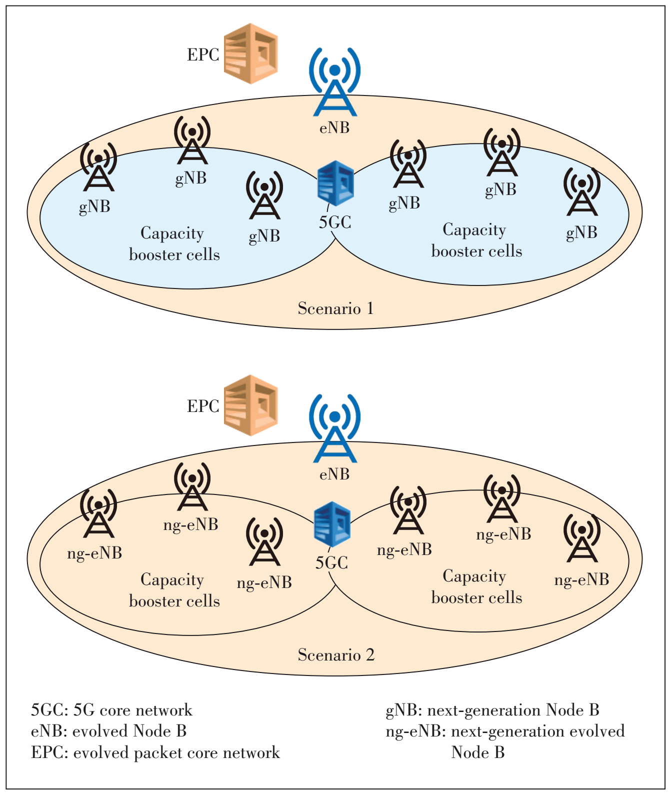


▲ **Figure 8. Users in China continued to move from 2G and 3G networks to 4G networks**

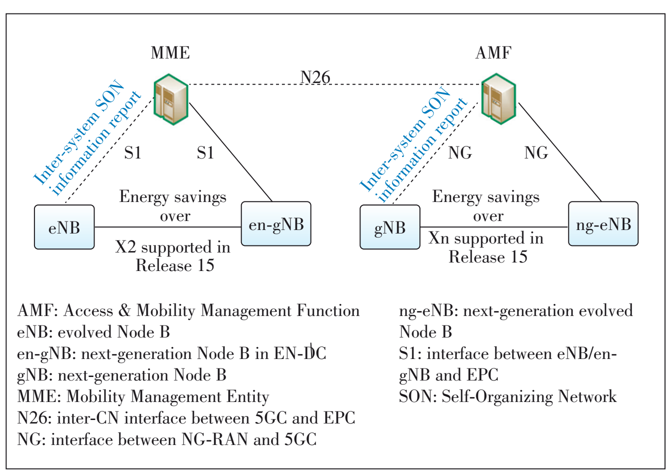
▼**Table 2. The inter-system ES scenarios of 4G and 5G systems (involving EPC and 5GC)**

Scenario	Coverage Provider	Capacity Booster Provider
1	eNB connected with EPC	gNB connected with 5GC
2	eNB connected with EPC	ng-eNB connected with 5GC

5GC: 5G core network
eNB: evolved Node B
EPC: evolved packet core network
ES: energy saving
gNB: next-generation Node B
ng-eNB: next-generation evolved Node B



▲**Figure 9. Inter-system energy saving of 4G and 5G systems**



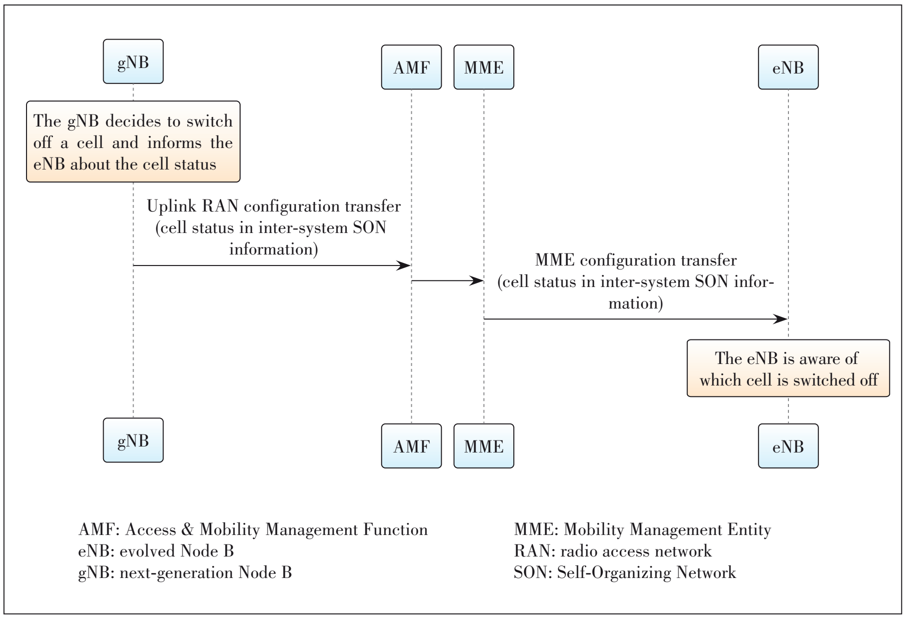
▲ **Figure 10. Inter-system energy saving is not supported by signaling defined by 3GPP Release 15** vate the NR cell, the gNB should be informed by the LTE eNB via the NG/S1 message. In R16, the inter-system Self-Organizing Network (SON) configuration transfer Information Element (IE) is introduced over both the S1 and NG interfaces in the following messages:

- eNB Configuration Transfer (TS36.413[4]);
- Mobility Management Entity (MME) Configuration Transfer (TS36.413[4]);
- Uplink RAN Configuration Transfer (TS38.413[5]);
- Downlink RAN Configuration Transfer (TS38.413[5]).

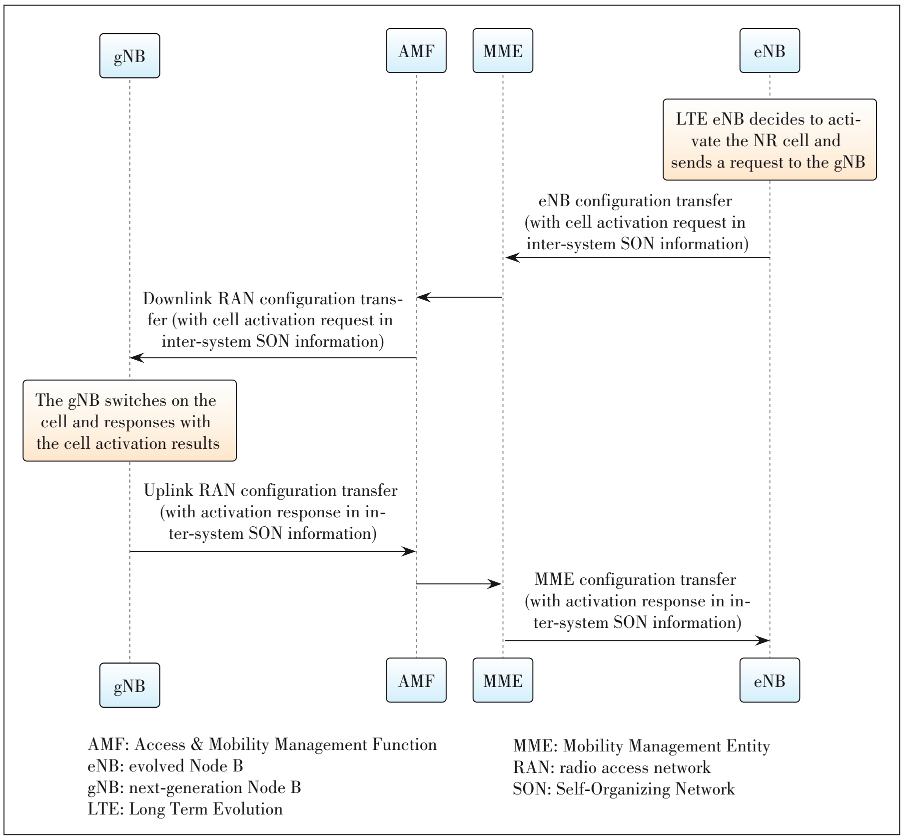
The messages of SON configuration transfer procedures, e.g., the eNB/MME configuration transfer message over S1 and uplink/downlink RAN configuration transfer message over NG, can be reused for R16 inter-system energy saving, as the inter-system SON Information IE in such messages is extended to support inter-system cell status transfer and cell activation request and response between the eNB and NG-RAN node. The detailed signaling flows are shown in **Figs. 11** and **12** respectively.

In Fig. 11, the NG-RAN node owns a capacity booster cell and can autonomously switch off this cell to the dormant state. The switch-off decision is typically based on cell load information and consistent with the configured information. This decision may also be taken by the O&M entity. The NG-RAN node indicates either the switch-off or switch-on actions to the eNB over the NG and S1 interfaces.

In Fig. 12, the eNB providing basic coverage may request an NG-RAN node's cell reactivation based on its own cell load information or neighbor cell load information, and the switch-on decision may also be taken by O&M. The eNB requests an NG-RAN node's cell reactivation and receives the NG-RAN node's cell reactivation reply from the NG-RAN node over the S1 and NG interfaces.



▲Figure 11. Next-generation radio access network (NG-RAN) node connected with 5GC informs cell status to Long Term Evolution (LTE) eNB connected with evolved packet core (EPC) network



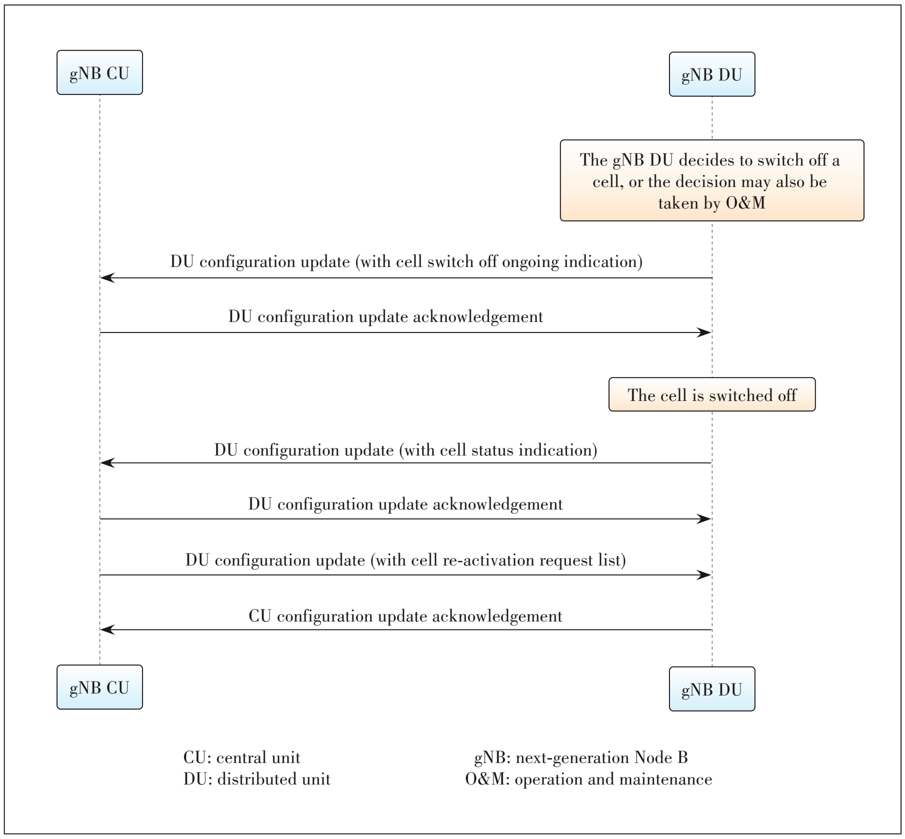
▲Figure 12. LTE eNB connected with EPC requests to activate an NR CELL connected with 5GC

5 Energy Saving in CU/DU Split Architecture

For the intra-system ES described in Section 3 and inter-system ES described in Section 4, if a gNB is deployed with CU/DU split architecture, the F1 interface shall be enhanced to support the cell reactivation procedure and cell status exchange (**Fig. 13**).

When the booster gNB with CU/DU split decides to switch off cell(s) to the dormant state, the decision is typically made by the gNB-DU based on cell load information or by the O&M entity. Before the cell in the gNB-DU enters into the dormant mode, the gNB-DU will send the gNB-DU configuration update message to the gNB-CU to indicate that the gNB-DU will switch off the cell after some time. During the switch-off period, the gNB-CU shall offload the UE to a neighboring cell and simultaneously not accept any incoming UE towards this switch-off ongoing cell. After the cell at gNB-DU enters into the dormant mode, the gNB-DU sends a new gNB-DU configuration update message to inform the "inacitve" status of this cell to the gNB-CU. The gNB-CU needs to inform the updated cell status to the coverage provider node.

When the gNB-CU receives the cell activation request/EN-DC cell activation request from a coverage provider node over the Xn or X2 interface, or the gNB-CU decides to activate the dormant cell by itself, it will trigger the gNB-CU configuration update message to the gNB-DU with a list of the cells to be activated.



▲Figure 13. CU/DU energy saving signaling support over F1 interface

6 Energy Efficiency KPI

EE Key Performance Indicator (KPI) shows data energy efficiency in NG-RAN. The EE KPI is defined as the data volume (in kbits) divided by energy consumption (in kWh) of the considered network elements. The unit of this KPI is bit/J[6].

$$EE = \frac{\sum_{Samples}(DRB.PdcpSduVolumnUL + DRB.PdcpSduVolumnDL)}{\sum_{Samples} PEE.Energy} \quad \text{(for non−split gNBs)}. \tag{1}$$

$$EE = \frac{\sum_{Samples} \left[(F_1uPdcpSduVolumeUL + XnuPdcpSduVolumeUL + X_2uPdcpSduVolumeUL) + (F_1uPdcpSduVolumeDL + XnuPdcpSduVolumeDL + X_2uPdcpSduVolumeDL) \right]}{\sum_{Samples} PEE.Energy} \quad \text{(for split gNBs)}. \tag{2}$$

For non-split gNBs (the gNBs without CU/DU split), the defined $DRB.PdcpSduVolumnUL$ in Eq. (1) is the measured data volume of Packet Data Convergence Protocol (PDCP) Service Data Unit (SDU) of a DRB in the uplink, delivered from the PDCP layer to Service Data Adaptation Protocol (SDAP) layer; $DRB.PdcpSduVolumnDL$ in the equation is the measured data volume of PDCP SDU of a DRB in the downlink, delivered to the PDCP layer. The total data volume (in kbit) is obtained by measuring the amount of uplink and downlink PDCP SDU bits of all DRBs of the non-split gNBs over the measurement period.

For gNBs with CU/DU split, the defined $F_1uPdcpSduVolumeUL$ in Eq. (2) is the measured data volume of PDCP SDU in the uplink, delivered to gNB-CU-UP (gNB-CU-User Plane entity) from gNB-DU via F1-U (F1 User plane interface), $XnuPdcpSduVolumeUL$ is that in the uplink delivered from external gNB-CU-UP via Xn-U (Xn User plane interface), and $X_2uPdcpSduVolumeUL$ is that in the uplink delivered from external eNB via X2-U; the defined $F_1uPdcpSduVolumeDL$ in the equation is the measured data volume of PDCP SDU in the downlink, delivered from GNB-CU-UP to GNB-DU via F1-U, $XnuPdcpSduVolumeDL$ is that in the downlink delivered to external gNB-CU-UP via Xn-U, and $X_2uPdcpSduVolumeDL$ is that in the downlink delivered to external eNB via X2-U. The total data volume (in kbit) is obtained by measuring the amount of uplink and downlink PDCP SDU bits of all interfaces (F1-U, Xn-U and X2-U) of the split gNBs over the measurement period.

The energy consumption (in kWh) is obtained by measuring the Power, Energy and Environmental (PEE) of the considered network elements over the same period of time.

7 Machine Learning Based Energy Saving

In this paper, we introduce 3GPP energy saving schemes by cell switching on/off in ultra-dense networks. However, the traditional cell switching on/off relying on real-time load information is not accurate enough, the inappropriate switching-off of the cells may seriously deteriorate network performance because other active cells have to serve some extra traffic. In order to solve the potential issues of existing ES methods and to achieve intelligent ES, we propose a machine learning (ML) based scheme that relies on the load prediction to save energy by switching off the selected cell[7]. As the data used for load prediction have various features, such as the current and history load and the neighbor cells' load, different techniques are used for the different features of load prediction. The auto-regression integrated moving average (ARIMA), Prophet, Random Forest (RF), Long Short-Term Memory (LSTM), ensemble learning model, and linear regression are used as the input models for load prediction in this paper.

1) ARMIA: It is a time series analysis model, which is fitted to time series data either for better understanding the data or for predicting future points in the time series. When the trend change (T), cyclic change (C), seasonal change (S) and irregular change (I) are used to characterize the time features, the time sequences can be described as

$$Y_t = f(T_t, C_t, S_t, I_t) = T_t + C_t + S_t + I_t. \tag{3}$$

As the ARIMA model has certain requirements for data stability, if considerable changes happen in the load distribution, the model may cause the forecast deviation. Hence, the data could be filtered based on the data stability, so as to select the cells with better response to ARIMA, thereby improving the accuracy of prediction. ARIMA is generally denoted as

$$ARIMA(p, d, q), p, q \in \{0,1,2,3\}, d \in \{0,1\}, \tag{4}$$

where p is the order of the autoregressive model, q is the order of the moving-average model, p and q are determined by the lowest Bayesian information criterion (BIC), and d is the degree of differentiation to make the data stationary.

2) Prophet: The Prophet model, which is similar to ARIMA mode, is expressed as

$$Y_t = f(g_t, s_t, h_t, e_t), \tag{5}$$

where g_t denotes non-periodic changes, such as linear growth or logical growth; s_t is cyclic changes, like seasonality; h_t is irregular changes caused by users; e_t is the error used to describe the abnormal changes in the model.

3) RF: The preliminary of the random forest prediction model is the decision tree learning that segments the features based on their characteristics. Combining the random subspace method with the decision tree, the RF model selects the features to enhance the prediction, increasing the correlation among the selected features. The model exploits the historical loads to predict future loads; in this way, loads in the past and neighbor loads are taken into consideration when constructing the model.

4) LSTM: It is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Different from the standard convolution neural network, LSTM could process the single data points like images, as well as sequences of data such as video or speech. The prediction system in this paper is composed of three layers, two LSTM layers and one fully connection layer.

5) Ensemble learning: This mode combines multiple learning algorithms to achieve better performance for a particular intelligence problem. In other words, ensemble learning can combine several weak models that get poor prediction to produce a strong learning model. While some simple models only learn part of the data, the ensemble method can strategically divide the data set into small data sets, train them separately, and then combine them with certain strategy.

In order to compare the algorithms mentioned above, the mean absolute error (MAE) is used to measure the difference between the forecast and the real load. It can be described as

$$MAE = \frac{\sum_{i=1}^{N} \left| P_i - R_i \right|}{N} = \frac{\sum_{i=1}^{N} \left| e_i \right|}{N}, \tag{6}$$

where N is the number of points, P is the predicted load output by an algorithm, and R is the real load. The intuitive meaning of the function MAE is quite clear: the greater the distance between the predicted value P and the true value R, the larger the loss, and vice versa.

6) Linear regression: It is a linear algorithm to map the relationship between a scalar response and one or more explanatory variables. In the linear regression, unknown model parameters are also estimated from the data. If the goal is to predict or forecast the state, the linear regression is able to fit a predictive model to an observed data. Given a data set $\{y_i, x_{i1}, \ldots, x_{ip}\}_{i=1}^{n}$, the linear regression model can be expressed as:

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i,\ i = 1, \ldots, n, \tag{7}$$

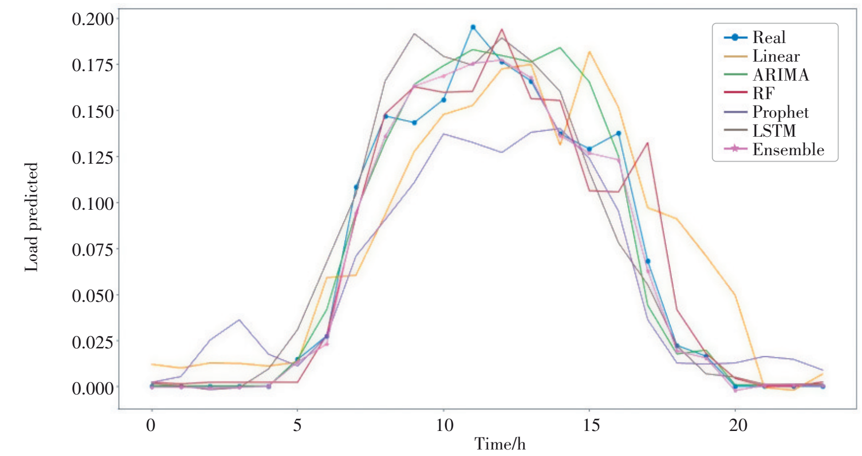
where T denotes the transpose, $\mathbf{x}_i^T \boldsymbol{\beta}$ is the inner product between vectors $\mathbf{x}_i$ and $\boldsymbol{\beta}$.

Fig. 14 compares the load prediction in one cell with different prediction models. The simulation results of load prediction are based on the physical resource block (PRB) utilization in 50 cells. It can be seen that the ensemble learning model has further improved the prediction accuracy compared to each independent sub-model. The average MAE of the ensemble learning method is reduced by an average of 0.008.

The comparison and analysis results of the machine learning models mentioned above is listed in **Table 3**. These different load prediction models are suitable for dealing with different radio access networks.

We evaluate the application of machine learning techniques with real scenarios, and time efficiency is taken into account, ARIMA is implemented in our ML based ES scheme. There are 1 089 cells and 329 base stations (BSs) in the test area. Different switch-off strategies including the symbol switch-off, channel switch-off and carrier switch-off are applied for different groups of the measured BSs. A cell is considered as the switch-off as the cell carrier(s) is/are all switched-off, and the BS shall indicates the switch-off action to the BS providing basic coverage. In the real deployment, different BS types may cause different power consumption, so power saving would be averaged to all cells of the measured BS groups. As expected, the artificial intelligence (AI) energy saving scheme predicts load accurately and switch off the cell in time to achieve better performance on energy saving. In addition, the actual energy saving of each cell per day is also significantly increased. **Fig. 15** shows that the AI based power saving could reach up to 1.24 kWh each cell per day, and no matter what switch-off strategy is used, AI based ES is a better solution to power saving.

Table 4 shows the power consumption and electricity charge saving with different kinds of ES methods and without ES. We can see that the power consumption is totally 25 988 kWh every week if any ES method is not used, while the power consumption with the AI ES methods is 22 304 kWh. Electricity charge saving with the AI ES methods increases more than that with the

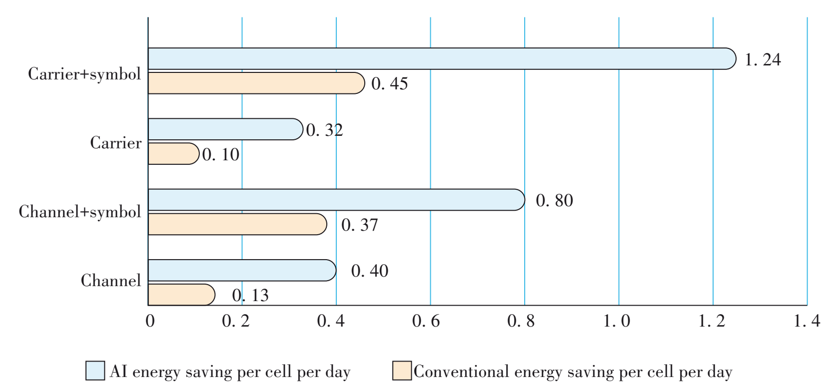


▲Figure 14. Load prediction by using different models

▼Table 3. Comparison and analysis of the machine learning models

Model	Accuracy	Speed	Complexity
ARIMA	Medium	Fast	Low
Prophet	Medium	Fast	Low
LSTM	High	Slow	High
RF	High	Slow	High
Ensemble	High	Extremely slow	High

ARIMA: auto-regression integrated moving average　LSTM: Long Short-Term Memory　RF: Random Forest



▲Figure 15. Statistics of the power saving (kWh)

▼Table 4. Comparison of power consumption and electricity charge saving with/without ES methods

Switch-off Strategy	Number of Measured Cells	Power Consumption of Measured Cells (kWh/Week)			Electricity Charge Saving of Measured Cells (CNY/Week)		
		No ES	Conventional ES	AI ES	Conventional ES	AI ES	Increase
Carrier	8	382	377	364	5	18	13
Carrier+symbol	7	366	344	305	22	61	39
Channel	633	16 853	16 265	15 872	588	981	393
Channel+symbol	327	8 387	7 541	6 555	846	1 832	986
Total	975	25 988	24 527	22 304	1 461	3 684	2 223

AI: artificial intelligence　ES: energy saving conventional ES, and the saving is totally increased by 2 223 CNY every week.

8 Conclusions and Future work

In this paper, we introduce the 3GPP energy saving schemes by switching on/off cells in ultra-dense networks. In order to achieve intelligent ES, we also propose a machine learning based ES scheme by switching off the cell selected based on load prediction. How to apply AI into the 5G network is a new topic for 3GPP and future works might focus on potential solutions to smart energy saving based on AI and the corresponding 3GPP standard impacts on data collection and interface between NG-RAN nodes.

References

[1] 3GPP. Technical specification—energy efficiency of 5G release 16: TS28.310 V16.2.0 [S]. 2020
[2] ZTE. Consideration on inter-RAT energy saving: 3GPP R3-191453 [R]. 2019
[3] Qualcomm. Inter-system inter-RAT energy saving: 3GPP R3-204802 [R]. 2020
[4] 3GPP. Technical specification—S1 application protocol release 16: TS36.413 V16.2.0 [S]. 2020
[5] 3GPP. Technical specification—NG application protocol release 16: TS38.413 V16.3.0 [S]. 2020
[6] 3GPP. Technical specification—5G end to end key performance indicators release 16: TS 28.554 16.5.0 [S]. 2020
[7] GAO Y, CHEN J, LIU Z, et al. Machine learning based energy saving scheme in wireless access networks [C]//16th International Wireless Communications and Mobile Computing Conference (IWCMC). Limassol, Cyprus: IEEE, 2020: 1573–1578

Biographies

LIU Zhuang (liu.zhuang2@zte.com.cn) received the master's degree in computer science from Xidian University, China in 2003. He is currently a 5G senior 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.

GAO Yin received the master's degree in circuit and system from Xidian University, China in 2005. She has been engaged in the study of 4G/5G technology since 2005 and is currently a wireless expert and project manager at the R&D center of ZTE Corporation and the State Key Laboratory of Mobile Network and Mobile Multimedia, China. She has authored or co-authored about hundreds of proposals for 3GPP meetings and journal papers in wireless communications and has filed more than 200 patents. In August 2017, she was elected as 3GPP RAN3 Vice Chairman.

LI Dapeng received the master's degree in computer science from University of Electronic Science and Technology of China in 2003. He is currently a senior researcher at the R&D center of ZTE Corporation and mainly focuses on research and implementation of wireless access network systems.

CHEN Jiajun received the master's degree in electronics and communications engineering from Shanghai University, China in 2019. He has been a technology pre-research engineer at the R&D center of ZTE Corporation. His research interests include next-generation radio access network and deep learning.

HAN Jiren received the master's degree in wireless communication systems from University of Sheffield, UK in 2016. He is currently a technology pre-research engineer at the R&D center of ZTE Corporation. His research focuses on next-generation radio access networks.