Chinese Journal of Electronics
Vol.XX, No.X, Jan. 20XX

Real-World Wireless Network Modeling and Optimization: From Model/Data-Driven Perspective

LI Yang[1,2], ZHANG Shutao[3,1], REN Xiaohui[4], ZHU Jianhang[5], HUANG Jiajie[5], HE Pengcheng[6], SHEN Kaiming[3], YAO Zhiqiang[4], GONG Jie[5], CHANG Tsung-Hui[3,1], SHI Qingjiang[6,1], and LUO Zhi-Quan[3,1,2]

(1. *Shenzhen Research Institute of Big Data, Shenzhen 518172, China*)

(2. *Pengcheng Lab, Shenzhen 518055, China*)

(3. *School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen 518172, China*)

(4. *College of Mathematics and Computational Sciences, Xiangtan University, Xiangtan 411105, China*)

(5. *School of Computer Science and Engineering, Sun Yat-Sen University, Guangzhou 510275, China*)

(6. *School of Software Engineering, Tongji University, Shanghai 201804, China*)

Abstract — With the rapid development of the fifth-generation wireless communication systems, a profound revolution in terms of transmission capacity, energy-efficiency, reliability, latency, and connectivity is highly expected to support a new batch of industries and applications. To achieve this goal, wireless networks are becoming extremely dynamic, heterogeneous, and complex. The modeling and optimization for the performance of real-world wireless networks are extremely challenging due to the difficulty to predict the network performance as a function of network parameters, and the prohibitively huge number of parameters to optimize. The conventional network modeling and optimization approaches rely on drive test, trial-and-error, and engineering experience, which are labor intensive, error-prone, and far from optimal. On the other hand, while the research community has spent significant efforts in understanding the fundamental limits of radio channels and developing physical layer techniques to operate close to it, very little is known about the performance limits of wireless networks, where millions of radio channels interact with one another in complex manners. This paper reviews the very recent mathematical and learning based techniques for modeling and optimizing the performance of real-world wireless networks in five aspects, including channel modeling, user demand and traffic modeling, throughput modeling and prediction, network parameter optimization, and IRS empowered performance optimization, and also presents the corresponding notable performance gains.

Key words — Mathematical methods, Learning based methods, Network performance modeling, Network performance optimization.

I. Introduction

The performance of a real-world wireless network depends on not only the capabilities of its hardware (e.g., base stations, mobile handsets), but also how it is configured. A wireless network serving a metropolitan area consists of large numbers of base stations, each covering several cells and providing various communication services for users in this area. The service range, signal strength, antenna settings, and many other configurations of the base stations greatly impact the communication quality and user experience [1].

In real-word wireless networks, the surrounding environment (such as terrains, building distributions, crowd densities, user distributions and trajectories, and network traffic loads) of different base stations are also diverse. Therefore, millions of tunable parameters need to be optimized to improve the overall performance of the wireless networks. Only by setting them to match the local

*Manuscript Received June XX, 20XX; Accepted July XX, 20XX. This work was supported by the National Key R&D Program of China under grant 2022YFA1003900, the National Natural Science Foundation of China (No. 62101349, No. 62001411, No. 62171481, No. 62231019), the Special Support Program of Guangdong (No. 2019TQ05X150), and the Natural Science Foundation of Guangdong Province (No. 2021A1515011124).

 DOI:10.1049/cje.20XX.0X.0XX

radio environment and user traffic patterns can the network reach its optimal performance.

Traditional optimization methods rely on human knowledge. Professional network optimization engineers monitor and analyze the radio environment and log data of each base station, locate the problems, and propose the optimization solution based on the expert experience. However, in a real-world wireless network, the optimization for millions of parameters is extremely costly and time consuming, making the conventional approaches only suitable for very small-scale networks [2]. On the other hand, while significant efforts have been made to understand the fundamental limits of radio channels, it is still challenging to develop efficient methods to model and optimize the performance of the real-world wireless networks. What are the performance gains that can be brought by adjusting the parameters of a network? Can mathematical/learning models defeat the best network engineers and fully reap the capacity of a large-scale wireless network?

Effective approaches for modeling and optimizing the network performance are greatly needed. In particular, some fundamental problems are still open with many challenges, such as:

1) Lack of Historical Data: Most parameters in wireless networks are configured to be default values, and there is no diversified historical data;

2) Large Network Fluctuation: The fast-varying propagation environment and user behaviors make the network performance non-stationary and hard to predict;

3) Large Search Space: The number of network parameter combinations is huge, reaching sizes of $10^{2,000,000}$ for a metropolitan 5G wireless network;

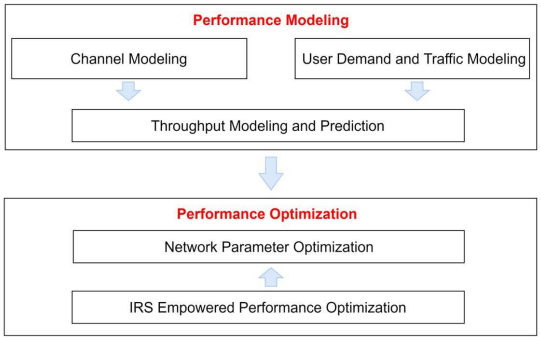


Fig. 1 The relations among the five aspects in performance modeling and optimization of real-world wireless networks.

4) Complex Interactions: Parameter configuration for a single cell affects its own performance but also that of the neighbor cells, inducing complex interactions.

Faced with these challenges, this paper reviews the very recent mathematical and learning based techniques, including various cutting-edge technologies such as deep learning, reinforcement learning, and black-box optimization to effectively improve the service quality and user experience of the real-world wireless networks. Specifically, we focus on five aspects in the performance modeling and optimization for real-word wireless networks, which consist of channel modeling, user demand and traffic modeling, throughput modeling and prediction, network parameter optimization, and an advanced intelligent reflecting surface (IRS) empowered performance optimization. The relations among the five aspects are illustrated in Fig. 1. Firstly, channel modeling provides the accurate estimate of the wireless environment. On the other hand, user demand and traffic modeling establishes a precise prediction for the dynamic data traffics. Based the the well established channel models, user demand and traffic models, the mapping function from the parameters in channels and data traffics to the overall performance of the wireless networks can be investigated through the throughput modeling and prediction. Once the mapping function is obtained, various parameters can be optimized to maximize the overall performance of the wireless networks. In particular, some advanced scenarios such IRS involve a lot of special continuous and/or discrete parameters. By judiciously designing these parameters, the overall performance can be much improved.

The rest of the paper is organized as follows. Section II reviews state-of-the-art channel modeling methods. Section III introduces the main traffic prediction and traffic models. In Section IV, we discuss how to achieve the throughput modeling and prediction through the cutting-edge mathematical and learning based technologies. After the performance modeling, we further move to the network parameter optimization in Section V, and introduce the optimization methods for an advanced IRS empowered scenario in Section VI. Finally, the conclusions are included in Section VII.

II. Channel Modeling

In this section, we introduce the 5G channel modeling methods and the channel models commonly used in existing network optimizers. Then, we propose our design of data-driven localized statistical channel modeling.

1. 5G channel modeling methods

According to the modeling methods, channel modeling can be classified into deterministic manner and stochastic manner. The deterministic channel models usually depend on the map information of environment, and describe the propagation paths by solving the Maxwell's equations or approximated propagation equations. As for the stochastic channel models, they characterize the channel parameters by using certain probability distributions,

and can be applied to various scenarios with relatively low accuracy. Generally speaking, the deterministic channel models cost much higher computational complexity than the stochastic channel models.

The new techniques and applications in 5G mobile communication systems have introduced new challenges to the channel modeling [3]. A wide frequency range and broad bandwidths of channel measurement need to be conducted for millimeter wave (mmWave) communication in 5G systems. Taking massive MIMO for example, due to the large dimension of antenna array, the distance between the transceiver and the clusters could be smaller than the Rayleigh distance, resulting in that the assumption of plane wavefront does not hold. Besides, the array non-stationarity would occur. The clusters may appear and disappear from the viewpoint of one antenna element to the next one, which means the path parameters such as power and delay can drift over different antennas. These two properties set new requirements to the channel modeling for massive MIMO. In [4], the spherical wavefront with an ellipsoid model is considered in the massive MIMO channel model. The appearance and disappearance of the clusters over the array is modeled by a birth-death process with cluster generation rate λ_G and cluster recombination rate λ_R.

2. Channel models in the network optimizer

As channel models are indispensable for system design and performance evaluation in the network optimizer, here we focus on two types of channel models that are exploited in the network optimizer:

1) Empirical path loss model. Coverage and capacity optimization (CCO) is a typical application of network optimization by tuning the tilt angle and azimuth angle of antenna array. To describe the channel characteristics, the simplified empirical path loss model (e.g., COST231-Hata model) can be exploited in the CCO. However, such ideal one-dimensional path loss model is not accurate enough and fails to embrace the actual complexity and randomness of practical 5G networks. The only one-dimensional path loss model needs to be extended into multi-dimensional channel model, because the massive MIMO and beamforming in the 5G networks yield high-resolution multi-path channels in the angular domain.

2) Neural network channel model. A neural network enabled wireless channel model framework is proposed in [7]. The input parameters are transmitter (Tx) and receiver (Rx) coordinates, Tx-Rx distance, and carrier frequency, while the output parameters are channel statistical properties, including root mean square (RMS) delay spread (DS), and RMS angle spread (AS). Datasets used to train and test the neural network are collected from both real channel measurements and a geometry based stochastic model (GBSM). Different datasets from different scenarios will show different channel statistics.

3. Localized statistical channel modeling

We propose a novel data-driven localized statistical channel modeling (LSCM) for the network optimizer, which is capable of sensing the physical geographical structures of the targeted cellular environment [5]. The proposed channel modeling solely relies on the reference signal receiving power (RSRP) of the user equipment, unlike the traditional methods which use full channel impulse response matrices. The key is to build the relationship between the RSRP and the channel's angular power spectrum. Based on it, we formulate the task of channel modeling as a sparse recovery problem where the non-zero entries of the spare vector indicate the channel paths' powers and angles of departure (AoD). By taking advantage of the spatial consistency of channel, LSCM is able to be constructed in a manner of multiple grids [6].

Similar to the 3GPP's technical report [8], we focus on the tilt AoD, azimuth AoD, and channel gain from the base station to the user equipment with one receive antenna, ignoring the arrival angle and delay. Suppose the uniform rectangular array of the base station contains $N_T = N_x \times N_y$ antennas. The CIR of antenna (x, y) from the base station to the user equipment is given as

$$h_{x,y}(t) = \sum_{i=1}^{N_V}\sum_{j=1}^{N_H}\sqrt{\alpha_{i,j}(t)} \times g_{i,j} \times e^{-j2\pi\frac{d_x x}{\lambda}\cos\theta_i \sin\varphi_j} \times e^{-j2\pi\frac{d_y y}{\lambda}\sin\theta_i} \times e^{-j\omega_{i,j}(t)-j\omega_{x,y}(t)}. \quad (1)$$

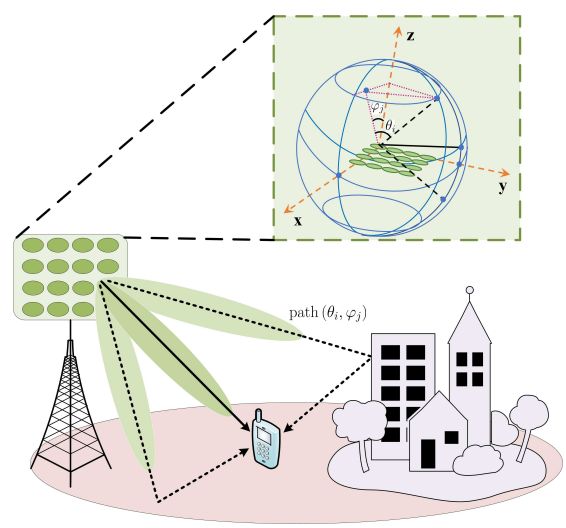


Fig. 2 Angular discretization of the channel model in the downlink.

We consider the channel modeling of the massive MIMO downlink communication system with beamforming, where the signal is transmitted from a base station to the user equipment through several propagation paths, as shown in Fig. 2. The tilt and azimuth AoD over the free space are discretized into N_V and N_H angles, respectively. If there does not exist a path in the angles (θ_i, φ_j), the corresponding channel gain $\alpha_{i,j}(t)$ is zero, otherwise $\alpha_{i,j}(t) > 0$ for the path (θ_i, φ_j) in the channel. It is readily known that the channel gain $\alpha_{i,j}(t)$ is sparse, and contains multiple propagation paths in the angular domain.

The channel gain $\alpha_{i,j}(t)$ consists of path loss and the shadowing effect. Path loss is determined by the physical environment (distance, carrier frequency, buildings) which is assumed to be relatively static, while shadowing is caused by the obstacles and is usually modeled as the log-normal distribution. Thus, we assume $\alpha_{i,j}(t)$ follows the log-normal distribution, with its mean representing path loss and its covariance representing the shadowing effect. Notcie that $\omega_{i,j}(t)$ is the random phase error between different angles caused by the reflection, diffraction, and scattering effect of electromagnetic waves, while $\omega_{x,y}(t)$ is the random phase error of different antennas caused by the imperfect hardware of the antenna array. Inspired by [8], we assume $\omega_{i,j}(t)$ follows the uniform distribution between $-\pi$ and π, and $\omega_{x,y}(t)$ follows the Gaussian distribution with zero mean and variance σ^2.

The channel measurement of our proposed LSCM is the RSRP measured from multiple beams. In the downlink of 5G cellular systems, the RSRP can be measured from synchronization signals block (SSB) beams or channel state information-reference signal (CSI-RS) beams. Denote the precoding matrix of the k_{th} beam as $\boldsymbol{W}^{(m)} \in \mathbb{C}^{N_x \times N_y}$, whose size is the same as that of the antenna array. The entry of $\boldsymbol{W}^{(m)}$ is $\left(w_{x,y}^{(m)}\right) = \left(e^{j\phi_{x,y}^{(m)}}\right)$. Denote the CIR matrix from the base station to the user equipment as $\boldsymbol{H} \in \mathbb{C}^{N_x \times N_y}$, where the entry of $\boldsymbol{H}$ is $h_{x,y}(t)$ in Eq.(1). The RSRP of the m_{th} beam at time t is defined as

$$rsrp_m(t) = P\left|\sum_{x,y} h_{x,y}(t) w_{x,y}^{(m)}\right|^2 = P\left|\operatorname{tr}\left(\boldsymbol{H}^T \boldsymbol{W}^{(m)}\right)\right|^2, \tag{2}$$

where P denotes the transmit power. The quality of the channel can be represented through $rsrp_m(t)$, and a larger value of $rsrp_m(t)$ reflects better channel quality.

Notice that $rsrp_m(t)$ is a random variable, since $h_{x,y}(t)$ contains three independent random variables, i.e., $\omega_{x,y}(t)$, $\omega_{i,j}(t)$, and $\alpha_{i,j}(t)$. We want to show the relationship between the first-order statistics of $rsrp_m(t)$ and $\alpha_{i,j}(t)$, i.e., their expectation $\mathbb{E}\left(rsrp_m(t)\right)$ and $\mathbb{E}\left(\alpha_{i,j}(t)\right)$, implying the statistical relationship of the RSRP and angular information. Suppose the expectation of $rsrp_m(t)$ is $\mathrm{RSRP}_m \triangleq \mathbb{E}\left(rsrp_m(t)\right)$, and the expectation of channel gain is $X_{i,j} \triangleq \mathbb{E}\left(\alpha_{i,j}(t)\right)$, then we have

$$\mathrm{RSRP}_m = \sum_{i=1}^{N_V}\sum_{j=1}^{N_H} \mathrm{A}_{i,j}^{(m)} X_{i,j}, \tag{3}$$

where

$$\mathrm{A}_{i,j}^{(m)} \triangleq P g_{i,j}^2 \Bigg(N_x N_y \left(1 - e^{-\sigma^2}\right) + e^{-\sigma^2} \sum_{x,y}\sum_{x',y'} \cos\left(\psi_{i,j,x,y}^{(m)} - \psi_{i,j,x',y'}^{(m)}\right) \Bigg).$$

The statistical relationship between the expectation of RSRP measurements RSRP_m and the first-order statistics of channel gain $X_{i,j}$ is connected by the coefficient $\mathrm{A}_{i,j}^{(m)}$. The angular power spectrum (APS) can be expressed by $X_{i,j}$, where the subscripts i and j indicate the tilt angle θ_i and azimuth angle φ_j, respectively. According to this finding, it is possible to extract the angular information from beam-wise RSRP if we can infer $X_{i,j}$ from RSRP_m.

The channel statistics of APS from the proposed channel model are consistent with those of the true propagation environment, and can provide enough information for the network optimization simulator to generate channels similar to reality.

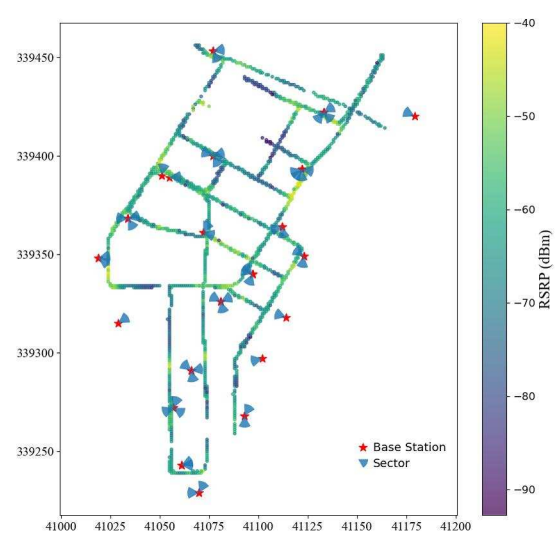


Fig. 3 RSRP distribution in real-world drive testing.

As shown in Fig. 3, it is an example of collecting the real-world RSRP measurement by using the drive-test in the street of Chengdu city, China. The carrier frequency of the cellular network is 2.6 GHz with 100 MHz bandwidth. The number of antennas in the base station is

32. Compared with the traditional network optimization which needs many rounds of drive testing, the proposed LSCM only need one round of drive testing, and thus, saving a lot of money and enabling the assessment of network optimization. Depending on the velocity of the driving car, dozens of samples are collected when passing through a 10×10 square meters grid in the interval of several seconds. The samples of RSRP measurements are averaged to alleviate the effect of fast fading. If the number of samples is smaller, then the noise effect will be dominant which leads to a higher mean absolute error.

III. User Demand and Traffic Modeling

Optimizing the performance of a real-world wireless network is extremely challenging because of the difficulty to predict the network performance as a function of network parameters, and the prohibitively large problem size. Therefore,traffic prediction and modeling plays a fundamental role in the overall performance optimization of wireless networks.

Industry and industry communication field attach great importance to traffic prediction. Huawei predicted that mobile network traffic will increase a hundredfold in 2030 during the Global Mobile Broadband Forum 2021 (MBBF2021), saying that wireless networks are an important pillar for moving towards an intelligent world. Ericsson not only predicted that the traffic in 2026 would increase by 4.5 times compared with 2020, but also proposed that there would be 3.5 billion 5G users in the world by 2026. On the other hand, Nokia Bell Labs made predictions on the infrastructure and interconnection equipment in the future. Among them, 4G expenditure will account for 80% of the infrastructure expenditure of operators in the next few years. It is estimated that there will be 50 billion interconnection devices in the world by 2025. Prediction of network traffic is being rapidly advanced by the world's leading research institutions.

Therefore, how to predict the traffic with low cost and high accuracy under the rapidly changing propagation environment, user environment as well as huge number of network parameter combinations. It has become an urgent international open problem in the industry, with great market demand and academic value. It is also a very attractive research hotspot.

1 Traffic models.

Our goal is to accurately characterize and predict the stochastic behaviour of the real world 4G/5G wireless network. High precision traffic prediction method has become a research hotspot in the industry and academia. In order to improve the prediction accuracy, great efforts have been made in model establishment and prediction methods, and fruitful results have also been achieved, which has promoted the progress of traffic prediction and parameter estimation.

The prediction models used in the existing research are roughly categorized into the following three types:

1) Statistical models

At present, the statistical models used in the research are mainly time series, probability estimation and particle filter models. Some representative models include Holt Winters (HW) [9] and ARIMA[10]. ARIMA(p, d, q) model is an extension of ARMA(p, q) model, which can be expressed as:

$$(1 - \sum_{i=1}^{p} \lambda_i B^i)(1 - B)^d y_i = (1 - \sum_{i=1}^{q} \theta_i B^i)\epsilon_t, \quad (4)$$

where B is the lag operator, $\sum_{i=1}^{p} \lambda_i B^i$ and $\sum_{i=1}^{q} \theta_i B^i$ are autoregressive coefficient polynomial and moving average coefficient polynomial respectively. ϵ_t is zero mean white noise sequence, $d \in \mathbb{Z}, d > 0$.

Network slicing is the key technology of the fifth generation (5G) mobile network. In order to meet the demand of network operators for dynamic deployment of network slicing, scholars proposed to use HW model [9] to reduce the cost of network operators by 15% and energy consumption by 13%. The utilization rate of resources also increased by more than 6 per cent. Clemente et al. [11] used the method of combining naive Bayesian classifier and Holt Winter to accurately predict cell traffic, and limited the prediction RMSE to 5%.

For short-term flow forecast represented by seasonal problems, Wang [10] et. al used MSE and MAE to evaluate the performance of the proposed ARMA model, and proved that short-term prediction has high prediction accuracy. The literature [12] proved that the prediction accuracy of conditional probability estimation model is higher, and the daily seasonal model error is 9.92%.

ARIMA model and HW model are univariate time series models. In addition to these two models, there are ES method [13] and particle filter method based on sequence Monte Carlo (SMC) [14] for traffic demand prediction. As the classical method, the prediction accuracy of statistical model is much less than machine learning and deep learning model since most of them are based on the linear relationship between input values and output values. Considering statistical models are not suitable for large-scale problems and are difficult to be trained in parallel, scalability is also its drawback.

2) Machine learning models

Common machine learning prediction models are mainly tree based, such as random forest(RF) [15] and

LightGBM [16], and others are based on Gaussian process or other mechanisms. Decision trees are the most basic concepts in tree based models, which can be used to solve classification or regression problems. Predict y_i by giving x_i to get the objective function:

$$Obj(\Theta) = L(\theta) + \Omega(\Theta). \tag{5}$$

where L represents the training error function, and Ω represents the regularization term to control the complexity of the model and prevent over fitting. For logical regression, the most commonly used loss function is the Logistic function, which is expressed as:

$$L(\theta) = \sum_i [y_i ln(1 + e^{-\hat{y}_i}) + (1 - y_i) ln(1 + e^{\hat{y}_i})]. \tag{6}$$

The regularization is expressed as:

$$\Omega(\Theta) = \theta_j (1 - \alpha \frac{\lambda}{m}) - \frac{\alpha}{m} \sum_i^m [(h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}]. \tag{7}$$

where λ is constraint multiplier and $h_\theta(x)$ represents the prediction function.

In the literature [17], Prophet model and GPR model are combined to predict single cell traffic, and RMSE, MAE and MAPE are used to verify the superiority of prediction accuracy and explain the inherent space-time correlation of traffic data. In order to realize ultra reliable low delay communication (URLLC) in large-scale machine communication networks, the Tree based ML model was proposed by Weerasinghe et. al [18].

In order to improve users' life comfort, Abozariba et. al [19] proposed ML model to accurately predict users' traffic demand. Performance index analysis shows that the prediction results can provide higher resource utilization. In addition, in order to improve the service quality of smart cities and understand the distribution of service demand in time and space, a spatio-temporal Bayesian hierarchical learning method [20] is used to learn and predict the distribution of MEC resource demand in time and space, and the resource allocation efficiency predicted by the model is higher.

Compared with deep learning models, these models have the characteristics of "shallow" structure. Machine learning models usually perform better than statistical models, but only in a few cases can obtain better prediction results than deep learning models.

3) Deep learning models

The main advantages of deep learning model include high precision prediction performance and high scalability. The deep learning models used in the study include Convolutional Neural Network (CNN)[21]-[24], Graph Convolutional Network (GCN)[25]-[32], Long Short-Term Memory (LSTM)[33]-[45]. The GCN model is expressed as:

$$f(H^{(l)}, A) = \sigma(\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}), \tag{8}$$

where input layer is $H(0) = X$, output layer is $H(L) = Z$, and L is the number of layers. W and σ represent linear transformation and nonlinear transformation respectively.

Due to the wide coverage of deep neural network, it has been used for deep learning and has achieved great success in a series of prediction problems in the past decade, such as intelligent transportation system [26], urban traffic prediction [36] and cellular network prediction [45]. In addition, there are Generative Adversarial Network (GAN) [28], Deep Neural Network (DNN) [36], Attention-based Periodic-Temporal neural Network (APTN) [37], Multi-Stage Attention Spatial-Temporal Graph Networks (MASTGN) [39], Neural Architecture Search (NAS) [46].

CNN can extract local spatial features of data, but it needs a large sample size, high computational complexity, high redundancy, and no memory. Therefore, it is suitable for nonlinear time series prediction with large sample size, high prediction accuracy, and time delay insensitivity. RNN can flexibly capture the time dependence of data, and may also cause gradient disappearance or gradient explosion. It is suitable for dynamic short correlation time series prediction. DNN can learn deep nonlinear feature transformation, but it can not use historical information to assist the current task. LSTM can better deal with the influence of large time scale data, but its convergence speed is slow, the parameters cannot be directly determined, and it is easy to fall into local optimization. It is suitable for time series prediction with large number of samples and long-term dependence. Considering the structural correlation between nodes, GCN is very effective in processing graph data and has over smoothing problem. Because GCN has the function of a low-pass filter, the characteristics tend to be the same after multiple iterations. It is suitable for spatial prediction scenarios. GCN-GAN can predict plateau, single engine, and double engine traffic. The model is suitable for large-scale burst traffic prediction, and can predict three types of traffic.

In particular, GCNs and LSTM are employed to model the spatial and temporal aspects [32]. To get a better understandings of deep learning model, we discuss several works in Table. 1.

Table. 1 List of papers on deep learning traffic prediction

Reference	Year	Spatial/Temporal	Proposed model	Application	Short/Long Term	Evaluation metrics
21	2020	Both	MDL, CNN	Traffic Flow	Both	RMSE, MAE
22	2020	Spatial	CNN, GCN	Cell traffic	Long-term	MAE, MSE
23	2021	Both	CNN	ITS	short-term	RMSE, MAPE
24	2017	N/A	CNN, LSTM	Mobile traffic	Both	RMSE, MAE, MA
25	2020	Both	LSTM, GCN	Intelligent Route Planning	Both	RMSE, MAE
26	2022	Both	LSTM, GCN	ITS	Both	MAE,RMSE,MAPE
27	2020	N/A	GCN-GAN	Burst Events	Long-term	MSE
28	2020	N/A	ML, GCN	Intelligent Transportation Systems	Short-term	N/A
29	2021	Both	LSTM, MASTGN, GCN	ITS	Both	RMSE,MAE,MAPE
30	2020	Temporal	T-GCN	ITS	Short-term	RSME,MAE
31	2021	N/A	GCN-GAN	Elastic Optical Networks	Short-term	MSE
32	2018	Both	LSTM, GCN	Cellular Networks	Both	MAE
33	2021	N/A	LSTM, DL	Transportation Planning	Long-term	TSMAPE,TRMSE
34	2019	N/A	LSTM	Optical Networks	Both	N/A
35	2017	N/A	LSTM	Big Date Oriented Networks	Short-term	N/A
36	2022	N/A	DNN, LSTM	Road Network	Both	N/A
37	2020	Both	LSTM, APTN	ITS	Both	RMSE,MAE,MASE
38	2021	N/A	LSTM	Cellular Network	Both	NRMSE
39	2021	Temporal	LSTM	Cellular Network	Both	RMSE
40	2021	Spatial	LSTM	Cellular Traffic	Both	RMSE,MAE
41	2020	N/A	LSTM	Mobile Traffic	Short-time	RMSE
42	2020	N/A	LSTM,ANFIS	Network Traffic	Both	MSE,RMSE,MAE
43	2019*a*	N/A	LSTM,ARIMA	Cellular Traffic	Both	RMSE
44	2019*b*	N/A	LSTM,ARIMA	Cellular Traffic	Both	RMSE

2. Prediction problems

1) Predicting network behaviours at the millisecond timescale is both technically infeasible and practically unnecessary, considering the stochasticity and non-stationarity of networks caused by the rapidly changing propagation environment and user behaviour;

2) How to clean, store, analyze and calibrate the prediction model in the case of the gigantic problem size and massive data sets of network parameters;

3) How to provide an accurate, efficient, yet inexpensive assessment of network performance for any network configuration.

3. Discussion

In the past decade, researchers have designed many network traffic prediction algorithms based on deep learning and made great progress. Network traffic prediction has made a great breakthrough from the traditional method to the application of deep learning technology. However, the current network traffic prediction research is faced with the problems of few public data sets, privacy of traffic data, and limited disclosure.

We forecast the temporal and spatial traffic distribution in the beam-space. It is the basis for characterizing the time/frequency/spatial resource utilization of each cell, which in turn, helps quantifying the perceived interference per beam-space cluster.

To deal with the intrinsic difficulty of the traffic forecasting task in cellular networks, GCN is first used to leverage the spatial correlation of the traffic in neighboring beam-space clusters. The output of the GCN is then used by a LSTM neural network, which captures the temporal behaviour of the data traffic, and enables accurate forecasting. We present the following data-driven spatial and temporal traffic forecast diagram in Fig. 4.

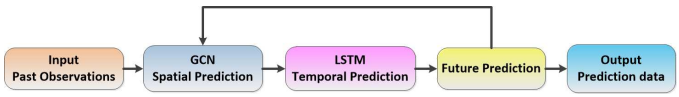


Fig. 4 Forecast diagram.

Therefore, the influence of spatial and temporal factors on the prediction results should be fully considered when making traffic prediction, and GCN and LSTM are combined to predict. In addition, traffic modeling should be human-centric, fully consider the needs of users, and conduct traffic modeling under the conditions of high security, confidentiality and privacy.

IV. Throughput Modeling and Prediction

With the development of machine learning, deep neural networks are widely used in wireless communication systems for modeling and prediction. Neural networks have powerful data fitting capability suitable for complex multi-factor communication scenarios. The downlink Internet Protocol (IP) throughput, defined as the payload data volume on IP level per elapsed time unit on the U-u interface, is an important performance metric for the quality of service experienced by the end user. In this

section, we propose a deep neural network-based modeling approach to predict the downlink IP throughput. Real-trace data of cellular systems in South Africa, i.e., user-uploaded data including physical layer measurement, user scheduling information, user throughput and so on, are used for model training and testing. The experimental results show that our proposed model performs well for downlink IP throughput prediction.

The downlink IP throughput, defined as the payload data volume on IP level per elapsed time unit on the U-u interface, is an important performance metric for the quality of service experienced by the end user. To make sure that only impacts from the RAN is included in this measurement, time units to be included in "elapsed time unit on the Uu interface" shall only be the ones where there is data in the buffer to be transmitted e.g., in application data flows such as a web session, there are times when there is no data to transmit by the eNodeB due to bursty traffic pattern, then this "eNodeB idle time" shall not be included in "elapsed time unit on the Uu interface".

To achieve a throughput measurement that is independent of file size it is important to remove the samples where one time and temperature indicator (TTI) on the radio interface is not utilized.

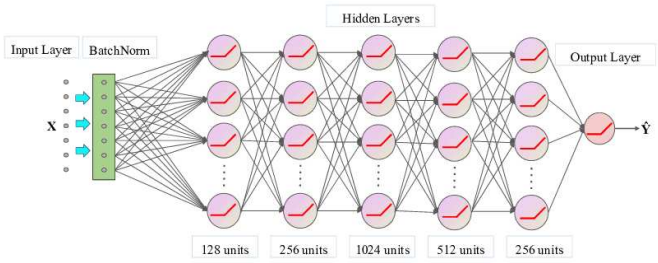


Fig. 5 The net architecture of proposed model.

1. Dataset introduction

By enabling cell history record (CHR) subscription at the 5G base station side, we can get the information reported by the user base station measurements in event blocks. The recording rule is either periodically triggered (e.g. 5s) or event triggered (user switching between cells). A total of seven event blocks exist in our dataset, the contents of which are shown in Table. 2. The first four events are user measurements, where all users in the cell report cumulative measurements every 5s. The fifth event file is the base station measurement file, where the base station counts the beam IDs of all beams uplinked in a time period as well as the RSRP. The last two events are records of users cutting in and out of the base station. These two event blocks are event triggered and therefore much smaller in volume than the other five event blocks. Our dataset includes measurements from 3 base stations for 31 consecutive days.

2. Neural network-based rate prediction model

We propose a data-driven network rate prediction model. The model uses user and base station measurements collected from real environment as input and uses the real measured experience rate as output. Thanks to the huge amount of data collected by the network operators and the powerful data fitting ability of deep learning, many channel modeling studies combining deep learning have emerged. Unlike traditional methods, the neural network-based approach can fully explore the deep coupling relationship among a large number of measurement metrics, and can facilitate the analysis of the impact of network parameters on communication performance. On the one hand, IP throughput is different from traditional rate metrics and it is not clear which parameters are of higher impact on this metric. The small time scale feature makes IP throughput prediction quite challenging. These two factors require us to fully explore the relationship between different measurements and IP throughput from massive data, so we choose the NN-based prediction model.

Table. 2 Dataset event block list

5G CHR event block list	Description
PERIOD INTRA FREQ MEASUREMENT	Downlink Service Cell SSB RSRP
PERIOD PRIVATE UE MEASUREMENT	User scheduling data, including MCS, RNK, etc.
PERIOD PRIVATE THROUGHPUT MEASUREMENT	User flow measurement data
PERIOD PRIVATE UE MIMO MEASUREMENT	User MIMO related data
PERIOD PRIVATE UL BEAM INFO	BS measurement uplink beam level SRS RSRP
PRIVATE HO IN	Cell switch-in events
PRIVATE INTRA RAT HO OUT	Cell switch-out events

1) Input measurement data. After the CHR is turned at the base station we can get the complete daily measurement data for all cells at the base station. We match the data with user IDs by time index to obtain the

complete data. Currently, the basic network measurement data required for our prediction network include:

a) Signal power per beam: 5G BSs can transmit and receive their reference signals through multiple radiating beams. BSs and UEs typically record measurements for the beam with the maximum signal power. Such measurements include, among others, RSRP and SINR estimates for the synchronization signal block (SSB), the channel state information reference signal (CSI-RS), and the detection reference signal (SRS).

b) User connection information: For each UE connection, BS also records its experience by storing statistical information, such as the amount of data transmitted, resources allocated, class, modulation and coding schemes (MCSs), packet failures, retransmissions, and data rates.

c) Base station statistics: The user experience rate is also related to the bandwidth resources that users can allocate during the count transmission time, which is related to the average number of activated users in the cell, thus the higher the number of users, the lower the user experience rate. We use not only the average number of activated users over a single measurement period, but also the number of tied activated users over a larger time range (e.g., over 10 and 30 minutes). This information allows capturing more temporal information, such as user behavior habits throughout the day.

2) Framework for the Model. We use a multilayer neural network to build the rate prediction model, and the network structure is shown in Fig. 5. The network contains a total of 5 hidden layers, and the number of neurons in each hidden layer is indicated in Fig. 5. We use Rectified Linear Unit (ReLU) as the activation function to increase the nonlinearity. In addition, we add a BathNorm layer between the input layer and the first hidden layer to normalize the data samples. The loss function is Mean Absolute Percentage Error (MAPE).

3. Experimental results

1) Data filtering. The experience rate is calculated by removing the tail packet traffic and tail packet count transmission time. Predicting the rate is a difficult task for data with a relatively large percentage of tail packet traffic and tail packet count transmission time. Therefore, we consider predicting the data with a larger percentage first. We set two fixed thresholds a and b to filter the data set, and select the sample data with the tail packet traffic ratio exceeding a and the tail packet transmission time ratio exceeding b for prediction.

In Table. 3, we use different thresholds ($a = b$) to filter the data, and count the remaining percentage of the filtered data and the remaining percentage of data transmission traffic. It can be seen that, according to this method, even a large amount of data is filtered out, the proportion of the remaining traffic is still very large, which is acceptable in practical engineering, so we will use the threshold $a = b = 0.2$ for experiments.

Table. 3 Statistical of different filtering thresholds

Threshold	Data items remaining	Traffic remaining
0.1	54%	94%
0.2	39%	86%
0.3	30%	77%
0.37	25%	70%

2) Results. We use MAPE as a test indicator. Set the filtering threshold $a = b = 0.2$. The results obtained from the experiments on base stations $2343, 3434, 3955$, and the three base stations combined are shown in Table. 4.

Table. 4 Multi-base station multi-day experiment

Base Station	Days	Test MAPE
3955	1∼20	21.50%
2343	1∼20	25.72%
3434	1∼20	25.00%
2343, 3434, 3955	1∼20	23.50%

We can see that the 3955 base station has the best performance. Its test MAPE can reach 21.5%, while the other two base stations perform slightly worse, but can also reach about 25%. The result of combining the three base stations together is about the average of the respective results of the three base stations. In general, the MAPE between the predicted value and the actual value can reach below 25% which meets the needs of the industry.

The following experiments are aimed at the verification of generalization performance.

First, we verify whether it is generalizable for different days. For the three base stations respectively, the data of the first 20 consecutive days is selected as the training data, and then the trained model is tested with the data of the first 20 days and the last 10 days, and the results are shown in Table. 5.

Table. 5 Generalization experiment of day

Base station	Train days	Test days	Test MAPE
3955	1∼20	1∼20	**21.50%**
	1∼20	21∼30	23.13%
2343	1∼20	1∼20	**25.72%**
	1∼20	21∼30	37.14%
3434	1∼20	1∼20	**25.00%**
	1∼20	21∼30	30.83%

We can see that the performance of the base station 3955 is still the best. The ratio of MAPE tested on different days increase by about 2%, but the performance of the other two base stations is not ideal. The performance of base station 3434 increases compared to the original. At first glance, the trained model does not generalize for simply dividing the days continuously.

Table. 6 Generalization experiment of day (cycle on week)

Base Station	Train days	Test days	Test MAPE
3955	3, 10, 17	3, 10, 17	***24.17%***
	3, 10, 17	24	26.41%
2343	3, 10, 17	3, 10, 17	***23.99%***
	3, 10, 17	24	45.02%
3434	3, 10, 17	3, 10, 17	***29.91%***
	3, 10, 17	24	38.94%
3955	6, 13, 20	6, 13, 20	***18.68%***
	6, 13, 20	27	24.80%
2343	6, 13, 20	6, 13, 20	***25.05%***
	6, 13, 20	27	34.84%
3434	6, 13, 20	6, 13, 20	***23.67%***
	6, 13, 20	27	32.00%

Take into account that traffic may have weekly periodicity as traffic usage will vary between weekdays and weekends. We divide the days into weeks, selected Wednesday (day 3, 10, 17, 24) and Sunday (day 6, 13, 20, 27) for the experiment. From Table. 6, it can be seen that whether it is Wednesday or Sunday, the MAPE tested by each base station on different days is about 10% higher than that on the original day, which is not much different from the effect of continuous day division. That is to say, the model we trained based on the provided measured data does not have generalization to days.

Table. 7 Generalization experiment of base station

Train Base Station	Test Base Station	Days	Test MAPE
3955	3955	1∼20	**21.50%**
	2343	1∼20	40.45%
	3434	1∼20	56.28%
2343	2343	1∼20	**25.72%**
	3955	1∼20	39.49%
	3434	1∼20	50.64%
3434	3434	1∼20	**25.00%**
	3955	1∼20	35.83%
	2343	1∼20	40.72%

Next, we verify its generalizability for different base stations. The first 20 days are selected for the range of days, and the data of the three base stations are cross-validated respectively. From Table. 7, it can be seen that the effect of testing the model with different base stations is far worse than the test results of the same base station, which shows that the differences between different base stations may be huge due to differences in geographical location and number of users and so on. The model trained by one base station cannot be used to predict the experience rate of other base stations.

4. Discussion

In this section, we combine machine learning and wireless communication systems modeling and prediction, to predict downlink IP throughput in cellular networks with deep neural networks. Our proposed method models the real-trace data of the South African cellular system, and reaches the MAPE between the predicted downlink IP throughput and the measured value less than 25%, indicating the effectiveness of our method. In addition, our method filters most of the data, which are all small packets. Although filtering out small packets has little impact on the data transmission traffic, it occupies for the majority of the data so its value cannot be ignored. It is difficult to predict small packets with the original method.

V. Network Parameter Optimization

As mentioned before, the performance of a real-world network depends not only on the capabilities of its hardware, but also on how it is configured. Thus, careful network optimization is exceedingly desirable to ensure efficient radio resource management and high quality of service (QoS). Traditional optimization methods, which are heavily dependent upon the experience of professional network engineers, are often costly and time-consuming. In light of this, a self-organizing network (SON) has been proposed as a promising solution for increasing the network's efficiency and, hence, for maintaining and/or increasing the network's QoS while saving capital and operational costs at the same time. Various methods for network optimization have been proposed in existing literature which can be categorized into three classes, i.e. optimization-based methods, learning-based methods, and meta-heuristic methods.

1. Optimization-based methods

The computation of the objective function is usually a time-consuming and non-differentiable process due to the complex relationships between network parameters and key performance indicators (KPIs). Therefore, many

existing approaches are zero-order methods which only rely on the computation of the objective values. Niemela et al [47] studied the impact of antenna downtilt on the performance of cellular WCDMA network by utilizing a radio network planning tool. The optimal downtilt angles of antennas can be obtained based on solid simulation experiments by varying the downtilt angles inside a given range. Since the objective function values for all the combinations of the tuning of parameters are obtainable, the exhaustive method is supposed to always reach the optimal in theory. However, it is usually unfeasible because the solution space is too large. To reduce the time complexity, rule-based algorithm [48], greedy algorithm [49], Taguchi's method [50], and Nelder-Mead approach[51] are adopted to search the solution space to find a near-optimal solution. Line search algorithms are also proposed, which choose a search direction and then find a better solution along that direction iteratively. The coordinates are typically selected as the search directions, which is referred to as Coordinate Descent (CD) [52]-[54]. These search methods can efficiently reach the optimal or a near-optimal point in small-scale network with accurate network model, but are not applicable in large-scale network since the search space grows exponentially with the number of parameters.

In coping with large-scale network parameter optimization, line search methods in the optimal direction, i.e., the gradient direction or its approximate version, are often better than methods in the random or preordered coordinate directions. However, the objective function, especially the coverage, is usually discrete-valued and thus non-differentiable. One popular approach is to convert the non-differentiable objective function to a smoothed approximation to make the gradient-driven methods or other convex optimization approaches feasible [55],[56]. The approximation is usually based on simplified path loss model and antenna radiation pattern proposed by 3GPP. In addition to the fact that the performance of these methods depends largely on the accuracy of the approximation model, they need to be further validated in real scenarios. Another approach is to estimate the gradient by applying subtle changes to each network parameter [57],[58], which can be carried out on accurate network simulators. However, this kind of gradient estimator is also time-consuming and inefficient in large scale network. Li et al. [5] proposed a zeroth-order continuation method which is able to estimate the gradient with only two objective calculations. However, the gradient descent over the $[0, 1]$-valued coverage ratio is severely inefficient and the convergence of the algorithm can only be guaranteed under very strict assumptions. In addition to these gradient-driven methods, Engles et al. [59] formulated the CCO problem as a mixed-integer linear program (MILP) by introducing a low-complexity interference approximation model. They utilized state-of-the-art MILP solvers such as CPLEX or Gurobi Optimizer to compute (optimal) solutions. Partov et al. [60] reformulated the non-convex utility fairness problem in a convex manner by proposing a linear approximation of the antenna gain model, and solved it using a primal-dual approach.

In general, most of the optimization-based algorithms are designed to be executed offline, based on the assumption that factors such as the physical environment and user distribution are constant during the optimization process, or based on statistical network environments. Only the network configuration that is the solution of the offline optimization is then applied to the real network. These methods are usually highly effective, but require accurate modeling of the network to be optimized, i.e., the locations of all users need to be known and the received power of users with different parameter settings needs to be precisely evaluated.

2. Learning-based methods

Unlike optimization-based methods that require a large amount of prior knowledge to accurately model the system, the learning-based methods have the advantage that, in general, no prior knowledge or only sparse prior knowledge about the system's behavior is required. Moreover, learning-based approaches are often used for online network optimization, which makes it necessary to balance not only the various KPIs, but also the immediate gains and long-term performance. Therefore, many studies resort to multi-armed bandit (MAB) theory to address the resulting exploration and exploitation trade-offs. [61]-[63]. Dhahri et al. [61] proposed a novel multi-player MAB framework to model the trade-off in the CCO problem and a Pareto search framework to deal with the multi-objective optimization. In order to accelerate the convergence to the optimal selection, Shen et al. [62] proposed a novel multi-armed bandit model called generalized global bandit, which allows for the modeling of similarities across arms. They developed a series of greedy algorithms to achieve the optimal trade-off between sufficient small cell coverage and limited macro-leakage without prior knowledge of the deployment environment. In order to exploit the prior knowledge of the system, Wang et al [63] proposed the MAB algorithm with Bayesian principle on the small base station transmit power allocation problem. They incorporated performance correlations between similar power values and considered a power switching penalty to discourage frequent variations.

In addition to MAB, reinforcement learning is also widely used in online network optimization [64]-[68]. Dandarov et al. [64] proposed an RL approach awarded by the sum data rate normalized to the sector capacity and the number of satisfied users normalized to the potential total number of served users to address CCO problem. The network can predict the optimal antenna tilts for a particular user distribution. But the employment of explore-then-commit algorithm makes it highly inefficient for large-scale networks. The authors of [65] treat the coverage and capacity objectives as black-box functions, with no analytical formula and no gradient observations. They identified the set of Pareto optimal solutions through Bayesian optimization (BO) and the deep deterministic policy gradient algorithm (DDPG). These methods are still difficult to adapt to large-scale networks because the search space and data volume increase dramatically with the number of parameters. Therefore, Bouton et al. [66] proposed a coordinated RL approach modeling cellular networks as coordination graphs. Message passing and parameter sharing across the graph edges enabled the coordinated RL to operate in distributed manner and scale to large-scale networks with more than two hundred agents. Naderializadeh et al. [67] proposed a distributed RL algorithm for resource management and interference mitigation in wireless networks. In this framework, agents were able to make decisions simultaneously in a distributed manner and the DNN structure did not vary with the actual size of the wireless network, which made the algorithm scalable to large-scale networks. In addition, safety issues are also an important factor that has limited the real-world deployment of RL methods for network optimization. The primary source of unsafety in RL methods can be imputed to the agent;s exploration, resulting in undesired network performance degradation. Vannella et al. [68] formulated the remote electrical tilt (RET) optimization problem in the safe reinforcement learning (SRL) framework. The SRL aimed at solving RL problems in which a minimum performance level must be guaranteed during learning and deployment and thus undesired network performance degradation was avoided.

In order to address the curse-of-dimensionality, hybrid approaches, known as fuzzy reinforcement learning (FRL) framework, that combine Fuzzy Logic (FL) and RL have also been developed [69]-[73]. The fundamental idea is to address the curse-of-dimensionality by utilizing FL to encode a threshold-based discrete state space and then utilize it in RL algorithms. Razavi et al. [69] proposed an FRL method in which an FL module models the intrinsic uncertainty of the states in cellular networks, and then uses a Q-learning approach based on this encoded state. In [70], the author proposed a fuzzy Q-Learning (FQL) algorithm for self-optimization of the tilt angle in an LTE network scenario, which can operate in a fully distributed, asynchronous and autonomous fashion without any need for a priori information for the network conditions or any human interventions. Fan et al. [71] used a fuzzy neural network architecture in a cooperative Q-learning approach to execute the joint optimization of sector-edge and sector-center performance indicators. Sparse sampling is another technique that can handle the-curse-of-dimensionality problem. Thampi et al. [72] applied a reinforcement learning approach for the coverage self-optimization through antenna tilting and used sparse sampling algorithm to handle the the-curse-of-dimensionality problem. It has the ability to adapt to network environments without prior knowledge, handle large state spaces, perform self-healing and potentially focus on multiple coverage problems in LTE networks. In addition, Balevi et al. [73] designed a multi-agent RL framework to transform the weighted sum-rate optimization problem into a markov decision process (MDP). The two-step deep RL algorithm first aggregates inter-cell interference through mean-field theory to tackle the curse of dimensionality problem and then applies Q-learning with linear function approximation to utilize the single agent features.

The main disadvantage of learning-based solutions is the large number of iterations and measurement data, which keeps increasing with the size of the network. Besides, there is usually no convergence guarantee if they are applied to multi-agent systems. Generally, parameters must be set very precisely and are often based on an application, which means they must be customized for each system. Furthermore, theoretically, the more knowledge of the system is effectively utilized, the better the performance of the corresponding method will be. Therefore, the solution quality of learning-based methods is usually inferior to that of optimization-based methods.

3. Meta-heuristic methods

Meta-heuristic methods mainly include, genetic algorithm (GA) [74]-[78], particle swarm optimization (PSO) [79]-[81], simulated annealing (SA) [82], Tabu search [83], multi-objective evolution algorithm (EA) [84] and ant colony algorithm (ACA) [85], [86].

Yoon et al. [74] proposed an efficient GA based on quotient space property of their optimization problem to maximize the coverage deployment in wireless sensor networks. Yin et al. [75] solved the outage compensation problem by continually optimizing the antenna tilt based on GA to reduce the coverage holes. Liu et al. [76]

took the geometric distribution of the candidate sites into consideration and proposed a geometry-induced GA to efficiently maximize the cellular system coverage. Lakshminarasimman et al. [77] introduced a modified non-dominated sorting GA to solve various parameters of cellular base station placement problem such as site coordinates, transmitting powers, heights, and antenna tilts. Zhang et al. [78] proposed a hybrid two-layer optimization framework to enhance the network capacity and coverage and the genetic programming (GP) approach was exploited for the eNB operation at small time granularity.

Sousa et al. [79] implemented a PSO algorithm to optimize areas of low coverage and high interference simultaneously, through the adjustment of the antenna tilts and/or antenna orientation. Huang et al. [80] modified the PSO algorithm by employing a heuristic power control scheme to guide the algorithm to search for the global optimal solution. Qin et al. [81] investigated the metric structure of quotient subspace of the solution space and further proposed the metric-guided PSO for the cellular coverage problem.

The authors of [82] optimized the antenna tilt and azimuth network-wise, with the objective of minimizing the CPICH power consumption by simulated annealing algorithm. Hurley et al. [83] introduced a mathematical model for the automated design of fixed wireless access networks through the automatic selection and configuration of base station sites, and adopted the Tabu search method to generate the fixed wireless access network infrastructure design. Mai et al. [84] proposed a BS planning model based on TD-LTE system and designed a evolutionary algorithm with local search to solve this model. The model aimed at reducing the co-channel interference, expanding network capacity, and saving the network construction cost at the same time. Rui et al. [85] utilized ACA to find the optimal pilot power of each small cell to obtain the optimal coverage. Bo et al. [86] applied an ant colony optimizing algorithm for balancing the network load based on the ground user density by adjusting the handover parameters.

In all of these methods, the solution space is searched extensively in a fixed or stochastic manner in order to find the optimal or near-optimal solution. As a result, they are difficult to adapt to large-scale network parameter optimization and lack performance guarantees.

VI. IRS Empowered Performance Optimization

We now turn to a frontier technology called intelligent reflecting surface (IRS) that has brought a completely new perspective on how to improve the performance of wireless systems. Rather than optimizing the network infrastructure (including base stations, remote radio heads, and user terminals, etc.), IRS aims to improve the wireless environment. By manipulating the phase shift between the incident signal and the reflected signal on each reflective element of IRS, we can render the multipath propagations add up constructively at the target receiver, thereby augmenting the signal-to-noise ratio (SNR).

Clearly, in the IRS configuration, the key event of interest is *passive beamforming*, i.e., how to coordinate phase shifts across the reflective elements in order to achieve the signal focusing effect at the destination. For ease of discussion, let us restrict our attention to a single-input single-ouput (SISO) setup with an IRS comprising N reflective elements. We use $h_0 \in \mathbb{C}$ to denote the direct channel from the transmitter to the receiver, $h_n \in \mathbb{C}$ the reflected channel induced by the nth reflective element, $\theta_n \in [0, 2\pi)$ the phase shift of the nth reflective element, Z the additive white Gaussian noise at the receiver. The relationship between the transmit signal $X \in \mathbb{C}$ and the receive signal $Y \in \mathbb{C}$ is given by

$$Y = h_0 X + \sum_{n=1}^{N} h_n e^{j\theta_n} X + Z. \tag{9}$$

The SNR boosting goal can now be characterized as maximizing the following objective function $f(\cdot)$ over the passive beamforming vector $\boldsymbol{\theta} := (\theta_1, \ldots, \theta_N)$:

$$f(\boldsymbol{\theta}) = \frac{\mathbb{E}[|Y - Z|^2]}{\mathbb{E}[|Z|^2]} = \left| h_0 + \sum_{n=1}^{N} h_n e^{j\theta_n} \right|^2. \tag{10}$$

Inspection of $f(\boldsymbol{\theta})$ shows that the optimal strategy is to align each $h_n e^{j\theta_n}$ with h_0 by setting $\theta_n^\star = \angle h_0 - \angle h_n$, where $\angle \cdot$ represents the phase of a complex number. Nevertheless, the practical implementation of passive beamforming poses two significant challenges. First is the discrete constraint on θ_n as elaborated in Section VI-1; second is channel acquisition as elaborated in Section VI-2.

1. Passive beamforming for IRS with CSI

As compared to the other wireless devices, IRS bears a distinguishing feature that it does not generate any new signals but rather modifies the incident ones. Such feature has triggered a wave of research interest in incorporating the passive trait of IRS into the conventional beamforming paradigm. Although the link-level beamforming problem can be readily solved as stated under (2), a general system-level beamforming in the presence of multiple users is much more difficult to tackle because of nonconvexity. Among a variety of sophisticated mathematical tools developed for passive beamforming, semidefinite relaxation (SDR) [87] and fractional programming

(FP) [88], [89] have been most extensively applied in the literature to date. For instance, the authors of [90]-[96] find SDR particularly useful because the passive beamforming problem can be rewritten as a quadratic program, while the fractional structure of the passive beamforming problem (due to the ratio terms such as SNR or SINR) motivates the use of FP in [97]-[104]. Other nonconvex optimization approaches for passive beamforming include successive convex approximation (SCA) [105], [106], alternating direction method of multipliers (ADMM) [94], [107], minorization-maximization (MM) [108]-[110].

So far we treat the passive beamforming for IRS as a continuous problem wherein each phase shift θ_n can be arbitrarily chosen in $[0, 2\pi)$. However, this is not the case in the real world. On the contrary, the prototype realizations of IRS [111]-[116] typically limit the phase shift choices to a small set Φ_K with K discrete values, i.e.,

$$\Phi_K = \{\omega, 2\omega, \ldots, K\omega\}, \tag{11}$$

where

$$\omega = \frac{2\pi}{K}. \tag{12}$$

There are three reasons for the above restriction. First, the hardware cost increases with the number of phase shift choices. Second, the real-time configuration is difficult for continuous beamforming. Third, the reflection loss becomes higher when more PIN diodes are integrated into each reflective element for the continuous beamforming purpose. This discrete constraint on $\boldsymbol{\theta}$, however, results in a huge challenge in optimizing phase shifts. Because of the wide belief that the discrete IRS beamforming problem is NP-hard, those works aimed at the global optimum resort to the exponential-time algorithms such as the exhaustive search [107] and the branch-bound method [108], [109]. But the recent work [110] shows a somewhat surprising result that the binary beamforming problem of the single-user case can be globally solved in linear time.

In the existing literature [111]-[114], [118]-[120], a popular approach is to first relax the discrete beamforming problem as the continuous and then round the solution to the closest point in the discrete set Φ_K, but the resulting error is difficult to analyze. Another approach is to optimize one phase shift θ_n at a time while holding the rest phase shifts fixed [121], [125], [127], namely block coordinate descent. A penalty term is added to the relaxed problem in [126] to enforce the discrete constraint. Moreover, [120] not only shows that SDR works for discrete beamforming, but also develops a novel approximation algorithm called APX with a strictly better performance guarantee than SDR and the aforementioned closest point projection (CPP) method, as illustrated in Fig. 6.

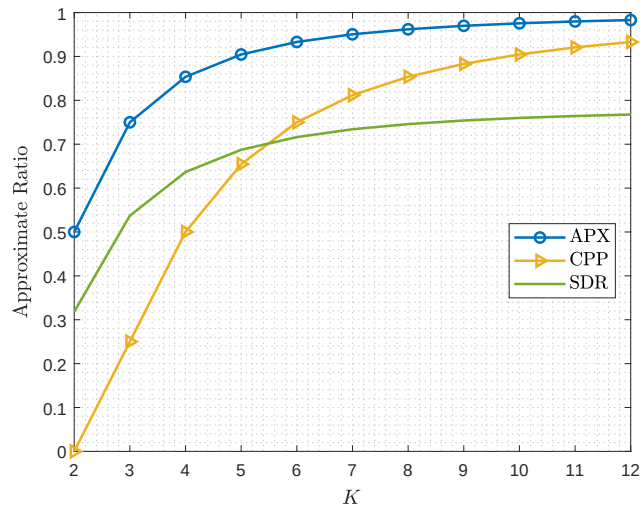


Fig. 6 The approximation ratios, i.e., $\frac{achieved\ SNR\ boost}{global\ optimal\ SNR\ boost}$, of the different algorithms when there are K phase shift choices for each reflective element.

2. Passive beamforming for IRS without CSI

The above algorithms, either for continuous beamforming or for discrete beamforming, all heavily depend upon CSI, so channel acquisition in IRS-aided systems has attracted considerable research interests over the past few years. We are mainly faced with the following challenges in this area:

1. Each reflected channel h_n alone can be much weaker than the direct channel h_0, so h_n is difficult to estimate accurately.

2. It requires the access to the in-phase and quadrature components stored in the communication chip, but this is not supported by the current 5G protocol.

3. The channel estimation for the cascaded channels enhances the time complexity and the overhead cost of passive beamforming.

The prior studies mostly concentrate on challenge #1. An early work [108] suggests estimating one single h_n at a time, assuming that the rest reflective elements are OFF and could absorb their incident signals; the more recent works [131], [132] further develop this ON-OFF policy by estimating a group of h_n's simultaneously in order to address the weak reflection issue as noted earlier. The state-of-the-art estimation method is based on the discrete Fourier transform (DFT) matrix [118], [122], [133], [134]. The main idea is to generate a pilot sequence by setting the beamforming vector to each row of the DFT matrix; in particular, [128] shows that the Cramér-Rao lower bound can be achieved for the least-squares estimation. Moreover, a line of studies [104], [117], [119], [125]

examine the channel estimation task from a compressed sensing perspective, aiming to find a sparse representation of the cascaded channels. Notice that none of the above methods accounts for challenges #2 and #3. To the best of our knowledge, not any channel estimation methods have been implemented in a real-world 5G network.

Because of the aforementioned bottlenecks in channel acquisition, the area of IRS has branched out to a new front wherein passive beamforming is performed without any channel information. For instance, [87], [88] suggest a random rotation strategy that does not require instantaneous CSI. There is also a deep learning based approach. In contrast to [111], [117] that apply the deep neural network to channel estimation and [135]-[137], [143] that apply the deep neural network to the IRS configuration given CSI, the recent work [138] proposes using the deep neural network to learn the direct mapping from the received pilot signal to the passive beamforming solution, thus sidestepping the channel estimations stage. Moreover, a cluster of works [89],[130]-[141] are based on beam training, the main idea of which is to sweep all possible directions of the reflected beam. However, the beam training method is limited to the millimeter/terahertz frequency bands with sharp beams.

Alternatively, we could exploit the statistics of received signal. The so-called RFocus method in [142] aims to decide the ON-OFF state of each reflective element; it simply tries out different ON-OFF combinations and then chooses the state for each reflective element according to the average performance. This statistical approach is further developed in [112] to account for the discrete phase shift $\theta_n \in \Phi_K$. Specifically, we first generate T random samples of $\boldsymbol{\theta}(t)$ over the discrete set Φ_K, where the sample index $t = 1, \ldots, T$, and then measure the corresponding received signal power $|Y(t)|^2$ with respect to each $\boldsymbol{\theta}(t)$. Furthermore, consider a subset $Q_{nk} \subseteq \{1, \ldots, T\}$ for each reflective element n and each phase shift option k, which comprises the indices of all those random trials with the nth phase shift chosen to be $k\omega$, i.e.,

$$\mathcal{Q}_{nk} = \{t : \theta_n(t) = k\omega\}. \tag{13}$$

We now compute the conditional sample mean of the received signal power within the above subset:

$$\widehat{\mathbb{E}}[|Y|^2|\theta_n = k\omega] = \frac{1}{|\mathcal{Q}_{nk}|} \sum_{t \in \mathcal{Q}_{nk}} |Y(t)|^2. \tag{14}$$

Intuitively, the conditional sample mean $\widehat{\mathbb{E}}[|Y|^2|\theta_n = k\omega]$ quantifies the average performance of setting $\theta_n = k\omega$, so it is natural to choose each phase shift to maximize the conditional sample mean, i.e.,

$$\theta_n = k_n^\star \omega, \tag{15}$$

where

$$k_n^\star = \arg\max_k \widehat{\mathbb{E}}[|Y|^2|\theta_n = k\omega]. \tag{16}$$

Aside from the theoretical justification that the above method guarantees a quadratic SNR boost in the number of reflective elements, [112] further demonstrates through field tests the practical effectiveness of the above conditional sample mean algorithm.

We conclude this section by remarking the implementation of the proposed conditional sample mean method in practice. The proposed method does not require any assistance or collaboration from the base station side, so it works in a plug-and-play fashion. The random sampling in (9) can be performed by using one or more sensor devices to measure the received signal quality in the vicinity of the target user terminal(s). Notice that the sensors used in our method are much simpler and cheaper than most user terminals because they only measure the received signal power, not even requiring the phase information. After receiving the random sample data from the sensors, the IRS then locally decides the passive beamforming vector according to (10) and (11), which, according to the field test in [112], takes merely a few seconds in total.

VII. Conclusions

This paper has reviewed state-of-the-art mathematical and learning based methods for performance modeling and optimization of real-world wireless networks. The current cutting-edge techniques from five different perspectives, including channel modeling, user demand and traffic modeling, throughput modeling and prediction, network parameter optimization, and IRS empowered performance optimization, have been elaborated. With the principles and properties of different methods explained and illustrated, we hope that this paper will facilitate the suitable choice of methods for further research on performance modeling and optimization of future wireless networks involving advanced techniques such as IRS.

References

[1] W. Xu et al., "Ten key ICT challenges in the post-Shannon era (in Chinese)", *Sci. Sin. Math.*, Vol.51, No.7, pp.1095–1138, 2021.

[2] Huawei Technologies, "Huawei 5G wireless network planning solution white paper", 2018, https://www-file.huawei.com/- /media/corporate/pdf/white%20paper/2018/5g_wireless_network_planing_solution_en.pdf?la=en-ch.

[3] C.-X. Wang, J. Bian, J. Sun, W. Zhang, and M. Zhang, "A Survey of 5G channel measurements and models", *IEEE Communications Surveys & Tutorials*, Vol.20, No.4, pp.3142–3168, 2018.

[4] J. Bian, C.-X. Wang, X. Gao, X. You, and M. Zhang, "A general 3D non-stationary wireless channel model for 5G and beyond", *IEEE Transactions on Wireless Communications*, Vol.20, No.5, pp.3211–3224, 2021.

[5] S. Zhang, X. Ning, X. Zheng, Q. Shi, T. Chang, and Z. Luo, "A Physics-based and Data-driven Approach for Localized Statistical Channel Modeling", in *IEEE International Teletraffic Congress (ITC)*, Sept. 2022.

[6] X. Ning, S. Zhang, X. Zheng, T. Chang, "Multi-Grid-Based Localized Statistical Channel Modeling: A Radio Map Approach", in *IEEE International Symposium on Wireless Communication Systems (ISWCS)*, Oct. 2022.

[7] J. Huang, C.-X. Wang, L. Bai, J. Sun, Y. Yang, J. Li, O. Tirkkonen, and M. Zhou, "A big data enabled channel model for 5G wireless communication systems", *IEEE Transactions on Big Data*, Vol.6, No.2, pp.211–222, 2020.

[8] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz (Release 16)", *Technical Report (TR) 38.901*, 2020, version 16.1.0.

[9] J. Zhou, W. Zhao, S. Chen, "Dynamic network slice scaling assisted by prediction in 5G network," IEEE Access, vol. 8, pp. 133700-133712. 2020.

[10] L. N. Wang, C. R. Zang, Y. Y. Cheng. The short-term prediction of the mobile communication traffic based on the product seasonal model. SN Applied Sciences, 2020, 2(3): 1-9.

[11] D. Clemente, G. Soares, D. Fernandes, et al. Traffic forecast in mobile networks: Classification system using machine learning//2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall). IEEE, 2019: 1-5.

[12] J. Huang, M. Xiao, Mobile network traffic prediction based on seasonal adjacent windows sampling and conditional probability estimation. IEEE Transactions on Big Data, 2020.

[13] Q. T. Tran, L. Hao, Q. K. Trinh. A comprehensive research on exponential smoothing methods in modeling and forecasting cellular traffic. Concurrency and Computation: Practice and Experience, 2020, 32(23): e5602.

[14] A. Perveen, R. Abozariba, M. Patwary, et al. Dynamic traffic forecasting and fuzzy-based optimized admission control in federated 5G-open RAN networks. Neural Computing and Applications, 2021: 1-19.

[15] Y. Yamada, R. Shinkuma, T. Sato, et al. Feature-selection based data prioritization in mobile traffic prediction using machine learning//2018 IEEE Global Communications Conference (GLOBECOM). IEEE, 2018: 1-6.

[16] H. Xia, X. Wei, Y. Gao, et al. Traffic prediction based on ensemble machine learning strategies with bagging and lightgbm//2019 IEEE International Conference on Communications Workshops (ICC Workshops). IEEE, 2019: 1-6.

[17] C. Liu, T. Wu, Z. Li, et al. Individual traffic prediction in cellular networks based on tensor completion. International Journal of Communication Systems, 2021, 34(16): e4952.

[18] T. N. Weerasinghe, I.A. M. Balapuwaduge, F. Y. Li. Supervised learning based arrival prediction and dynamic preamble allocation for bursty traffic//IEEE INFOCOM 2019-IEEE conference on computer communications workshops (INFOCOM WKSHPS). IEEE, 2019: 1-6.

[19] R. Abozariba, M. K. Naeem, M. Asaduzzaman, M. Patwary. Uncertaintyaware RAN slicing via machine learning predictions in next-generation networks. In 2020 IEEE 92nd vehicular technology conference (VTC2020-Fall) (pp. 1-6). IEEE,2020.

[20] L. Ale, N.Zhang, S. A. King, J. Guardiola, Spatio-temporal Bayesian learning for mobile edge computing resource planning in smart cities. ACM Transactions on Internet Technology (TOIT), 21(3), 1C21,2021.

[21] J. Zhang *et al*, Flow prediction in spatiotemporal networks based on multitask deep learning, *IEEE Trans. Knowl. Data Eng.*, Vol. 32, No. 3, pp. 468-478, 2020.

[22] K. Guo *et al*, Optimized graph convolution recurrent neural network for traffic prediction, *IEEE Trans. Intell. Transp. Syst.*, 2020.

[23] S. Zhan, L.Yu, Z. Wang, Y. Du, Y. Yu, Q. Cao, et. al. Cell traffic prediction based on convolutional neural network for software-defined ultra-dense visible light communication networks. Security and Communication Networks, 2021.

[24] C. W. Huang, C. T. Chiang, Q. Li. A study of deep learning networks on mobile traffic forecasting[C]//2017 IEEE 28th annual international symposium on personal, indoor, and mobile radio communications (PIMRC). IEEE, 2017: 1-6.

[25] R. Huang *et al*, LSGCN: Long short-term traffic prediction with graph convolutional networks, *in Proc. 29th Int. Joint Conf. Artif. Intell.*, pp. 2355C2361, 2020.

[26] X. Y. Yin *et al*, "Deep learning on traffic prediction: methods, analysis, and future directions," *IEEE*.

[27] C. Vinchoff *et al*, "Traffic prediction in optical networks using graph convolutional generative adversarial networks," *In Proceedings of the International Conference on Transparent Optical Networks*, pp. 3C6, 2020.

[28] A. Boukerche and J. Wang, Machine learning-based traffic prediction models for intelligent transportation systems, *Comput. Netw.*, Vol. 181, No.107530, 2020.

[29] X. Yin *et al*, Multi-stage attention spatial-temporal graph networks for traffic prediction, *Neurocomputing*, Vol. 428, pp. 42C53, 2021.

[30] L. Zhao *et al*, T-GCN: A temporal graph convolutional network for traffic prediction, *IEEE Trans. Intell. Transp. Syst.*, Vol. 21, No. 9, pp. 3848C3858, 2020.

[31] M. Aibin *et al*, "On Short-and Long-Term Traffic Prediction in Optical Networks Using Machine Learning," *In Proceedings of the 25th International Conference on Optical Network Design and Modelling*, 2021.

[32] L. Fang *et al*, Mobile demand forecasting via deep graph-sequence spatiotemporal modeling in cellular networks, *IEEE Internet of Things Journal*, Vol. 5, No. 4, pp. 3091-3101, 2018.

[33] Z. Wang *et al*, "Long-Term Traffic Prediction Based on LSTM Encoder-Decoder Architecture," *in IEEE Transactions on Intelligent Transportation Systems*, Vol. 22, No. 10, pp. 6561-6571, 2021.

[34] M. Aibin, "Deep learning for cloud resources allocation: long-short term memory in EONs," *In Proceedings of the International Conference on Transparent Optical Networks*, pp. 8C11, 2019.

[35] F. Morales *et al*, "Virtual network topology adaptability based on data analytics for traffic prediction," *J. Opt. Commun. Netw.* 2017.

[36] D. A. Tedjopurnomo *et al*, "A survey on modern deep neural network for traffic prediction: trends, methods and challenges," *in IEEE Transactions on Knowledge and Data Engineering*, Vol. 34, No. 4, pp. 1544-1561, 2022.

[37] X. Shi *et al*, A spatial-temporal attention approach for traffic prediction, *IEEE Trans. Intell. Transp. Syst.*, 2020.

[38] Y. Zhu, S. Wang. Joint traffic prediction and base station sleeping for energy saving in cellular networks[C]//ICC 2021-IEEE International Conference on Communications. IEEE, 2021: 1-6.

[39] A. M. Nagib, H. Abou-Zeid, H. S. Hassanein, et al. Deep learning-based forecasting of cellular network utilization at millisecond resolutions[C]//ICC 2021-IEEE International Conference on Communications. IEEE, 2021: 1-6.

[40] T. Kuber, I. Seskar, N. Mandayam. Traffic prediction by augmenting cellular data with non-cellular attributes[C]//2021 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2021: 1-6.

[41] G. L. Santos, P. Rosati, T. Lynn, et al. Predicting short-term mobile Internet traffic from Internet activity using recurrent neural networks. International Journal of Network Management, 2022, 32(3): e2191.

[42] T. H. H. Aldhyani, M. Alrasheedi, A. A. Alqarni, et al. Intelligent hybrid model to enhance time series models for predicting network traffic[J]. IEEE Access, 2020, 8: 130431-130451.

[43] A. Azari, P. Papapetrou, S. Denic, et al. Cellular traffic prediction and classification: A comparative evaluation of LSTM and ARIMA[C]//International Conference on Discovery Science. Springer, Cham, 2019: 129-144.

[44] A. Azari, P. Papapetrou, S. Denic, et al. User traffic prediction for proactive resource management: Learning-powered approaches[C]//2019 IEEE Global Communications Conference (GLOBECOM). IEEE, 2019b: 1-6.

[45] Y. Zang *et al*, "Wavelet transform processing for cellular traffic prediction in machine learning networks," *in proceedings of the 2015 IEEE China Summit and International Conference on Signal and Information Processing*, pp. 458-462, 2015.

[46] T. Li *et al*, "AutoST: efficient neural architecture search for spatio-temporal prediction," *in Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, pp. 794C802, 2020.

[47] Victor Buenestado, Matias Toril, Salvador Luna-Ramirez, Jose Maria Ruiz-Aviles, and Adriano Mendo, "Self-tuning of remote electrical tilts based on call traces for coverage and capacity optimization in LTE" *IEEE Transactions on Vehicular Technology*, 2017.

[48] S. Berger, M. Soszka, A. Fehske, P. Zanier, I. Viering and G. Fettweis, "Joint throughput and coverage optimization under sparse system knowledge in LTE-A networks," 2013 International Conference on ICT Convergence (ICTC), 2013, pp. 105-111, doi: 10.1109/ICTC.2013.6675318.

[49] Bahar Partov, Douglas J. Leith, and Rouzbeh Razavi. Utility fair optimization of antenna tilt angles in LTE networks. In IEEE/ACM Transactions on Networking, 2015.

[50] H. Eckhardt, S. Klein and M. Gruber, "Vertical Antenna Tilt Optimization for LTE Base Stations," 2011 IEEE 73rd Vehicular Technology Conference (VTC Spring), 2011, pp. 1-5, doi: 10.1109/VETECS.2011.5956370.

[51] Alexander Engels, Michael Reyer, Xiang Xu, Rudolf Mathar, Jietao Zhang, and Hongcheng Zhuang. Autonomous self-optimization of coverage and capacity in LTE cellular networks. IEEE Transactions on Vehicular Technology, 2013.

[52] S. Berger, A. Fehske, and G. Fettweis, Force field based joint optimization of strictly monotonic KPIs in wireless networks, in IFIP/IEEE Wireless Days (WD), 2012.

[53] Yaxi Liu, Wei Huangfu, Haijun Zhang, and Keping Long. An effcient stochastic gradient descent algorithm to maximize the coverage of cellular networks. IEEE Transactions on Wireless Communications, 2019.

[54] W. An, Y. Liu and W. Huangfu, "A Fast Coordinate Descent Algorithm Improving both Coverage and Capacity in Cellular Networks," 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC), 2020, pp. 43-48, doi: 10.1109/ICEIEC49280.2020.9152354.

[55] Pablo A. Sanchez Ordonez, Salvador Luna-Ramirez, and Matias Toril. A computationally efficient method for QoE-driven self-planning of antenna tilts in a LTE network. IEEE Access, 2020.

[56] Chaima Dhahri and Tomoaki Ohtsuki. Antenna parameters optimization in self-organizing networks: Multi-armed bandits with Pareto search. In IEEE 86th Vehicular Technology Conference (VTC Fall), 2017.

[57] Z. Wang and C. Shen, "Small Cell Transmit Power Assignment Based on Correlated Bandit Learning," in IEEE Journal on Selected Areas in Communications, vol. 35, no. 5, pp. 1030-1045, May 2017, doi: 10.1109/JSAC.2017.2679660.

[58] Cong Shen, Ruida Zhou, Cem Tekin, and Mihaela van der Schaar. Generalized global bandit and its application in cellular coverage optimization. IEEE Journal of Selected Topics in Signal Processing, 2018.

[59] Maxime Bouton, Hasan Farooq, Julien Forgeat, Shruti Bothe, Meral Shirazipour, and Per Karlsson. Coordinated reinforcement learning for capacity optimization in mobile networks. In Private Communication, 2012.

[60] Weisi Guo, Siyi Wang, Yue Wu, Jonathan Rigelsford, Xiaoli Chu, and Tim O'Farrell. Spectral and energy-efficient antenna tilting in a hetnet using reinforcement learning. In IEEE Wireless Communications and Networking Conference (WCNC), 2013.

[61] Eren Balevi and Jeffrey G. Andrews. Online antenna tuning in heterogeneous cellular networks with deep reinforcement learning. IEEE Transactions on Cognitive Communications and Networking, 2019.

[62] Nikolay Dandanov, Hussein Al-Shatri, Anja Klein, and Vladimir Poulkov. Dynamic self-optimization of the antenna tilt for best trade-off between coverage and capacity in mobile networks. Wirell Pers. Commun., 2017.

[63] Ryan M. Dreifuerst, Samuel Daulton, Yuchen Qian, Paul Varkey, Maximilian Balandat, Sanjay Kasturia, Anoop Tomar, Ali Yazdan, Vish Ponnampalam, and Robert W. Heath. Optimizing coverage and capacity in cellular networks using machine learning. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021.

[64] Sascha Berger, Albrecht Fehske, Paolo Zanier, Ingo Viering, and Gerhard Fettweis. Online antenna tilt-based capacity and coverage optimization. IEEE Wireless Communications Letters, 2014.

[65] Rouzbeh Razavi, Siegfried Klein, and Holger Claussen. A fuzzy reinforcement learning approach for self optimization of coverage in LTE networks. In Bell Lab. Tech. J. John Wiley & Sons, Inc., 2010.

[66] R. Razavi, S. Klein, and H. Claussen, "Self-optimization of capacity and coverage in LTE networks using a fuzzy reinforcement learning approach," in Proc. 21st Annu. Int. Symp. Pers., Indoor Mobile Radio Commun., 2010, pp. 1865-1870.

[67] Fan Shaoshuai, Hui Tian, and Cigdem Sengul. Self-optimization of coverage and capacity based on a fuzzy neural network with cooperative reinforcement learning. EURASIP Journal on Wireless Communications and Networking, 2014.

[68] A. Thampi, D. Kaleshi, P. Randall, W. Featherstone, and S. Armour, A sparse sampling algorithm for self-optimisation of coverage in LTE networks, in Proc. Int. Symp. Wireless Commun. Syst., Aug. 2012, pp. 909C913.

[69] H. Zhang, J. Du, J. Cheng, K. Long, and V. C. M. Leung, Incomplete CSI based resource optimization in SWIPT enabled heterogeneous networks: A non-cooperative game theoretic approach, IEEE Trans. Wireless Commun., vol. 17, no. 3, pp. 1882C1892, Mar. 2018.

[70] J. Chen, H. Jiang, and M. Peng, Adaptive configuration of antenna down tilt for optimizing coverages in IMT-Advanced systems, in Proc. Int. Conf. Commun. Technol. Appl., Oct. 2011, pp. 495C499.

[71] U. Turke and M. Koonert, Advanced site configuration techniques for automatic umts radio network design, in Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st, vol. 3. IEEE, 2005, pp. 1960C1964.

[72] J. Niemela, T. Isotalo, and J. Lempiainen, Optimum antenna downtilt angles for macro-cellular WCDMA network, EURASIP J. Wireless Commun. Netw., vol. 2005, no. 5, 2005, Art. no. 610942.

[73] A. Awada, B. Wegmann, I. Viering, and A. Klein, Optimizing the radio network parameters of the long term evolution system using Taguchi's method, IEEE Trans. Veh. Technol., vol. 60, no. 8, pp. 3825C3839, Oct. 2011.

[74] Y. Yoon and Y.-H. Kim, An efficient genetic algorithm for maximum coverage deployment in wireless sensor networks, IEEE Trans. Cybern., vol. 43, no. 5, pp. 1473-1483, Oct. 2013.

[75] Y. Mengjun, F. Lei, L. Wenjing, Y. Peng, and Q. Xuesong, Cell outage compensation based on CoMP and optimization of tilt, J. China Univ. Posts Telecommun., vol. 22, no. 5, pp. 71-79, 2015.

[76] Y. Liu, W. Huangfu, H. Zhang, H. Wang, W. An, and K. Long, An efficient geometry-induced genetic algorithm for base station placement in cellular networks, IEEE Access, vol. 7, pp. 108604C108616, 2019.

[77] N. Lakshminarasimman, S. Baskar, A. Alphones, and M. Willjuice Iruthayarajan, Evolutionary multiobjective optimization of cellular base station locations using modified NSGA-II, Wireless Netw., vol. 17, no. 3, pp. 597-609, 2011.

[78] J. Zhang, C. Sun, Y. Yi, and H. Zhuang, A hybrid framework for capacity and coverage optimization in self-organizing LTE networks, in Proc. IEEE 24th Int. Symp. Pers. Indoor Mobile Radio Commun., Sep. 2013, pp. 2919-2923.

[79] M. Sousa, A. Martins, and P. Vieira, Self-optimization of low coverage and high interference in real 3G/4G radio access networks, Inst. Eng. Lisbon Academic J. Electron. Telecommun. Comput., vol. 3, no. 1, p. 12, 2018.

[80] L.Huang,Y.Zhou,J.Hu,X.Han,andJ.Shi, "Coverage optimization for femtocell clusters using modified particle swarm optimization," in Proc. IEEE Int. Conf. Commun., Ottawa, ON, Canada, Jun. 2012, pp. 611-615.

[81] Y. Qin, W. Huangfu, H. Zhang, W. Liu, and K. Long, Accelerated coverage optimization with particle swarm in the quotient space characterizing antenna azimuths of cellular networks, IEEE Access, vol. 7, pp. 86252C 86264, 2019.

[82] I. Siomina, P. Varbrand, and D. Yuan, Automated optimization of service coverage and base station antenna configuration in UMTS networks, IEEE Wireless Commun., vol. 13, no. 6, pp. 16C25, Dec. 2006.

[83] S. Hurley, S. Allen, D. Ryan, and R. Taplin, Modelling and planning fixed wireless networks, Wireless Netw., vol. 16, no. 3, pp. 577C592, 2010.

[84] W. Mai, H.-L. Liu, L. Chen, J. Li, and H. Xiao, Multi-objective evolutionary algorithm for 4G base station planning, in Proc. IEEE Int. Conf. Comput. Intell. Secur., Leshan, China, Dec. 2013, pp. 85C89.

[85] R. Han, C. Feng, H. Xia, and Y. Wu, Coverage optimization for dense deployment small cell based on ant colony algorithm, in Proc. 80th Veh. Technol. Conf., 2014, pp. 1C5.

[86] W. Bo, S. Yu, Z. Lv, and J. Wang, A novel self-optimizing load balancing method based on ant colony in lte network, in Wireless Communications, Networking and Mobile Computing (WiCOM), 2012 8th International Conference on. IEEE, 2012, pp. 1C4.

[87] Z.-Q. Luo, W.-K. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems", *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20C34, May 2010.

[88] K. Shen and W. Yu, "Fractional programming for communication systemsPart I: Power control and beamforming", *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616C2630, May 2018.

[89] , "Fractional programming for communication systemsPart II: Uplink scheduling via matching", *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2631C2644, May 2018.

[90] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming", *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394C5409, Nov. 2019.

[91] G. Zhou, C. Pan, H. Ren, K. Wang, M. D. Renzo, and A. Nallanathan, "Robust beamforming design for intelligent reflecting surface aided MISO communication systems", *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1658C1662, Oct. 2020.

[92] B. Zheng, C. You, and R. Zhang, "Double-IRS assisted multi-user MIMO: Cooperative passive beamforming design", *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4513C4526, Jul. 2021.

[93] H. Xie, J. Xu, and Y.-F. Liu, "Max-min fairness in IRS-aided multi-cell MISO systems with joint transmit and reflective beamforming", *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1379C1393, Feb. 2021.

[94] S. Huang, Y. Ye, M. Xiao, H. V. Poor, and M. Skoglund, "Decentralized beamforming design for intelligent reflecting surface-enhanced cell-free networks", *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 673C677, Mar. 2021.

[95] W. Ni, X. Liu, Y. Liu, H. Tian, and Y. Chen, "Resource allocation for multi-cell IRS-aided NOMA networks", *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4253C4268, Jul. 2021.

[96] M. Zeng, X. Li, G. Li, W. Hao, and O. A. Dobre, "Sum rate maximization for IRS-assisted uplink NOMA", *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 234C238, Jan. 2021.

[97] Z. Zhang, L. Dai, X. Chen, C. Liu, F. Yang, R. Schober, and H. V. Poor, "Active RIS vs. passive RIS: Which will prevail in 6G?", 2021, [Online]. available at *https://arxiv.org/abs/2103.15154*.

[98] Y. Cao, T. Lv, and W. Ni, "Intelligent reflecting surface aided multi-user mmWave communications for coverage enhancement", in *IEEE Ann. Int. Symp. Personal Indoor Mobile Radio Commun. (SPAWC)*, Aug. 2021.

[99] J. Hu, Y.-C. Liang, and Y. Pei, "Reconfigurable intelligent surface enhanced multi-user MISO symbiotic radio system", *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2359C2371, Apr. 2021.

[100] J. Zhu, Y. Huang, J. Wang, K. Navaie, and Z. Ding, "Power efficient IRS-assisted NOMA", *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 900C913, Feb. 2021.

[101] Z. Zhang and L. Dai, "A joint precoding framework for wideband reconfigurable intelligent surface-aided cell-free network", *IEEE Trans. Signal Process.*, vol. 69, pp. 4085-4101, Jun. 2021.

[102] Y. Cao, T. Lv, Z. Lin, and W. Ni, "Delay-constrained joint power control, user detection and passive beamforming in intelligent reflecting surface-assisted uplink mmWave system", *IEEE Trans. Cognitive Commun. Netw.*, vol. 7, no. 2, pp. 482-495, Jun. 2021.

[103] K. Feng, X. Li, Y. Han, S. Jin, and Y. Chen, "Physical layer security enhancement exploiting intelligent reflecting surface", *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 734-738, Mar. 2021.

[104] T. Shafique, H. Tabassum, and E. Hossain, "Optimization of wireless relaying with flexible UAV-borne reflecting surfaces", *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 309-325, Jan. 2021.

[105] X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, "Exploiting intelligent reflecting surfaces in NOMA networks: Joint beamforming optimization", *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6884-2020, Oct. 2020.

[106] M.-M. Zhao, A. Liu, Y. Wan, and R. Zhang, "Two-timescale beamforming optimization for intelligent reflecting surface aided multiuser communication with QoS constraints", *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 6179-6194, Sep. 2021.

[107] B. Ning, Z. Chen, W. Chen, and J. Fang, "Beamforming optimization for intelligent reflecting surface assisted MIMO: A sum-path-gain max- imization approach", *IEEE Wireless Commun. Lett.*, vol. 9, no. 7, pp. 1105C1109, Jul. 2020.

[108] H. Shen, W. Xu, S. Gong, C. Zhao, and D. W. K. Ng, "Beamforming optimization for IRS-aided communications with transceiver hardware impairments", *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 1214-1227, Feb. 2021.

[109] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO communications relying on intelligent reflecting surfaces", *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218-5233, Aug. 2020.

[110] X. Yu, D. Xu, Y. Sun, D. W. K. Ng, and R. Schober, "Robust and secure wireless communications via intelligent reflecting surfaces", *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2637-2652, Nov. 2020.

[111] V. Arun and H. Balakrishnan, "RFocus: Beamforming using thousands of passive antennas", in *USENIX Symp. Netw. Sys. Design Implementation (NSDI)*, pp. 1047-1061, Feb. 2020.

[112] S. Ren, K. Shen, Y. Zhang, X. Li, X. Chen, and Z.-Q. Luo, "Configuring intelligent reflecting surface with performance guarantees: Blind beamforming", [Online]. available at *https://kaimingshen.github.io/doc/BlindBeamforming.pdf*, 2021.

[113] X. Pei, H. Yin, L. Tan, L. Cao, Z. Li, K. Wang, K. Zhang, and E. Björnson, "RIS-aided wireless communications: Prototyping, adaptive beamforming, and indoor/outdoor field trials", *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8627-8640, Dec. 2021.

[114] N. M. Tran, M. M. Amri, D. S. Kang, J. H. Park, M. H. Lee, D. I. Kim, and K. W. Choi, "Demo: Demonstration of reconfigurable metasurface for wireless communications", in Proc. *IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2020, pp. 1-2.

[115] K. M. Y. Kishiyama, "Research of transparent RIS technology toward 5Gevolution & 6G", *NTT DOCOMO*, 2021.

[116] P. Staat, S. Mulzer, S. Roth, V. Moonsamy, A. Sezgin, and C. Paar, "IRShield: A countermeasure against adversarial physical-layer wireless sensing", *arXiv preprint arXiv:2112.01967*, 2021.

[117] W. Cai, M. Li, and Q. Liu, "Practical modeling and beamforming for intelligent reflecting surface aided wideband systems", *IEEE Commun. Lett.*, vol. 24, no. 7, pp. 1568C1571, Jul. 2020.

[118] X. Yu, D. Xu, and R. Schober, "Optimal beamforming for MISO communications via intelligent reflecting surfaces", in *IEEE Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, May 2020.

[119] B. Di, H. Zhang, L. Song, Y. Li, Z. Han, and H. V. Poor, "Hybrid beamforming for reconfigurable intelligent surface based multi-user communications: Achievable rates with limited discrete phase shifts", *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1809-1822, Aug. 2020.

[120 Y. Zhang, K. Shen, S. Ren, X. Li, X. Chen, and Z.-Q.Luo, Configuring intelligent reflecting surface with performance guarantees: Optimal beamforming, in *IEEE J. Sel. Topics Signal Process.*, to be published.

[121] J. Qiao and M.-S. Alouini, "Secure transmission for intelligent reflecting surface-assisted mmWave and terahertz systems", *IEEE Commun. Lett.*, vol. 9, no. 10, pp. 1743-1747, Oct. 2020.

[122] C. You, B. Zheng, and R. Zhang, "Channel estimation and passive beamforming for intelligent reflecting surface: Discrete phase shift and progressive refinement", *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2604-2620, Nov. 2020.

[123] J. Qiao and M.-S. Alouini, "Secure transmission for intelligent reflecting surface-assisted mmWave and terahertz systems", *IEEE Trans. Wireless Commun.*, vol. 9, no. 10, pp. 1743-1747, Oct. 2020.

[124] B. Zheng, Q. Wu, and R. Zhang, "Intelligent reflecting surface-assisted multiple access with user pairing: NOMA or OMA?", *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 753-757, Apr. 2020.

[125] Q. Wu and R. Zhang, "Beamforming optimization for wireless networks aided by intelligent reflecting surface with discrete phase shifts", *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838-1851, Mar. 2020.

[126] M.-M. Zhao, Q. Wu, M.-J. Zhao and R. Zhang, "Exploiting amplitude control in intelligent reflecting surface aided wireless communication with imperfect CSI", *IEEE Trans. Commun.*, vol. 69, no, 6, pp. 4216-4231, Jun. 2021.

[127] A. Al-Hilo, M. Samir, M. Assi and S. Sharafeddine, "Reconfigurable intelligent surface enabled vehicular communication: Joint user scheduling and passive beamforming", *IEEE Trans. Veh. Technol.*, vol. 71, no.3, pp.2333-2345, Mar. 2022.

[128] M.-M. Zhao, A. Liu and R. Zhang, "Outage-constrained robust beamforming to intelligent reflecting surface aided wireless communication", *IEEE Trans. Signal Process.*, vol. 69, pp. 1301-1316, Feb. 2021.

[129 D. Li, "Fairness-aware multiuser scheduling for finite-resolution intelligent reflecting surface-assisted communication", *IEEE Commun. Lett.*, vol. 25, no. 7, pp. 2395-2399, Jul. 2021.

[130] Z. Feng, B. Clerckx and Y. Zhao, "Waveform and beamforming design for intelligent reflecting surface aided wireless power transfer: Single-user and multi-user solutions", *IEEE Trans. Commun.*, 2022, to appear.

[131] L. You, J. Xiong, D. W. K. Ng, C. Yuen, W. Wang, and X. Gao, "Energy efficiency and spectral efficiency tradeoff in RIS-aided multiuser MIMO uplink transmission", *IEEE Trans. Signal Process.*, vol. 69, pp. 1407C1421, Mar. 2021.

[132] W. Zhang, J. Xu, W. Xu, D. W. K. Ng, and H. Sun, "Cascaded channel estimation for IRS-assisted mmWave multi-antenna with quantized beamforming", *IEEE Commun. Lett.*, vol. 25, no. 2, pp. 593-597, Feb. 2021.

[133] H. Wang, N. Shlezinger, Y. C. Eldar, S. Jin, M. F. Imani, I. Yoo, and D. R. Smith, "Dynamic metasurface antennas for MIMO-OFDM receivers with bit-limited ADCs", *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2643C2659, Apr. 2021.

[134] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent-reflecting surface: Practical phase shift model and beamforming optimization", *IEEE Trans. Commun.*, vol. 68, no. 9, pp. 5849C5863, Sep. 2020.

[135] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surface for energy efficiency in wireless communication", *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157C4170, Aug. 2019.

[136] J. Gao, C. Zhong, X. Chen, H. Lin, and Z. Zhang, "Unsupervised learning for passive beamforming", *IEEE Commun. Lett.*,

vol. 24, no. 5, pp. 1052C1056, May 2020.

[137] C. Liu, X. Liu, D. W. K. Ng, and J. Yuan, "Deep residual network empowered channel estimation for IRS-assisted multi-user communication systems, in *Proc. IEEE Int. Commun. Conf. (ICC)*, Jun. 2021.

[138] G. T. de Araújo, A. L. F. de Almeida, and R. Boyer, "Channel estimation for intelligent reflecting surface assisted MIMO systems: A tensor modeling approach", *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 3, pp. 789-801, Apr. 2021.

[139] D. Mishra and H. Johansson, "Channel estimation and low-complexity beamforming design for passive intelligent surface assisted MISO wireless energy transfer", in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2019.

[140] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization", *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 518-522, Apr. 2020.

[141] T. L. Jensen and E. de Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator", in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, May 2020.

[142] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network", *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106-112, Jan. 2020.

[143] C. Huang, G. C.Alexandropoulos, C. Yuen, and M. Debbah, "Indoor signal focusing with deep learning designed reconfigurable intelligent surfaces", in *IEEE Int. Workshop Sig. Process. Advances Wireless Commun. (SPAWC)*, Aug. 2019.

LI Yang received the Ph.D. degree in Department of Electrical and Electronic Engineering from The University of Hong Kong in 2019. From 2019 to 2020, he has been a Senior Research Engineer in Huawei Noah's Ark Lab. He is the winner of the 2020 Innovation Pioneer Award of Huawei. Currently, he is a Research Scientist with Shenzhen Research Institute of Big Data. His research interests include radio resource management, learning to optimize, and large-scale optimization. (Email: liyang@sribd.cn)

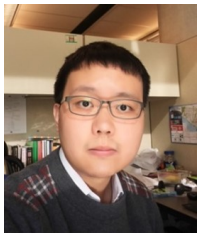
ZHANG Shutao received the Bachelor's degree in communication engineering and the Master's degree in electronics and communication engineering from Sun Yat-sen University, Guangzhou, China, in 2018 and 2020, respectively. He is currently pursuing the Ph.D. degree with The Chinese University of Hong Kong, Shenzhen. He is also enrolled in the Joint Education Program of the Shenzhen Research Institute of Big Data. His research interests include sparse signal processing and wireless channel modeling. (Email: shutaozhang@link.cuhk.edu.cn)

REN Xiaohui was born in 1996. She received the B.S. degree from School of Mathematics and Information Science, Hebei University, in 2022. She is now a Ph.D. candidate in school of Mathematics and Computational Sciences of Xiangtan University since 2022. Her research interests include network optimization, traffic prediction and modeling. (Email: 2654966935@qq.com)

ZHU Jianhang was born in Henan, China, in 1999. He received the B.E. degree in communication engineering from the School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China, in 2021, where he is currently pursuing the M.E. degree with the School of Computer Science and Engineering. His research interests include age of information, edge computing, and the Internet of Things. (Email: zhujh26@mail2.sysu.edu.cn)

HUANG Jiajie was born in Guangzhou, China, in 1999. He received the B.E. degree in communication engineering from the School of Electronics and Communication Engineering, Sun Yat-sen University, Guangzhou, China, in 2021, where he is currently pursuing the M.E. degree with the School of Computer Science and Engineering. His research interests include age of information, intelligent network, and the Internet of Things. (Email: huangjj7@mail2.sysu.edu.cn)

HE Pengcheng was born in 1998. He received the B.S. degree from School of Software Engineering, Tongji University, in 2020. He is now a Ph.D. candidate in the School of Software Engineering, Tongji University since 2020. His research interests include optimization, machine learning and signal processing. (Email: steven_he@tongji.edu.cn)



SHEN Kaiming received the B.Eng. degree in information security and the B.S. degree in mathematics from Shanghai Jiao Tong University, Shanghai, China in 2011, then the M.A.Sc. and Ph.D. degrees in electrical and computer engineering from University of Toronto, Ontario, Canada in 2013 and 2020, respectively. Since 2020, he has been an Assistant Professor with the School of Science and Engineering at The Chinese University of Hong Kong (Shenzhen), China. His main research interests include optimization, wireless communications, and information theory. Dr. Shen received the IEEE Signal Processing Society Young Author Best Paper Award in 2021. (Email: shenkaiming@cuhk.edu.cn)

YAO Zhiqiang received the M.S. and Ph.D. degrees from the School of Electronic and Information Engineering, South China University of Technology, Guangzhou, China, in 2004 and 2010, respectively. He was a Postdoctoral Fellow at the Chinese University of Hong Kong, Shenzhen, China, in Prof. Zhi-Quan Luo's research group. Since 2010, he has been with Xiangtan University, where he is currently a full Professor and the Dean in the College of Automation and Electronic Information. His research interests include signal processing, communication, localization, and optimization. He is Senior Member of the Chinese Institute of Electronics and IEEE. (Email: yaozhiqiang@xtu.edu.cn)

GONG Jie received his B.S. and Ph.D. degrees in the Department of Electronic Engineering in Tsinghua University, Beijing, China, in 2008 and 2013, respectively. From Jul. 2012 to Jan. 2013, he visited the Institute of Digital Communications, University of Edinburgh, Edinburgh, UK. From Jul. 2013 to Oct. 2015, he worked as a postdoctorial scholar in Tsinghua University. He is currently an associate professor in the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China. He served as Editor for IEEE Transactions on Green Communications and Networking, Workshop Co-chair for IEEE/CIC ICCC 2022 and Publicity Co-chair for IEEE WCNC workshop since 2018. He was a co-recipient of the Best Paper Award from IEEE Communications Society Asia-Pacific Board in 2013. His research interests include green communications and networking, energy harvesting technology, mobile edge computing and age of information. (Email: gongj26@mail.sysu.edu.cn)

CHANG Tsung-Hui received the B.S. degree in electrical engineering and the Ph.D. degree in communications engineering from the National Tsing Hua University (NTHU), Hsinchu, Taiwan, in 2003 and 2008, respectively. He currently is an Associate Professor with the School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, China. Prior to being a Faculty Member, he held research positions with NTHU, from 2008 to 2011, and the University of California, Davis, CA, USA, from 2011 to 2012. His research interests include signal processing and optimization problems in data communications, machine learning, and Big Data analysis. Dr. Chang is an Elected Member of IEEE Signal Processing Society (SPS) Signal Processing for Communications and Networking Technical Committee (SPCOM TC), the Funding Chair of IEEE SPS Integrated Sensing and Communication Technical Working Group (ISAC TWG), and the IEEE SPS Regional Director-at-Large of Region 10. He was on the Editorial Board for main SP journals, including an Associate Editor (2014C2018) and Senior Area Editor since February 2021 of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS (2015C2018) and IEEE OPEN JOURNAL OF SIGNAL PROCESSING since January 2020. Dr. Chang was the recipient of the Young Scholar Research Award of National Taiwan University of Science and Technology in 2014, IEEE ComSoc Asian-Pacific Outstanding Young Researcher Award in 2015, Outstanding Faculty Research Award of SSE, CUHKSZ, in 2021, and IEEE SPS Best Paper Award in 2018 and 2021. (Email: tsunghui.chang@ieee.org)

SHI Qingjiang received his Ph.D. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2011. From September 2009 to September 2010, he visited Prof. Z.-Q. (Tom) Luo's research group at the University of Minnesota, Twin Cities. In 2011, he worked as a Research Scientist at Bell Labs China. From 2012, He was with the School of Information and Science Technology at Zhejiang Sci-Tech University. From Feb. 2016 to Mar. 2017, he worked as a research fellow at Iowa State University, USA. From Mar. 2018, he is currently a full professor with the School of Software Engineering at Tongji University. He is also with the Shenzhen Research Institute of Big Data. His interests lie in algorithm design and analysis with applications in machine learning, signal processing and wireless networks. So far he has published more than 80 IEEE journals and filed about 40 national patents.

Dr. Shi was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was the recipient of IEEE Signal Processing Society Best Paper Award in 2022, the Huawei Technical Cooperation Achievement Transformation Award (2nd Prize) in 2022, the Huawei Outstanding Technical Achievement Award in 2021, the Golden Medal at the 46th International Exhibition of Inventions of Geneva in 2018, the First Prize of Science and Technology Award from China Institute of Communications in 2017, the National Excellent Doctoral Dissertation Nomination Award in 2013, the Shanghai Excellent Doctorial Dissertation Award in 2012, and the Best Paper Award from the IEEE PIMRC'09 conference. (Email: shiqj@tongji.edu.cn)

LUO Zhi-Quan (corresponding author) received the B.S. degree in applied mathematics from Peking University, China, and the Ph.D. degree in operations research from MIT in 1989. From 1989 to 2003, he held a faculty position with the ECE Department of McMaster University, Canada. He held a tier-1 Canada Research Chair in information processing from 2001 to 2003. After that, he has been a full professor at the ECE Department, University of Minnesota and held an endowed ADC Chair in digital technology. Currently, he is the Vice President (Academic) of The Chinese University of Hong Kong (Shenzhen) and the director of Shenzhen Research Institute of Big Data (SRIBD). Prof. Luo is a Fellow of IEEE and SIAM. He was elected to Fellow of Royal Society of Canada in 2014 and a Foreign Member of the Chinese Academy of Engineering (CAE) in 2021. He received four best paper awards from the IEEE Signal Processing Society, one best paper award from EUSIPCO, the Farkas Prize from INFORMS and the prize of Paul Y. Tseng Memorial Lectureship in Continuous Optimization as well as some best paper awards from international conferences. In 2021, he was awarded 2020 ICCM Best Paper Award by International Consortium of Chinese Mathematicians. He has published over 350 refereed papers, books and special issues. Prof. Luo has served as an Associate Editor for many internationally recognized journals and the Editor-in-Chief for IEEE Transactions on Signal Processing. His research mainly addresses mathematical issues in information sciences, with particular focus on the design, analysis and applications of large-scale optimization algorithms. (Email: luozq@cuhk.edu.cn)