Integration of Massive MIMO and Machine Learning in the Present and Future of Power Consumption in Wireless Networks: A Review

Sampson E. Nwachukwu, Maurine Chepkoech, Albert A. Lysko, Kehinde Awodele, Joyce Mwangama
Department of Electrical Engineering
University of Cape Town
Cape Town, South Africa
nwcsam001@myuct.ac.za, chpmau001@myuct.ac.za, albert.lysko@uct.ac.za, kehinde.awodele@uct.ac.za, joyce.mwangama@uct.ac.za

Albert A. Lysko, Chris R. Burger
NextGen Enterprises and Institutions
Council For Scientific and Industrial Research (CSIR)
Pretoria, South Africa
alysko@csir.co.za, *crburger@csir.co.za*

***Abstract*— The steady increase in data traffic rates and systems' complexity have contributed to the information and communication technologies (ICT) sector's increased energy consumption and CO_2 emissions. These pose a significant challenge to the telecommunication industry and the environment. This challenge has necessitated considering energy efficiency as a critical design pillar in 5G and future wireless networks. As a result, current research efforts for future wireless networks focus on minimising energy usage and improving efficiency. This work investigates several energy optimisation techniques in the present and future wireless networks, their contributions, advantages, and limitations. Based on the review of different techniques, we discuss the architecture of the massive MIMO (mMIMO) technique, including its operation and requirements. We also present the performance evaluation of mMIMO using different precoding algorithms, which is crucial for energy efficiency in future wireless networks. We further review incorporating intelligence using a Machine Learning (ML) approach in switching off underused mMIMO arrays to minimise energy usage. Finally, we discuss several critical open research issues in mMIMO and ML that make future research and implementation possible in next-generation wireless networks.**

Keywords—Multiple Input Multiple Output, MIMO, massive MIMO, mMIMO, Advanced Sleeping Mode, ASMs, ML, Small Cells, 5G, energy efficiency, power consumption, energy optimisation.

I. INTRODUCTION

The telecommunication industry has experienced a sharp increase in network users, demanding improved network capacity, increased peak data rate, and a better quality of service (QoS) [1]. With the evolution of mobile networks, the fifth generation (5G) mobile networks have been developed with enhanced network capacity, and improved quality of service (QoS), compared to the fourth generation (4G). Thus, to support critical trends such as 99.9999% availability and reliability and less than 1 ms end-to-end latency, modern networks have evolved towards more functionality and architectural complexity, and increased energy requirements, leading to increased operational costs and an added global emission of CO_2 [2]. These pose a challenge to sustainable network deployments [3]. Therefore, to achieve better, greener, and more efficient future networks, the energy consumption of network technology in 5G and beyond must be improved to reduce network operating costs and global carbon emissions [4].

Existing energy efficiency techniques focus on lowering energy consumption at the base station through efficient hardware design, especially for the power amplifier, air conditioning, and signal processing [4]. Additionally, various researchers in the telecommunication industry now focus on multi-RAT (Radio Access Technologies) deployments to allow various technologies to share the same infrastructure, reducing operational and ownership costs in 5G and future wireless networks [5]. For example, Lopez-Perez et al. [6] and Ghosh et al. [7] suggested the use of massive MIMO (mMIMO), small cells, Artificial Intelligence (AI), Advanced Sleeping Mode (ASM) and the cooperation of mMIMO and Cognitive Femtocells, to achieve optimised energy consumption in 5G networks.

However, integrating various sophisticated components to build a reliable network increases complexity of the network. This leads to the need for integrating autonomous intelligence enabling rapid decision making, control, traffic offloading, energy level control, and seamless allocation of network resources to diverse service requests [4]. Machine Learning (ML) is a promising technique to achieve the expected autonomy in spectrum use, interference control, user tracking, and collaborative functioning of the diverse technologies to determine appropriate power transfer levels and energy-aware transmissions in future wireless networks [8].

In this paper, the motivation for energy optimisation and various techniques for energy efficiency in 5G networks is considered. The key objectives of this work are to:

- Examine critical existing energy efficiency (EE) techniques, including their benefits and constraints.
- Discuss the most reliable and viable technique that network operators can adopt in optimising energy consumption in future wireless networks.
- Discuss the performance evaluation of the mMIMO technique using different precoding algorithms.

- Explore adoption of ML for intelligence in switching on/off power in antenna arrays of base stations according to anticipated traffic patterns.
- Outline anticipated challenges and recommend future research directions towards energy optimisation in wireless networks.

II. SAMPLE ENERGY EFFICIENCY TECHNIQUES

This section discusses various existing energy efficiency techniques in the present and future networks, their benefits, and constraints.

A. MIMO and mMIMO

Multiple-input multiple-output (MIMO) is a signal-improvement approach that has been implemented in several wireless communications systems, including 4G/LTE, to meet increasing user traffic requirements. MIMO uses antenna arrays at the transmitter and receiver to utilise "spatial diversity" and overcome multipath interference. As a result, the network capacity can be improved without increasing the overall bandwidth requirements [9].

As per the International Telecommunications Union (ITU), the telecommunications sector is experiencing exponential traffic growth due to network subscribers' increasing number and needs. This increase raises the demand for increased data rates and an improved QoS. As a result, the MIMO technique was extended to massive MIMO (mMIMO), integrated into the 5G networks to cater for the high rate of data rates of about 20 Gbps at the downlink and 10 Gbps at the uplink [10]. mMIMO uses a wide range of antennas integrated with the base station (BS) to improve the data rate, energy efficiency, reliability, user tracking, and spectral efficiency of 5G networks while supporting more subscribers [10], [11].

The antenna beams in the mMIMO system are usually narrow and spatially targeted at the intended users [10]. Having narrow beams improves signal quality for subscribers by reducing the possibility of interference between the adjacent beams. In addition, the gain obtained from combining several antennas reduces the radiated power and improves the system's QoS [12].

In 5G, the beamforming is extended from azimuthal-only 2D beamforming to elevation-inclusive (3D) beamforming and beam tracking [12]. As a result, users can be tracked as they move so that more directional beams are delivered to them for better signal quality and higher data rates. In addition, mMIMO allows for high degrees of freedom to accommodate more demanding user traffic, thus simultaneously improving the capacity, QoS and spectral efficiency [12].

Massive IoT and better mobile broadband are among the user cases that next-generation network (5G) is expected to provide. Also, "licensed bands, unlicensed bands, shared bands, and lower sub-6 GHz to millimeter-wave bands" are some of the varieties of bands and spectrums 5G should be able to function on to guarantee spectrum accessibility. However, delivering the most efficient interface to fulfil spectrum and user case needs is a major problem for mMIMO antenna designs [13].

B. Advanced Sleeping Modes (ASMs)

Traditionally, communication systems have been developed to deliver higher data rates, coverage, and reliable connection and availability. They must thus be available at all times and dimensioned to provide good quality of service (QoS) during peak hours and large-scale events. However, this provisioning leads to an underused and over-dimensioned mobile network, impacting both capital and operational costs. Such is particularly inefficient at off-peak hours and during the night, when traffic demand is low. Therefore, the Sleep Modes (SMs) technique is adopted to turn off the inactive base station components for a predetermined time. This technique can gradually shut idle BS components to minimise energy consumption [14].

According to Salem et al. [14], four distinct degrees of SMs known as Advanced Sleep Modes (ASMs) are presented in the "IMEC Power Model Tool." The tool is used to determine the amount of energy consumed by the BS's components based on the SMs. The BS components are integrated by the SMs using comparable transition timings, specified as the durations of activation and deactivation. The SMs considered include "Sleep Mode 1 (SM_1), Sleep Mode 2 (SM_2), Sleep Mode (SM_3), and Sleep Mode 4 (SM_4)."

SM_1 level is the least sleep mode with a 71 μs transition duration. The Power Amplifier and certain processing components are deactivated in this SM level, making it the quickest among other SMs. Also, SM_2 is an intermediate level of BS deactivation, with a time frame of 1 ms. In SM_2, more components are deactivated. Furthermore, in SM_3 level, most of the BS components are put to sleep, except the clock generator, and it has a time frame of 10 ms. Finally, The BS stops operating in SM_4 except for the backhaul that remains on standby for reactivation purposes. SM_4 usually takes a short time frame of 1 s [14],[15].

In implementing the sleep modes, it is important to consider three stages, shutting down, sleep and waking up. Sleep length defines when the BS components are in an "inactive" state, while the activation period defines the amount of time to return an idle component to an active state [14],[15].

Deeper sleep levels like SM_3 and SM_4 can save more energy. However, they consume more time due to the component's reactivation delay, which also affects network users. As a result, a compromise must be made between energy savings and delay. Furthermore, compared to the threshold situation, when all consumers are served simultaneously, the buffering delay, which depends on the sleep mode level, reduces the system capacity [15].

C. Artificial Intelligence (AI)

The increasing complexity of 5G and future communication systems and the need to perform repeated activities necessitate new automation solutions that take advantage of "Artificial Intelligence (AI)/machine learning (ML)" methods to enhance overall system efficiency [16].

According to [17], "AI is a technology that allows a machine to mimic human behaviour. At the same time, ML is a branch of AI that allows a machine to learn from prior experience without programming the data explicitly." These approaches promise to provide more accurate energy optimisation based on network traffic, time of day, and other factors. Unlike the traditional deterministic methods, ML allows considering a nearly arbitrary large number of inputs, outputs and dependencies. The subclasses of ML, namely Supervised, Reinforcement, and Unsupervised Learning, can help make educated decisions on complex situations, such as dynamic wireless network conditions [18].

Supervised Learning is well suited for channel issues, including detection, learning behaviour, and channel estimation to forecast future outcomes. Supervised Learning generates output from gathered data based on previous events. In contrast, Reinforcement Learning (RL) is best suited for new and unknown network conditions by constantly adapting and changing to achieve the desired outcomes through learning from the results and optimising the judgments [19].

With all these exciting prospects of AI-based techniques, there are still some challenges barring adoption. Firstly, the techniques often lack the data sets necessary for training the models. Secondly, all the data must be appropriately matched and should be made free from errors before training. Finally, gathering data sets and error corrections consume a lot of time and effort during the training process. Therefore, there is a need for future research to focus on the differences between model simplification and efficient wireless network-specific ML, particularly in places where energy efficiency is crucial [19].

III. RELATED WORK

Several authors have focused on solving the wireless network's energy challenge.

Mowla et al. [20] carried out research that focused on integrating both passive optical network (PON) and mmWave techniques to provide a 5G wireless network with an optimised energy solution. In addition, the authors noted the need for dynamic backhauling to support low power usage with fluctuating traffic volumes. Their paper proposed an integrated approach to address backhauling challenges for the 5G heterogeneous network. The result obtained in their work shows that the combination of the mmWave and PON can help to save the energy in the 5G network by 32%.

The coordinated multipoint beamforming (CMBF) scheme can be vital in enhancing energy efficiency, especially in heterogeneous networks (HetNets) with mMIMO. This approach was used by Yinghui et al. [21], where the authors integrated CMBF, mMIMO, and small cells to enhance two-tier HetNets' energy efficiency. Their work considered emitted energy consumption and static hardware parameters needed by the mMIMO to minimise power consumption while providing the required coordinated schemes and solving QoS challenges. To obtain the best solution concerning energy efficiency, they used a zero-forcing (ZF) precoding technique and a CMBF algorithm with little complexity. The obtained simulation and analytical results indicate that HetNets with mMIMO can improve the wireless network's energy efficiency.

Furthermore, Hawasli and Colak [22] proposed an approach that is able to adjust to the changes in the functionality of the small base stations (SBSs) active/sleep (on/off) based on traffic load fluctuation. The general optimum power reduction problem for HetNets, which demands relatively high computing complexity, was the driving force behind their work. Their work provided practical answers concerning the time interval and the specific SBs that should be turned on/off, the factors that determine when to turn them on/off, and the best and worst algorithms for turning the BS on/off. The proposed novel methods performed similarly to the best approach while reducing overall energy usage in HetNets, providing the necessary QoS and having significantly lower computing complexity.

Future wireless networks are evolving towards autonomy and intelligence. As a result, ML is an integral part of large and complex network operations. For example, Sharma et al. [23] adopted a Reinforcement Learning approach in HetNets to enable the BS on/off mechanism to minimise BS energy consumption during off-peak hours. As a result, an 82% drop in energy consumption was achieved in the network. Such learning is significant in mapping out traffic demands in the network, hence controlling energy supply and consumption by intelligently switching off idle BS components.

IV. MMIMO ARCHITECTURE

According to Khwandah et al. [12], mMIMO is predicted to provide better energy and spectral efficiency than alternative energy optimisation techniques we discussed in Section 3. mMIMO offers high degrees of freedom and accommodates more significant throughput for users, thus increasing the system's energy efficiency, spectral efficiency and capacity (up to 50 times greater in 5G than with 4G MIMO), while minimising the interuser interference [12]. From these, mMIMO is a significant driver for implementing next-generation fixed and mobile networks.

Unlike the conventional MIMO, which serves only a single user at a time [27] and, in 4G, supports only eight antennas at a base station, mMIMO employs more than a hundred antennas at the transmitter to provide service to numerous user equipment (UE) at a similar time-frequency resource [24].

mMIMO typically uses more antennas in a cell than the number of UE supported. If the expected number of UE in a cell increases, the BS is improved to ensure that the capacity increases linearly with the transmit antennas, as long as the number of receiver antennas is proportional to the number of transmit antennas.

Additionally, 5G mMIMO works in the "Time-Division Duplex (TDD)" system to minimise the acquisition overhead of "Channel State Information (CSI)" [38]. The BS also uses various orthogonal channels to communicate with different users [24]. Architecturally, 5G mMIMO further needs a power amplifier (PA) and disintegrated transceiver cascading for an individual antenna; An increase in the number of antennas will result in a proportional decrease in the radiated power of each antenna [25].

mMIMO further depends on spatial multiplexing and requires substantial information about the channel from the BS on both the uplink and downlink channels. The uplink can be realised by allowing the terminals to forward the pilot signal (i.e., signal transmitted mainly for synchronisation or reference purposes [26]), which the base station uses to analyse each terminal's channel responses [11]. One of the ways to lower power consumption in future networks is to minimise the need for transmitting synchronisation signals.

V. mMIMO PERFORMANCE EVALUATION

This section examines and compares the performance of linear precoders and massive multiuser-MIMO (MU-MIMO) downlink systems.

MU-MIMO is a crucial technology in 5G communication standards that use a large antenna array (massive MIMO) [24], [39]. Also, signal processing methods used in base stations (e.g., linear and nonlinear precoding algorithms) are crucial for developing 5G technology. The precoding (i.e., beamforming) first determines the incoming signals and shares them correctly with many integrated antenna components [27].

The comparison of linear precoders and MU-MIMO performance informs of the impacts of antenna configuration on the overall system's performance and the relationship between the system's achievable data rates and energy efficiency (EE) and spectral efficiency (SE) for various linear precoding algorithms. The linear precoding algorithms that are considered here include Minimum Mean Square Error (MMSE), Maximum Ratio Transmission (MRT), and Zero Forcing (ZF).

The analysis and derivation of various precoders are discussed in the following subsections, while the EE and SE evaluation follow afterwards.

A. *Minimum Mean Square Error (MMSE)* Algorithm

The MMSE precoding algorithm depends on the method of Mean Square Error (MSE) [29]. In the MU-MIMO downlink system, it is referred to as optimal linear precoding. This precoder algorithm can be determined using the Lagrangian optimisation equations below [24].

$$\boldsymbol{A}_{MMSE} = \frac{1}{\beta} H^* \left(H^T H^* + \frac{K}{P_{tr}} I_K \right)^{-1}, \tag{1}$$

$$\beta = \sqrt{\frac{tr(BB^H)}{P_{tr}}}, \text{ and} \tag{2}$$

$$\mathrm{B} = H^* \left(H^T H^* + \frac{K}{P_{tr}} I_K \right)^{-1}, \tag{3}$$

where, "P_{tr} is the transmit power; I_K is the identity matrix of the user equipment, *K*; *K* is the number of users equipment" [24]; and *H* is the Channel knowledge [30].

B. *Zero Forcing (ZF) Algorithm*

In this approach, the interuser interference for an individual user is cancelled out [29]. This cancellation is based on the assumption that precoding implements the channel matrixes pseudo inverse [24]. The ZF can be represented as:

$$A_{ZF} = \frac{1}{\beta} H^H (HH^*)^{-1} \tag{4}$$

where, $B = H^*(HH^*)^{-1}$, *and* (5)

β (beta) is the Wiener Filter's scalar.

C. *Maximum Ratio Transmission (MRT) Algorithm*

MRT performs effectively in MU-MIMO systems with low transmitted power at the base station [46]. However, the SNR is minimised in this approach [24]. Mathematically, the MRT is given as:

$$\boldsymbol{A}_{MRT} = \frac{1}{\beta} H^*, \tag{6}$$

where, $\mathrm{B} = H^*$ (7)

D. *Rayleigh Fading Channel*

Due to multipath propagation in an environment, it is challenging to determine accurate channel parameters in wireless communication. However, to estimate the system's performance, channel models are used. "Rayleigh Fading Channel" is one of such channel models used to simulate the fading channel [24], with real *c* and imaginary *d* parts:

$$h_{Rayleigh} = c + \mathrm{j} \cdot d \tag{8}$$

Two Gaussian variables are used in equation (8) to determine the channel coefficients (*c* and *d*) of the Rayleigh fading. These variables have a 0.5 variance and zero mean [24].

E. *Achievable Rate*

Generally, the Shannon theorem characterises maximum supported capacity [31]. Shannon's theorem determines the "maximum rate the transmitter can transmit over the channel" [31]. Based on the MU-MIMO system under study, the MRT, ZF and achievable rate are described assuming that total downlink power is constant and distributed evenly across all users. Over an "Additive White Gaussian Noise (AWGN) channel" [31], the channel capacity can be determined from Shannon's theorem as:

$$R = log_2(1 + SNR) \quad \text{(bits/s/Hz)} \tag{9}$$

where, *SNR* is the ratio of signal-to-noise powers.

Since the interference of the MU-MIMO system is comprised of interuser interference and additive noise, and the transmitter uses a perfect Channel State Information (CSI) to create a link with all the users, for the K^{th} user, achievable rate, R_k is a function of expectation E [24]:

$$R_k = \mathrm{E}[log_2(1 + SNR_k)] \quad \text{(bits/s/Hz)} \tag{10}$$

F. *Spectral Efficiency (SE)*

According to [24], "the spectral efficiency is determined with a single cell massive MU-MIMO system", as shown:

$$R_p = \sum_{k=1}^{K} R_k \quad \text{(bits/s/Hz)}, \tag{10}$$

where, R_p is the spectral efficiency (sum-rate); R_p is the achievable rate of a *K*th user.

G. *Energy Efficiency (EE)*

Similarly, the energy efficiency can be expressed as [24]:

$$\eta = \frac{R_p}{P_{tr}} \quad \text{(bits/s/J/Hz)}, \tag{11}$$

where, P_{tr}= BS's total transmit power (in W).

VI. EVALUATING SPECTRAL EFFICIENCY (SE) AND ENERGY EFFICIENCY (EE)

This section evaluates and compares the performance of various precoding algorithms analysed above in a single cell massive MU-MIMO system for a "Rayleigh fading channel". This evaluation process is based on the work of Alsabbagh and Gelgor [29] and Selvan et al. [24]. Different numbers of users (K), the number of base station antennas (M) and the signal-to-noise ratio per user (<u>*SNRu*</u>) are used in this performance analysis. The main focus is the scenario where $M >> K$, which is described as the MU-MIMO system.

In the work of Selvan et al. [24], the performance of the achievable rate against the *SNRu* at $K = 10$ and $M = 20$ were compared. The results obtained show that the performance of MRT is better when the value of *SNR* is low compared to when the *SNRu* is high. In contrast, the performance of the *ZF* is better when the *SNR* is high compared to when the *SNRu* is low. MMSE gives the best achievable rate across a different range of *SNR*, whether high or low. Also, an improvement in the system's performance was observed when the number of antennas, M, is increased from 20 to 40. This increases the achievable rate of the precoders, except that when the *SNRu* is moving from 0 to -5 dB, *ZF* performs better than MRT.

Alsabbagh and Gelgor [29] also analysed the impact of the active BS antennas on spectral efficiency at *SNRu* = -5 dB and $K = 10$. The results show that when the number of BS antennas (M) increases, spectral efficiency also increases for all precoding algorithms. MMSE achieves the best spectral efficiency. Likewise, an *SNRu* of 10 dB is used in [52]. The results show significant improvement in SE for all precoders due to increasing M. Furthermore, the research evaluates the impact of K on SE. Setting $M = 100$, *SNRu* = -5 dB and increasing the number of UEs shows in Figure 5.1 that in ZF schemes, the spectral efficiency directly increases when the number of users K increases from 10 to 70.

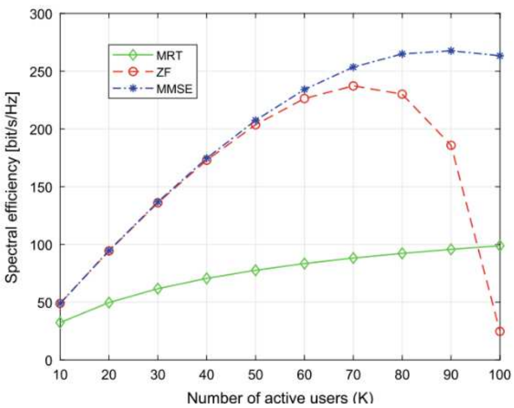

Figure 5.1: Spectral efficiency at M = 100 [29].

Conversely, a sharp deterioration was observed when $K > 80$. This decline is because of increased interuser interference as the distance between users becomes smaller. Similarly, when the MRT precoder is employed, the SE increases slowly with respect to K and its performance is worse than ZF and MMSE. In MMSE, the SE increases until $K = 80$. However, MMSE delivers better results, and its performance is more effective at increased K than MRT and ZF.

Energy efficiency is another essential criterion for MU-MIMO performance evaluation. With linear precoding, the system's performance regarding energy efficiency against spectral efficiency is evaluated using a similar scenario where $K = 10$ and $M = 50$ or 100 [29]. Based on equation (9), MRT outperforms ZF at low SE and high EE, whereas ZF outperforms MRT at low EE. Overall, MMSE outperforms both MRT and ZF in delivering the desired result, and by increasing the antenna count M, the energy efficiency increases significantly.

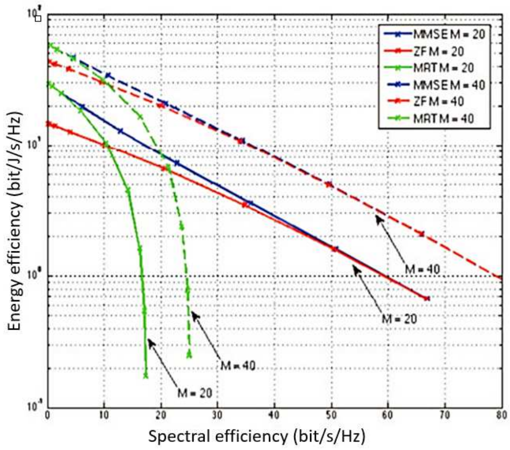

Figure 5.2: EE versus SE at M = 20 and 40 [24].

For example, in reference [24], when the BS antennas $M = 20$, the EE is 10 bits/J/Hz for MRT precoding, at SE of 10 bits/s/Hz. Likewise, when the BS antenna $M = 40$, EE increases (from 10 bits/J/Hz to 30 bits/J/Hz, as shown in Figure 5.2). This increase demonstrates that most MU-MIMO's theoretical benefits can be achieved under real-world conditions.

VII. mMIMO AND MACHINE LEARNING (ML) INTEGRATION FOR ENERGY OPTIMISATION

Due to the need to optimise energy use while serving the intended users, it is valuable to incorporate intelligence in switching on/off underused arrays to optimise the overall energy consumption in a mMIMO network. It is expected that the Machine Learning approaches can achieve this incorporation efficiently. The energy optimisation will, in turn, maximise profit for the network operators by reducing operational expenses.

Machine Learning (ML) gives machines the capacity to solve problems without being provided explicit instructions in solving the specified problems. Instead, the learning process consists of feeding a machine learning algorithm with the sampled parameters of the problem to be solved and allowing the machine to uncover inferences and different ways that maximise decision-making based on predefined goals specified by the user [32].

Reinforcement Learning (RL) and Deep RL are ML techniques. In RL, appropriate decisions are made based on mapping events to actions and assessing the best to maximise the reward. On the other hand, Deep RL learns about the human brain operation and uses that knowledge in developing models based on artificial neural networks integrated with multiple layers of neurons [33].

The capabilities of ML necessitate the need for its adoption in solving the future wireless network's dynamic and complex problems to improve efficiency and ensure the achievement of critical requirements, including user and profit maximisation for the operators. Traditional explicit algorithms are often deemed unsustainable because of high complexity, computational costs and rigidity. In contrast, ML is ideal for dynamic resource allocation tasks because of its ability to perform adaptive learning and deployment of energy efficiency techniques (e.g., in mMIMO) based on channel load without affecting QoS [34].

Recently, ML was used to improve the operation of base stations. ML works by monitoring the operation of the selected BS, gathering the historical traffic load and user habits and anticipating the BS traffic load based on on-site coverage [18]. ML can be applied in mMIMO to enhance coverage in a multi-cell situation while accounting for inter-site contamination across several 5G mMIMO cell sites [35].

Sanguinetti et al. [36] applied a deep learning algorithm that allows the network transmission systems to learn about users' location and adequately inform on the directionality of the beam and allocated downlink power of mMIMO networks. Hammoudeh [8] applied distributed Q-learning (an RL technique) controls in the HetNets with multi-agent systems comprising distributed small cells, an individual BS and a macro-BS. The serving BS can be switched on or off depending on the network's traffic load requirements. The learning implemented on the BS provides the small cells with an essential capability to sufficiently determine on/off times both for current and future operations. Miozzo et al. [37] used the distributed Q-Learning algorithm on an offline training period BS on/off algorithm. The algorithm is used to set up initial switch on/off policies. This training offers the BS a background of expected traffic patterns, reducing the initial exploration phase and quickening the setup phase to improve overall system performance.

VIII. DISCUSSIONS

While mMIMO and ML address many challenges in conventional communication systems, they also raise a slew of new issues. These issues need to be investigated during the design and implementation of energy-efficient mMIMO for next-generation wireless networks [10], [11].

Firstly, mMIMO systems use many antennas to decrease the effects of interference, fading, and channel noise. Unfortunately, this array of antennas skyrockets the cost of components and the system's complexity. These challenges can perhaps be solved by developing a mMIMO system with cheaper components. However, one of the disadvantages of this affordable equipment would be hardware impairment like in-phase and quadrature (IQ) imbalance, amplifier distortion, and phase noise. Although removing the hardware constraint will be difficult, its impacts can be reduced by employing suitable compensatory techniques [10].

The precoding methods of mMIMO systems improve throughput and decrease interference. However, they can raise the entire system's computational complexity. As a result, it is often more practicable to employ simple and highly efficient precoders with mMIMO. So, finding effective precoding techniques for mMIMO is also an important research area to explore.

Finally, ML algorithms require accurate datasets that match the requirements of dynamic future wireless networks. However, gaps in the availability of datasets for training and related proprietary concerns around traffic patterns by mobile operators limit the development of dynamic and accurate algorithms [28]. Therefore, there is a need to evaluate the balance between complexities introduced by integrating efficient and robust ML algorithms and simplifying energy-efficient models for mMIMO.

IX. CONCLUSION

This paper has reviewed the current state-of-the-art for solving energy efficiency problems in the next-generation wireless networks. Based on the observed benefits and limitations of the reviewed techniques, massive MIMO (mMIMO) is seen as the most viable option for achieving improved energy and spectral efficiency in 5G and beyond networks. A comparison of the performance of linear precoders in massive multiuser (MU) MIMO is made to understand the impacts of antenna configuration on the overall system's performance and the relationship between the system's achievable rate and its spectral and energy efficiency for various linear precoding algorithms. Due to future wireless networks' dynamic nature and complexity, several research now focus on using the machine learning (ML) approach in solving energy efficiency issues associated with 5G networks.

Although mMIMO and ML techniques provide significant benefits for 5G and beyond, several deployment problems, such as high cost of components, system's and computational complexity, and gaps in the availability of data set for training, are yet to be overcome. Nevertheless, some of these challenges offer network operators and researchers new opportunities, a new path to reach them and research problems to tackle in the future.

REFERENCES

[1] H. Kour and R. K. Jha, "Power Optimisation using Spectrum Sharing for 5G Wireless Networks", 2019 11th International Conference on Communication Systems & Networks (COMSNETS), 2019, pp. 395-398, doi: 10.1109/COMSNETS.2019.8711443. [Accessed 2021-06-16].

[2] O. Blume, D. Zeller, and U. Barth, "Approaches to energy-efficient wireless access networks," *Final Progr. Abstr. B. - 4th Int. Symp. Commun. Control. Signal Process. ISCCSP 2010*, pp. 2–6, 2010, doi: 10.1109/ISCCSP.2010.5463328. [Accessed 2021-06-16].

[3] S. Zhang, X. Cai, W. Zhou, and Y. Wang, "Green 5G enabling technologies: An overview," *IET Commun.*, vol. 13, no. 2, pp. 135–143, 2019, doi: 10.1049/iet-com.2018.5448. [Accessed 2021-06-16].

[4] Feng, S. Mao and T. Jiang, "Base Station ON-OFF Switching in 5G Wireless Networks: Approaches and Challenges", in IEEE Wireless Communications, vol. 24, no. 4, pp. 46-54, Aug. 2017, doi: 10.1109/MWC.2017.1600353.

[5] P. Bhuvaneswari and L. Nithyanandan, "Improving energy efficiency in backhaul of LTE-A network with base station cooperation," *Procedia Comput. Sci.*, vol. 143, pp. 843–851, 2018, doi: 10.1016/j.procs.2018.10.378.

[6] D. Lopez-Perez *et al.*, "A Survey on 5G Radio Access Network Energy Efficiency: Massive MIMO, Lean Carrier Design, Sleep Modes, and Machine Learning," pp. 1–96, 2021, [Online]. Available: http://arxiv.org/abs/2101.11246. [Accessed 2021-06-16].

[7] Ghosh, S., De, D. & Deb, P. Energy and Spectrum Optimization for 5G Massive MIMO Cognitive Femtocell Based Mobile Network Using Auction Game Theory. *Wireless Pers Communication* **106,** 555–576 (2019). https://doi.org/10.1007/s11277-019-06179-3.

[8] Hammoudeh *et al.*, "A concise introduction to reinforcement learning," from *Big Data Cogn. Comput.*, vol. 2, no. 4, pp. 1–11, 2018, [Online]. Available: https://www.researchgate.net/profile/Ahmad_Hammoudeh3/publication/323178749_A_Concise_Introduction_to_Reinforcement_Learning/li

nks/5a84c09d4585159152b80ae3/A-Concise-Introduction-to-Reinforcement-Learning.pdf
[9] S. Phule and A. Y. Kazi, “MIMO antenna system for LTE (4G),” *Proc. 20121st Int. Conf. Emerg. Technol. Trends Electron. Commun. Networking, ET2ECN 2012*, no. 2, pp. 1–6, 2012, doi: 10.1109/ET2ECN.2012.6470110.
10] Chataut R, Akl R. Massive MIMO Systems for 5G and Beyond Networks-Overview, Recent Trends, Challenges, and Future Research Direction. Sensors (Basel). 2020 May 12;20(10):2753. doi: 10.3390/s20102753. PMID: 32408531; PMCID: PMC7284607. [Accessed 2021-07-01].
[11] E. G. Larsson, O. Edfors, F. Tufvesson and T. L. Marzetta, Massive MIMO for next-generation wireless systems," in IEEE Communications Magazine, vol. 52, no. 2, pp. 186-195, February 2014, doi: 10.1109/MCOM.2014.6736761. [Accessed 2021-08-02].
[12] Sinan A. Khwandah, and John P. Cosmas et al., “Massive MIMO Systems for 5G Communication,” A Journal on Wireless Communication, April 2021, no. 0123456789. [Online]. Available: https://link.springer.com/content/pdf/10.1007/s11277-021-08550-9.pdf. [Accessed 2021-07-02].[13] Ishteyaq and K. Muzaffar, “Multiple input multiple output (MIMO) and fifth generation (5G): an indispensable technology for sub-6 GHz and millimeter wave future generation mobile terminal applications,” *International Journal of Microwave and Wireless Technologies*, pp. 1–17, 2021.
[14] Fatma Ezzahra Salem et al., “Advanced Sleep Modes and their impact on the flow-level performance of 5G networks,” IEEE Vehicular Technology Conference, p 1-7, Sept. 2018. [Online]. Available:https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8288125 [Accessed 2021-07-03].
[15] Fatma Ezzahra Salem et al., “Optimal Policies of Advanced Sleep Modes for Energy-Efficient 5G networks,” IEEE Vehicular Technology Conference, p 1-7, Sept. 2019. [Online]. Available: 1909.09011.pdf (arxiv.org). [Accessed 2021-07-03].
[16] Michael Gold, “Artificial intelligence will maximise the efficiency of 5G network operations,” computerweekly.com. para. 2, 2021-04-28. [Online]. Available: https://www.computerweekly.com/opinion/Artificial-intelligence-will-maximise-efficiency-of-5G-network-operations [Accessed August. [Accessed 2021-08-01].
[17] Satavisa Pati, “The Difference between Artificial Intelligence and Machine Learning,” analyticsinsight.net. para. 2, August . [Online]. Available: https://www.analyticsinsight.net/the-difference-between-artificial-intelligence-and-machine-learning/ [Accessed November 2021].
[18] ITU-T, “Smart Energy Saving of 5G Base Station: Based on AI and other emerging technologies to forecast and optimise the management of 5G wireless network energy consumption.” ITU-T Focus Group on Environmental Efficiency for Artificial Intelligence and other Emerging Technologies (FG-AI4EE), March 2021. [Online]. Available: https://www.itu.int/en/ITU-T/focusgroups/ai4ee/Documents/TR-D.WG3_02-Smart%20Energy%20Saving%20of%205G%20Base%20Station%20Based%20on%20AI%20and%20other%20emerging%20technologies_Tan.pdf. [Accessed 2021-08-01]
[19] Amna Mughees et al., “Towards Energy Efficient 5G Networks Using Machine Learning: Taxonomy, Research Challenges, and Future Research Directions.” IEEE Access, p. 187498-187522, July 2020, Vol. 8. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9218920. [Accessed 2021-08-02].
[20] M. M. Mowla, I. Ahmad, D. Habibi, and Q. V. Phung, “Energy Efficient Backhauling for 5G Small Cell Networks,” IEEE Transactions on Sustainable Computing, vol. 4, no. 3, pp. 279-292, July 11, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8361042. [Accessed 2021-06-17].
[21] Yinghui Zhang, Yang Liu, and Jing Gao, “Optimal design of energy-efficient two-tier HetNets with massive MIMO by coordinated multipoint beamforming,” Physical Communication, vol. 28, pp 183-189, June 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1874490718300569?via%3Dihub [Accessed 2021-08-02].
[22] M. Hawasli and S. A. Çolak, “Toward green 5G heterogeneous small-cell networks: power optimisation using load balancing technique,” AEU - International Journal of Electronics and Communications, pp 474 – 485, vol. 82, October 16, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1434841117314917. [Accessed 2021-08-10].
[23] Sharma, S., Darak, S.J., & Srivastava, A., “Energy saving in heterogeneous cellular network via transfer reinforcement learning based policy,” *International Conference on Communication Systems & Networks IEEE,* 2017.
[24] V. P. Selvan, M. S. Iqbal and H. S. Al-Raweshidy, “Performance analysis of linear precoding schemes for very large Multiuser MIMO downlink system,” Fourth edition of the International Conference on the Innovative Computing Technology (INTECH 2014), 2014, pp. 219-224, doi: 10.1109/INTECH.2014.6927765.
[25] F. Gregorio, G. González, C. Schmidt, J. Cousseau, “Signal Processing Techniques for Power Efficient Wireless Communication Systems,” Practical Approaches for RF Impairments Reduction. Springer, 2020. [E-Book]. Available: http://link.springer.com/10.1007/978-3-030-32437-7. [Accessed 2021-08-02].
[26] Wikipedia, "Pilot Signal," Wikipedia.org. para. 2, 2021-08-18. [Online]. Available: https://en.wikipedia.org/wiki/Pilot_signal.
[27] D. Subitha and R. Vani, “Analysis of Linear Precoding Techniques for Massive MIMO-OFDM Systems under various scenarios,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1084, no. 1, p. 12053, 2021, doi: 10.1088/1757-899x/1084/1/012053.
[28] Ge, Z., Haiyan, W. Linear precoding design for massive MIMO based on the minimum mean square error algorithm. *J Embedded Systems* **2017,** 20 (2017). https://doi.org/10.1186/s13639-016-0064-4.
[29] S. Alsabbagh and A. Gelgor, “Performance Analysis for Massive MIMO Systems Based on Quadriga Channel Model.” International Youth Conference on Electronics, Telecommunications and Information Technologies, YETI 2020. Chapter 49. Vol. 255, 2021-01-17. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-58868-7_49. [Accessed 2021-08-02].
[30] D. Subitha and R. Vani, “Analysis of Linear Precoding Techniques for Massive MIMO-OFDM Systems under various scenarios,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1084, no. 1, p. 012053, 2021, doi: 10.1088/1757-899x/1084/1/012053.
[31] T. Parfait, Y. Kuang and K. Jerry, “Performance analysis and comparison of ZF and MRT based downlink massive MIMO systems,” 2014 Sixth International Conference on Ubiquitous and Future Networks (ICUFN), 2014, pp. 383-388, doi: 10.1109/ICUFN.2014.6876818.
[32] N. Ivanova, V. Gugleva, M. Dobreva, I. Pehlivanov, S. Stefanov, and V. Andonova, “We are IntechOpen, the world ' s leading publisher of Open Access books Built by scientists , for scientists TOP 1 %,” *Intech*, vol. i, no. tourism, p. 13, 2016.
[33] J. Kaur, M. A. Khan, M. Iftikhar, M. Imran and Q. Emad Ul Haq, “Machine Learning Techniques for 5G and Beyond,” in IEEE Access, vol. 9, pp. 23472-23488, 2021, doi: 10.1109/ACCESS.2021.3051557.
[34] H. Fan, L. Zhu, C. Yao, J. Guo and X. Lu, “Deep Reinforcement Learning for Energy Efficiency Optimization in Wireless Networks,” IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), 2019, pp. 465-471, doi: 10.1109/ICCCBDA.2019.8725683.
[35] Omkar Dharmadhikari, “Leveraging Machine Learning and Artificial Intelligence for 5G.” cablelabs.com. para. 2, June 18, 2019. [Online]. Available: https://www.cablelabs.com/leveraging-machine-learning-and-artificial-intelligence-for-5g. [Accessed 2021-08-02].
[36] L. Sanguinetti, A. Zappone, and M. Debbah, “Deep Learning Power Allocation in Massive MIMO,” *Conf. Rec. - Asilomar Conf. Signals, Syst. Comput.*, vol. 2018-October, pp. 1257–1261, 2019, doi: 10.1109/ACSSC.2018.8645343.
[37] M. Miozzo, L. Giupponi and RossI, “Switch-On/Off Policies for Energy Harvesting Small Cells through Distributed Q-Learning,” in *IEEE Wireless Communication and Networking Conference Workshops (WCNCW)*, an Francisco, CA, USA, 2017.
[38] Emil Björnson, Jakob Hoydis and Luca Sanguinetti “Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency,” Foundations and Trends R in Signal Processing, vol. 11, pp 154–655. [Online]. Available: https://www.nowpublishers.com/article/Details/SIG-093. [Accessed 2021-08-02].
[39] M. Sahoo and H. K. Sahoo, “Multiuser Massive MIMO Channel Estimation and BER Analysis Using Convex Combination Based Algorithms,” *Wirel. Pers. Commun.*, vol. 123, no. 4, pp. 3025–3049, 2022, doi: 10.1007/s11277-021-09275-5