Integrated Cellular and Cell-Free Communication Systems Toward Global Connectivity: Motivations, Challenges, and Research Roadmap

Imtiaz Ahmed, *Howard University, Washington, DC, 20059, USA*

Md. Zoheb Hassan, *Université Laval, Quebec, QC, G1V 0A6, Canada*

Majumder Haider, *Howard University, Washington, DC, 20059, USA*

Kamrul Hasan, *Tennessee State University, Nashville, TN, 37209, USA*

Ensuring global connectivity and bridging the digital divide among urban, rural, and remote communities are the fundamental visions of 6G networks. Although various technologies, such as nonterrestrial networks, small cells, and wireless backhaul, are envisioned to enable global connectivity, their coexistence with the conventional cellular network architecture requires investigations. Meanwhile, 6G architecture is expected to accommodate a user-centric cell-free network, thanks to its robustness against inter-cell interference and offering macro-diversity. In this article, we propose a convergence of the conventional cellular and evolving cell-free communication networks to provide seamless coverage over vast geographical areas. The paper makes the following contributions. It introduces an architecture for integrated cell-free and cellular (ICFC) networks, incorporating digital twin technology and context-aware design. The paper emphasizes artificial intelligence-driven methods for managing radio resources and ensuring security. Additionally, we explore research avenues to enhance ICFC networks for seamless connectivity in the 6G era.

The evolution of cellular networks is marked by notable discrepancies in service quality among individual cells, leading to inconsistent network performance, particularly evident in the reduced peak data rates experienced at cell peripheries.[1,2] While approximately 80% of cell interiors can achieve high data rates, this falls short of meeting the demands of an increasingly wireless-dependent society, as highlighted in Akyildiz et al.,[3] Tong and Li,[4] Yang et al.,[5] and Chowdhury et al.,[6] emphasizing the necessity for comprehensive and reliable network coverage to support future communication paradigms.

1520-9202 © 2024 IEEE
Digital Object Identifier 10.1109/MITP.2024.3433511
Date of current version 3 September 2024.

Cell-free massive multiple-input, multiple-output (CF-mMIMO) technology is a pivotal technology, eliminating traditional cell boundaries and thereby enhancing connectivity and macro-diversity.[1,7,8,9,10] The concept of a distributed antenna system (DAS), introduced initially in Zhou et al.,[11] strategically places antennas across a geographical area to bolster user support. Subsequent studies, such as Zhang and H. Dai[12] and Zhang et al.,[13] explore cooperative downlink processing with multi-antenna access points (APs) for nonlinear coprocessing, laying the groundwork for cooperative multipoint (CoMP) access. These efforts advocate for a user-centric approach to network infrastructure assignment, where users are served by the nearest distributed antennas and joint processing by multiple and colocated base stations (BSs).

CF-mMIMO integrates technological advancements in mMIMO, DAS, and CoMP systems. Although such a setup enhances energy and spectral efficiency while offering multiplexing and array gains, it exhibits its limited scalability. Since only 10%–20% of APs are effectively utilized by a given user due to path loss,[7] a scalable solution allows each user to be supported by a selected subset of APs. While the network-centric user-AP clustering approach reintroduces boundaries, the user-centric approach eliminates predefined cells.[14,15,16] Nevertheless, the user-centric approach requires frequent adjustment of user clusters and high control signals. Developing a practical user-centric approach that balances scalability and power efficiency remains a significant research endeavor in CF-mMIMO systems.

However, a complete replacement of cellular networks with CF networks is impractical. Instead, a more feasible approach is facilitating cellular and CF infrastructures to coexist, resulting in integrated cell-free and cellular (ICFC) communication systems. The key aspect of this approach entails dividing large geographic areas into nonoverlapping regions served by either cellular or CF systems, fostering cooperation between these networks. For instance, macro-cellular networks can serve urban areas, while CF systems, incorporating aerial APs, serve rural and remote areas. To ensure the seamless integration of legacy and new technologies and accommodate heterogeneous APs distributed across horizontal and vertical planes, novel network planning, resource optimization, and security within ICFC networks are essential. To this end, this article makes the following three contributions. First, we present an architecture of the ICFC network and explain the context-aware design of such a network while leveraging digital twins (DTs). Second, we present several artificial intelligence (AI)-based approaches to meet the unique radio resource management (RRM) and security of the ICFC networks. Finally, we present several research opportunities to enhance the ICFC network's performance. To the best of our knowledge, this is the first article to systematically explore the challenges and potential solutions toward converging cellular and cell-free communications architecture for meeting the connectivity and coverage demands of 6G architecture.

PROPOSED ICFC COMMUNICATION FRAMEWORK

We propose an integrated network illustrated in Figure 1, combining cellular and CF communication systems in urban and rural areas, respectively. This integration addresses two main issues: Urban networks are typically optimized for urban regions, leaving rural areas underserved, and CF systems can enhance service quality in rural areas due to their resilience to inter-cell interference and macro-diversity gain. Both networks support

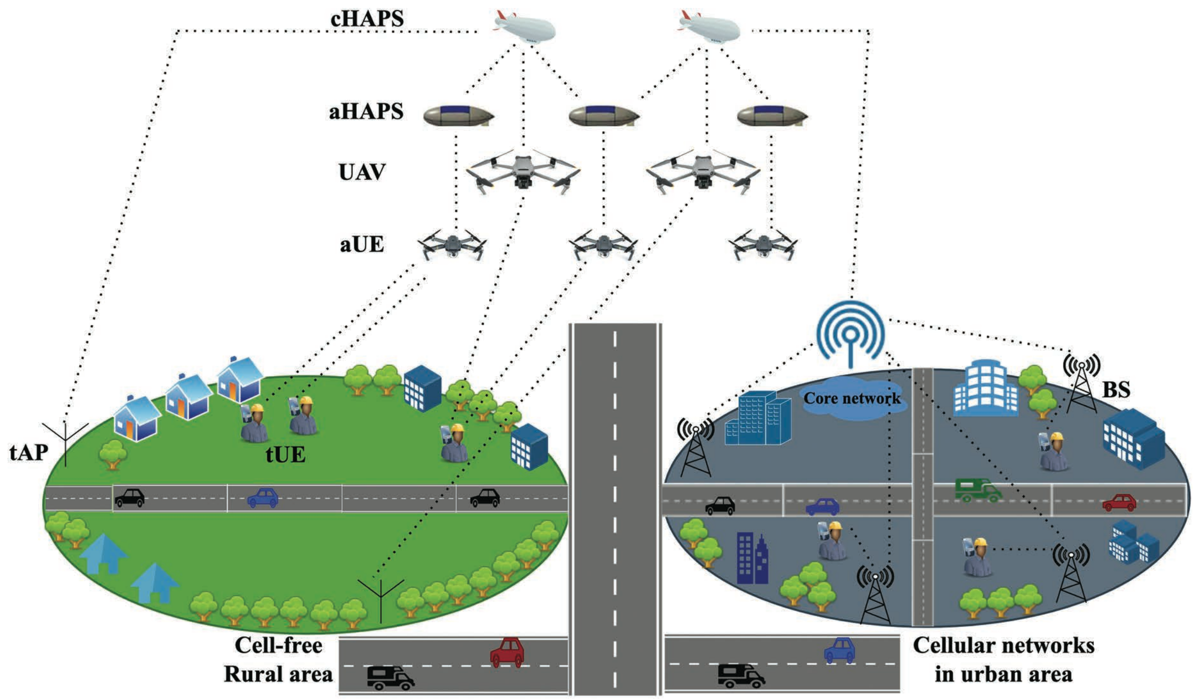

FIGURE 1. Proposed ICFC architecture.

terrestrial and aerial users seamlessly. Urban areas utilize traditional high-power BSs mounted on buildings or towers connected to centralized infrastructure, while rural areas are covered by a CF-mMIMO network using aerial APs and distributed CPUs to ensure extensive coverage across challenging landscapes. A high-altitude platform station (HAPS) is a super macro-BS, offering extended coverage, favorable channel conditions, and low latency.[17] Thus, HAPS are leveraged as airborne CPUs and a subset of aerial APs in CF systems. The ICFC network's CF system consists of the following platforms:

- *CPU HAPS (cHAPS)*: These are dedicated airborne processing units implemented over HAPS and used for centralized precoding and combining in the CF system. The cHAPS are edge-computing HAPS (EC-HAPS)[18] and connected to the shared core network via high-speed backhaul links.
- *Aerial APs (AAPs) and Terrestrial APs (tAPs)*: Two types of AAPs are considered: high-altitude AAPs, such as non-edge competing HAPS (NEC-HAPS),[18] and low-altitude AAPs, namely unmanned aerial vehicles. The former offers broad coverage on the ground with a stable footprint, while the latter provides on-demand and swift deployment capability for offloading hotspot traffic. AAPs handle precoded data traffic exchange with the cHAPS for downlink and uplink communications. tAPs are also strategically placed in rural locations. Both tAPs and AAPs are linked to cHAPSs through fronthaul connections.

The ICFC network's cellular system comprises traditional terrestrial BSs serving terrestrial and aerial users. In alignment with standards, HAPS-enabled CF networks can utilize frequency bands ranging from 700 to 900 MHz, 1.7 GHz, and 2.6 GHz. Cellular networks may operate in the same or different bands, leading to various operational scenarios and challenges. When both CF and cellular systems use the same band, effective resource management strategies are necessary to mitigate interference, particularly in border regions between urban and rural areas. When CF and cellular systems use different bands, seamless handover mechanisms must be developed to facilitate uninterrupted user mobility between urban and rural areas.

ICFC NETWORK PLANNING: DT-ENABLED APPROACH

The CF system within the ICFC network requires meticulous planning to ensure effective collaboration with existing cellular infrastructure while enhancing service quality for rural and remote users. This planning involves the following four key optimizations:

- *P1. Deployment Optimization*: Excessive densification of AAPs and cHAPS increases deployment costs and interference. Optimizing the number and 3-D deployment locations of AAPs and cHAPS, and the ground coverage of AAPs, are required to minimize costs and network outages.
- *P2. Backhaul Optimization*: Efficient fronthaul links between AAPs and cHAPS and backhaul links between cHAPS and core networks are essential. Integrated access and backhaul technology offer promise but require careful management of interference among active links through bandwidth allocation.
- *P3. Connectivity Optimization*: Given the limited computational processing capability of cHAPS and the large number of AAPs, establishing connections between them is costly. This optimization focuses on building a connected aerial topology by minimizing the overall connection cost.
- *P4. Frequency Optimization*: Channel allocation to the access and backhaul links of the CF system is required to maximize the system capacity while minimizing interference with the cellular system.

However, optimizing large-scale ICFC networks is confronted by the increased control overhead and the difficulty of dynamically updating live network setups without degrading cellular system performance. DTs, virtual representations of complex physical systems, can address these challenges. DTs capture the entire environmental context and ICFC network features and interface with the physical network. This enables operators to conduct various "what-if" analyses without impacting the network's performance through trial and error. *DT-assisted ICFC network planning* involves three steps.

- *S1. Context Acquisition*: This step aims to acquire contexts that describe the network's states. The context factors include the deployment locations of BSs, AAPs, and cHAPs, communication parameters, transceiver models, channel models, network coverage, blockages, weather, users' positions, density, mobility, and transceiver models. A context acquisition module saves these contexts in a database, which periodically updates the dynamic ones through feedback from the physical network.
- *S2. DT Construction*: This stage aims to construct a DT using static and dynamic contexts from the context database. The DT consists of virtual

network topologies and various models for communication within the ICFC network. Virtual topologies can be built with tools like Sionna and Wireless Insite. Models such as communication, traffic, and key performance monitoring are integrated with this virtual topology to capture RF propagation, QoS variations, and key performance indicators of the ICFC networks.

S3. Algorithm Execution: This stage aims to optimize network planning (P1–P4) by using DT. Optimization models (heuristic algorithms or machine learning models) are created for network planning tasks. These models update and validate ICFC network planning decision variables using DT information. Once an optimal outcome is achieved, the network planning decisions are implemented in the physical ICFC network.

For optimal network planning, it is crucial to maintain a high similarity between the DT and the physical ICFC networks. A context monitoring unit tracks context variables as the ICFC network undergoes temporal and dynamic changes. Network planning steps are repeated when significant changes are detected.

RADIO RESOURCE MANAGEMENT FOR ICFC NETWORKS

Integrating cellular and CF networks brings several RRM challenges, including network selection, user clustering, pilot assignment, beamforming, power control, and interference management. Traditional optimization approaches cannot handle these tasks efficiently for large-scale ICFC networks. Integrating DT and AI offers promising solutions to address these RRM challenges efficiently.

Network Selection

One of the key RRM challenges for the ICFC network is to select a suitable system for the users in intersection regions of urban and rural areas, such as users in the suburban zone. To this end, network-assisted clustering can be devised by leveraging collaboration among cellular and CF systems. This clustering problem aims to determine each user's suitable cluster (cellular or CF) based on user and network-specific features, such as traffic characteristics, QoS, users' and APs' locations, and network loading factor. Unsupervised ML models, such as density-based spatial clustering of applications with noise and expectation-maximization, can be exploited to solve such a clustering problem. Since practical networks only allow deploying trained models, the clustering model can be trained using DT in an edge cloud before deployment. Given the temporal variability in user density and activity, these models must be periodically updated, potentially utilizing a transfer-learning framework.

Network-Assisted User Clustering and Channel Estimation

Large-scale CF networks face the crucial challenge of channel estimation as the number of users and APs increase. Traditional user-centric clustering fails to consider site-specific factors because of using empirical path loss models and suffers from erroneous channel estimation due to pilot reusing and contamination. To address these challenges, we propose a network-assisted user clustering and channel estimation scheme leveraging DT technology. An edge-computing empowered centralized controller estimates user channels and determines optimization variables by considering local dynamic factors. Figure 2 depicts our proposed approach, which has the following three engines.

- *Ray Tracing Engine*: A ray tracing engine computes channel gain by tracing rays while considering BS, AAP, user positions, and obstacles' locations, which are obtained from the DT. Off-the-shelf ray tracing tools can be leveraged for this engine.
- *Channel Prediction Engine*: The channel computed by off-the-shelf ray tracing tools differs from the actual channel gains between a user and an AP. A deep learning (DL)-aided channel prediction engine is used to mitigate such error. DL models are trained offline and updated periodically to track network changes.
- *Resource Prediction Engine*: The resource prediction engine computes user-AP clusters, precoding, and combining vectors. Based on the estimated channel gains, it first determines user-AP clusters using a bipartite matching algorithm. Subsequently, it determines each AP's downlink precoding and uplink combining vectors for the associated users via the maximum ratio processing method.

The proposed approach executes four steps at each time slot. First, users' locations are updated based on feedback from the physical ICFC network. Second, the ray-tracing engine computes the channel gain of each user-AP link by adding ray-tracing profiles. Third, the channel gains calculated by the ray tracing engine are applied to the stored DL model to estimate user-AP links' channel gain. Finally, the resource prediction engine determines user-AP clusters, precoding, and combining vectors and sends the decision variables to the APs over reliable control channels. This approach has several advantages over existing schemes. First, it avoids pilot contamination, making channel estimation more reliable. Second, it reduces overhead by eliminating the need for periodic pilot assignments. Third, user locations can be accurately inferred using sensing-assisted models, allowing for proactive prediction of channel estimation and resource allocation.

Power Control and Interference Management

In the ICFC network, users experience 1) multi-user interference in the CF system since APs simultaneously transmit multiple users' precoded data and 2) cross-tier interference between CF and cellular systems. Transmit power control is crucial in managing such interference and maximizing the system's capacity. Traditional methods of controlling transmit power are often inadequate due to high signaling overhead and time-consuming optimization processes. Fortunately, reinforcement learning (RL) can help overcome these challenges. By incorporating an RL agent within each BS or AP and using local CSI as state variables, optimal power allocation decisions can be determined. These RL agents can be trained using edge computing and implemented in live networks. Additionally, transfer RL can be used for post-deployment updates of these models.

SECURITY FOR ICFC NETWORKS: A ZERO-TRUST APPROACH

To ensure secure operation of the ICFC communication systems, it is essential to establish trust through authentication. Trust management between aerial and ground nodes is critical for secure message transmissions. To achieve this, a zero-trust (ZT) protocol needs to be developed with a robust authentication mechanism and continuous user monitoring. In this context, HAPSs can be utilized to enhance security. To design the system, deep RL and latent Dirichlet allocation (LDA) can be used in tandem. Developing an efficient, lightweight, and distributed ZT protocol for the ICFC communication system requires two investigation phases.

- *Phase 1: Initial Authentication*: Development of a robust authentication mechanism for the ICFC network.
- *Phase 2: Behavior Analysis*: Continuous monitoring for anomaly detection in the ICFC network.

Phase 1: Initial Authentication

The implementation of ZT authentication mechanisms is pivotal to ensure a secure ICFC network. Advanced cryptographic techniques, public key infrastructure, and real-time behavioral analytics must be utilized to verify every access request. In light of the

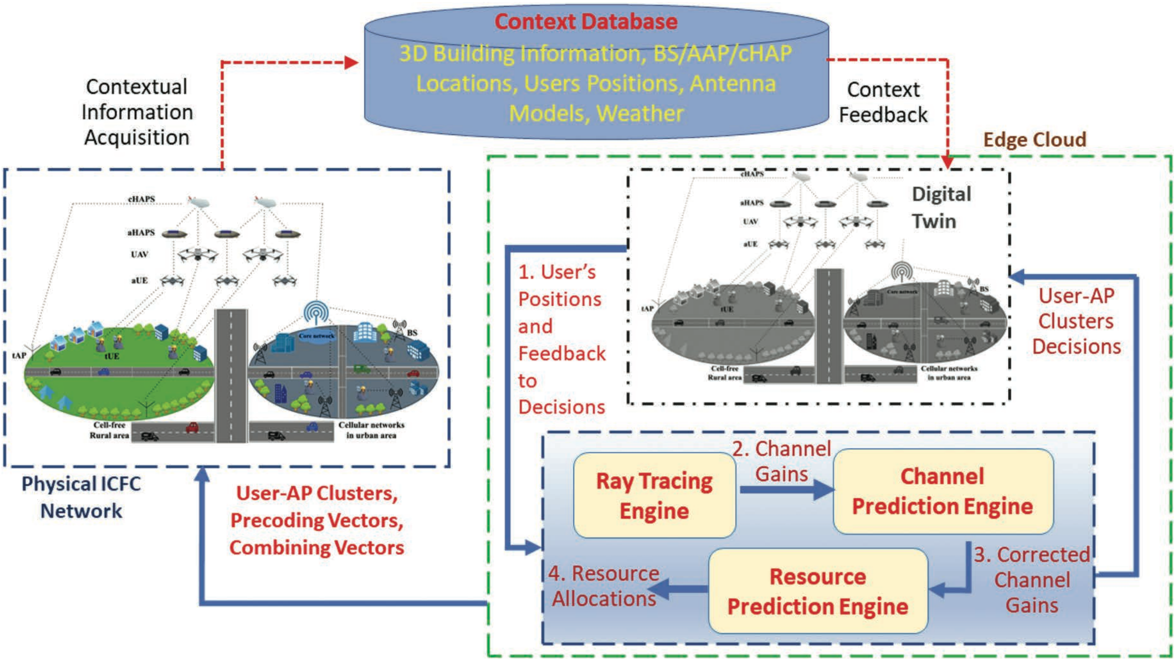

FIGURE 2. Schematic diagram of the proposed network-assisted user clustering and channel estimation scheme.

network's heterogeneity, a multimodal authentication protocol that incorporates traditional credentials, context-aware, and biometric factors is necessary. This protocol should be designed to identify the behavioral patterns of cellular and CF network users, facilitating a secure transition across the urban–rural divide.

Phase 2: Behavior Analysis

This phase aims to develop a sophisticated, data-based approach that integrates LDA for identifying behavioral patterns and DRL for detecting adaptive anomalies. Particularly, it addresses the risk posed by attackers who have successfully bypassed the initial authentication process. The machine language (ML)/AI framework considers the unique behaviors of users in cellular and cell-free environments, enabling the system to identify anomalies that are specific to each network type. The DRL component of the system enhances its detection abilities as the network evolves, ensuring that it remains robust and secure in the long run. This continuous learning loop, grounded in robust AI methodologies, will provide immediate security benefits and contribute to the long-term resilience and integrity of the ICFC network.

CHALLENGES AND FUTURE DIRECTIONS

- *Handover for ICFC communications*: To successfully implement an ICFC network that incorporates both ground and aerial APs with cellular networks, it is essential to reexamine the handover protocol. The protocol must account for various factors such as limited onboard energy, transmit power, load balancing, and interference to effectively minimize common issues like signal fluctuation, congestion, and interference.
- *ICFC network design for open radio access network (O-RAN)*: O-RAN revolutionizes wireless communications by separating hardware and software components, fostering flexibility and innovation. In CF component of ICFC networks, the O-RAN distribution unit (O-DU) and O-RAN radio access unit (O-RU) will represent the cHAPs and APs, respectively. Likewise, O-DU and O-RU will be implemented in BSs for the cellular network. Meanwhile, real-time and non-real-time RAN intelligent controllers (RICs) implemented in the core network will manage both systems of ICFC networks. Designing RIC for ICFC systems necessitates detailed research for network intelligence and programmability.
- *Interference management and seamless support for ground and aerial users*: ICFC offers ground and aerial nodes network coverage. Nevertheless, managing interference presents a significant challenge, as aerial APs can interrupt ground nodes, and vice versa. Utilizing AI tools to comprehend and anticipate interference behavior becomes crucial in creating effective interference cancellation algorithms and protocols.
- *Coexistence with enabling technologies of beyond 5G*: The advent of 6G cellular systems involves integrating cutting-edge technologies like terahertz communications, intelligent reflecting surfaces, joint sensing and communications, edge computing, and blockchain-enabled resource management. Further investigations are needed to integrate these technologies into the ICFC networks harmoniously.

CONCLUSION

This article introduced the integration of existing cellular communication networks with evolving CF communications, proposing the ICFC network infrastructure for B5G communications. We presented a systematic, realistic, robust, resilient DT framework for the proposed ICFC network. A network-wide resource management scheme was proposed that can lead to efficient approaches for network selection, user clustering, channel estimation, and power and interference management. We emphasized building secure communications by adopting a ZT network architecture. Several open challenges and future research directions were proposed to ensure an efficient deployment of the ICFC network for the next generation of cellular communication systems.

REFERENCES

1. J. Zhang et al., "Cell-free massive MIMO: A new next-generation paradigm," *IEEE Access*, vol. 7, pp. 99,878–99,888, 2019, doi: 10.1109/ACCESS.2019.2930208.
2. D. Niyato and E. Hossain, "Call admission control for QoS provisioning in 4G wireless networks: Issues and approaches," *IEEE Netw.*, vol. 19, no. 5, pp. 5–11, Sep./Oct. 2005, doi: 10.1109/MNET.2005.1509946.
3. I. F. Akyildiz et al., "6G and beyond: The future of wireless communications systems," *IEEE Access*, vol. 8, pp. 133,995–134,030, 2020, doi: 10.1109/ACCESS.2020.3010896.
4. W. Tong and G. Y. Li, "Nine challenges in artificial intelligence and wireless communications for 6G," *IEEE Wireless Commun.*, vol. 29, no. 4, pp. 140–145, Aug. 2022, doi: 10.1109/MWC.006.2100543.
5. P. Yang et al., "6G wireless communications: Vision and potential techniques," *IEEE Network*, vol. 33, no. 4, pp. 70–75, Jul./Aug. 2019, doi: 10.1109/MNET.2019.1800418.

6. M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 957–975, 2020, doi: 10.1109/OJCOMS.2020.3010270.
7. H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017, doi: 10.1109/TWC.2017.2655515.
8. E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 77–90, Jan. 2020, doi: 10.1109/TWC.2019.2941478.
9. E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4247–4261, Jul. 2020, doi: 10.1109/TCOMM.2020.2987311.
10. S. Elhoushy, M. Ibrahim, and W. Hamouda, "Cell-free massive MIMO: A survey," *IEEE Commun. Surv. Tut.*, vol. 24, no. 1, pp. 492–523, Firstquarter 2021, doi: 10.1109/COMST.2021.3123267.
11. S. Zhou, M. Zhao, X. Xu, J. Wang, and Y. Yao, "Distributed wireless communication system: A new architecture for future public wireless access," *IEEE Commun. Mag.*, vol. 41, no. 3, pp. 108–113, Mar. 2003, doi: 10.1109/MCOM.2003.1186553.
12. H. Zhang and H. Dai, "Cochannel interference mitigation and cooperative processing in downlink multicell multiuser MIMO networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2004, pp. 1–14, Dec. 2004.
13. H. Zhang, N. B. Mehta, A. F. Molisch, J. Zhang, and S. H. Dai, "Asynchronous interference mitigation in cooperative base station systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 155–165, Jan. 2008, doi: 10.1109/TWC.2008.060193.
14. S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: User-centric approach," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 706–709, Dec. 2017, doi: 10.1109/LWC.2017.2734893.
15. S. Chen et al., "A survey on user-centric cell-free massive MIMO systems," *Digit. Commun. Netw.*, vol. 8, no. 5, pp. 695–719, 2022, doi: 10.1016/j.dcan.2021.12.005.
16. S. Buzzi, C. D'Andrea, A. Zappone, and C. D'Elia, "User-centric 5G cellular networks: Resource allocation and comparison with the cell-free massive MIMO approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1250–1264, Feb. 2019, doi: 10.1109/TWC.2019.2952117.
17. M. S. Alam, G. K. Kurt, H. Yanikomeroglu, P. Zhu, and N. D. Đào, "High altitude platform station based super macro base station constellations," *IEEE Commun. Mag.*, vol. 59, no. 1, pp. 103–109, Jan. 2021, doi: 10.1109/MCOM.001.2000542.
18. M. Ke et al., "An edge computing paradigm for massive IoT connectivity over high-altitude platform networks," *IEEE Wireless Commun.*, vol. 28, no. 5, pp. 102–109, Oct. 2021, doi: 10.1109/MWC.221.2100092.
19. HAPS Alliance. [Online]. Available: https://hapsalliance.org/blog/advancing-global-connectivity-the-expanded-horizon-of-the-haps-spectrum/
20. H. Jelodar et al., "Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey," *Multimedia Tools Appl.*, vol. 78, no. 11, pp. 15,169–15,211, 2019, doi: 10.1007/s11042-018-6894-4.

IMTIAZ AHMED is an assistant professor with the Department of Electrical Engineering and Computer Science, Howard University, Washington, DC, 20059, USA. His research interests include artificial intelligence-aided physical layer designs, cell-free communications, integration of aerial and terrestrial communication networks, and communications with terahertz bands and intelligent reflecting surfaces. Ahmed received his Ph.D. degree in electrical and computer engineering from The University of British Columbia. He is the corresponding author of this article. Contact him at imtiaz.ahmed@howard.edu.

MD. ZOHEB HASSAN is an assistant professor in the Department of Electrical and Computer Engineering, Université Laval, Quebec, QC, G1V 0A6, Canada. His research interests include radio resource optimization, interference management, spectrum sharing, and optical wireless communications. Hassan received his doctorate degree in electrical and computer engineering from the University of British Columbia. Contact him at md-zoheb.hassan@gel.ulaval.ca.

MAJUMDER HAIDER is a Ph.D. candidate in electrical engineering at Howard University, Washington, DC, 20059, USA. His current research interests include wireless communication systems design and optimization applying artificial intelligence/machine learning techniques, intelligent reflecting surface, massive multiple-input, multiple-output communications, and industrial Internet of Things networks. Haider received his M.Sc. degree in electronics and communication systems engineering from City University of Bremen. Contact him at majumder.haider@bison.howard.edu.

KAMRUL HASAN is an assistant professor of electrical and computer engineering at Tennessee State University, Nashville, TN, 37209, USA. His research interests include explainable artificial intelligence (AI), secure wireless communications, computer networks, and AI-driven cybersecurity. Hasan received his Ph.D. degree in computational engineering from Old Dominion University. Contact him at mhasan1@tnstate.edu.