

Review article

Approximate computing in B5G and 6G wireless systems: A survey and future outlook

Hans Jakob Damsgaard^{a,*}, Aleksandr Ometov^a, Md Munjure Mowla^b, Adam Flizikowski^b, Jari Nurmi^a

^a Tampere University, Korkeakoulunkatu 1, Tampere, 33720, Finland

^b IS-Wireless, Puławska 45B, Piaseczno, 05-500, Poland

ARTICLE INFO

Keywords:

Approximate computing
Beyond 5G
Mobile networks

ABSTRACT

As modern 5G systems are being deployed, researchers question whether they are sufficient for the oncoming decades of technological evolution. Growing numbers of interconnected intelligent devices put these networks under tremendous pressure, demanding their development. Paving the way for beyond 5G and 6G systems, commonly denoted by B5G herein, therefore means seeking enablers to increase efficiency from different perspectives. One novel look on this is the application of inexact computations where *nine 9s* reliability is not needed, for example, in non-critical mobile broadband traffic. The paradigm of Approximate Computing (AxC) focuses on such areas where constrained quality degradation results in savings that benefit the users and operators. This paper surveys the state-of-the-art publications on the intersection of AxC and B5G systems, identifying and emphasizing trends and tendencies in existing work and directions for future research. The work highlights resource allocation algorithms as particularly mesmerizing in the former, while research related to Intelligent Reflective Surfaces appears the most prominent in the latter. In both, problems are often NP-hard and, thus, only solvable using heuristics or approximations, Successive Convex Approximation and Reinforcement Learning are most frequently applied.

1. Introduction

Modern wireless cellular systems constantly evolve in response to new applications being developed and in preparation for other envisioned use cases [1,2]. This has recently led to systems switching from 4G to 5G, with such transitions happening roughly once a decade. With the well-founded status of 5G systems only waiting for widespread deployment, researchers have turned their attention to the technological enablers for next-G – beyond-5G or 6G–systems [3–5]. For simplicity, we refer to all next-G systems, including beyond-5G and 6G, as B5G in the remainder of the paper. This transition has already begun. However, as the popularity of smartphones and Internet of Things (IoT) devices continues to grow, so does the amount of data being transferred [6,7], and although 5G has yet to be fully deployed, some already question whether it is sufficient for the next ten years [2].

The B5G systems are expected to enable new classes of applications, tying the physical world closer to the digital one by delivering near-instant responsiveness, previously unmatched reliability, and better scalability [8]. Specifically, researchers envision much greater numbers of wearables, implants, and other cyber–physical systems; fully

autonomous cars; wireless cameras and sensors with Artificial Intelligence (AI)-based data processing; robots in homes and health care; and a greater reliance on ultra-low latency communications in the industry [8–10]. Expectedly, these applications will change our lives drastically. However, they also imply enormous challenges for the underlying networking infrastructure, especially in terms of throughput, latency, reliability, security, and scalability [11].

Solving these challenges requires further research into several technologies identified as crucial for the success of B5G. Related technologies include AI, cell-free communications, dynamic network slicing, beamforming, Intelligent Reflective Surfaces (IRSs), big data analytics, massive Multiple-Input Multiple-Output (MIMO), extremely high-frequency air interfaces, and new security primitives [12,13]. Most of these are computationally heavy, yet to keep with current trends and avoid enormous increases in overall energy consumption, they must be implemented optimally and efficiently. However, as we have reached or gotten near the end of helpful historical trends such as Moore's law and Dennard scaling, we can no longer expect to match these requirements simply by microprocessors becoming faster every

* Corresponding author.

E-mail addresses: hans.damsgaard@tuni.fi (H.J. Damsgaard), aleksandr.ometov@tuni.fi (A. Ometov), m.mowla@is-wireless.com (M.M. Mowla), a.flizikowski@is-wireless.com (A. Flizikowski), jari.nurmi@tuni.fi (J. Nurmi).

<https://doi.org/10.1016/j.comnet.2023.109872>

Received 20 April 2023; Received in revised form 2 June 2023; Accepted 8 June 2023

Available online 14 June 2023

1389-1286/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

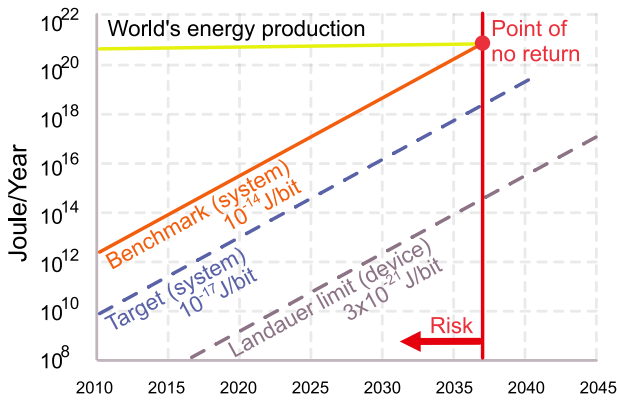


Fig. 1. Energy consumption in computing vs. World energy production.

year [14] – modern devices with computing capabilities are expected to require more energy than the energy resources can provide by 2040, according to Semiconductor Industry Association [15], see Fig. 1. Radical changes are needed to address this issue.

In this direction, Approximate Computing (AxC) appears as an alluring technology [16]. In AxC, approximations are applied at different system levels to reduce computational complexity, transforming into savings in various system metrics [17,18]. The paradigm includes a wide range of techniques that fall into two categories [19]:

Computing on approximate data means applying approximations at the application or architecture level. The textbook example is the quantization of floating-point operations into simpler fixed-point equivalents. Yet, despite being a powerful and frequently applied technique, quantization alone usually fails to fully exploit applications’ error resilience. Supplementing it with orthogonal techniques like loop perforation and skipping, neural approximation, inexact arithmetic, etc., is often possible and may increase savings [18,20,21].

Computing on unreliable hardware means introducing approximations at the circuit level by using unintentionally faulty circuits or purpose-designed ones. Putting faulty chips that produce acceptable errors to use can increase production yield [22], while purpose-designed ones may bring significant savings from, e.g., over-scaled operating voltage or frequency, reduced refresh rate Dynamic RAM, approximately synthesized logic, etc. [18,20,23].

Using techniques from either or both of these categories enables interesting time-energy-accuracy trade-offs. Generally, applications that produce outputs evaluated on their *acceptability* rather than correctness are likely candidates for approximation. The most well-known examples of such are from the AI and Big Data analysis fields. Nevertheless, many algorithms within B5G systems pertain the same feature, including scheduling, resource allocation, user-channel pairing, and other optimization algorithms, as illustrated in Fig. 2. When handling non-critical applications that do not demand *nine 9s* reliability like Enhanced Mobile Broadband (eMBB) traffic, these algorithms may be approximated to induce benefits for users and operators. The potential gains thereof are plentiful and include:

- **Improved energy efficiency** as lower-precision computations require fewer operations and may, for example, be carried out at lower operating voltages or frequencies to reduce energy consumption [18, 20]. This can be crucial for wireless devices with limited energy available.
- **Greater throughput** resulting from approximate operations typically having lower latencies than their accurate equivalents. Exploiting this even without varying operating voltage or frequency may increase system throughput [18].

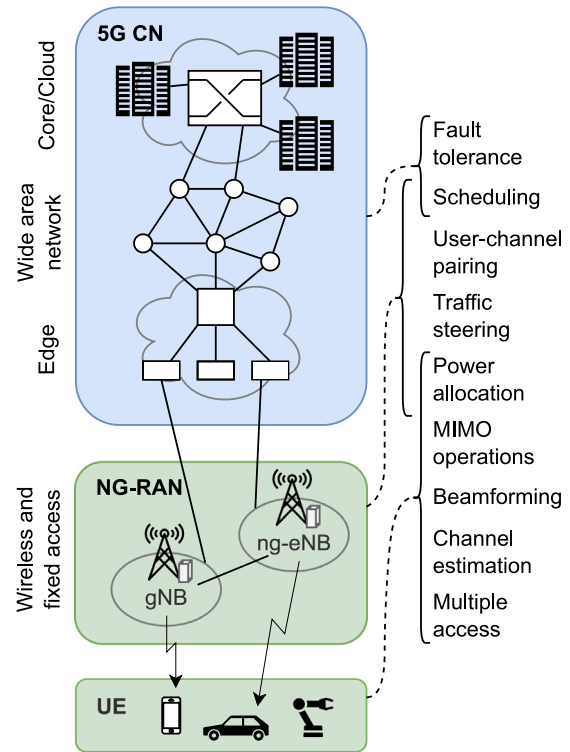


Fig. 2. Illustration of resources [24, Figure 2-1] commonly used to implement the 3GPP-standardized 5G architecture [25, Figure 4.1-1] combined with approximable applications highlighted in the present survey. Acronyms cf. [25].

- **Enhanced robustness** as systems integrating AxC inherently tolerate a certain amount of imprecision in their computations, they also show greater resilience to faults and errors [26]. This is particularly important in the context of wireless networks that may be subject to interference, signal degradation, and other sources of noise.
- **Strengthened flexibility and adaptability** from reducing the complexity and compute requirements of the components of B5G algorithms, improving overall system scalability. With AxC, a system can be tailored more closely to the accuracy needs of different applications, enabling customization of computing solutions to suit diverse application requirements [27]. This in turn may enable deploying larger and more complex wireless networks or addressing the demands of B5G use cases better [28].
- **Increased cost efficiency** arising from the aforementioned reduced hardware and processing requirements. These reductions may not only bring savings in hardware and power consumption, but also in system maintenance as known faulty circuits may continue to operate so long as they deliver acceptable accuracy levels [19,22], making AxC an attractive option for deploying cost-effective B5G systems [13,29].

In this paper, we keep these benefits in mind and survey the state-of-the-art in AxC applied to B5G systems, focusing on techniques applied at the application level. We do not provide a comprehensive survey of approximate computing techniques nor cover all details expected from B5G systems. For those interested in these two topics, please refer to [17,18,20,23,30–32] and [2,8,13,24,33,34], respectively.

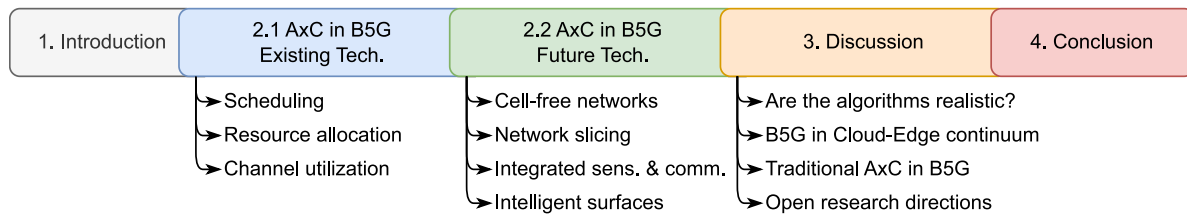


Fig. 3. Structure of this paper and the content of its sections.

The main contributions of this paper are:

- A literature survey on algorithms for B5G systems that lend themselves to any aspect of approximation;
- An overview of problem formulations, algorithms, and KPIs used in the literature;
- Identification of future research directions and the potential role of AxC in these.

We have followed the modified PRISMA guidelines [35] for reproducibility purposes. As such, we defined the following search expression with keywords found in relevant articles: (approximat* OR inexact OR inaccurate OR ‘good enough’) AND (B5G OR ‘beyond 5g’ OR 6g). Using this expression, Scopus [36] returned 269 results, which we filtered in two rounds: first, a coarse filtering by paper titles, and second, a fine filtering by abstracts. In both rounds, we applied the following exclusion criteria: **C1** no relation to AxC or B5G systems; **C2** work not applying approximations at runtime; **C3** reliance on generally inaccessible technologies like airplanes and Unmanned Aerial Vehicles; and **C4** full text not available. We also filter clearly iterative work to include only the most recent publication and limit inclusions to direct search results. This results in the set of 80 papers presented in the following.

The remainder of this paper is structured as illustrated in Fig. 3. First, Section 2 covers the main technical part of the survey, reviewing and summarizing work on existing and future technologies. Next, Section 3 discusses overall observations of the survey combined with ideas for future work, and last, Section 4 concludes the paper.

2. Approximate computing in B5G systems

Having motivated our survey and presented the strategy for finding relevant work, let us first consider a set of publications that do not naturally fall into the two categories above. We identify three papers that present hardware architectures relevant to the physical layer of future wireless communication systems.

Specifically, Madanayake et al. [37] propose several approximate Discrete Fourier Transform implementations suitable for beamforming antenna arrays. Their designs avoid costly multipliers, utilizing sparse matrices with signed unit-valued entries, only of which is complex. Doing so allows for significant area savings with negligible impact on beamforming performance. Idrees et al. [28] similarly focus on the Digital Signal Processing (DSP) side of wireless networks, applying inexact arithmetic in the Base Station (BS)’s pulse shaping filters and the User Equipment (UE)’s decoders. As a result, they reduce power consumption drastically at the expense of increased Bit Error Rate (BER). Finally, Morsali et al. [38] focus on massive MIMO transceivers and propose a novel approach to construct Radio Frequency-based analog Deep Neural Network (DNN)s, enabling them to implement the transceiver with a digital DNN for signal processing and an analog one for beamforming. Being limited to Downlink (DL) transmissions, the resulting architecture achieves the same performance as a fully digital equivalent while being more adaptive to transmitter/receiver mappings and consuming less hardware.

All three papers clearly highlight the potential benefits of applying AxC at the architectural level. However, they are limited in their scope to the DSP parts of communication systems. Broadening the scope to the application level opens many more opportunities for approximations. As such, we will first consider such publications that apply AxC to well-established technologies, most of which are already in use in 5G and prior systems.

2.1. Existing technologies

The remaining work is challenging to categorize; thus, there may be better ways to group it than by key topics as in the following. The publications are characterized by mostly *not* applying approximations in the notion common to AxC. Rather, they propose various approximate algorithms for problems that are difficult, if not nearly impossible, to solve optimally in the field with brute force algorithms. Within the scope of this survey, we consider these to be approximations at the application level. Our survey reveals that such approximations are frequently applied to resource allocation algorithms, as we will see later. We begin by surveying papers on scheduling algorithms.

2.1.1. Scheduling

Scheduling refers to the problem of determining when to perform particular actions based on resource availability, system constraints, and user expectations, and it is vital for perceived Quality of Experience in multi-user systems. We include five papers on this topic that focus on mostly different scenarios, as highlighted in Table 1. First, Li et al. [39] consider channel and queuing-aware scheduling of a multi-user DL. They evaluate a time-multiplexed scheme that first schedules the user with the best channel conditions in each time slot, applying approximations by considering users individually and modeling their number of queued packets with Markov chains. Their model matches the simulated performance well. Similarly, Salameh et al. [40] consider opportunistic scheduling of non-contiguous Orthogonal Frequency-Division Multiple Access (OFDMA) Resource Blocks in a cognitive radio network, improving spectrum utilization flexibility and efficiency. They formulate it as a Binary Linear Programming (BLP) problem and propose an approximate polynomial-time algorithm to find the solution, claiming double the performance of contiguous time and contiguous frequency algorithms.

Second, Almekhlafi et al. present work on scheduling Ultra-Reliable Low-Latency Communication (URLLC) traffic onto resources occupied by ongoing eMBB traffic through puncturing [41] or superposition [42], i.e., by either overriding eMBB symbols (forcing them to be re-scheduled) or transmitting them simultaneously with differently modulated URLLC symbols. Either technique is essential to satisfy the URLLC packets’ strict latency requirements. In the former work, they propose to minimize interference on URLLC symbols using symbol similarity, sought after by one of three linear-time routines. In the latter, they formulate the problem as a Mixed Integer Non-Linear Programming (MINLP) problem and propose polynomial-time approximate algorithms to solve it. Both report improvements in Symbol Error Rate (SER) and the number of users served.

And last, Beitollahi et al. [43] consider a Federated Learning (FL) environment in which many devices must communicate their updated

Table 1

Overview of the type of approximate algorithm used to solve the surveyed scheduling problems, and their related KPIs.

Class	Ref.	Problem	Formulation	Algorithm	KPIs
Scheduling	[39]	Channel and queue status-aware user scheduling	Non-convex optimization	Mean-field approximation	Latency
	[41]	Similarity-based eMBB and URLLC scheduling with puncturing	—	Linear approximation	SER, latency, reliability
	[43]	Scheduling FL communication in a wireless network	MINLP	Outer approximation	Latency
	[42]	Combining eMBB and URLLC traffic with superposition	MINLP	Hungarian method	Loss, reliability, latency
	[40]	Resource block scheduling in cognitive radio	BLP	Relaxation and rounding	Throughput, fairness

Table 2

Overview of the approximate algorithms used to solve the surveyed resource allocation problems and their related KPIs.

Class	Ref.	Problem	Formulation	Algorithm	KPIs
Power Allocation	[44]	Energy efficiency optimization in delta-OMA	MBNLP	Outer approximation	Energy/spectral efficiency
	[45]	Total power optimization in OFDMA-NOMA	Non-convex optimization	SCA	Throughput, transmit power
	[46]	Massive MIMO performance optimization with imperfect hardware	Non-convex optimization	SCA	Sum rate
	[47]	Massive MIMO performance optimization for FL users	Non-convex optimization	SCA	Throughput
	[48]	Throughput optimization in hybrid NOMA	MINLP	Outer approximation	Admissions, associations, power/user, throughput
	[49]	Energy efficiency optimization in hybrid NOMA	Concave fractional optimization, MINLP	Outer approximation	Admissions, associations, power/user, throughput
	[50]	Power allocation in heterogeneous multi-carrier NOMA	MINLP	SCA, iterative approximation	Throughput
	[51]	Bandwidth allocation in multi-user Multiple-Input Single-Output	Non-convex optimization	SCA	Sum rate, transmit power
User-Channel Pairing	[52]	Clustering in user-centric networks	Iterative optimization	DNN	Sum rate
	[53]	Channel allocation in URLLC	Non-convex optimization	Inner approximation	(Secrecy) sum rate
	[54]	User clustering in ultra-dense networks	Non-convex optimization	Variable relaxation, SCA	Energy efficiency
	[55]	User association in in-X networks	Centralized graph coloring	DNN	Loop failure probability
	[56]	Beamforming and channel allocation in multi-user URLLC	Non-convex optimization	SCA	Decoding error probability
	[57]	User pairing in NOMA communications	—	Evolutionary algorithm	Sum rate
	[58]	Channel allocation in hybrid OFDMA-NOMA	Non-convex optimization	SCA	Sum rate
Traffic Steering	[59]	Traffic steering in latency-guaranteed virtualized networks	Mixed Integer Linear Programming (MILP)	SCA, RL	Energy efficiency, computation time
	[60]	Traffic steering in open-RAN networks	MINLP	Variable relaxation, SCA	Throughput, latency
IAB	[61]	Association and power allocation in IAB	Non-convex optimization	SCA	Sum rate
	[62]	Bandwidth allocation in IAB	MILP	Iterative approximation, GNN	Spectral efficiency
Random Access	[63]	Active device detection	Odd total squared correlation	Approximate message passing	SNR, perfect recoveries, missed detections, false alarms
	[64]	Distributed congestion prediction for latency optimization	Non-convex optimization	RL	Latency
Fault Tolerance	[65]	Hypervisor adaptation optimization	k -pathwise disjoint cover	Heuristic, ILP	Acceptances, migrations
	[66]	Deterministic fault tolerant service function chaining	k -pathwise disjoint cover	Heuristic	Bandwidth, survivability
Others	[67]	Network service federation admission control	Markov decision process	RL	Optimality gap
	[68]	Spectrum assignment to service providers	—	—	Capacity, energy/ spectral/ cost efficiency
	[69]	Task scheduling and power allocation in Fog computing	Non-convex optimization	SCA	Latency
	[70]	Disparate groups QoS optimization	Markov decision process	—	QoS

models to a centralized controller. They propose making scheduling decisions for several update iterations – frames – centrally, taking transmission errors into account, and formulate this as a MINLP problem, solved using an approximate algorithm. The approach outperforms traditional round-robin scheduling.

2.1.2. Resource allocation

While scheduling may be seen as a branch of resource allocation, it is far from the only one. As we display in this section, resource allocation is central to many parts of B5G systems. We group papers focusing on similar topics, but readers may notice some disparity between proposals within the groups. Their main points are listed in Table 2.

Power allocation is a valuable technique for two purposes: (1) managing or limiting total transmit power and (2) enabling transmit power-based Non-Orthogonal MA (NOMA) schemes [71]. The latter appears particularly attractive and is described as a way to increase spectral efficiency in B5G systems by Hashemi et al. [44], who analyze the performance of Uplink (UL) delta-Orthogonal MA (OMA) and optimize its energy efficiency. They model the situation as a Mixed Binary

Non-Linear Programming (MBNLP) problem and apply the outer approximation to solve it, leading it to outperform NOMA and OFDMA alternatives under certain conditions. Yin et al. [50] instead jointly optimize device assignment and power allocation in a NOMA system using iterative approximation and SCA, reporting significant improvements in network throughput.

Minimizing total transmit power is the focus of two other papers. Firstly, Salameh et al. [45] focus on a hybrid OFDMA-NOMA network, modeling the minimization problem as a non-convex optimization. Given user activity, rate demand, and power consumption constraints, they apply Successive Convex Approximation (SCA) to solve this problem. The resulting system reduces transmit power over OFDMA drastically. Secondly, Vu et al. [51] propose a scheme to collaboratively maximize user fairness and minimize transmit power in an over-loaded multi-user system. The problem is non-convex and solved using SCA, like [45]. Supplemented by a user clustering policy, the scheme outperforms other state-of-the-art algorithms.

Ghafoor et al. present two very similar approaches to optimizing NOMA heterogeneous networks' throughput [48] and energy efficiency [49]. In both cases, they formulate MINLP problems and apply the outer approximation to estimate their solutions, reporting improved

network performance in the number of admissions, associations, energy efficiency, and throughput over a macro BS architecture.

Wang et al. [46] and Farooq et al. [47] aim to optimize massive MIMO systems for different cases, respectively, when using imperfect hardware and under Federated Learning (FL)-related traffic. Like [45], both use non-convex optimization models and apply SCA to solve them, achieving improved sum rates.

User-channel pairing is relevant for allocating channel resources to users in BSs, a problem closely linked to power and bandwidth allocation optimizations. The publications on this topic are diverse, focusing on many different network configurations. Starting with ultradense networks, Zhu et al. [54] model the collaborative optimization of user association and channel allocation as a non-convex problem. They apply variable relaxation to convert the problem to a convex one, which they proceed to solve, nearly doubling system energy efficiency.

Dai et al. [53] and Li et al. [56] both present approximate algorithms for non-convex optimization of channel allocation in networks supporting URLLC traffic; Dai et al. apply a game theory approach and inner approximation, and Li et al. use SCA instead. The authors report improved (secrecy) sum rates.

Two other papers consider NOMA-based networks. Narottama et al. [57] seek to optimize user pairings. They propose an evolutionary algorithm inspired by quantum systems and show how it outperforms the typically applied random algorithm. Abdel et al. [58] seek to maximize the sum rate in hybrid OFDMA-NOMA by clustering users and assigning them to transmit powers. As in [56], the problem is non-convex and solved by SCA, outperforming conventional algorithms.

A final pair of papers both propose using DNNs in place of compute-heavy algorithms. Specifically, Liu and Zhang [52] focus on user-centric networks and use a DNN to approximate a weighted minimum mean square error algorithm for user clustering and allocation. Adeogun et al. [55] replace a costly graph coloring algorithm for channel selection in so-called in-X networks (e.g., in-vehicles) with a DNN. The papers claim comparable performance at two orders of magnitude lower complexity and minimal input data, respectively.

Traffic steering is considered in another pair of publications. Pham [59] and Kavehmadavani et al. [60] consider virtualized networks, the former on maximizing cost efficiency in service function chaining while guaranteeing delay, and the latter on simultaneously maximizing eMBB throughput and minimizing URLLC latency. Both model their problem as a non-convex optimization and solve it using Reinforcement Learning (RL) and SCA, respectively, showing improvements in the targeted Key Performance Indicators (KPIs).

Integrated Access and Backhaul is a central concept in B5G systems, ensuring low-latency connectivity by letting next-Generation Node Bs wirelessly relay traffic between one another over frequency bands otherwise used to service UEs rather than directly to a wired backhaul link. This problem is considered in two papers. Zhang et al. [61] jointly optimize user association and power allocation, like [48, 49], utilizing a novel backhaul capacity-aware utility function. Much like the related work, the problem is a non-convex optimization solved using SCA, resulting in a greatly improved network sum rate. Orthogonally, Harris et al. [62] consider bandwidth allocation and propose an iterative approximate algorithm and a Graph Neural Network (GNN)-based algorithm to solve it. Their experiments reveal that combining the two is a practical solution that improves spectral efficiency.

Random access operations must be handled rapidly in B5G systems to satisfy their availability and latency requirements [2]. This operation, however, may be tackled from two sides: the UE's and the BS's. Liu et al. [63] focus on the latter, presenting a scheme targeting massive Machine Type Communication. They optimize the active device detection stage of envisioned grant-free random access using the correlation of sets of received binary signatures, reducing the

complexity with approximate message passing. The resulting algorithm outperforms the commonly applied Gaussian matrix technique used in compressed sensing.

Bekele and Choi [64] focus on the UE's perspective and propose a distributed RL algorithm to predict Access Point (AP) congestion. The authors show how selecting a low-congestion AP can vastly reduce random access latency and increase the probability of successful access, and display that their algorithm outperforms 3GPP's default algorithm.

Fault tolerance becomes ever more critical as requirements for network availability increase beyond seven nines [2], challenging both physical and virtualized network cores. Two papers focus on the latter, presenting algorithms for improving reliable service function chaining. Specifically, Mogyorosi et al. [65] focus on hypervisors, minimizing their overheads and maximizing their fault tolerance while ensuring support for instantaneous changes in traffic, while Zheng et al. [66] propose a fault-tolerant function slicing approach. Both formulate their cases as k -pathwise disjoint cover problems and use heuristics to solve them, improving service acceptance and survivability.

Others include work that does not naturally fit the above categories. Bakhshi et al. [67] consider admission control in network service federation across providers and, thus, administrative domains. They formulate the problem as a Markov decision process but propose using Q-learning or R-learning to solve it in practice, revealing R-learning as the superior algorithm that achieves near-optimal performance.

Orthogonally, Fei et al. [69] consider a Fog computing environment [72] and jointly optimize task scheduling with UL/DL power allocation. Like much of the above work on power allocation [45–47,51], the authors formulate the problem as a non-convex optimization and solve it using SCA. The results indicate improved user fairness and reduced system latency. Iyer [68], again orthogonally, proposes a dynamic scheme for spectrum assignment across different service providers according to their instantaneous number of subscribers. Through simulation, they show improved capacity and efficiency of the spectrum, energy, and cost.

Last, Ahmed et al. [70] model the Quality of Service (QoS) in heterogeneous networks with two classes of users: privileged and shadowed. They describe two use cases with different user class mixtures, formulate them as Markov models, and evaluate different admission restrictions. Yet, they do not propose a way to optimize for QoS.

2.1.3. Channel utilization

Having surveyed much work on resource allocation, let us shift our focus to more channel-related topics. We begin by considering work on various techniques for massive MIMO communications: precoding, demodulation, and data detection, and continue with publications on beamforming, channel estimation, multiple access, and security. While some of these topics are part of the above publications, they are not their main topics. Once again, we tabulate the related publications in Table 3.

Massive MIMO is expected to improve signal strength for individual devices and spectral efficiency in general at the expense of much higher computational demands. Meeting these is important for enabling the use of higher frequency bands [8]. As such, we survey some relevant algorithms.

Firstly, four papers consider precoding algorithms used to minimize inter-user interference. Yan et al. [73] and Busari et al. [74] present hybrid digital–analog precoding schemes whose linear algebra is approximated, showing how doing so minimally affects their performance. Similarly, Ribeiro et al. [76] propose a fully digital precoding scheme that combines user grouping with mathematical approximations, achieving near-optimal performance at much lower complexity. Orthogonally, Li et al. [75] consider a rate-splitting scenario in which multiple parallel transmissions of the same data serve to reduce interference, improving overall channel performance. They optimize the achievable rate by modeling different precoder designs as a non-convex

Table 3

Overview of the approximate algorithms used to solve the surveyed channel-related problems and their related KPIs.

Class	Ref.	Problem	Formulation	Algorithm	KPIs	
Multiple-Input Multiple-Output	[73]	Hybrid precoding for THz communications	—	Math. approximation	Spectral efficiency	
	[74]	Hybrid precoding for THz communications	—	Math. approximation	Spectral efficiency	
	[75]	Rate-splitting optimization	Non-convex optimization	SCA	Sum rate	
	[76]	Digital precoding for extra-large MIMO	—	Math. approximation	SNR, sum rate	
	[77]	MIMO Demodulation	—	Math. approximation	FER	
	[78]		—	Math. approximation	FER	
	[79]		—	Expectation propagation, GNN	SER	
	[80]		—	Gaussian approximation, DNN	BER	
	MIMO Data Detection	[81]	—	Coordinate descent	BER	
		[82]	—	Variable relaxation	BER	
Beamforming		[83]	Hybrid beamforming for mmWave communications	Non-convex optimization	SCA, linear approximation	Energy efficiency
		[84]	Hybrid beamforming for mmWave communications	Non-convex optimization	RL	Sum rate, energy spectral efficiency
	[85]	Beamforming and user-cloud association optimization	Non-convex optimization	Fractional program, inner approximation	Energy efficiency	
Channel Estimation	[86]	Activity detection and channel estimation for grant-free random access	MSE minimization	DNN, approximate message passing	MSE	
	[87]	Channel and phase noise estimation for GFDM communications	Least squares	Linear approximation	Sum rate, MSE, BER, SNR	
	[88]	Channel estimation for high mobility OFDM communications	Compressed sensing	DNN	MSE	
	[89]	CSI prediction	Least squares	DNN, quantization	NMSE	
	[90]	CSI prediction	Least squares	DNN	Beamforming gain, NMSE, BER	
	[91]	Channel estimation for OFDM communications	MSE minimization	CNN, residual NN, RNN	NMSE, latency	
Multiple Access	[92]	OMA and NOMA schemes for orbital angular momentum channels	Non-convex optimization	SCA, iterative approximation	Channel capacity,	
	[93]	CSI-informed beamforming approach for rate splitting MA	Non-convex optimization	Sample average approximation	Sum rate	
	[29]	Multi-dimensional MA protocol to match UE QoS constraints	Non-convex optimization	SCA	Throughput	
	[94]	Precoding, stream splitting and transmit power optimization for rate splitting	Non-convex optimization	SCA	Sum rate	
Security	[95]	Secrecy rate optimization	Non-convex optimization	SCA, DNN	Secrecy rate, latency	
	[96]	Cost and secrecy rate maximization in femtocell BSs	Game theory, Bernstein approximation	Iterative approximation	Cost, secrecy rate	
	[97]	Attack rate detection	—	Math. approximation	Latency, security level	
	[98]	Key length impact on execution time	—	Math. approximation	Security level, time	
	[99]	Qualitative analysis on approaches of hardware-embedded security approximation	—	—	Execution time	
	[100]	Cryptographic primitive execution time	—	—	Latency, security level	

problem and assuming joint decoding, commonly applied in rate splitting. Applying SCA, their results show that rate splitting outperforms traditional unicast transmissions.

Secondly, two publications consider demodulation, i.e., identifying subscriber information in the masses of radio signals received at a MIMO antenna. In both their papers, Bakulin et al. [77,78] apply a non-Gaussian approximation of the prior distribution of parameters for demodulation and use Newton's method to optimize them. They report that the proposed schemes reduce the frame error rate under different channel conditions.

Lastly, another four papers present algorithms for data detection, i.e., detecting (UL) traffic in multi-user, multi-antenna interference-filled systems. Essentially tackling the same problem, they propose four different detector designs: (1) Kosasih et al. [79] combine expectation propagation with GNNs for a design that works well under strong multi-user interference; (2) Bicaïs et al. [80] propose two designs using Gaussian approximation or DNNs; (3) Hasan et al. [81] use coordinate descent (another approach to solving convex optimizations) to design an adaptive detector; (4) Kumar et al. [82] propose using QR-decomposition and maximum likelihood combined with several mathematical estimations. All four papers report improved SER or BER over traditional schemes at lower computational complexity.

Beamforming is essential in ensuring good signal strength despite issues with the attenuation of high-frequency signals. Minimizing related computations can improve its performance, especially in

multi-user scenarios. Three publications consider this scenario, modeling it as a non-convex optimization but maximizing its energy efficiency with different different approximate algorithms. Specifically, Kaushik et al. [83] combine SCA and linear approximations, Fozi et al. [84] use RL with per-user QoS constraints and describe a novel training algorithm based on approximate message passing, while Reifert et al. [85] jointly optimize energy efficiency and user-Cloud association with fractional programming and inner approximation. All three report that their algorithms outperform traditional approaches.

Channel estimation refers to computing the propagation-related parameters of a wireless channel in a General Frequency-Division Multiplexing (GFDM) system. It is an inherently approximate process and, thus, an obvious place to apply AxC. While none of the surveyed work uses typical AxC techniques, they propose approximate algorithms, as seen above. Four publications present different algorithms specifically for this purpose.

Qiang et al. [86] jointly perform activity detection and channel estimation with a purpose-designed deep learning algorithm based on approximate message passing. Mohammadian and Tellambura [87] jointly perform channel and phase noise estimation with an iterative approximation algorithm and propose another two algorithms for data detection and phase noise compensation. Li et al. [88] formulate the situation as a compressed sensing problem in a high mobility case and use a DNN to solve it. Lastly, Shi et al. [91] propose using various types of NNs before converging on an Long Short-Term Memory-based architecture to solve the problem for both stationary and non-stationary

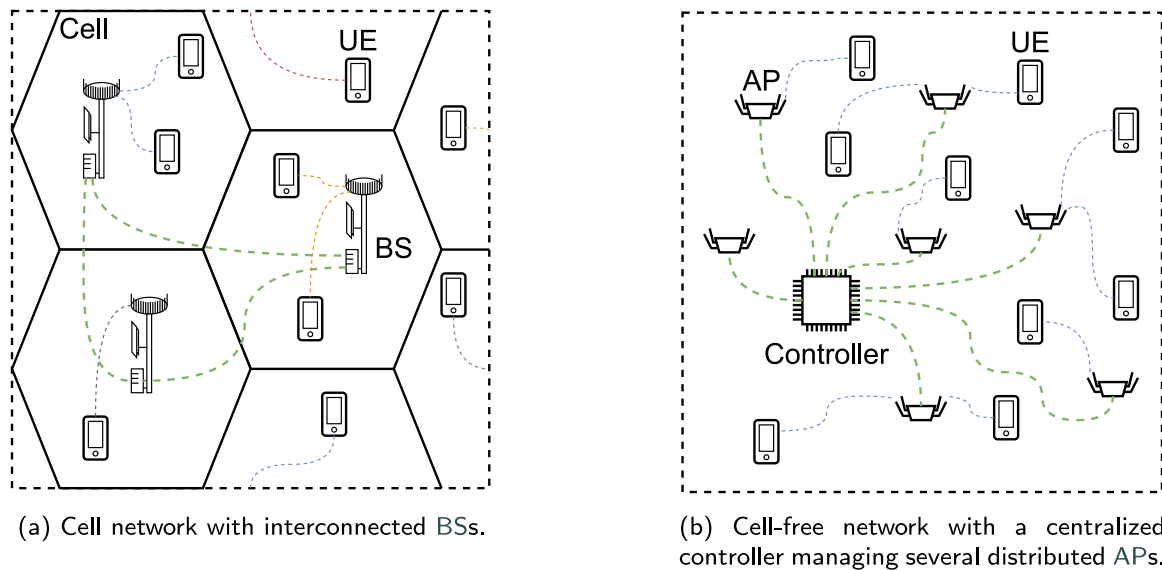


Fig. 4. Figurative description of (4(a)) a traditional cell network and (4(b)) an envisioned cell-free network [101]. Only the network fronthaul is shown.

environments. In their respective use cases, all four papers report improved Mean SE (MSE) or Normalized Mean SE (NMSE).

Another two papers present techniques to predict Channel State Information (CSI), otherwise communicated as a response from UE to BS with the overheads that follow. Both publications use deep learning techniques to solve this, but with different details. Zhang et al. [89] present a novel complex-valued Convolutional Neural Network (CNN) architecture, while Belgiovine et al. [90] present a simpler Multi-Layer Perceptron architecture; both report improved error rates over more complex least squares-based algorithms.

Multiple Access schemes may also be optimized in various ways, often requiring large amounts of computation. Four publications do precisely this, but each for very different cases. Specifically, Jian et al. [92] consider improving channel capacity with so-called orbital angular momentum encoding and explore its functionality under OMA and NOMA schemes. They optimize for optimal mode division in OMA and sum capacity in NOMA using SCA and report improved capacity over simpler schemes. Ahmad et al. [93] instead consider a rate splitting scheme for Cloud networks, optimizing its sum rate using beamforming, user clustering, and available CSI. The problem is non-convex, as in [92], and solved using SCA. The results show improved sum rate.

Mei et al. [29] propose dynamically switching between MA schemes depending on individual UE needs. They combine metrics from interference cancellation and resource allocation to cluster UEs with different requirements that can be served simultaneously. The problem is non-convex and solved by SCA, outperforming both rate-splitting and MIMO-based NOMA. Lastly, Khisa et al. [94] focus on a MIMO-based rate splitting scheme. They jointly optimize precoding, stream splitting, and transmit power in a non-convex problem, again solved using SCA. The proposed scheme outperforms its state-of-the-art alternatives.

Security is the focus of the last group of papers, which seek to optimize the secrecy rate of two different network types in environments with eavesdroppers and untrustworthy relay nodes. Bastami et al. [95] jointly optimize beamforming and transmit power, modeled as a non-convex problem and solved using SCA. As low latency is required in practice, they also propose using a lower-complexity DNN trained using their original model. Xie et al. [96] optimize cost and secrecy rate assuming femtocell BSs using a game framework in combination with Bernstein approximation. As a separate subgroup of papers not directly related to AxC, lightweight security on wearables and resource constrained devices appears. The main concepts addressed are trade-offs between the needed levels of security/privacy vs. cryptography primitive strength through, e.g., the length of asymmetric/symmetric keys

used for data encryption/decryption [97,98]. Those aspects directly affect the execution time and energy consumption and can be considered as approximation approaches on a higher level of abstraction. Finally, the book chapter [99] delves deeper into general applications of AxC in hardware security, e.g., approximate arithmetic circuits. The majority of papers report several simulation/measurement results but, commonly, do not provide any comparisons to other algorithms.

2.2. Future technologies

Having surveyed approximate algorithms applied to well-established technological enablers, we now focus on technologies that have yet to see an application in practice but are expected to become essential parts of future B5G systems. We classify work into four categories: cell-free networks, network slicing, Integrated Sensing and Communication (ISAC), and Intelligent Reflective Surfaces (IRSs) or Reconfigurable Intelligent Surfaces (RISs). Again, the publications are listed in Table 4.

2.2.1. Cell-free networks

Despite implementing massive MIMO and beamforming, B5G systems are expected to provide excellent, broad coverage in areas not coverable by traditional means. This is not always achievable, and thus, some network architects are considering a cell-free architecture in which, rather than segmenting areas into cells covered by a single BS, the cells are covered by numerous APs connected to a centralized controller, as illustrated in Fig. 4. This introduces new challenges and scalability issues that must be solved to ensure reliable, high-performance operation [102].

Four papers consider such matters. First, Shao et al. [103] propose a scheme to enable multiple APs to collaboratively perform activity detection. Their algorithm is based on the covariance between so-called *device state vectors* of neighboring APs and forward-backward splitting training. It achieves lower error rate than centralized schemes and traditional cell-based massive MIMO networks. Second, Shaik et al. [104] consider soft detection in UL massive MIMO, describing a distributed algorithm that passes approximate bit probabilities between APs until predictions are certain enough to facilitate decoding. The authors, unfortunately, fail to provide an evaluation of their scheme. Third, Peng et al. [105] devise a coordination scheme for DL traffic relaying, aiming to maximize the sum rate of clustered NOMA users. This optimization is non-convex but solvable with an iterative convex approximation. Owing to its dynamic nature, the scheme outperforms its

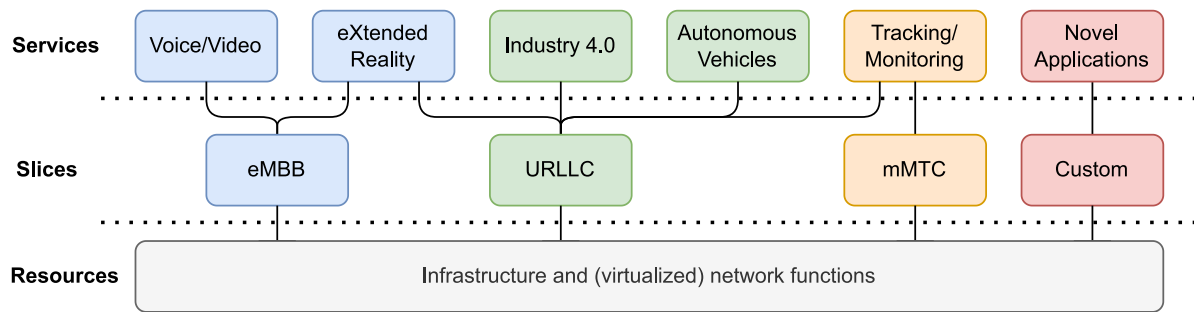


Fig. 5. Conceptual illustration of single-tier network slicing. Advanced implementations may allow for sub-slices of the shown slice types, for example, creating a standalone eMBB sub-slice for eXtended Reality.

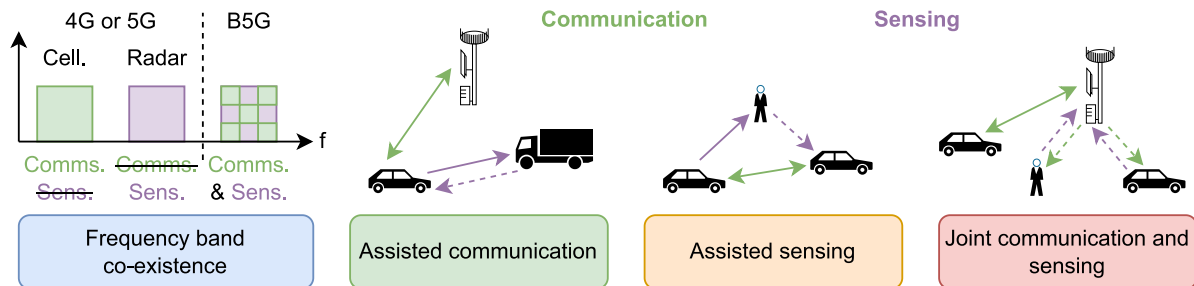


Fig. 6. The concept of ISAC. The B5G systems' support for frequency band co-existence permits assisting either communication or sensing by their counterpart.

static counterparts. And fourth, Conceicao et al. [106] aim to optimize the fairness in UL cell-free networks without relying on costly convex approximations. They prioritize the spectral efficiency of UEs with poor channel conditions. They use metaheuristics to minimize their experienced error rates, demonstrating low latency and scalability.

2.2.2. Network slicing

Virtual and software-defined networks may be sliced into specialized compartments with different functions tailored for specific applications [107], as shown in Fig. 5. For example, this kind of slicing may enable support for high-throughput video streaming in one slice and low-throughput grant-free IoT traffic in another, managed concurrently on the same network [8]. In practice, slices may be hierarchical and provide access to several different services. Reminiscent of the resource allocation publications above, four publications consider this topic.

Han et al. [108] model a queue-based access scheme for managing network slices, a kind of slice-as-a-service scheme. Modeling the situation as a Markov decision process and incoming requests according to different stochastic processes, they propose using Machine Learning (ML) algorithms to optimally provide access but leave such solutions as future work. Utilizing the same kind of model, Xiao et al. [109] focus on minimizing over-provisioning in network slices, giving higher priority to URLLC traffic. They propose using RL to solve this with dynamic adaptability, demonstrating how the algorithm improves over time.

To solve a similar problem, Chergui et al. [110] criticize existing algorithms for requiring a lot of monitoring data passed between domains. They propose avoiding this by using FL, bringing analysis closer to data generation. Doing so drastically reduces monitoring traffic and reduces its resource overheads.

Lastly, Cao et al. [111] present an approximate algorithm for resource allocation in network slices. Their scheme implements several different algorithms and selects between them depending on the type of incoming requests. Doing so reduces latency and improves resource utilization over state-of-the-art algorithms.

2.2.3. Integrated sensing and communication

ISAC denotes a new technology in which radio resources are used simultaneously for communication and environmental sensing [124], as

shown in Fig. 6. The latter will enable the detection of objects in space and high-accuracy localization and positioning. Three publications focus on the optimization of this technology.

Xu et al. [112] consider a radar-communication system, applying rate splitting MA, previously seen in [29,75,93], due to its good interference characteristics. They jointly optimize the sum rate and beam pattern error, modeling it as a non-convex problem and solving it using an Alternating Direction Method of Multipliers (ADMM) algorithm. The resulting performance is better than comparable schemes. Huang et al. [113] focus on a similar system and jointly optimize target detection and channel estimation, combining it with a pilot-based target search. The problem is non-convex, so the authors utilize several mathematical approximations to make it easier to solve. Their results show near-optimal performance. Lastly, Li et al. [114] use beamforming to maximize worst-case Signal to Noise Ratio (SNR) within a region. Like [112] and [113], the authors model this as a non-convex optimization but apply SCA to solve it. Plots indicate that the algorithm achieves its goal, covering the targeted region well.

2.2.4. Intelligent surfaces

Much like massive MIMO is needed to operate current high-frequency networks, moving to even higher frequency bands in B5G systems will further the troubles of signal attenuation. While beamforming can aid in this, it cannot compensate for potential multi-path or no-line-of-sight problems in urban environments. Mitigating such issues is possible using so-called IRSs; electrical surfaces designed to reflect (essentially relay) incoming radio signals with programmable phase shift and beam characteristics [2,120], illustrated in Fig. 7. Effectively managing these is crucial to ensure their performance but costly in terms of computations [8]. An IRS controller manages this by choosing between performing its computations locally [122] or offloading them, e.g., to the Cloud [120]. We survey nine publications on this topic with widely varying focus points.

Two papers use IRS as an enabler for other techniques. Firstly, returning to the topic of ISAC, Xing et al. [116] consider IRS-aided environmental sensing systems and propose a CNN-based correction scheme for such. They describe the situation as a compressed sensing problem and solve it using generalized approximate message passing. Their

Table 4
Overview of the approximate algorithms used to solve the surveyed future technology problems and their related KPIs.

Class	Ref.	Problem	Formulation	Algorithm	KPIs
Cell-Free Networks	[103]	Cooperative activity detection for grant-free random access	Non-convex optimization	Forward-backward splitting	Error rate
	[104]	Decentralized detection in UL massive MIMO	Log-likelihood ratio maximization	Max-log approximation	—
	[105]	NOMA-based DL relaying	Non-convex optimization	Iterative approximation	Sum rate
	[106]	Max–min fairness optimization in UL massive MIMO	Non-convex optimization	Metaheuristic	Spectral efficiency, latency
Network Slicing	[108]	Slice admission with heterogeneous tenant requests	Markov decision process, ILP	(Future work:) ML, heuristic	—
	[109]	Acceptance ratio maximization, prioritizing URLLC functions	Markov decision process	RL	Acceptance ratio, resource utilization
	[110]	Monitoring data minimization	Non-convex optimization	Game theory, FL	Spectral efficiency, resource utilization, NMSE, latency
	[111]	Slice resource allocation	—	—	Acceptance ratio, resource utilization, latency
ISAC	[112]	Combining radar communication with rate splitting MA	Non-convex optimization	ADMM	Sum rate, beampattern
	[113]	Target detection and channel estimation for radar systems	Non-convex optimization	Mathematical approximation	MSE, false alarms
	[114]	Beamforming optimization for sensing	Non-convex optimization	SCA	Beampattern
IRSs/RISs	[115]	Channel estimation for mmWave MIMO systems	Non-convex optimization	ADMM, approximate message passing	MSE
	[116]	IRS-based environmental sensing correction	—	CNN	MSE, recover ratio, false alarms
	[117]	Resource allocation in RIS-supported networks	MINLP	Outer approximation	Throughput, energy efficiency
	[118]	UL beamforming and compression optimization	Non-convex optimization	SCA	Sum rate
	[119]	Phase shift matrix optimization	Non-convex optimization	SCA, ADMM	Sum rate, energy efficiency
	[120]	Rate splitting and user clustering in Cloud-RAN systems	Non-convex optimization	SCA	Energy efficiency
	[121]	Combining UL NOMA with over-the-air FL	Non-convex optimization, MINLP	SCA, semidefinite relaxation	Optimality gap, sum rate, accuracy
	[122]	Joint precoding and phase shift matrix optimization	Non-convex optimization	RL	Sum rate
	[123]	Joint power allocation, user clustering, and phase shift matrix optimization	Non-convex optimization	SCA, iterative approximation	Sum rate

results show significant error reductions when applying the proposed correction. Secondly, Ni et al. [121] propose an IRS-supported over-the-air FL scheme that aggregates models by exploiting superposition on the wireless channel. They aim to maximize model convergence given uncertain channel conditions, modeling this as a MINLP problem. They solve this using SCA and demonstrate near-optimal performance.

The remaining seven papers seek to optimize the IRS' operation. Hu et al. [115] propose extending passive IRSs with deeply quantized channel estimators using generalized approximate message passing. This optimization is non-convex and solved using ADMM. The scheme outperforms its state-of-the-art alternatives while requiring fewer receivers. Orthogonally, Ahsan et al. [117] propose a resource allocation scheme for IRS-supported networks that maximizes overall energy efficiency. They formulate this as a MINLP problem and use outer approximation to solve it, demonstrating better results over a comparable scheme.

Centered around the idea of Cloud-Radio Access Network (RAN), Zhang et al. [118] jointly optimize user transmission beamforming, passive IRS beamforming, and data compression to maximize the UL sum rate. Weinberger et al. [120] maximize network energy efficiency by employing a combination of rate splitting, user clustering, and passive IRS beamforming. Both formulate their problems as non-convex optimizations and apply SCA in combination with other mathematical approximations to solve them, and both report improvements in their metrics.

Finally, Pereira et al. [122] jointly optimize the MIMO precoding and IRS phase shift matrices to maximize the network sum rate. As the problem is non-convex, they propose solving it dynamically with a combination of RL to learn policies and DNNs to approximate functions. This approach achieves comparable performance to a more complex counterpart. Along the same lines, but only focusing on phase shift matrix optimization, Nguyen et al. [119] formulate another non-convex problem that they apply ADMM and SCA to solve. The result is an improved network sum rate. Elhattab et al. [123] instead jointly optimize power allocation, user clustering, and IRS phase shift matrices.

Like [119], they use alternating optimization and SCA. Their scheme outperforms both NOMA- and OMA-based alternatives.

3. Discussion and potential future directions

While the previous section has revealed a vast amount of work already done in proposing approximate algorithms for various problems in future wireless systems, many challenges remain. The surveyed papers directly point some out, while others arise from patterns in their proposals. We structure this discussion slightly differently than the above, as some publications from different categories point out similar future work.

An observation originating from the above is that most publications focus on optimizing one or more network components to improve various KPIs, including BER, accuracy, latency, sum rate, throughput, and energy, spectral, or cost efficiency. However, while some papers share goals in these directions, they rarely focus on achieving them in the same scenario, leading to disparate proposals. Combining these could lead to more generally applicable algorithms. Furthermore, it could introduce interesting new trade-offs between generality and performance. Such a direction is worth considering!

3.1. Are these algorithms realistic?

It is well known that academic research results generated from narrow-scoped simulations, as in the majority of the surveyed papers, often fail to address critical constraints pertaining to real-world deployment [125]. This gap arises as the industry develops and deploys proprietary solutions to practical problems leaving academia to reverse-engineer or speculate on these. When such speculation becomes the foundation for further research, the gap gradually grows and the practical feasibility of proposed solutions shrinks. Albeit qualitative, we find it relevant to review this here.

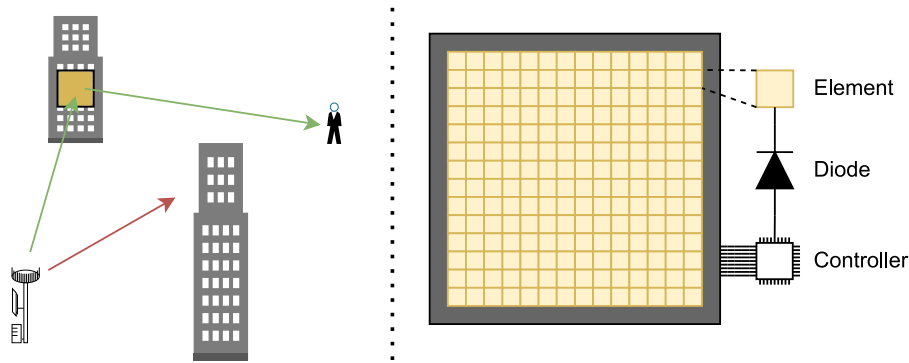


Fig. 7. Conceptual view of an IRS and its internal components.

Within the scope of this survey, we are particularly interested in whether the surveyed algorithms can be deployed to practical networks. A sign of this relevance is whether a paper has authors from the industry. This is the case in nine of the 80 surveyed papers: [38,41,54,55,65,88,92,108,113], in none of which industrial authors represent a majority of the author lists.

Furthermore, judging by Tables 1, 2, 3, and 4, the majority of the surveyed papers propose approximate algorithms to solve otherwise NP-hard problems. And while algorithms such as SCA, outer approximation, RL, and DNNs might produce approximate solutions in polynomial time to these exponential-scale problems, they still suffer from long latency. As networks become more distributed, the need for executing these algorithms closer to the BSs or APs grows, implying a need for distributed computing units [8,55,90]. However, such devices typically have extremely constrained computing resources, likely rendering them incapable of reaching sufficient throughput to execute many of the proposed algorithms in real time [8]. This is a challenge that is also highlighted by the 6G Flagship partners [126] that must be overcome to render common visions of 6G realistic [8]. Thus, research into good, reduced-complexity alternatives, such as ML, to established algorithms is a promising direction [127].

3.2. B5G systems in the Cloud-Edge Continuum

An interesting development happening in parallel with the development of B5G systems is the expansion of the so-called Cloud-Edge continuum [8,11,13]. Denoting a more continuous computing paradigm than traditional Cloud computing, this continuum involves the distribution of compute power across the network's layers reaching out to its geographical edge [6]. It is enabled by the increase in compute performance of low-power devices and driven by a need for more localized handling of the growing (cellular) traffic amounts that double every two years [7] and are expected to reach 226 EB/month in 2026 [85] and as high as 5016 EB/month in 2030 [2], as well as the increasing number of users and devices exceeding 5.3 billion and 13.1 billion by 2023 [6]. The increased degree of distribution is interesting for a multitude of reasons: firstly, end devices retain the benefits of traditional offloading; secondly, users may experience lower latency communications as data or computations are provided at a shorter distance to them; and thirdly, data needs to travel over fewer links and through fewer shared network nodes, potentially decreasing contention and improving privacy [18,128,129].

A greater distribution of computations, particularly at the Edge, is expected to be at the core of B5G systems [2,33,34,126,128]. Once implemented, it will enable rapid network optimization and reconfiguration [90], new classes of low-latency or high-bandwidth services (e.g., Augmented Reality) [7,128] and end-to-end network slicing [33] (as described above) needed to handle the predicted massive connectivity communication demands [6,13,93]. Combined with AxC, this implies new interesting energy-latency-accuracy trade-offs achievable

by applying approximations across network layers; that is, the data-producing devices, their communication links, and the node that carries out their processing whether it be Edge or Cloud-located. Analog to applying approximations across software and hardware in isolated systems, this requires tools for quality management, error modeling, and formal verification [18]. Additional design space exploration may also be needed to identify (near-)optimal combinations of AxC techniques for different applications with diverse quality and reliability constraints.

Some of the surveyed work already proposes algorithms specifically targeting this Cloud-Edge continuum, relying on it for performing heavy computations or exploiting its centralized nature to optimize network functionality. The former is needed for the AI-based scheduling in [4] and the ML-based channel estimation in [90]. The latter is more commonly represented and used for message clustering [85] and transmission coordination [105,120] as well as for exploiting spatial locality in accessed data [93]. Of these, only Bonati et al. [4] describe that Cloud-based network architectures may suffer from long latency if their involved data transfers are not minimized. Unfortunately, this latency is critical and may impact, among other factors, the practicality or user experience of, e.g., the aforementioned AR applications [7,130].

To illustrate this criticality, we reproduce simulation results of [131] for a number of different execution scenarios. The setting involves a wearable device and a smartphone, as well as an Edge server and the Cloud. The smartphone may execute tasks offloaded from the wearable itself or simply serve as a relay or gateway node to propagate these and return their results to and from the Edge or the Cloud. Fig. 8 shows the results thereof, for which the execution scenarios are enumerated as follows: (1) local execution on the wearable, (2) offloaded execution on the gateway (smartphone), (3) offloaded execution in the Cloud at a fixed distance of 50 m to the nearest BS, and (4a-d) offloaded execution to the Edge at distances of 50 m, 100 m, 200 m, and 400 m to the nearest BS. The simulation assumes an AR-like workload where the wearable is connected to the gateway via Wi-Fi, that tasks are atomic and indivisible, and that the received results are orders of magnitude smaller than the transmitted data rendering their effects negligible [132]. The results show the latency benefits of offloading regardless of the distance to the nearest BS, while there are energy benefits only when the BS is less than 200 m away.

Interestingly, and yet to be explored, these results may change if AxC is integrated into the system. Keeping our focus on application-level approximations, using AxC to reduce algorithmic complexity may render local execution beneficial energy-wise, while the latency will likely remain in favor of offloading. Approximating the wireless communication part instead will most likely merely enhance the already clear benefits of offloaded processing both in energy and latency. Advanced systems may integrate approximations in both processing and communication by, for example, using inexact pre-processing or lossy compression locally, deliberately neglecting errors during transmission, and approximately processing data at the Edge. Predicting the energy-latency relations for such systems is highly complex and requires further analysis.

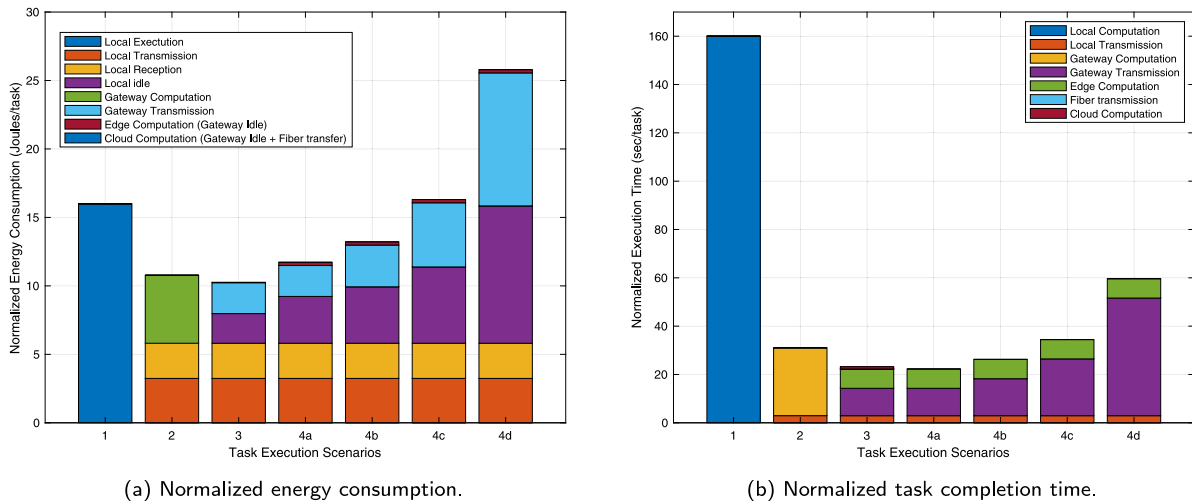


Fig. 8. Simulation results for (8(a)) normalized energy consumption and (8(b)) normalized task completion time for local and offloaded execution of tasks at 20 mb/task.

3.3. Traditional approximate computing in B5G

Another observation made through this survey is that only four of the 80 surveyed papers apply traditional AxC techniques: [28,37,38,115]; traditional in the sense that they are commonly reviewed or applied in other work [17,18,20,30]. Yet, as many of the algorithms involved in making B5G systems perform optimally can sustain approximations at the application level, we find it likely they can also benefit from approximations at the architecture or even the circuit level, also pointed out in [28]. While the arithmetic precision is undefined in most of the reviewed publications, their reliance on Integer Linear Programming (ILP) solvers, SCA algorithms, and various ML models hints at floating-point being used [49,84,89,91,133]. When these solvers' convergence criteria permit such, one may consider switching to fixed-point formats within them before evaluating further approximations.

In other contexts, particularly related to so-called Network on Chips used to interconnect embedded devices (even wirelessly), approximations have been introduced both by voltage over-scaling links [134], truncating packets in network interfaces [135], or simply skipping the transmission of certain low-significance packets [136]. Such techniques could be adapted to a broader networking scope and find use in, e.g., low-significance IoT data collection systems or multimedia streaming. Related work already applies filtering and lossy data compression, albeit to reduce effects of network latency fluctuations [137]. A more generic application of such techniques is yet to be explored.

3.4. Open research directions

Despite the amount of work already carried out and published in the field, many research questions remain open. We summarize these in Table 5 and provide more details here. Most of the papers directly suggest extending their schemes in different ways. Such extensions may involve taking more metrics into account in the proposed algorithm [44,59,67,70,87,123] or focus on widening the scope of the proposals. Within the latter, we find extending DNN transceivers to UL traffic [38], using puncturing in coded streams [41], and adding support for massive MIMO [88], device-to-device communication [54], or simultaneous information and power transfer networks [105]. While such extensions make the algorithms relevant to more use cases, they often increase their overheads. Constraining or optimizing the algorithms thus becomes relevant again [59,62,81,87,90,106]. Once fully optimized, making algorithms work in practical environments entails a need for acceleration and distribution that needs addressing [55,90].

While extensions in the same direction are one way to optimization, changing algorithms is another. Another set of papers pursues this. Such proposals include using learning to reduce optimality search complexity [41], exploring alternative routing, bandwidth allocation, and content caching algorithms [61], and developing new algorithms for resource allocation in IAB [62]. Adapting already proposed algorithms for resource allocation [93] and IRS-based relaying [119] to distributed computing systems is another direction that needs more work. Finally, further attention should be paid to making systems aware of their spectral efficiency [42] and to optimize beamforming and fronthaul compression to maximize their energy efficiency [118].

Yet another set of publications propose future work in more general terms. Related to ML, such directions include low-complexity RL models and other algorithms for non-convex optimization [64,108], generation and publication of more open datasets [90], and techniques for tackling security and mobility issues in over-the-air FL [121]. As ML is expected to become an essential component in B5G systems, these challenges should be addressed as soon as possible [8]. Unrelated to this, two papers point out particular fields that need further work: RAN slicing and THz communications, which require further perfecting [111] or technological advances in precoding, beamforming, and tracking [73], respectively.

Several proposals require further evaluation. This need ranges from evaluating optimality [51]; over applying schemes to more dynamic or challenging use cases [106] with hypervisor migration [65] or different data distributions and realistic mobility scenarios [121]; to evaluating different metrics, particularly energy efficiency [74] and spectral efficiency and their trade-off [54]. Another proposes to evaluate their proposed scheme in a completely different setting (namely, one with multiple queues) to identify the proper balance between multiplexing gain (i.e., assigning several tasks to a slice) and overbooking (i.e., assigning too many tasks to one slice) in network slicing [108].

4. Conclusion

As the number of smart connected and intelligent devices increases, the cellular network is put under tremendous pressure demanding its development. 5G systems have been standardized for some years and are currently being deployed, but researchers already question whether they are sufficient for the next ten years. As a result, technologies for B5G systems are undergoing research. These next-generation networks are expected to enable entirely new classes of intelligent applications and support growing needs for data traffic. Optimizing the networks for various objectives comes at high computational costs, as the underlying algorithms tend to be complex. Minimizing these costs is essential for

Table 5
Summarized directions of future work extracted from surveyed works.

Topic	Refs.	Challenges	Topic	Refs.	Challenges
Extension	[38]	DNN transceivers for UL traffic.	Design Space Exploration	[41]	Reduced optimality search complexity.
	[41]	Puncturing for coded streams.		[61]	Alternative routing, allocation, and caching algorithms.
	[88]	Support for massive MIMO.		[62]	New IAB resource allocation algorithms.
	[54]	Support for device-to-device communication.		[119]	Distribution of IRS algorithms.
	[105]	Support for simultaneous information and power transfer.		[118]	Beamforming and fronthaul compression for energy efficiency.
Machine Learning	[64], [108]	Low-complexity RL and non-convex optimization algorithms.	Further Evaluation	[51], [65], [106], [108]	Assess optimality of proposals in other more (realistic) environments.
	[121]	Tackling security and mobility issues in over-the-air FL.		[54], [74]	Analyzing trade-offs between energy and spectral efficiency.
	[90]	Publication of more open datasets.			
AxC in B5G and 6G	[20] [18], [30]	Explore circuit-level AxC techniques applicability in wireless networking systems.			
	[18]	Adapt approximate NoC techniques to larger networks.			
	[138], [90], [139], [109], [122], [140]	Consider using neural approximation in place of more algorithms than, e.g., channel estimation, network slicing, IRSs, and network switches.			
	[122], [140]				
	[131]	Identify benefits and drawbacks of AxC to Cloud-Edge task offloading scenarios.			

ensuring their practical deployment. One approach to do so for non-critical applications which require less than nine 9s reliability is to apply AxC for reduced algorithmic complexity, latency, and energy savings.

In this paper, we have surveyed publications on the state-of-the-art in the intersection of AxC and B5G systems, identifying and highlighting trends and tendencies in existing work and directions for future research. We have considered publications in two categories: existing and future technologies. Within the former, we find that resource allocation algorithms are particularly popular, while research related to IRSs appears the most prominent in the latter. In both, problems are often NP-hard and, thus, only solvable using heuristics or approximations, SCA and RL being most frequently applied.

Many of the surveyed papers point out application-specific directions for future work in extending existing algorithms, adapting them to other use cases, developing lower-complexity alternatives, or simply extending their evaluations. Others highlight more general open research directions, including low-overhead ML and RL algorithms and FL security primitives. In addition, the survey has shown a clear likelihood of the proposed algorithms not being practical due to their computational demands, requiring further adaptation before making it into real-world systems. As the algorithms are error-resilient, one such adaptation may be to introduce approximations with AxC techniques. This direction may bring large performance gains but is largely unexplored till now.

Acronyms

3GPP Third Generation Partnership Project
ADMM Alternating Direction Method of Multipliers
AI Artificial Intelligence
AP Access Point
AR Augmented Reality
AxC Approximate Computing
BER Bit Error Rate
BLP Binary Linear Programming
BS Base Station
CNN Convolutional Neural Network
CSI Channel State Information
DNN Deep Neural Network
DRAM Dynamic RAM
DFT Discrete Fourier Transform
DL Downlink
DSP Digital Signal Processing
eMBB Enhanced Mobile Broadband
ER Error Rate
FER Frame Error Rate

FL Federated Learning
FPGA Field-Programmable Gate Array
GNN Graph Neural Network
GPU Graphics Processing Unit
GFDM General Frequency-Division Multiplexing
gNB next-Generation Node B
IAB Integrated Access and Backhaul
ILP Integer Linear Programming
IoT Internet of Things
IRS Intelligent Reflective Surface
ISAC Integrated Sensing and Communication
KPI Key Performance Indicator
LDPC Low-Density Parity-Check
LSTM Long Short-Term Memory
MILP Mixed Integer Linear Programming
MSE Mean SE
MA Multiple Access
MBNLP Mixed Binary Non-Linear Programming
MIMO Multiple-Input Multiple-Output
MINLP Mixed Integer Non-Linear Programming
ML Machine Learning
MLP Multi-Layer Perceptron
mMTC massive Machine Type Communication
mmWave millimeter Wave
NMSE Normalized Mean SE
NOMA Non-Orthogonal MA
NN Neural Network
NoC Network on Chip
OMA Orthogonal MA
OFDMA Orthogonal Frequency-Division Multiple Access
OFDM Orthogonal Frequency-Division Multiplexing
PU Processing Unit
QoE Quality of Experience
QoS Quality of Service
RNN Recurrent Neural Network
RAM Random Access Memory
RAN Radio Access Network
RB Resource Block
RF Radio Frequency
RIS Reconfigurable Intelligent Surface
RL Reinforcement Learning
SER Symbol Error Rate
SCA Successive Convex Approximation
SE Soft Error
SIMD Single Instruction Multiple Data

SNR Signal to Noise Ratio
 UAV Unmanned Aerial Vehicle
 UE User Equipment
 UL Uplink
 URLLC Ultra-Reliable Low-Latency Communication
 Wi-Fi Wi-Fi
 XR eXtended Reality

CRedit authorship contribution statement

Hans Jakob Damsgaard: Conceptualization, Visualization, Writing – original draft. **Aleksandr Ometov:** Supervision, Writing – review & editing, Methodology, Validation. **Md Munjure Mowla:** Conceptualization, Writing – review & editing. **Adam Flizikowski:** Conceptualization, Supervision, Writing – review & editing. **Jari Nurmi:** Supervision, Writing – review & editing, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Hans Jakob Damsgaard reports financial support was provided by EU Framework Programme for Research and Innovation Marie Skłodowska-Curie Actions.

Data availability

No data was used for the research described in the article

Acknowledgments

The authors gratefully acknowledge funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska Curie grant agreement No. 956090 (APROPOS).

Appendix. Note on LDPC decoding

The survey has not returned any results on Low-Density Parity-Check (LDPC) decoders; an otherwise well-known opportunity for optimization through approximation [141]. Typical LDPC decoders implement probabilistic, successive cancellation algorithms that are highly parallel and require little arithmetic precision internally. Such codes, and related hardware, are already used in 5G, and some variant of them will likely be part of future B5G systems too. As a result, it is worth researching the benefits of applying AxC to this field. One must note that due to its computational requirements, LDPC decoding is usually carried out in hardware [142]. Yet, there exist publications reporting software versions with SIMD or GPU acceleration [143,144] that achieve impressive throughput merely due to exploited parallelism.

To motivate further research on this topic, we present a few papers that apply AxC to LDPC decoders. Note that this list is far from comprehensive.¹ The common feature is that the designs trade off increases in bit error rate for reductions in decode time, chip area, or energy consumption, as is typical for AxC-equipped designs [18,20].

¹ Inquery (approximat* OR inexact OR inaccurate OR ‘good enough’) AND (ldpc OR ‘low-density parity-check’) to Scopus and limiting results to computer science returns 752 results of which many appear to be relevant.

Algorithmic approximations

We consider six publications presenting various algorithmic approximations. Proposed decoders are typically built around so-called variable nodes and check nodes that communicate probabilities of their current predicted bit values between one another. They are most frequently approximated in their update logic. Such an implementation is used by Li et al. [145] who propose a reduced-complexity *minimum-sum* way to update the check nodes, which halves the number of multipliers and removes the need for adders at the expense of some shifters. Despite these savings, their algorithm achieves the same bit error rate as its exact counterparts. Sun and Jiang [146] use a similar implementation and propose a hybrid approach combining minimum-sum and linear approximation update rules to the check nodes. The algorithm outperforms traditional normalized minimum-sum algorithms (like that of [145]) and performs comparably to conventional belief propagation at a significantly lower computational complexity.

Honghao et al. [147] propose another layered minimum-sum algorithm with reduced-complexity check node updates. They optimize correction factors, internal thresholds, and evaluation order of individual nodes to maximize decoding performance, and report comparable performance to belief propagation with less than half as many iterations. Wu et al. [148] present a more involved algorithm that simultaneously applies minimum-sum check node updates with NN-based internal coefficient adjustments. By passing the check node values through the network, some correction factors are adapted to current channel conditions. This enables their algorithm to outperform belief propagation under certain conditions while they do not declare the computational overheads it entails.

Sun et al. [149] focus instead on so-called Bose-Chaudhuri-Hocquenghem-constrained codes for URLLC communications, which improves error correction performance at the cost of a lower data rate. They consider short block-length codes and apply approximated belief propagation with reduced-complexity *maximum-logarithm* check node updates. Their algorithm outperforms the error correction capabilities of the polar codes used in 5G, albeit at greater computational complexity.

Watanabe et al. [150] consider the bit reliability-based decoding algorithm, which, much like the other layered algorithms, is split between check and variable nodes but differs in what they pass between each other. They propose a “deep learning”-like approach in which the edges between nodes are weighted, and the overall decoder resembles a DNN. Training the weights for different noise ranges on the input signals presents an interesting trade-off between generality and error correction performance under specific conditions. Despite reduced complexity compared with minimum-sum algorithms, the reported performance is comparable.

Hardware approximations

Another five works present various hardware decoder architectures. Zhou et al. [151] describe a layered decoder design. Rather than approximating the algorithm, as in [145–147], they employ inexact truncated arithmetic to implement the variable update logic, reducing circuit area and delay. They demonstrate their design for codes used in various short-range wireless networks, showing comparable performance with exact fixed-point designs.

Tsatsaragkos and Paliouras [152] describe a decoder design based on reduced-complexity minimum-sum algorithms. They implement parallel variable node processors, a centralized check node unit, a concatenation unit to generate the final result, and a control unit to manage the decoding. In addition, they propose several logarithmic barrel shifters for interconnecting the variable node processors with their respective check nodes. The design is evaluated on Wi-Fi-related codes and shows improved throughput at a reduced circuit area. Patel and Engineer [153] describe a vastly simpler design that operates

as an accelerator in a FPGA. They also implement a minimum-sum algorithm whose approximations are limited to quantized fixed-point arithmetic, achieving similar performance to a full-precision software implementation.

Lopez et al. [154] also present an architecture implementing an approximate minimum-sum algorithm. Specifically, instead of computing two minima in each variable node, they compute one and estimate the other by adding a constant factor to the former. Their architecture resembles [152], but the approximations make their check node units and interconnect networks simpler. They demonstrate their technique for a Wi-Fi-related code and show improved area efficiency over related works.

Yang et al. [155] propose another, more involved than [153], FPGA-based decoder design. They particularly consider very long codes suitable for quantum key-based cryptography schemes. Like [153], they quantize internal values to relatively large fixed-point formats, apply second-order approximation to their otherwise logarithmic message-passing function, and optimize their design for the resources available on a FPGA. These design choices allow their design to function at a much lower SNR than related works.

References

- [1] S. Yrjölä, P. Ahokangas, M. Matinmikko-Blue, Visions for 6G Futures: A Causal Layered Analysis, in: Proceedings of Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), IEEE, 2022, pp. 535–540.
- [2] M.Z. Chowdhury, M. Shahjalal, S. Ahmed, Y.M. Jang, 6G Wireless Communication Systems: Applications, Requirements, Technologies, Challenges, and Research Directions, *IEEE Open J. Commun. Soc.* 1 (2020) 957–975.
- [3] N. Chukhno, A. Orsino, J. Torsner, A. Iera, G. Araniti, 5G NR Sidelink Multi-Hop Transmission in Public Safety and Factory Automation Scenarios, *IEEE Netw.* (2023).
- [4] L. Bonati, S. D'Oro, M. Polese, S. Basagni, T. Melodia, Intelligence and Learning in O-RAN for Data-Driven NextG Cellular Networks, *IEEE Commun. Mag.* 59 (10) (2021) 21–27.
- [5] F. Rinaldi, S. Pizzi, A. Molinaro, A. Iera, G. Araniti, Cooperative Resource Allocation in Integrated Terrestrial/Non-Terrestrial 5G and Beyond Networks, in: Proceedings of IEEE Global Communications Conference, GLOBECOM, IEEE, 2020, pp. 1–6.
- [6] C. and/or its affiliates, Cisco Annual Internet Report (2018–2023), Cisco, 2020, URL <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>.
- [7] P. Cerwal, et al., Ericsson Mobility Report, Tech. Rep. EAB-22:010742, Ericsson, 2022, URL <https://www.ericsson.com/en/reports-and-papers/mobility-report/reports/november-2022>.
- [8] H. Viswanathan, P.E. Mogensen, Communications in the 6G Era, *IEEE Access* 8 (2020) 57063–57074.
- [9] B. Yang, X. Cao, K. Xiong, C. Yuen, Y.L. Guan, S. Leng, L. Qian, Z. Han, Edge Intelligence for Autonomous Driving in 6G Wireless System: Design Challenges and Solutions, *IEEE Wirel. Commun.* 28 (2) (2021) 40–47.
- [10] A. Shashin, A. Belogaev, A. Krasilov, E. Khorov, Adaptive Parameters Selection for Uplink Grant-free URLLC Transmission in 5G Systems, *Comput. Netw.* 222 (2023) 109527.
- [11] B. Ji, Y. Han, S. Liu, F. Tao, G. Zhang, Z. Fu, C. Li, Several Key Technologies for 6G: Challenges and Opportunities, *IEEE Commun. Stand. Mag.* 5 (2) (2021) 44–51.
- [12] A. Jagannath, J. Jagannath, T. Melodia, Redefining Wireless Communication for 6G: Signal Processing Meets Deep Learning with Deep Unfolding, *IEEE Trans. Artif. Intell.* 2 (6) (2021) 528–536.
- [13] W. Jiang, B. Han, M.A. Habibi, H.D. Schotten, The Road Towards 6G: A Comprehensive Survey, *IEEE Open J. Commun. Soc.* 2 (2021) 334–366.
- [14] H. Esmailzadeh, E. Blem, R.S. Amant, K. Sankaralingam, D. Burger, Dark Silicon and the End of Multicore Scaling, in: Proceedings of 38th International Symposium on Computer Architecture, ISCA, ACM, 2011, pp. 365–376.
- [15] Semiconductor Industry Association and/or its affiliates, Rebooting the IT Revolution: A Call for Action, Tech. rep., Semiconductor Industry Association and the Semiconductor Research Corporation, 2018, URL <https://www.semiconductors.org/wp-content/uploads/2018/06/RITR-WEB-version-FINAL.pdf>.
- [16] A. Ometov, J. Nurmi, Towards Approximate Computing for Achieving Energy vs. Accuracy Trade-offs, in: Proceedings of Design, Automation & Test in Europe Conference & Exhibition, DATE, IEEE, 2022, pp. 632–635.
- [17] S. Venkataramani, S.T. Chakradhar, K. Roy, A. Raghunathan, Approximate Computing and the Quest for Computing Efficiency, in: Proceedings of 52nd Design Automation Conference, DAC, ACM, 2015, pp. 1–6.
- [18] H.J. Damsgaard, A. Ometov, J. Nurmi, Approximation Opportunities in Edge Computing Hardware: A Systematic Literature Review, *ACM Comput. Surv.* 55 (12) (2022) 1–49.
- [19] L. Anghel, M. Benabdenbi, A. Bosio, M. Traiola, E.I. Vatajelu, Test and Reliability in Approximate Computing, *J. Electron. Test.* 34 (4) (2018) 375–387.
- [20] S. Mittal, A Survey of Techniques for Approximate Computing, *ACM Comput. Surv.* 48 (4) (2016) 1–33.
- [21] J. Han, M. Orshansky, Approximate Computing: An Emerging Paradigm for Energy-Efficient Design, in: Proceedings of 18th European Test Symposium, ETS, IEEE, 2013, pp. 1–6.
- [22] K. Palem, A. Lingamneni, Ten Years of Building Broken Chips: The Physics and Engineering of Inexact Computing, *ACM Trans. Embed. Comput. Syst. (TECS)* 12 (2s) (2013) 1–23.
- [23] S. Amanollahi, M. Kamal, A. Afzali-Kusha, M. Pedram, Circuit-Level Techniques for Logic and Memory Blocks in Approximate Computing Systems, *Proc. IEEE* 108 (12) (2020) 2150–2177.
- [24] Ö. Bulakci, M. Gramaglia, et al., View on 5G Architecture, Tech. Rep., 5GPPP Architecture Working Group, 2021, URL <https://5g-ppp.eu/white-papers/>.
- [25] J. Korhonen, et al., Technical Specification Group Radio Access Network; NR; NR and NG-RAN Overall Description; Stage 2, Tech. Rep. TS 38.300 V17.4.0, 3rd Generation Partnership Project, 2023, URL <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3191>.
- [26] G.S. Rodrigues, J. Fonseca, F. Benevenuti, F. Kastensmidt, A. Bosio, Exploiting Approximate Computing for Low-Cost Fault Tolerant Architectures, in: Proceedings of 32nd Symposium on Integrated Circuits and Systems Design, SBCCI, IEEE, 2019, pp. 1–6.
- [27] M. Gao, Q. Wang, M.T. Arafin, Y. Lyu, G. Qu, Approximate Computing for Low Power and Security in the Internet of Things, *Computer* 50 (6) (2017) 27–34.
- [28] M. Idrees, M.M. Maqbool, M.K. Bhatti, M.M.U. Rahman, R. Hafiz, M. Shafique, An Approximate-Computing Empowered Green 6G Downlink, *Phys. Commun.* 49 (2021) 101444.
- [29] J. Mei, W. Han, X. Wang, H.V. Poor, Multi-Dimensional Multiple Access With Resource Utilization Cost Awareness for Individualized Service Provisioning in 6G, *IEEE J. Sel. Areas Commun.* 40 (4) (2022) 1237–1252.
- [30] Q. Xu, T. Mytkowicz, N.S. Kim, Approximate Computing: A Survey, *IEEE Des. Test* 33 (1) (2015) 8–22.
- [31] V.K. Chippa, S.T. Chakradhar, K. Roy, A. Raghunathan, Analysis and Characterization of Inherent Application Resilience for Approximate Computing, in: Proceedings of 50th Design Automation Conference, DAC, ACM, 2013, pp. 1–9.
- [32] V.K. Chippa, S. Venkataramani, S.T. Chakradhar, K. Roy, A. Raghunathan, Approximate Computing: An Integrated Hardware Approach, in: Proceedings of Asilomar Conference on Signals, Systems and Computers, IEEE, 2013, pp. 111–117.
- [33] N.B.L. and/or its affiliates, 6G Orchestration: A Scalable Approach to Application-Aware Network Optimization, Tech. Rep. CID 213047, Nokia Bell Labs, 2023, URL <https://www.bell-labs.com/institute/white-papers/6g-orchestration-a-scalable-application-driven-network-optimizing-approach/>.
- [34] G. Wikström, P. Persson, S. Parkvall, G. Mildh, E. Dahlman, B. Balakrishnan, P. Öhlén, E. Trojer, G. Rune, J. Arko, Z. Turányi, D. Roeland, B. Sahlin, W. John, J. Halén, H. Björkegren, 6G – Connecting a Cyber-Physical World, Tech. Rep. GFTL-20 001402, Ericsson, 2022, URL <https://www.ericsson.com/en/reports-and-papers/white-papers/a-research-outlook-towards-6g>.
- [35] M. Page, J. McKenzie, P. Bossuyt, I. Boutron, T. Hoffmann, C. Mulrow, L. Shamseer, J. Tetzlaff, E. Akl, S. Brennan, et al., The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews, *BMJ* 372 (2021) 11.
- [36] Elsevier, Scopus, 2022, <https://www.scopus.com/home.uri>.
- [37] A. Madanayake, R.J. Cintra, N. Akram, V. Ariyaratna, S. Mandal, V.A. Coutinho, F.M. Bayer, D. Coelho, T.S. Rappaport, Fast Radix-32 Approximate DFTs for 1024-Beam Digital RF Beamforming, *IEEE Access* 8 (2020) 96613–96627.
- [38] A. Morsali, A. Haghghat, B. Champagne, Deep Learning-Based Hybrid Analog-Digital Signal Processing in mmWave Massive-MIMO Systems, *IEEE Access* 10 (2022) 72348–72362.
- [39] C. Li, W. Chen, K.B. Letaief, Mean-Field Approximation based Scheduling for Broadcast Channels with Massive Receivers, in: Proceedings of Global Communications Conference, GLOBECOM, IEEE, 2021, pp. 1–6.
- [40] H.B. Salameh, H. Al-Obiedollah, R. Mahasees, Y. Jararweh, Opportunistic Non-Contiguous OFDMA Scheduling Framework for Future B5G/6G Cellular Networks, *Simul. Model. Pract. Theory* 119 (2022) 102563.
- [41] M. Almekhlafi, M. Chraïti, M.A. Arfaoui, C. Assi, A. Ghayeb, A. Alloum, A Downlink Puncturing Scheme for Simultaneous Transmission of URLLC and eMBB Traffic by Exploiting Data Similarity, *IEEE Trans. Veh. Technol.* 70 (12) (2021) 13087–13100.
- [42] M. Almekhlafi, M.A. Arfaoui, C. Assi, A. Ghayeb, Superposition-Based URLLC Traffic Scheduling in 5G and Beyond Wireless Networks, *IEEE Trans. Commun.* 70 (9) (2022) 6295–6309.
- [43] M. Beitollahi, N. Lu, Multi-Frame Scheduling for Federated Learning over Energy-Efficient 6G Wireless Networks, in: IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), IEEE, 2022, pp. 1–6.

- [44] R. Hashemi, H. Beyranvand, M.R. Mili, A. Khalili, H. Tabassum, D.W.K. Ng, Energy Efficiency Maximization in the Uplink Delta-OMA Networks, *IEEE Trans. Veh. Technol.* 70 (9) (2021) 9566–9571.
- [45] H.B. Salameh, R. Tashtoush, H. Al-Obiedollah, A. Alajlouni, Y. Jararweh, Power Allocation Technique with Soft Performance Guarantees in Hybrid OFDMA-NOMA Cognitive Radio Systems: Modeling and Simulation, *Simul. Model. Pract. Theory* 112 (2021) 102370.
- [46] M. Wang, D.-W. Yue, H.H. Nguyen, S.-N. Jin, Massive MIMO Relaying with Imperfect RF Chains and Coarse ADC/DAC in Beyond 5G Networks, *Trans. Emerg. Telecommun. Technol.* 33 (1) (2022) e4393.
- [47] M. Farooq, T.T. Vu, H.Q. Ngo, L.-N. Tran, Serving Federated Learning and Non-Federated Learning Users: A Massive MIMO Approach, in: 2022 IEEE 23rd International Workshop on Signal Processing Advances in Wireless Communication, SPAWC, IEEE, 2022, pp. 1–5.
- [48] U. Ghafoor, M. Ali, H.Z. Khan, A.M. Siddiqui, M. Naeem, Efficient Resource Allocation for Hybrid Nonorthogonal Multiple Access based Heterogeneous Networks Beyond Fifth-Generation, *Trans. Emerg. Telecommun. Technol.* 33 (12) (2022) e4630.
- [49] U. Ghafoor, H.Z. Khan, M. Ali, A.M. Siddiqui, M. Naeem, I. Rashid, Energy Efficient Resource Allocation for H-NOMA Assisted B5G HetNets, *IEEE Access* 10 (2022) 91699–91711.
- [50] W. Yin, L. Xu, P. Wang, Y. Wang, Y. Yang, T. Chai, Joint Device Assignment and Power Allocation in Multihoming Heterogeneous Multicarrier NOMA Networks, *IEEE Syst. J.* 16 (1) (2020) 671–682.
- [51] T.X. Vu, S. Chatzinotas, B. Ottersten, Dynamic Bandwidth Allocation and Precoding Design for Highly-Loaded Multiuser MISO in Beyond 5G Networks, *IEEE Trans. Wireless Commun.* 21 (3) (2021) 1794–1805.
- [52] J. Liu, H. Zhang, UcnPowerNet: Residual Learning based Clustering and Interference Coordination in 5G User-centric Networks, *Trans. Emerg. Telecommun. Technol.* (2021) e4324.
- [53] H. Dai, C. Zhang, J. Luo, C. Li, B. Wang, QoE-Driven Resource Allocation for Secure URLLC in 6G-Enabled IoT Networks, in: 2021 13th International Conference on Wireless Communications and Signal Processing, WCSP, IEEE, 2021, pp. 1–5.
- [54] L. Zhu, L. Yang, Q. Zhang, T. Zhou, J. Hua, Cluster-based Energy-Efficient Joint User Association and Resource Allocation for B5G Ultra-Dense Network, *Phys. Commun.* 46 (2021) 101311.
- [55] R. Adeogun, G. Berardinelli, P. Mogensen, Learning to Dynamically Allocate Radio Resources in Mobile 6G in-X Subnetworks, in: 2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC, IEEE, 2021, pp. 959–965.
- [56] W. Li, Y. Zhang, B. Bian, H. Yang, L. Li, Q. Wang, H. Zhang, J.-B. Wang, Joint Beamforming and Channel Allocation for Multi-User and Multi-Channel URLLC Systems, in: 2022 3rd Information Communication Technologies Conference, ICTC, IEEE, 2022, pp. 238–242.
- [57] B. Narottama, D.K. Hendraningrat, S.Y. Shin, Quantum-inspired Evolutionary Algorithms for NOMA User Pairing, *ICT Express* 8 (1) (2022) 11–17.
- [58] S. Abdel-Razek, H. Al-Obiedollah, H. Bany Salameh, Efficient User-Channel Pairing with Power-Domain Sum-Rate Maximization in Opportunistic Hybrid OFDMA-NOMA IoT Systems, *Cluster Comput.* 25 (4) (2022) 2501–2514.
- [59] T.-M. Pham, Traffic Engineering based on Reinforcement Learning for Service Function Chaining with Delay Guarantee, *IEEE Access* 9 (2021) 121583–121592.
- [60] F. Kavehmadavani, V.-D. Nguyen, T.X. Vu, S. Chatzinotas, Traffic Steering for eMBB and uRLLC Coexistence in Open Radio Access Networks, in: 2022 IEEE International Conference on Communications Workshops (ICC Workshops), IEEE, 2022, pp. 242–247.
- [61] Q. Zhang, W. Ma, Z. Feng, Z. Han, Backhaul-Capacity-aware Interference Mitigation Framework in 6G Cellular Internet of Things, *IEEE Internet Things J.* 8 (12) (2021) 10071–10084.
- [62] D. Harris, D. Raz, P. Sagiv, Bandwidth Resource Allocation in Integrated Access and Backhaul Networks, in: Proceedings of the ACM SIGCOMM Workshop on 5G and beyond Network Measurements, Modeling, and Use Cases, ACM, 2022, pp. 1–7.
- [63] B. Liu, Z. Zhou, Y. Yang, A.R. Adhikary, P. Fan, Constructions of Binary Signature Sets With Optimal Odd Total Squared Correlation and Their Application to Device Activity Detection, *IEEE Trans. Intell. Transp. Syst.* (2021) Early access.
- [64] Y.Z. Bekele, Y.-J. Choi, Random Access Using Deep Reinforcement Learning in Dense Mobile Networks, *Sensors* 21 (9) (2021) 3210.
- [65] F. Mogyorósi, P. Babarcsi, J. Zerwas, A. Blenk, A. Pašić, Resilient Control Plane Design for Virtualized 6G Core Networks, *IEEE Trans. Netw. Serv. Manag.* 19 (3) (2022) 2453–2467.
- [66] D. Zheng, C. Peng, B. Wang, X. Cao, Towards Deterministic Fault-Tolerant Service Function Slicing in Edge Networks, in: 2022 International Conference on Computer Communications and Networks, ICCCN, IEEE, 2022, pp. 1–8.
- [67] B. Bakhshi, J. Mangues-Bafalluy, J. Baranda, R-Learning-based Admission Control for Service Federation in Multi-Domain 5G Networks, in: Proceedings of Global Communications Conference, GLOBECOM, IEEE, 2021, pp. 1–6.
- [68] S. Iyer, Performance Analysis of a Dynamic Spectrum Assignment Technique for 6G, *IETE J. Res.* (2022) 1–9.
- [69] Z. Fei, Y. Wang, J. Zhao, X. Wang, L. Jiao, Joint Computational and Wireless Resource Allocation in Multicell Collaborative Fog Computing Networks, *IEEE Trans. Wireless Commun.* 21 (11) (2022) 9155–9169.
- [70] A. Ahmed, M. Zakarya, X. Liu, R. Khan, A. Ali, A.A. Khan, Quality Enhancement in a mm-Wave Multi-Hop, Multi-Tier Heterogeneous 5G Network Architecture, *Telecommun. Syst.* (2022) 1–19.
- [71] M. Vaezi, R. Schober, Z. Ding, H.V. Poor, Non-Orthogonal Multiple Access: Common Myths and Critical Questions, *IEEE Wirel. Commun.* 26 (5) (2019) 174–180.
- [72] F. Bonomi, R. Milito, J. Zhu, S. Addepalli, Fog Computing and Its Role in the Internet of Things, in: Proceedings of 1st Edition of the MCC Workshop on Mobile Cloud Computing, ACM, 2012, pp. 13–16.
- [73] L. Yan, C. Han, J. Yuan, Hybrid Precoding for 6G Terahertz Communications: Performance Evaluation and Open Problems, in: 2020 2nd 6G Wireless Summit (6G SUMMIT), IEEE, 2020, pp. 1–5.
- [74] S.A. Busari, S. Mumtaz, J. Rodriguez, Hybrid Precoding Techniques for THz Massive MIMO in Hotspot Network Deployment, in: 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), IEEE, 2020, pp. 1–6.
- [75] Z. Li, C. Ye, Y. Cui, S. Yang, S. Shamaï, Rate Splitting for Multi-Antenna Downlink: Precoder Design and Practical Implementation, *IEEE J. Sel. Areas Commun.* 38 (8) (2020) 1910–1924.
- [76] L.N. Ribeiro, S. Schwarz, M. Haardt, Low-Complexity Zero-Forcing Precoding for XL-MIMO Transmissions, in: 2021 29th European Signal Processing Conference, EUSIPCO, IEEE, 2021, pp. 1621–1625.
- [77] M. Bakulin, V. Kreindelin, D. Pankratov, A. Stepanova, Modified Newton's Method in Massive MIMO Demodulation with Non-Gaussian Approximation, in: 2022 Systems of Signals Generating and Processing in the Field of on Board Communications, IEEE, 2022, pp. 1–5.
- [78] M. Bakulin, V. Kreindelin, D.Y. Pankratov, A. Stepanova, Iterative Massive MIMO Demodulation Method with Non-Gaussian Approximation, *J. Commun. Technol. Electron.* 67 (6) (2022) 740–746.
- [79] A. Kosasih, V. Onasis, W. Hardjawana, V. Miloslavskaya, V. Andrean, J.-S. Leu, B. Vucetic, Graph Neural Network Aided Expectation Propagation Detector for MU-MIMO Systems, in: 2022 IEEE Wireless Communications and Networking Conference, WCNC, IEEE, 2022, pp. 1212–1217.
- [80] S. Bicaïs, A. Falempin, J.-B. Doré, V. Savin, Design and Analysis of MIMO Systems using Energy Detectors for Sub-THz Applications, *IEEE Trans. Wireless Commun.* 21 (6) (2021) 3678–3690.
- [81] M.K. Hasan, M.S. Hosain, T. Saha, S. Islam, L.C. Paul, S. Khatak, H.M. Alkhasawneh, E. Kariri, E. Ahmed, R. Hassan, Energy Efficient Data Detection with Low Complexity for an Uplink Multi-User Massive MIMO System, *Comput. Electr. Eng.* 101 (2022) 108045.
- [82] A. Kumar, S. Chakravarty, S. Suganya, H. Sharma, R. Pareek, M. Masud, S. Al-jahdali, Intelligent Conventional and Proposed Hybrid 5G Detection Techniques, *Alex. Eng. J.* 61 (12) (2022) 10485–10494.
- [83] A. Kaushik, E. Vlachos, C. Tsinos, J. Thompson, S. Chatzinotas, Joint Bit Allocation and Hybrid Beamforming Optimization for Energy Efficient Millimeter Wave MIMO Systems, *IEEE Trans. Green Commun. Netw.* 5 (1) (2020) 119–132.
- [84] M. Fozì, A.R. Sharafat, M. Bennis, Fast MIMO Beamforming via Deep Reinforcement Learning for High Mobility mmWave Connectivity, *IEEE J. Sel. Areas Commun.* 40 (1) (2021) 127–142.
- [85] R.-J. Reifert, A.A. Ahmad, H. Dahrouj, A. Chaaban, A. Sezgin, T.Y. Al-Naffouri, M.-S. Alouini, Joint Beamforming and Clustering for Energy Efficient Multi-Cloud Radio Access Networks, in: 2022 IEEE Wireless Communications and Networking Conference, WCNC, IEEE, 2022, pp. 608–613.
- [86] Y. Qiang, X. Shao, X. Chen, A Model-driven Deep Learning Algorithm for Joint Activity Detection and Channel Estimation, *IEEE Commun. Lett.* 24 (11) (2020) 2508–2512.
- [87] A. Mohamadian, C. Tellambura, Joint Channel and Phase Noise Estimation and Data Detection for GFDM, *IEEE Open J. Commun. Soc.* 2 (2021) 915–933.
- [88] Y. Li, X. Wang, R.L. Olesen, Unfolded Deep Neural Network (UDNN) for High Mobility Channel Estimation, in: 2021 IEEE Wireless Communications and Networking Conference, WCNC, IEEE, 2021, pp. 1–6.
- [89] Y. Zhang, J. Wang, J. Sun, B. Adebisi, H. Gacanin, G. Gui, F. Adachi, CV-3DCNN: Complex-valued Deep Learning for CSI Prediction in FDD Massive MIMO Systems, *IEEE Wirel. Commun. Lett.* 10 (2) (2020) 266–270.
- [90] M. Belgiovine, K. Sankhe, C. Bocanegra, D. Roy, K.R. Chowdhury, Deep Learning at the Edge for Channel Estimation in Beyond-5G Massive MIMO, *IEEE Wirel. Commun.* 28 (2) (2021) 19–25.
- [91] Q. Shi, Y. Liu, S. Zhang, S. Xu, V.K. Lau, A Unified Channel Estimation Framework for Stationary and Non-Stationary Fading Environments, *IEEE Trans. Commun.* 69 (7) (2021) 4937–4952.
- [92] M. Jian, Y. Chen, G. Yu, Improving Multiple-User Capacity through Downlink NOMA in OAM Systems, in: 2021 IEEE International Conference on Communications Workshops (ICC Workshops), IEEE, 2021, pp. 1–6.
- [93] A.A. Ahmad, Y. Mao, A. Sezgin, B. Clerckx, Rate Splitting Multiple Access in C-RAN: A Scalable and Robust Design, *IEEE Trans. Commun.* 69 (9) (2021) 5727–5743.
- [94] S. Khisa, M. Almekhlafi, M. Elhattab, C. Assi, Full Duplex Cooperative Rate Splitting Multiple Access for a MISO Broadcast Channel With Two Users, *IEEE Commun. Lett.* 26 (8) (2022) 1913–1917.

- [95] H. Bastami, M. Moradikia, H. Behroozi, R.C. de Lamare, A. Abdelhadi, Z. Ding, Secrecy Rate Maximization for Hardware Impaired Untrusted Relaying Network with Deep Learning, *Phys. Commun.* 49 (2021) 101476.
- [96] Y.-A. Xie, Z. Liu, S. Wang, J. Kang, Z. Xiong, K.Y. Chan, D. Niyato, Secure Information Transmission for B5G HetNets: A Robust Game Approach, *IEEE Internet Things J.* 9 (21) (2022) 21505–21519.
- [97] L. Hernández-Álvarez, J.J. Bullón Pérez, F.K. Batista, A. Queiruga-Dios, Security Threats and Cryptographic Protocols for Medical Wearables, *Mathematics* 10 (6) (2022) 886.
- [98] F. Mallouli, A. Hellal, N.S. Saeed, F.A. Alzahrani, A Survey on Cryptography: Comparative Study between RSA vs ECC Algorithms, and RSA vs El-Gamal Algorithms, in: Proc. of 6th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud) and 5th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom), IEEE, 2019, pp. 173–176.
- [99] W. Liu, C. Gu, G. Qu, M. O'Neill, Approximate Computing and Its Application to Hardware Security, in: *Cyber-Physical Systems Security*, Springer, 2018, pp. 43–67.
- [100] A. Ometov, K. Zeman, P. Masek, L. Balazevic, M. Komarov, A Comprehensive and Reproducible Comparison of Cryptographic Primitives Execution on Android Devices, *IEEE Access* 9 (2021) 54625–54638.
- [101] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, L. Hanzo, Cell-Free Massive MIMO: A New Next-Generation Paradigm, *IEEE Access* 7 (2019) 99878–99888.
- [102] H. He, X. Yu, J. Zhang, S. Song, K.B. Letaief, Cell-Free Massive MIMO for 6G Wireless Communication Networks, *J. Commun. Inform. Netw.* 6 (4) (2021) 321–335.
- [103] J. Shao, X. Chen, D.W.K. Ng, C. Zhong, Z. Zhang, Cooperative Activity Detection: Sourced and Unsourced Massive Random Access Paradigms, *IEEE Trans. Signal Process.* 68 (2020) 6578–6593.
- [104] Z.H. Shaik, E. Björnson, E.G. Larsson, Distributed Computation of A Posteriori Bit Likelihood Ratios in Cell-Free Massive MIMO, in: 2021 29th European Signal Processing Conference, EUSIPCO, IEEE, 2021, pp. 935–939.
- [105] X. Peng, X. Yuan, K. Zhang, Clustered NOMA-based Downlink Adaptive Relay Coordinated Transmission Scheme for Future 6G Cell-Free Edge Network, *Peer-To-Peer Netw. Appl.* 15 (1) (2022) 612–625.
- [106] F. Conceição, C.H. Antunes, M. Gomes, V. Silva, R. Dinis, Max-Min Fairness Optimization in Uplink Cell-Free Massive MIMO using Meta-Heuristics, *IEEE Trans. Commun.* 70 (3) (2022) 1792–1807.
- [107] J. Burke, What is Network Slicing and How Does it Work in 5G?, 2022, <https://www.techtarget.com/whatis/definition/network-slicing> (Accessed 25 October 2022).
- [108] B. Han, V. Sciancalepore, X. Costa-Perez, D. Feng, H.D. Schotten, Multiservice-based Network Slicing Orchestration with Impatient Tenants, *IEEE Trans. Wireless Commun.* 19 (7) (2020) 5010–5024.
- [109] D. Xiao, W. Ni, J.A. Zhang, R. Liu, S. Chen, Y. Qu, AI-Enabled Automated and Closed-Loop Optimization Algorithms for Delay-Aware Network, in: 2022 IEEE Wireless Communications and Networking Conference, WCNC, IEEE, 2022, pp. 806–811.
- [110] H. Chergui, L. Blanco, C. Verikoukis, Statistical Federated Learning for Beyond 5G SLA-Constrained RAN Slicing, *IEEE Trans. Wireless Commun.* 21 (3) (2021) 2066–2076.
- [111] H. Cao, J. Du, H. Zhao, D.X. Luo, N. Kumar, L. Yang, F.R. Yu, Toward Tailored Resource Allocation of Slices in 6G Networks With Softwarization and Virtualization, *IEEE Internet Things J.* 9 (9) (2021) 6623–6637.
- [112] C. Xu, B. Clerckx, S. Chen, Y. Mao, J. Zhang, Rate-Splitting Multiple Access for Multi-Antenna Joint Radar and Communications, *IEEE J. Sel. Top. Sign. Proces.* 15 (6) (2021) 1332–1347.
- [113] Z. Huang, K. Wang, A. Liu, Y. Cai, R. Du, T.X. Han, Joint Pilot Optimization, Target Detection and Channel Estimation for Integrated Sensing and Communication Systems, *IEEE Trans. Wireless Commun.* 21 (12) (2022) 10351–10365.
- [114] R. Li, Z. Xiao, Y. Zeng, Beamforming Towards Seamless Sensing Coverage for Cellular Integrated Sensing and Communication, in: 2022 IEEE International Conference on Communications Workshops (ICC Workshops), IEEE, 2022, pp. 492–497.
- [115] J. Hu, H. Yin, E. Björnson, mmWave MIMO Communication with Semi-passive RIS: A Low-Complexity Channel Estimation Scheme, in: Proceedings of Global Communications Conference, GLOBECOM, IEEE, 2021, pp. 01–06.
- [116] Z. Xing, F. Kong, Z. Zhang, Intelligent Reflecting Surface Aided Environment Sensing with CNN-based Correction, in: 2021 13th International Conference on Wireless Communications and Signal Processing, WCSP, IEEE, 2021, pp. 1–5.
- [117] M. Ahsan, S. Jamil, M.T. Ejaz, M.S. Abbas, Energy Efficiency Maximization in RIS-assisted Wireless Networks, in: 2021 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), IEEE, 2021, pp. 1–6.
- [118] Y. Zhang, X. Wu, H. Peng, C. Zhong, X. Chen, Beamforming and Fronthaul Compression Design for Intelligent Reflecting Surface aided Cloud Radio Access Networks, *Front. Inf. Technol. Electron. Eng.* 23 (1) (2022) 31–46.
- [119] T.V. Nguyen, T.P. Truong, T.M.T. Nguyen, W. Noh, S. Cho, Achievable Rate Analysis of Two-Hop Interference Channel with Coordinated IRS Relay, *IEEE Trans. Wireless Commun.* 21 (9) (2022) 7055–7071.
- [120] K. Weinberger, A.A. Ahmad, A. Sezgin, A. Zappone, Synergistic Benefits in IRS- and RS-enabled C-RAN with Energy-Efficient Clustering, *IEEE Trans. Wireless Commun.* 21 (10) (2022) 8459–8475.
- [121] W. Ni, Y. Liu, Y.C. Eldar, Z. Yang, H. Tian, STAR-RIS Integrated Nonorthogonal Multiple Access and Over-the-Air Federated Learning: Framework, Analysis, and Optimization, *IEEE Internet Things J.* 9 (18) (2022) 17136–17156.
- [122] D. Pereira-Ruisánchez, Ó. Fresnedo, D. Pérez-Adán, L. Castedo, Joint Optimization of IRS-assisted MU-MIMO Communication Systems through a DRL-based Twin Delayed DDPG Approach, in: 2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, BMSB, IEEE, 2022, pp. 1–6.
- [123] M. Elhattab, M.A. Arfaoui, C. Assi, A. Ghayeb, RIS-assisted Joint Transmission in a Two-Cell Downlink NOMA Cellular System, *IEEE J. Sel. Areas Commun.* 40 (4) (2022) 1270–1286.
- [124] A. Mueller, Integrated Communication & Sensing – The Killer Feature of 6G?, 2022, <https://one6g.org/download/2143/>. (Accessed 5 October 2022).
- [125] P. Pavon-Marino, J.-L. Izquierdo-Zaragoza, Net2plan: An Open Source Network Planning Tool for Bridging the Gap Between Academia and Industry, *IEEE Netw.* 29 (5) (2015) 90–96.
- [126] M. Juntti, R. Kantola, P. Kyösti, S. LaValle, C. Morais de Lima, M. Matinmikko-Blue, T. Ojala, A. Pouttu, A. Pärssinen, S. Yrjölä, et al., Key Drivers and Research Challenges for 6g Ubiquitous Wireless Intelligence, Tech. Rep. ISBN 978-952-62-2354-4, Finnish 6G Flagship at the University of Oulu, 2019, URL <https://www.6gflagship.com/key-drivers-and-research-challenges-for-6g-ubiquitous-wireless-intelligence/>.
- [127] S. Ali, W. Saad, D. Steinbach, White Paper On machine learning in 6G wireless Communication Networks, Tech. Rep. ISBN 978-952-62-2673-6, Finnish 6G Flagship at the University of Oulu, 2020, URL <https://www.6gflagship.com/white-paper-on-machine-learning-in-6g-wireless-communication-networks/>.
- [128] F. Siddiqui, N. Raman, H. Zaki, D. Winston, S. Prakash, The Edge Exponent: The Future of Edge Infrastructure, Partnerships, and the Impact on Communications Service Providers, Tech. Rep. CID 210911, Nokia Bell Labs Consulting, 2021, URL <https://www.bell-labs.com/institute/white-papers/the-edge-exponent-the-future-of-edge-infrastructure-partnerships-and-the-impact-on-communications-service-providers/>.
- [129] D. Alekseeva, A. Ometov, O. Arponen, E.S. Lohan, The Future of Computing Paradigms for Medical and Emergency Applications, *Comp. Sci. Rev.* 45 (2022) 100494.
- [130] D. Alekseeva, A. Ometov, E.S. Lohan, Towards the Advanced Data Processing for Medical Applications Using Task Offloading Strategy, in: 2022 18th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), IEEE, 2022, pp. 51–56.
- [131] W.B. Qaim, A. Ometov, C. Campolo, A. Molinaro, E.S. Lohan, J. Nurmi, Understanding the Performance of Task Offloading for Wearables in a Two-Tier Edge Architecture, in: Proceedings of 13th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops, ICUMT, IEEE, 2021, pp. 1–9.
- [132] Z. Cheng, P. Li, J. Wang, S. Guo, Just-in-Time Code Offloading for Wearable Computing, *IEEE Trans. Emerg. Top. Comput.* 3 (1) (2015) 74–83.
- [133] S.M. Perera, A. Madanayake, R.J. Cintra, Efficient and Self-Recursive Delay Vandermonde Algorithm for Multi-Beam Antenna Arrays, *IEEE Open J. Signal Process.* 1 (2020) 64–76.
- [134] G. Ascia, V. Catania, S. Monteleone, M. Palesi, D. Patti, J. Jose, Approximate Wireless Networks-on-Chip, in: Proceedings of Conference on Design of Circuits and Integrated Systems, DCIS, IEEE, 2018, pp. 1–6.
- [135] M.F. Reza, P. Ampadu, Approximate Communication Strategies for Energy-Efficient and High Performance NoC: Opportunities and Challenges, in: Proceedings of Great Lakes Symposium on VLSI, GLSVLSI, ACM, 2019, pp. 399–404.
- [136] S. Xiao, X. Wang, M. Palesi, A.K. Singh, L. Wang, T. Mak, On Performance Optimization and Quality Control for Approximate-Communication-enabled Networks-on-Chip, *IEEE Trans. Comput.* 70 (11) (2020) 1817–1830.
- [137] A. George, A. Ravindran, Scalable Approximate Computing Techniques for Latency and Bandwidth Constrained IoT Edge, in: *International Summit Smart City 360°*, Springer, 2020, pp. 274–292.
- [138] H. Esmailzadeh, A. Sampson, L. Ceze, D. Burger, Neural Acceleration for General-Purpose Approximate Programs, in: Proceedings of 45th International Symposium on Microarchitecture, MICRO, IEEE, 2012, pp. 449–460.
- [139] A. Gaydamaka, N. Yarkina, V. Khalina, D. Moltchanov, Comparison of Machine Learning Algorithms for Priority-Based Network Slicing in 5G Systems, in: 2021 13th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops, ICUMT, IEEE, 2021, pp. 72–77.
- [140] T. Swamy, A. Rucker, M. Shahbaz, I. Gaur, K. Olukotun, Taurus: A Data Plane Architecture for Per-Packet ML, in: Proceedings of 27th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, ASPLOS, ACM, 2022, pp. 1099–1114.
- [141] R. Gallager, Low-Density Parity-Check Codes, *IRE Trans. Inform. Theory* 8 (1) (1962) 21–28.
- [142] O. Ferraz, S. Subramanian, R. Chinthalala, J. Andrade, J.R. Cavallaro, S.K. Nandy, V. Silva, X. Zhang, M. Purnaprajna, G. Falcao, A Survey on High-Throughput Non-Binary LDPC Decoders: ASIC, FPGA, and GPU Architectures, *IEEE Commun. Surv. Tutor.* 24 (1) (2021) 524–556.

- [143] B. Le Gal, C. Jego, High-Throughput Multi-Core LDPC Decoders based on x86 Processor, *IEEE Trans. Parallel Distrib. Syst.* 27 (5) (2015) 1373–1386.
- [144] C. Tarver, M. Tonnemacher, H. Chen, J. Zhang, J.R. Cavallaro, GPU-based, LDPC Decoding for 5G and Beyond, *IEEE Open J. Circuits Syst.* 2 (2021) 278–290.
- [145] H. Li, J. Guo, C. Guo, D. Wang, A Low-Complexity Min-Sum Decoding Algorithm for LDPC Codes, in: 2017 IEEE 17th International Conference on Communication Technology, ICCT, IEEE, 2017, pp. 102–105.
- [146] K. Sun, M. Jiang, A Hybrid Decoding Algorithm for Low-Rate LDPC Codes in 5G, in: 2018 10th International Conference on Wireless Communications and Signal Processing, WCSP, IEEE, 2018, pp. 1–5.
- [147] C. Honghao, Y. Yingkun, Q. Yi, A Low Complexity Decoding Algorithm Design based on Quasi-Cyclic LDPC Codes, in: 2020 International Conference on Computing, Electronics & Communications Engineering, ICCECE, IEEE, 2020, pp. 45–50.
- [148] X. Wu, M. Jiang, C. Zhao, Decoding Optimization for 5G LDPC Codes by Machine Learning, *IEEE Access* 6 (2018) 50179–50186.
- [149] Z. Sun, Z. Chen, L. Yin, J. Lu, Design of LDPC Codes for Ultra Reliable Low Latency Communications, *IEEE Commun. Lett.* 25 (9) (2021) 2800–2804.
- [150] T. Watanabe, T. Ohseki, K. Yamazaki, Deep Learning-based Bit Reliability based Decoding for Non-binary LDPC Codes, in: 2021 IEEE International Symposium on Information Theory, ISIT, IEEE, 2021, pp. 1451–1456.
- [151] Y. Zhou, J. Lin, Z. Wang, Efficient Approximate Layered LDPC Decoder, in: Proceedings of International Symposium on Circuits and Systems, ISCAS, IEEE, 2017, pp. 1–4.
- [152] I. Tsatsaragkos, V. Paliouras, A Reconfigurable LDPC Decoder Optimized for 802.11 n/ac Applications, *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* 26 (1) (2017) 182–195.
- [153] D.J. Patel, P. Engineer, Design and Implementation of Quasi Cyclic Low Density Parity Check (QC-LDPC) Code on FPGA, in: 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), IEEE, 2017, pp. 181–185.
- [154] H. Lopez, H.-W. Chan, K.-L. Chiu, P.-Y. Tsai, S.-J.J. Jou, A 75Gb/s/mm^2 and Energy-Efficient LDPC Decoder based on a Reduced Complexity Second Minimum Approximation Min-Sum Algorithm, *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* 28 (4) (2019) 926–939.
- [155] S.-S. Yang, J.-Q. Liu, Z.-G. Lu, Z.-L. Bai, X.-Y. Wang, Y.-M. Li, An FPGA-Based LDPC Decoder with Ultra-Long Codes for Continuous-Variable Quantum Key Distribution, *IEEE Access* 9 (2021) 47687–47697.



Hans Jakob Damsgaard received the B.Sc. degree in electrical engineering in 2019 and the M.Sc. degree in computer science and engineering in 2021 as part of the Honours programme at the Technical University of Denmark, DTU. He is currently pursuing a Ph.D. in reconfigurable approximate accelerators for secure edge computing at Tampere University, TAU. He received the best paper award at NorCAS'21. His current research interests include hardware accelerators, networks-on-chip, computer architecture, and hardware verification.



Aleksandr Ometov received D.Sc.(Tech.) and M.Sx. degrees from the Tampere University of Technology, 2018 and 2016. He also holds a Specialist degree from Saint Petersburg State University of Aerospace Instrumentation, 2013. He is a Senior Research Fellow at Tampere University, Finland, and the coordinator of the CONVERGENCE of Humans and Machines research field. He is a Project and Training Manager of EU H2020 MCSA A-WEAR and APROPOS ITN projects. His research interests include wireless communications, information security, computing paradigms, blockchain technology, and wearable applications. He was recognized as a Publisher of the Year in 2022 by TAU.



Md Munjure Mowla works as an Expert R&D Engineer in IS-Wireless, Poland. He received a Ph.D. degree in Communication Engineering from Edith Cowan University, Australia. Dr. Mowla is experienced on developing and working on EU H2020 projects. He is author (and co-author) of more than 50 papers from different topics of wireless communications. His current research interests include Beyond 5G and 6G Wireless Communications, Radio Resource Management, Software Defined radio RAN, Edge Computing, AI/ML algorithms, cell free massive MIMO, etc.



Adam Flizikowski holds the position of Head of R&D at the ISRD Sp. z o.o., has 22 years of commercial experience in cellular systems (incl. 5G, LTE, WiMAX) and especially topics of RRM techniques (admission, congestion, NOMA). He is leading the research team, he also manages the ECSEL JU project called BRAINE (AI/ML supported micro edge DC) dealing with workload prediction and placement. He holds an M.Sc. from telecommunications and is pursuing Ph.D. on admission control in future wireless systems. He is author (and co-author) of more than 70 papers on different wireless communications, cyber security and adaptive multimedia system topics.



Jari Nurmi received the D.Sc.(Tech) degree in 1994. He works as a Professor at Tampere University, TAU (formerly Tampere University of Technology, TUT), Finland, since 1999. His research interests include embedded computing systems, System-on-Chip, approximate computing, wireless localization, positioning receivers, and software-defined radio and -networks. He held various research, education, and management positions at TUT since 1987 and was the Vice President of the SME VLSI Solution Oy from 1995 to 1998. He is the Coordinator of the European doctoral training network APROPOS.