Energy-Sustainable IoT Connectivity: Vision, Technological Enablers, Challenges, and Future Directions

ONEL L. A. LÓPEZ [1] (Member, IEEE), OSMEL M. ROSABAL [1] (Graduate Student Member, IEEE),
DAVID E. RUIZ-GUIROLA [1] (Graduate Student Member, IEEE),
PRASOON RAGHUWANSHI[1] (Student Member, IEEE),
KONSTANTIN MIKHAYLOV [1] (Senior Member, IEEE),
LAURI LOVÉN [2] (Senior Member, IEEE),
AND SRIDHAR IYER[3] (Senior Member, IEEE)

[1]Centre for Wireless Communications, University of Oulu, 90570 Oulu, Finland

[2]Center for Ubiquitous Computing, University of Oulu, 90570 Oulu, Finland

[3] Department of CSE (AI), KLE Technological University, Dr. M.S. Sheshgiri Campus, Belagavi 590008, India

CORRESPONDING AUTHOR: O. L. A. LÓPEZ (e-mail: Onel.AlcarazLopez@oulu.fi)

ABSTRACT Technology solutions must effectively balance economic growth, social equity, and environmental integrity to achieve a sustainable society. Notably, although the Internet of Things (IoT) paradigm constitutes a key sustainability enabler, critical issues such as the increasing maintenance operations, energy consumption, and manufacturing/disposal of IoT devices have long-term negative economic, societal, and environmental impacts and must be efficiently addressed. This calls for self-sustainable IoT ecosystems requiring minimal external resources and intervention, effectively utilizing renewable energy sources, and recycling materials whenever possible, thus encompassing energy sustainability. In this work, we focus on energy-sustainable IoT during the operation phase, although our discussions sometimes extend to other sustainability aspects and IoT lifecycle phases. Specifically, we provide a fresh look at energy-sustainable IoT and identify energy provision, transfer, and energy efficiency as the three main energy-related processes whose harmonious coexistence pushes toward realizing self-sustainable IoT systems. Their main related technologies, recent advances, challenges, and research directions are also discussed. Moreover, we overview relevant performance metrics to assess the energy-sustainability potential of a certain technique, technology, device, or network, together with target values for the next generation of wireless systems, and discuss protocol, integration, and implementation issues. Overall, this paper offers insights that are valuable for advancing sustainability goals for present and future generations.

INDEX TERMS Energy efficiency, energy harvesting, energy sustainability, energy transfer, green wireless communication, IoT, machine learning.

I. INTRODUCTION

THE VISION of a sustainable society is relentlessly pursued by industry, academic research, and regulatory bodies following the famous United Nations' sustainable development goals introduced in 2015.[1] In general,

1. Refer to https://sdgs.un.org/goals.

sustainability is stimulated by *simultaneously developing the economy, promoting social equity, and protecting the integrity of the environment for current and future generations* [1], as captured by the so-called sustainability triangle illustrated in Figure 1. Therefore, associating sustainability solely with environmentally conscious (also referred to as "green") practices is a widespread misconception
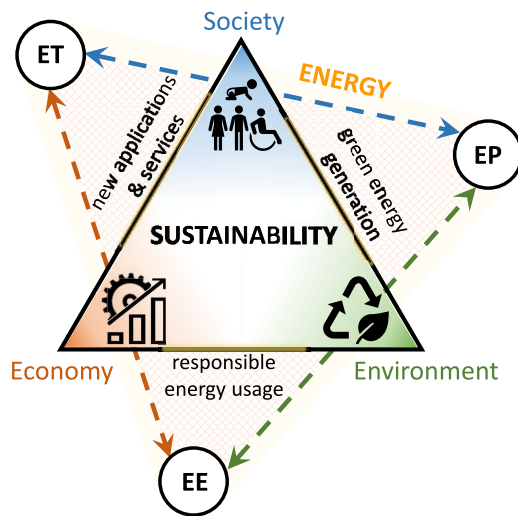
FIGURE 1. Sustainability triangle (with economy, environment, and society as corners) and energy sustainability supported by EP, ET, and EE processes and related technologies.

since sustainability comprises Economic, Societal (Equity), and Environmental factors (also known as the 3 E's of sustainability).

Quantifying/measuring and developing standardized metrics and benchmarks for sustainability is quite a challenging endeavor because i) sustainability is a subjective concept that can be assessed differently by various stakeholders; ii) the interdependencies, trade-offs, and feedback loops between the multiple interconnected systems and performance corners (economic, societal, and environmental) must be captured; and iii) pursuing specific long-term perspectives while measuring and tracking progress in the short term may not be easy and viable in general. This is why, despite numerous attempts over the years, e.g., [2], [3], [4], [5], [6], [7], [8], [9], none of them have gained significant traction.

Regardless of the diverse assessment criteria and methods, sustainability enablers are generally well-established and include Internet of Things (IoT) technologies [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26]. Specifically, IoT solutions promote i) economic development (by enabling automation, facilitating accurate decision-making via massive data collection and real-time processing, optimizing resource usage and efficiency, improving maintenance processes, and reducing systems' downtime); ii) social equity and well-being (by providing enhanced safety and security measures against potential safety hazards or security breaches and by facilitating flexibility and barrier-free experience to the end users via the ubiquitous provision of services and the exploitation of remote-controlled or completely autonomous systems); and iii) environment protection (by optimizing energy usage and reducing waste/pollution); thus advancing sustainability goals.

Notably, IoT systems are turning wireless connectivity into a basic utility, like water and electricity, while becoming increasingly massive [18], [20]. Specifically, the number of

IoT devices is forecast to almost triple from 9.7 billion in 2020 to more than 29 billion IoT devices in 2030[2] to support sustainable agriculture [13], [23], [27], [28], [29], transportation [10], [13], [14], manufacturing [11], [26], wireless charging [13], [15], [16], [21], [22], [26], e-health [12], [13], [23], [30], cities [12], [13], [14], [31], [32], [33], water/air/energy management [13], [14], [15], [23], [33], [34], [35], computing [10], [12], [14], [15], [36], [37], and many other use cases. Interestingly, this massive IoT growth poses sustainability challenges on its own mostly due to the increased maintenance operations, energy consumption (EC) to charge the IoT devices but also to store and process the massive amount of IoT data at the data and computing centers, and manufacturing/disposal of IoT devices, which may have long-term negative economical, societal, and environmental impacts. This calls for self-sustainable IoT ecosystems, which require minimal external resources or intervention and can function independently by utilizing renewable energy sources and recycling materials whenever possible. Notice that truly self-sustainable IoT approaches must consider sustainability aspects along the entire IoT product lifecycle, i.e., planning, manufacturing, deployment, operation (including the target use case), maintenance, and disposal.

A. ENERGY-SUSTAINABLE IOT ECOSYSTEM

Undoubtedly, one of the critical aspects of self-sustainable IoT systems is energy. Self-sustainable solutions must be energy-sustainable, which entails green[3] energy autonomy when considering solely the operation phase of an IoT product lifecycle. Energy-sustainable IoT operation is precisely the focus of this work, although our discussions sometimes extend to other sustainability aspects and IoT lifecycle phases. Readers interested in specific discussions on sustainability at planning, manufacturing, deployment, maintenance, and disposal phases are encouraged to refer to [11], [26], [33], [38], [39], [40], [41], [42], [43].

The main energy processes that are present in an energy-sustainable IoT ecosystem are illustrated in Figure 1 together with the sustainability triangle and discussed below.

Energy Provision (EP): refers to the charging process(es) exploiting green energy sources. Currently, the information and communication technology industry accounts for $1 - 1.5\%$ of the global electricity use [44], a figure that could rise to nearly 4% considering the increasing traffic demands at the transmission networks and the computation requirements at the data centers [45]. Consequently, the implementation of sustainable energy in this sector will have a significant impact on the environment considering that the current average global carbon intensity factor[4] is approximately 0.441 kg of CO_2 emissions per kWh. Furthermore, certain EP implementations enable the production of electricity at a local level, thereby making electricity (and also information and

---

2. Refer to https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide.

3. Green energy refers to energy from renewable sources.

4. Refer to https://ourworldindata.org/grapher/carbon-intensity-electricity.

communication technology services) accessible to remote communities.

Energy Transfer (ET): refers to the intentional movement of energy from one device/system to another. This is required, for instance, when green energy sources are not available for direct exploitation (or provide insufficient energy). In such cases, another device/system powered by green sources can transfer such required energy. The global market projections indicate a significant increase in the value of ET technologies, reflecting the growing interest in this field. As per [46], the global market of ET is projected to surge from USD $5, 705.1 million in 2020 to $35, 226.4 million by 2030, with a resulting average annual growth rate of 21.3%. Consequently, technological breakthroughs in ET will generate new business opportunities, enable previously infeasible IoT use cases, and facilitate the ubiquity of ET services across the globe.

Energy Efficiency (EE): refers to the ability of a device/system/process to perform its intended function with minimal energy and thus is a measure of how effectively energy is used to achieve a desired outcome [47]. Obviously, EE directly impacts the environment and economy sustainability corners as it pushes to minimize EC and processing, i.e., toward more responsible energy usage. Notably, an ambitious goal for the sixth generation (6G) of wireless systems is to realize $10-100$-fold EE gains with respect to the current fifth-generation (5G) networks [48]. Notice that EE designs are often low-complexity/cost, thus promoting sustainability in other IoT lifecycle phases as well.

The harmonious coexistence of the above processes ultimately pushes to zero dirty-EC, thus, realizing truly energy-sustainable IoT systems.

B. CONTRIBUTIONS & ORGANIZATION OF THIS WORK

The realization of energy-sustainable IoT connectivity depends critically on the technological advancements in the EP, ET, and EE processes, and their holistic integration. Notably, the related state-of-the-art discussions are sparse in the literature as most works focus on a specific enabling technology. Only a few other works, such as those listed in Table 1, have a wider scope, although still limited in terms of technology enablers, recent advances, key performance indicators (KPIs), challenges, and/or research directions, and their link to sustainability. Moreover, they generally lack clarity regarding the role or classification of the discussed technologies/approaches toward supporting (self-)sustainable IoT connectivity, e.g., by associating them with the corresponding EP, ET, and EE processes.

Our work aims to address the aforementioned research gap by making the following contributions:

- We provide a fresh look at energy-sustainable IoT. Moreover, we identify EP, ET, and EE as the three main energy-related processes whose harmonious coexistence pushes toward realizing self-sustainable IoT systems. This is mostly covered in Section I.

- We discuss the main technologies, recent advances, and challenges and associated research directions for EP (Section II), ET (Section III), and EE (including communication (Section IV) and learning/computation (Section V)-related aspects processes to support sustainable IoT connectivity.

- We present a set of energy-related performance metrics to assess the absolute/relative performance potential of a certain technique, technology, device, or network. They are classified as i) energy conversion and transfer, ii) energy storage and consumption, iii) EE ratio (EER), and iv) other metrics. Also, we overview related KPIs for the next generation of wireless systems. This is comprehensively covered in Section VI.

- We discuss the integration of the surveyed technologies, suitable protocols and cross-layer designs, scalability issues, tools and testing/validation frameworks, and real-world IoT implementations targeting energy-sustainable IoT ecosystems. This, and corresponding challenges and opportunities, are covered in Section VII.

In addition to these, we summarize the key relevant challenges and research directions both per technology and at the system level in Section VIII. Figure 2 illustrates how the structure of the paper covers the integration of the EP, ET, and EE processes. Finally, Table 2 lists the acronyms used throughout this article in alphabetical order.

II. ENERGY PROVISION (EP)

The need for a reliable energy supply to power the IoT has made EH technologies appealing to reduce/eliminate the need for batteries. There are readily available energy sources in our environment such as sunlight, wind, heat, electromagnetic (EM) radiation, and many others. Table 3 summarizes the advantages, limitations, and attractive use cases for the EH technologies covered in this section.

To power devices' electronics, EH circuits must transform the various energy forms into electric energy. For this, the architecture of a generic EH device, illustrated in Figure 3, consists of three main components:

- transducer, which turns a physical variable (or its variations) into electricity;

- power management unit, which accommodates the output signal of the transducer to power the devices' electronics and/or charge the storage element. One important function of this block is to dynamically match the output impedance of the transducer to a variable input impedance of a rectifying circuit or other device's peripherals;

- energy storage component, e.g., a battery or a supercapacitor, which buffers the variations of the ambient energy. We refer the reader to Section VI-B for more details on this component.

A. LIGHT-BASED EH

Light-based EH systems exploit the photovoltaic effect to generate electricity from light using photovoltaic cells (PVs).

TABLE 1. Representative survey/overview/tutorial papers related to energy-sustainable IoT connectivity (2017-2023).

Ref.	Energy Process			Technology Enablers	Main Contributions
	EP	ET	EE		
[10]	X		X	green mobile edge computing (MEC)	● survey of the state-of-the-art on joint radio-and-computational resource management in MEC ● discussion on MEC challenges and future research directions, including MEC system deployment, cache-enabled MEC, mobility management for MEC, green MEC, and privacy-aware MEC
[11]	X		X	energy harvesting (EH), smart grids, cloud computing, RFID, machine learning (ML)	● review of smart manufacturing and sustainable energy industry ● discussion of prominent technologies to support sustainable industrial IoT
[12]	X		X	EH, RFID, wake-up radio (WuR), cloud computing, data/communication reduction	● survey of potential technologies for greening the IoT and relevant research directions
[13]	X		X	smart microgrids, IoT for waste & energy management, ML, low-power wide-area networks, hardware (HW) and software (SW) duty cycling	● review of the state of the art research and advances on IoT technologies in/for energy management, environment monitoring, transportation, e-health, and smart cities
[14]	X		X	EH, low-power wide-area networks, Bluetooth low energy (BLE), MEC, smart grids	● review of applications of IoT in energy systems, especially in the context of smart grids ● discussion of IoT-enabling technologies and challenges in the energy sector, including privacy and security
[15]	X	X	X	EH, wireless ET (WET), smart grids, MEC, unmanned aerial vehicle (UAV)-assisted communications	● overview of advances/prospects on the utilization, redistribution, trading, and planning of harvested energy in future wireless networks ● review of wireless power and information transfer technologies
[16]	X	X		EH, WET	● proposal of a self-sufficient IoT architecture that adopts only single- and double-hop energy and data transitions to enable efficient energy sharing and reduced data traffic ● discussion of enablers and identification of future research directions
[22]	X	X		green radio-frequency (RF) WET, distributed ledger technology (DLT) for energy trading, ultra-low power receivers, enhanced energy transmitters, metasurface-assisted RF-WET	● discussion of RF-WET for enabling self-sustainable IoT ● overview of the main architectures, challenges, and techniques for efficient and scalable RF-WET ● outline of key research directions for realizing RF-powered 6G IoT
[23]	X		X	RFID, EH, TinyML HW, ML, MEC	● overview of enablers, architectures, technologies, energy models, and strategies for green IoT ● discussion of effective behavioral change models and strategies to create energy-awareness among users and IoT service providers
[25]	X		X	EH, energy-aware TinyML, MEC	● discussions on the incorporation of TinyML algorithms and application tasks on battery-less IoT devices ● study of several inference strategies, including local/cloud computing ● testing a battery-free person detection prototype
[26]	X	X	X	EH, RF harvesters, WET, backscattering	● overview of HW-related research trends, application use cases, and enabling technologies for sustainable IoT systems ● high-level discussion of eco-friendly manufacturing, sustainable WET, and low-power wireless connectivity solutions

Notice that sunlight can provide 100 mW/cm^2 of average power density (APD) outdoors (only during the daytime), whereas combined artificial light and indirect sunlight can illuminate indoors with 100 μW/cm^2 APD [49]. This opens the opportunity for indoor sensors illuminated by artificial light sources [50], a mixture of artificial and sunlight, e.g., in asset tracking [51], and for those in remote places exposed to sunlight with high probability, e.g., sensor and base stations (BSs) in smart agriculture applications [52]. However, these reference values can change depending on the geographic location, weather conditions, and indoor working hours. Increasing the conversion efficiency (CE) of PVs
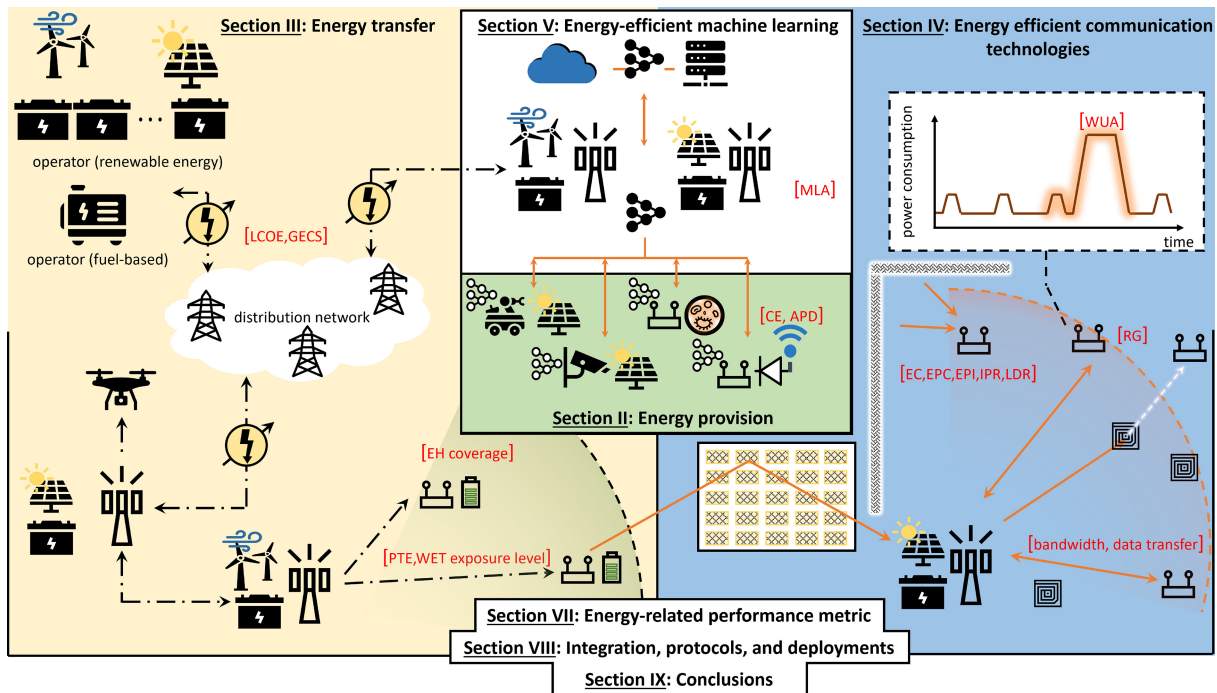
FIGURE 2. Integration of the EP, ET, and EE processes throughout the structure of the paper. The terms in brackets refer to KPIs, which are discussed in Section VI.

under different ambient light conditions is key for reducing the footprint, increasing the cost-effectiveness, and accelerating the adoption of light-based EH systems. In this front, multi-junction PVs, which, as shown in Figure 4a, are constructed by stacking multiple p-n junctions, become appealing as each layer is optimized for harvesting at different wavelength regions, thus, increasing total harvested energy [53]. However, multi-junction PVs are challenging and expensive to manufacture, which motivated researchers to propose the replacement of some silicon layers with perovskite PVs. Perovskite-based multi-junction cells facilitate tuning the optical properties of the PVs for improved response in the solar spectrum and lower the manufacturing costs considerably [54]. Unfortunately, perovskite PVs' instability in real-life conditions significantly degrades their CE over the long-term (compared to silicon PVs), which makes them (currently) a less favorable option for widespread use [55].

Concentrator photovoltaics (CPVs) are an alternate solution to boost the harvested energy. This system, as shown in Figure 4b, relies on a built-in light concentrator to focus the incoming light from multiple directions on a very small semiconductor area [56]. This eliminates the need for complex electro-mechanical tracking mechanisms, which increase the operational expenditure, aka OpEx, and limit the possible use cases, such as in tiny Bluetooth transmitters, due to the resulting heavier design and moving parts [57]. Note that a high concentration of sunlight increases the operating temperature of CPVs, thus decreasing their efficiency.

In general, the seamless integration of light-based EH with the corresponding application is required. This is favored

by CPVs, which allow the miniaturization of light-based EH systems to the millimeter and sub-millimeter scale with reduced manufacturing costs [58]. Recent developments even allow the implementation of flexible CPVs accommodating multiple PV modules in a small surface, leading to improved performance and seamless integration with a multitude of device designs [59]. Moreover, thin-film PVs improve the integration capabilities by allowing the deployment of light-based EH systems on complex-shaped surfaces such as cars' roofs and umbrellas [60]. In this regard, the recent introduction of organic and perovskite PVs permits printing PVs on practically any surface (e.g., check the IoT device in [61]) without impacting their mechanical properties or transparency [62]. Despite their current low efficiency, the abundance of required raw materials, low-cost and scalable manufacturing process, and the possibility of using biodegradable materials, make these cells a green choice for future EH IoT implementations.

B. HEAT-BASED EH
Heat-based EH exploits temperature changes in the surrounding environment [63]. Specifically, thermoelectric generators (TEGs) turn spatial gradients of temperature on the device's surface into electric energy. TEGs work on the principle of the Seebeck effect, due to which a pair of dissimilar materials exposed to different temperatures generate electricity [64]. The basic TEG's architecture is illustrated in Figure 5a and consists of a set of n- and p-type semiconductors connected electrically in series, and thermally in parallel. To maximize the output power, one side is connected to a heat sink which

TABLE 2. List of most common acronyms.

Acronym	Definition	Acronym	Definition
3GPP	3rd Generation Partnership Project	LOS	line-of-sight
5G/6G	fifth/sixth generation	MAE	mean absolute error
AHE	average harvested energy	MCU	microcontroller
APD	average power density	MDHE	meta distribution of the harvested energy
BC	backscatter communication	MDS	model size
BCAP	battery capacity	MEC	mobile edge computing
BL	battery lifetime	MFC	microbial fuel cell
BLE	Bluetooth low energy	MIMO	multiple-input multiple-output
BoI	bandwidth of influence	MIPS	million instructions per second
BS	base station	ML	machine learning
CA/CV	coverage area/volume	MLA	ML accuracy
CE	conversion efficiency	MPE	maximum level of exposure
CEE	component EE	NAS	neural architecture search
CEP	complex event processing	NHE	net harvested energy
CPV	concentrator photovoltaic	NFV	network functions virtualization
CSI	channel state information	NN, FANN	neural network, fast neural network
CU/DU	centralized/distributed unit	O-RAN	Open-RAN
DC	direct-current	PB/PC	power beacon/consumption
DL	deep learning	PFA	probability of false-alarm
DLT	distributed ledger technology	PHY	physical layer
DRL/DSL	deep RL/SL	PKD	peak demand
EC	energy consumption	PMD	probability of miss-detection
ED	energy density	PTE	power transfer efficiency
EE/EP/ET	energy efficiency/provision/transfer	PULP	parallel ultra-low power
EER	EE ratio	PV	photovoltaic cell
EH	energy harvesting	QoS	quality of service
EM, EMF	electromagnetic, electromagnetic field	RAN/RAT	radio access network/technology
ENO	energy-neutral operation	RF	radio-frequency
EOP	energy outage probability	RG	range
EPC/EPI	energy proportionality coefficient/index	RL	reinforcement learning
FL	federated learning	RMSE	root mean square error
FLOPS	floating-point operations per second	RU	radio unit
FN/FP	false negative/positive	SL	supervised learning
GECS	green EC share	STD	standard deviation
GLB	geographical load balancing	TEG	thermoelectric generator
GRS	grid reliability/stability	THP	throughput
HW/SW	hardware/software	TN/TP	true negative/positive
INFT	inference time	TX/RX	transmit/receive
IoT	Internet of Things	UAV	unmanned aerial vehicle
IPR	idle-to-peak power ratio	UE	user equipment
IRS	intelligent reflective surface	UL	unsupervised learning
KPI	key performance indicator	VEH	vibration-based EH
LA	localization accuracy	WET	wireless ET
LCOE	levelized cost of electricity	WUA	wake-up accuracy
LIS	large intelligent surface	WuR/WuS	wake-up radio/signal

increases the temperature difference with respect to the side exposed to the heat source. In practice, an array of thin-film TEGs becomes more appealing as one can accommodate more devices per unit volume, thus increasing the harvested energy [65]. Moreover, the miniaturization of the TEGs makes heat-based EH difficult since a very small surface area may not suffice to capture spatial temperature gradients.

TABLE 3. Comparison among EP technologies.

Source	CoC[a]	Limitations	Use cases
Light	M	Sensitive to blockage	Remote applications, indoor sensors
Heat	M-H	Sensitive to thermal stress	Wearables, infrastructure monitoring
Microbes	M-VH	Low stability/lifespan	Smart farming, wastewater treatment
Vibration	VH-H	Sensitive to mechanical fatigue	Wearables, infrastructure monitoring
Flow	H	Sensitive to weather	Medium to large EH systems
RF	L	Very low output power	Ultra-low power devices

[a] CoC refers to complexity and cost which is classified as very high (VH), high (H), moderate (M), low (L), and very low (VL)
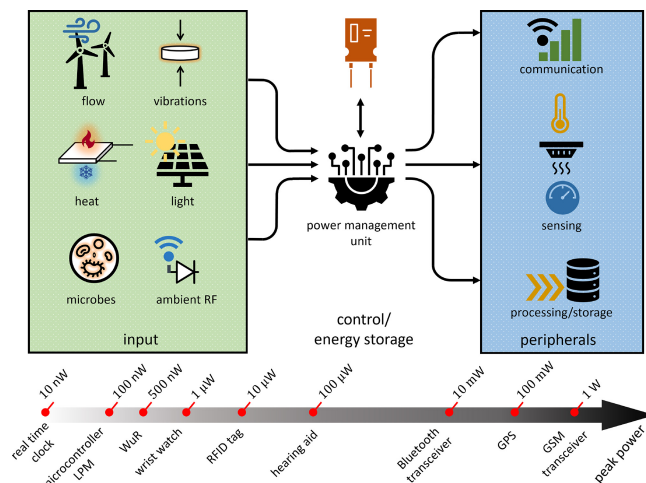

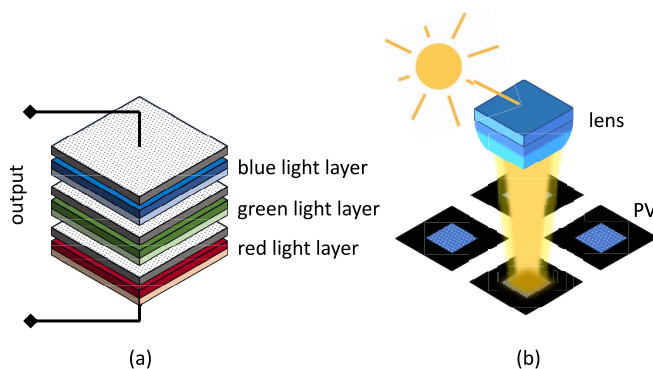**FIGURE 3.** Anatomy of an EH device.

FIGURE 4. Efficient light-based EH: (a) multi-junction PV and (b) CPVs.

Pyroelectric generators are an alternate implementation of heat-based EH that can transform temporal temperature variations into electricity. In principle, pyroelectric materials have an asymmetric crystallographic structure whose charges' disposition changes in response to a temperature variation [66]. Figure 5b illustrates the basic architecture
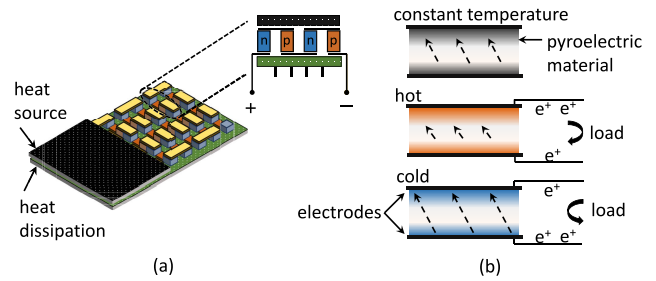

FIGURE 5. Basic architecture of (a) heat-based EH and (b) pyroelectric generator.

and operating principle of a pyroelectric generator. When the material is heated, its lattice structure expands creating more space for the charges to move, which weakens the polarization. On the contrary, the entire structure shrinks when the material is cooled, thus reinforcing the lattice asymmetry and increasing the spontaneous polarization of the material. In either case, the temperature variations will cause a flow of electrons between the electrodes to compensate for the new charges' disposition in the crystal.

Heat-based EH technologies are especially useful for waste heat recovery, e.g., in automotive applications [67], for powering wearables (using the temperature changes in the human body) [68], and in industrial facilities [69].

C. MICROBIAL FUEL CELLS (MFCS)

MFCs exploit bioelectrochemical conversion to harness the energy resulting from the metabolic processes of microorganisms to generate electricity. MFCs are useful in such scenarios where the devices interact with organic matter, such as compost, wastewater, and ponds [70].

Notice that MFCs require a continuous supply of fresh materials to prevent the microorganism from depriving the available organic matter and hence stopping the electricity generation. To solve this problem, plant MFCs leverage the plants' root exudates and rhizodeposits, resulting from the photosynthesis process, to continuously feed the microorganisms [71], as illustrated in Figure 6. Plant MFCs can be regarded as a form of converting the sunlight stored as chemical energy in plants into electricity. Factors such as plant species, weather conditions, soil nutrients, and microbial diversity heavily determine the lifetime and output power of this technology [71].

Further development directions focus on boosting the output power and lowering the manufacturing costs of current MFCs. In this regard, the authors in [72] showed that stacking multiple MFCs boosts the CE compared to individual MFCs. Moreover, the authors in [73], [74] discussed how to build cost-effective MFCs as well as key properties of the components such as high conductivity, large surface area, porosity, durability, and corrosion resistance.
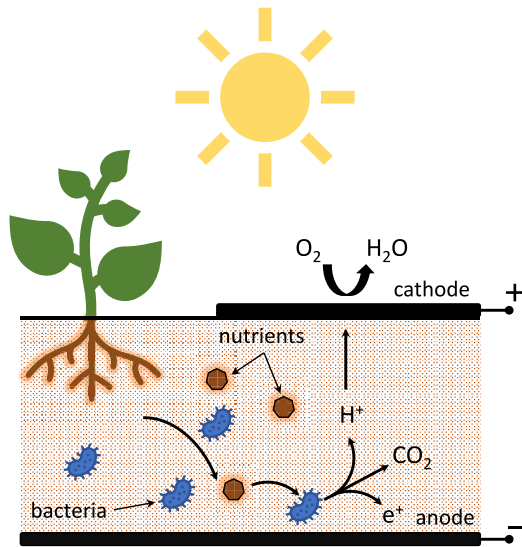
FIGURE 6. Plant MFC operating principle.

FIGURE 7. Basic working principle of VEH: (a) vibration EM, (b) piezoelectric, (c) triboelectric, and (d) electrostatic.

D. VIBRATION-BASED EH (VEH)

VEH is the process of converting mechanical energy in the form of vibration (including acoustic vibrations), impact, deformation, or friction into electric energy. As Figure 7 shows, there are four main types of mechanical-based EH:

- Piezoelectric energy harvesters, which exploit the direct piezoelectric effect, i.e., certain materials, such as quartz, generate electrical charges under mechanical stress [75]. The output power depends on the relative directions of the electric field and the stress/strain for the polarization direction of the piezoelectric material.
- Triboelectric energy harvesters, which exploit the triboelectric effect, i.e., certain materials generate electrical charges when they contact and separate from another material. Due to their relatively small form factor and high APD, these harvesters are widely used in wearables to harness energy from human motion [76].
- Electrostatic energy harvesters, which are variable capacitor structures. Specifically, their capacitance changes when the plates' overlap area or the gap distance varies in response to an external force. The basic operating principle consists of exciting the capacitor with an external power supply when the capacitance is maximum. Then, the energy is harvested when the potential energy stored in the capacitor increases in response to a smaller capacitance value [77]. Due to the high polarization voltage required in this method, some electrostatic energy harvesters incorporate electrets, which are quasi-permanent electric dipoles that can hold electrical charges for years.
- Vibration EM energy harvesters, which exploit Faraday's law of induction, i.e., the relative movement between a wire coil and a permanent magnet produces an electromotive force.
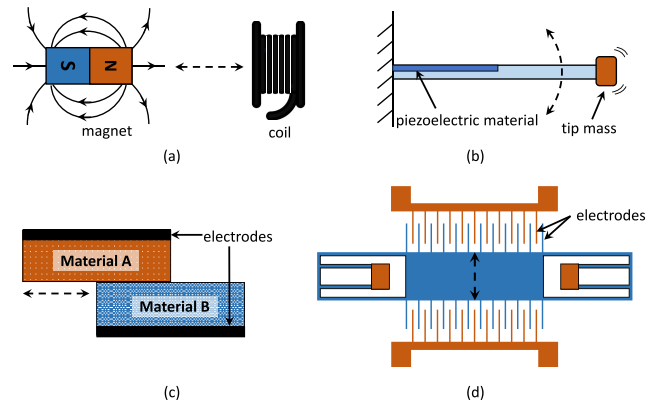
VEH presents a significant challenge: the resonant frequency of the EH device must match the input environmental frequency to maximize the output power [78]. However, ambient energy frequency can vary considerably, demanding the adjustment of the resonant frequency to ensure a reliable energy supply. There are various approaches in the literature to achieve this, including manual mechanical tuning [79], mechanical self-tuning [80], and electronic self-tuning [81]. Another solution is to design the device to resonate at multiple frequencies, thereby expanding its frequency response spectrum. For this, hybrid energy harvesters [82] or arrays of multiple harvesters [83] can be used, with each device resonating at a different frequency.

Another challenge of VEH is the relatively low frequency of ambient vibrations with respect to the resonant frequency of the devices. To address this issue, frequency up-conversion mechanisms have been proposed. For instance, by applying a strong external force for a brief period, e.g., using a free-moving mass, the system may vibrate at its natural resonant frequency [84]. Alternatively, one can lower the harvester's natural resonant frequency by reducing the stiffness of its moving parts. This can be achieved, for instance, by replacing them with fluids [85].

Finally, ambient VEH is often the result of multiple external forces acting in different directions. Thus, allowing devices to harvest energy from multiple directions, e.g., using two-dimensional [86] and three-dimensional [87] EH devices, can boost the total output power.

E. FLOW-BASED EH

Flow-based EH leverages the kinetic energy of naturally or artificially originated fluid flows, such as water streams and wind currents, to generate electricity. Notice that in most use cases, the underlying physical principles of VEH are also applicable to flow-based EH. Therefore, this subsection focuses on wind-based EH and, in particular, on their form factor, blade design, and energy-efficient design for medium- to large-scale electricity generation.
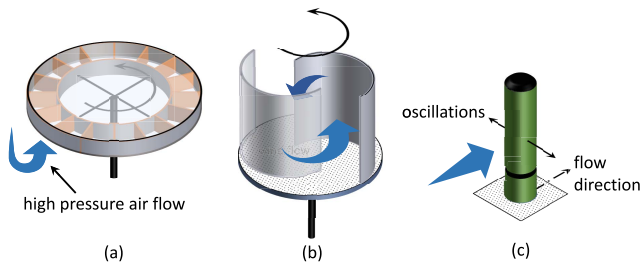
FIGURE 8. Wind turbine architectures: (a) Powerpods internal blades system, (b) Savonius turbines, and (c) Vortex Bladeless generator.

Traditional wind turbines leverage the kinetic energy of the wind to turn the rotor of an electric generator. Wind turbines are usually deployed in farms for large-scale electricity generation. However, massive IoT deployments are more likely to appear in cities where installing a huge wind turbine is not always feasible. Thus, research and industry efforts have been focused on building compact and cost-effective designs of wind-based EH systems. A significant issue with traditional wind turbine design is the potential threat to wildlife posed by the rotating blades. To address this, Halcium has created the PowerPods,[5] which are portable wind turbines designed for residential areas. The blades of the PowerPods are contained within the pod, as shown in Figure 8a, making them safe for people, pets, and wildlife in close proximity. The shell of the PowerPods captures wind from multiple directions and channels it through small exits, increasing the wind speed and therefore the harvested energy. Different from traditional wind turbines, PowerPods utilize vertical-axis blades whose low aerodynamic noise pollution and ability to operate under unstable wind flow (without needing a yawing mechanism) make them appealing for urban environments [88]. Vertical-axis turbines are also the basis of the so-called Savonius turbines (shown in Figure 8b) which are reliable and cost-effective systems that can operate under turbulent wind flows and stormy weather [89]. Notice that traditional wind turbines require a braking mechanism to operate under such conditions as high wind speeds can cause the system to shatter becoming a hazard for the nearby areas.

Another development area in wind-based EH involves blade-less turbines, e.g., Vortex Bladeless[6] shown in Figure 8c, which leverages the vortex shedding phenomenon [90]. Vortex shedding occurs when a fluid flow separates from a solid surface and creates vortices. In the case of the Vortex Bladeless, the vortex shedding effect causes a vertical mast to oscillate in resonance with the wind flow. One advantage of this design is the reduced maintenance costs, as there are no moving parts that can wear out through friction. However, the system can experience mechanical fatigue and stress at the base of the oscillating mast. Additionally, the smaller footprint of the Vortex Bladeless compared to traditional wind turbines allows for

5. For more information please check https://www.halcium.com/.
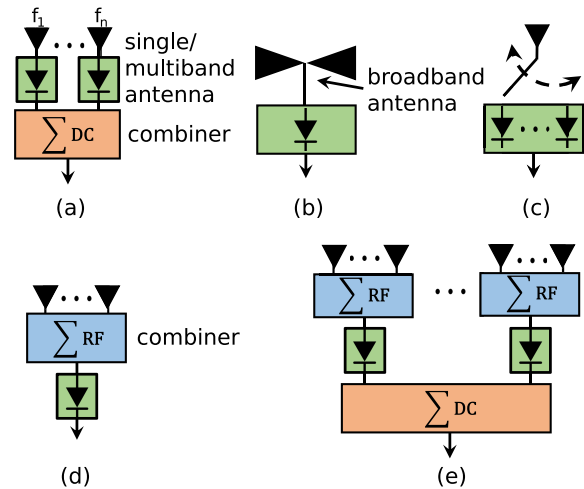6. For more information please check https://vortexbladeless.com/.

FIGURE 9. Architecture of (a) DC combining, (b) broadband, (c) extended dynamic range, (d) RF combiner, and (e) hybrid combiner.

the installation of multiple units in the same area, potentially compensating for their individual lower performance.

F. RF-EH

RF-EH technology utilizes a rectenna to convert EM waves into direct-current (DC) power, which can be harnessed to power and/or charge small electronic devices and batteries. The rectenna is composed of a receiving antenna and a rectifier plus a matching network to maximize the power transfer efficiency (PTE). The main sources of ambient RF energy are classified as [21]: i) static, which corresponds to energy transmissions that remain relatively stable over time, e.g., from television and radio transmitters, allowing long-term predictability of the energy supply, thus favoring the network planning, and ii) dynamic, which supplies time-varying energy over a certain region, either because of fluctuating transmit power levels or mobility, e.g., from WiFi/mobile access points or devices, thus should be adaptive and allow operation over multiple frequencies.

The main challenge of ambient RF-EH circuits lies in how to guarantee an appropriate performance given mostly unpredictable EM conditions. Notice that the receive signal parameters, e.g., carrier frequency, bandwidth, polarization, and antenna orientation determine the amount of harvested energy. Hence, key desirable features of RF-EH circuits include performing EH i) from multiple bands, which can be accomplished by using wideband, multi-band, or tunable receivers [21], ii) from multiple spatial directions, for which omnidirectional antennas are preferred, iii) from randomly polarized signals, which is mostly implemented by some sort of circularly polarized antenna; and iv) from a variety of input power levels, in which case the RF-EH adjust the sensitivity and saturation levels to obtain the most from the ambient energy source. Figs. 9a-c illustrate the basic architecture of some of the aforementioned receivers.

Multi-antenna RF-EH circuits, as shown in Figure 9a, Figure 9d, and Figure 9e, are another alternative to boost

TABLE 4. Comparison among WET technologies.

Technology	Efficiency	Complexity	Range (RG)	Scalability	Standardization
RF-WET	Medium	Moderate	Long	Very good	AirFuel Alliance [AirFuel RF], ARIB STD-T113, NFC Forum [Wireless Charging Specification 2.0]
Inductive	Very high	High	Medium	Moderate	Wireless Power Consortium [Qi,Ki standards], SAE [J2954,J2847/6], IEC 61980-3:2022, ISO 19363:2020, ISO 15118-20:2022, AirFuel Alliance [AirFuel resonant]
Capacitive	High	Moderate	Medium	Limited	ARIB STD-T113
Laser	Low	High	Long	Moderate	-
Acoustic	Very low	High	Very short	Very limited	-

the amount of harvested energy. In such case, the incoming signals can be combined i) in the RF domain; ii) in the DC domain, or iii) hybridly in both domains [91]. Multi-antenna rectennas also improve the spatial selectivity of the antenna in the direction of the maximum incident signal, if properly adjusted. However, for ambient RF-EH the optimal receive beamforming is difficult to attain due to the non-dedicated nature of the transmissions. To cope with this problem, several low-complexity solutions, including codebook-based beamforming, were proposed in [91] such that an ambient RF-EH device sweeps a phase shift table in an initial phase seeking the configuration that yields the best performance for exploitation in a second phase.

Magnetostrictive antennas are an alternative to rectennas. In such a case, the transducer is built with multiferroic structures that are able to sense the magnetic component of EM waves and, in response, produce vibrations. Then, a mechanically connected piezoelectric EH converts the vibrations into electrical energy. This allows for realizing an ultra-compact receiver, which is up to two orders of magnitude smaller than state-of-the-art rectennas [92]. Metamaterials-aided designs are another solution for realizing ultra-compact RF-EH implementations, due to the sub-wavelength periodicity of the unit cells within the lattice structure. Metasurfaces-aided RF-EH boosts the CE as it allows wider beamwidths, polarization-independent operation, built-in and less lossy matching networks, and higher antenna gains [93].

III. ENERGY TRANSFER (ET)
ET refers to the intentional transmission of energy using dedicated transmitters to power EH devices, and thus it is mostly/inherently wireless, leading to WET. In contrast to ambient EH, WET provides a controllable and predictable energy supply and the tools for increasing the end-to-end power CE. In Table 4, we compare the current development state of the WET technologies discussed in this section.

A. LASER-BASED WET
The most common implementations of laser-based WET are laser power beaming and distributed resonant beam charging. In the former case, shown in Figure 10a, a laser diode at the transmitter steers an optical beam towards a PV cell at
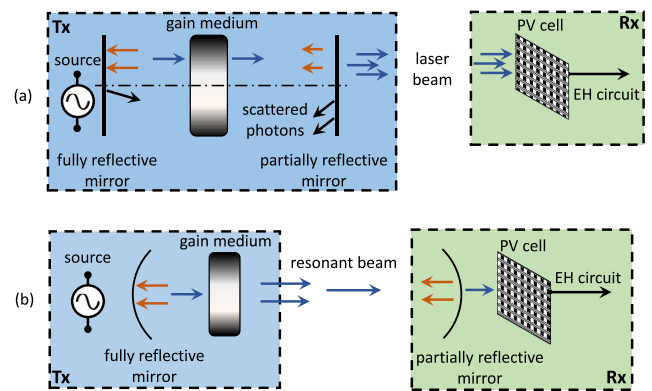

FIGURE 10. Laser-based WET architectures: (a) laser power beaming and (b) distributed resonant beam charging.

the receiver (similar working principle as in conventional solar power systems). Meanwhile, in the latter, the receiver bounces back a portion of the incident energy towards a gain medium at the transmitter, which amplifies the light and initiates a resonant beam [94], as illustrated in Figure 10b.

Although the above two implementations are regarded as long-RG WET technologies, laser power beaming provides superior coverage, in the order of several km [95], and may be key for intra-satellite WET applications [96], charging UAVs [97], and energy transmission from solar-powered satellites to terrestrial stations [98]. However, the performance of the laser power beaming heavily depends on the line-of-sight (LOS) and the atmospheric conditions. Besides, laser power beaming raises safety concerns due to the potential harm to the living species' tissue. Notably, lowering the transmission frequency is appealing to provide a hazardous-free system, although incurring high maintenance cost and lower efficiency (and therefore trading efficiency for safety) [95]. Alternatively, it is advisable to monitor the periphery of the beam path for detecting intruders [98].

Differently, resonant beam charging is inherently safe as any interruption of the link causes the resonance to stop immediately. In addition, the system is capable of supporting mobile users without the need for a tracking system as long as LOS conditions hold, thanks to its self-aligning capabilities which have been experimentally validated (e.g., cf. [99]). Moreover, the transmitter's ability to

generate simultaneous resonating beams also enables native multi-user support [100]. However, broadcasting energy to multiple receivers is typically more inefficient, often motivating time division-multiple device-charging protocols [101]. To overcome the LOS limitations, UAVs [94] or supporting reflectors [102] can relay energy transmissions to charge devices out of the coverage of the main transmitter.

B. ACOUSTIC WET

This technology leverages acoustic waves to charge EH devices by typically equipping transmitters and receivers with piezoelectric transducers [103]. Acoustic WET works in any medium capable of propagating pressure waves, e.g., metal, air, and human tissue, but it is particularly convenient in EM wave absorption-prone mediums, such as Faraday shielding structures [104] and water. Besides, for the same operating frequency, acoustic WET transmitters/receivers have a more compact form factor and achieve a higher directivity than those implementations based on EM WET [105].

Similar to other WET technologies, the channel attenuation, which varies significantly according to the medium, affects severely the system performance. Notably, one cannot increase deliberately the intensity of the acoustic waves as it may cause hearing impairments, body heating, and other unpleasant effects in humans and animals depending on the operating frequency [106]. Notice that the attenuation increases with the impedance mismatch of the medium between the transmitter and the receiver since traveling waves may encounter different materials in their path. This has motivated the use of glue or electromagnets to ensure a firm connection of both the transmitter and receiver to the matching layers in the path (when solid) and therefore reduce the propagation losses [104]. Moreover, when receivers are deployed inside a Faraday shielding structure but not connected to it, one can rely on hybrid WET approaches in which the last section of the transmission path relies on a different WET technology, such as inductive coupling [107].

Phased acoustic arrays, illustrated in Figure 11, which are composed of multiple piezoelectric transducers, also aid in overcoming medium attenuation by focusing the energy toward the receiver direction using sound beams. In fact, the performance of transmit phased acoustic arrays increases with the number of transducers given a form factor constraint [105], [108]. Further improvements can be achieved by also equipping the receiver with phased acoustic arrays and hence allowing different combining techniques to improve the PTE, similar to a traditional multiple-input multiple-output (MIMO) wireless link [109]. Notice that the achievable performance of a phased acoustic array also depends on the transducers' geometry and arrangement, and the array's aperture and diameter [110].

Finally, acoustic WET can also provide heterogeneous quality of service (QoS) to the end users. For instance, acoustic transmitters can use Lamb waves to create a pattern of peaks and valleys in solid structures. Therefore,
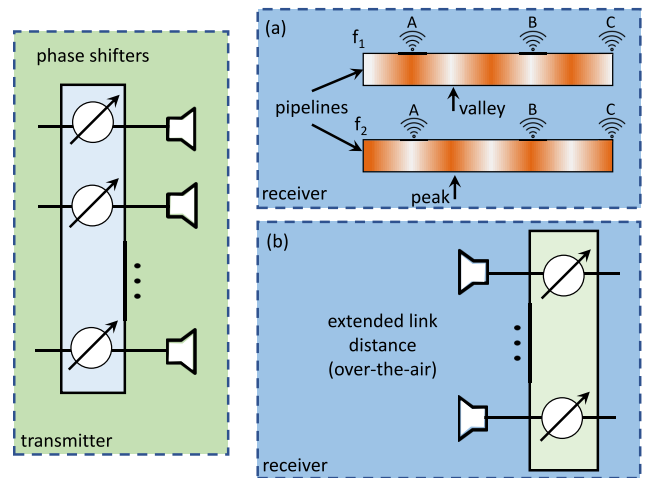

FIGURE 11. Acoustic phased-arrays architecture and use cases: (a) multiple piezoelectric transducers at the transmitter to power a sensor network embedded in pipelines and (b) phased-array at the receiver to extend the operating RG.

by changing the operating frequency, the resulting vibrations charge the devices deployed in different locations of the same structure with a different intensity [111]. Besides, adaptive acoustic beamforming also serves to discriminate which devices to charge depending on their locations and energy demands [112]. This might be especially useful in sensor networks embedded in structures such as buildings or bridges.

C. INDUCTIVE COUPLING-BASED WET

This technology exploits the inductive coupling phenomenon. The simplest setup consists of two wire coils coupled by a magnetic field such that the oscillating magnetic field in the transmitter's coil passes through the receiver's coil inducing an alternating current. This forms a highly efficient air-gap transformer whose performance depends on the operating frequency and the mutual inductance between both coils. However, this basic setup performs poorly when both coils are misaligned or too separated.

One can extend the charging coverage by tuning both coils to resonate at the same frequency. This is commercially known as magnetic resonant coupling, which compared to the basic inductive coupling, is more resilient to coils' misalignment. The most common magnetic resonant coupling methods include an external capacitor to compensate for the internal inductive reactance and extra coils for tuning and impedance matching. The first implementation is easier to realize, but the second one achieves a higher CE given that there are no power losses in external resonators [113].

Magnetic resonant coupling allows a transmitter to charge multiple nodes simultaneously. Here, the internal resistance of the resonators can be adjusted to control the mutual inductance among coils and hence boost the PTE. For instance, increasing the internal resistance of the nearest receivers may increase the harvested energy at the most distant receivers, although at a higher transmit power cost [114].

Deploying multiple transmitters also improves the system performance. On the one hand, the optimal transmitter deployment guarantees uniform power coverage and ensures a minimum available energy at the receivers regardless of their locations [115]. On the other hand, coordinated transmissions from multiple transmitters can result in a constructive combination of the magnetic fields at the receivers' locations, which is known as distributed magnetic beamforming [116].

D. CAPACITIVE COUPLING-BASED WET

This technology exploits the capacitance coupling phenomenon. The basic architecture consists of two pairs of transmit and receive plates each coupled by an electric field. One pair forwards the displacement current while the other provides the return path to close the circuit. Capacitive coupling systems come in two different flavors depending on how the plates are arranged. The horizontal capacitive coupler has the transmitter's (and receiver's) plates placed side by side in the same plane, while the vertical capacitive coupler has both the forward and return paths overlapped in the direction of the electric field. Although the former architecture has the highest coupling, it comes at the cost of being bulky and less reliable due to the large number of required components. For a fixed device size, one can enhance the coupling of the system by combining the geometry of the basic architectures and adding more plates [117].

Some critical challenges in capacitive coupling systems include the high excitation frequency/voltage due to the large capacitive reactance of the coupling and the control of the fringing electric fields, i.e., the non-uniform field at the edge of the plates, within safe limits. To cope with these issues, resonant matching networks can be added at both the transmit and receive sides. They compensate for the reactive losses in the system, hence reducing the required excitation voltage. Further, they can be designed as a transmit voltage gain stage to drop the displacement current of the plates, hence reducing the fringing field to the safety limits. Herein, the size of the inductors in the matching network is key to realizing compact designs. As an example, one can exploit the successive impedance transformation that provides multistage matching networks [118]. The required amount of gain and compensation of such a matching network depends on the ratio between the load and excitation currents, and the misalignment/distance between couplers [119].

Different from inductive coupling, capacitive coupling systems are more tolerant to the plates' misalignment, reduce the risk of interference to neighboring networks, and do not induce eddy currents in nearby metallic objects. Moreover, they are lighter, easy to integrate, and more mechanically robust. That is why they are regarded as promising for powering medical implants, vehicles, consumer electronics, and rotary electric machines [120], [121].

E. RF-WET

RF-WET technology relies on intentional RF transmissions to charge RF-EH devices. The main challenge of this technology is its limited coverage due to the channel attenuation and the regulations on the maximum radiating power of the RF energy transmitters, hereinafter referred to as power beacons (PBs). For this reason, directive antennas are preferred to increase the incident RF power in the desired direction without increasing the transmitter's radiating power. Notice that when serving multiple users, the PB can utilize an omnidirectional antenna to provide basic service guarantees and a directional antenna (for out-band) energy transmissions to meet more specific user equipment (UE) requirements [122]. However, in case the PB is equipped with non-reconfigurable antenna(s), a mechanical sweeping of the service area may be needed to charge those devices otherwise located in the minimum of the radiation pattern [123].

Another potential technique to extend RF-WET coverage is channel state information (CSI)-based energy beamforming. The main challenge here resides in the often unavoidable cost of instantaneous CSI acquisition, especially when powering a massive number of devices with strict HW constraints, which may null or even reverse the gains from accurate CSI-based transmit strategies. This has motivated alternative strategies relying on statistical CSI [124], received energy feedback [125], and positioning information [123].

Conveniently deploying the PBs is also key to overcoming channel attenuation, banning blind spots, and homogenizing the incident RF power according to the network requirements [22], [126], [127]. Notably, distributed single-antenna PBs offer better service than a single PB equipped with the same total number of antennas due to the reduced charging distance [127] and the reduced number of EH devices to be charged per PB [22]. Robotic WET, in which the PBs move [128] or fly [129], becomes also appealing not only to reduce the link distance but also to power the devices deployed in hard-to-reach places or to meet temporary service requirements during emergencies. Other potential technologies to boost the coverage of RF-WET are also discussed in Sections IV-B and IV-C.

The next generation of wireless systems envisions extremely large antenna arrays to compensate for the attenuation at high-frequency bands, increase reliability, and reduce interference. Consequently, many future IoT deployments may operate in the radiating near-field (or Fresnel region) of the transmit antennas, which contrasts with the traditional assumption of far-field operation. As Figure 12 illustrates, under such operating conditions, a PB can focus the energy in a particular location as opposed to what happens in far-field conditions, where the energy is steered towards a certain angular direction. As a consequence, near-field RF-WET will generate less RF pollution, thus interference, in both angle and distance domains and a reduced human RF exposure [130]. Moreover, the CE achievable in the near-field can be significantly high even when the
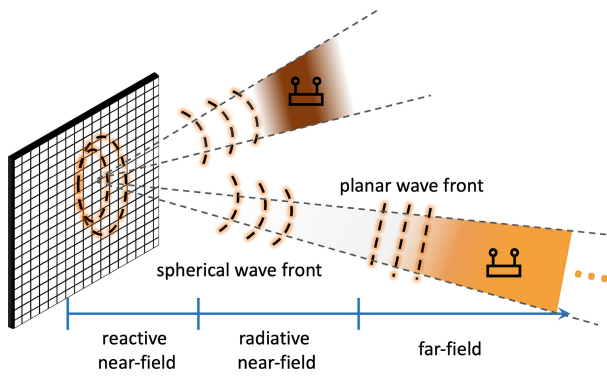
FIGURE 12. Near-field vs far-field beamforming.

energy receiver is not in the focal point of the transmitter's antenna [131].

Due to the above, RF-WET is usually regarded as a short-distance solution to charge low-power devices. However, some experiments have shown energy transmissions over distances greater than 1 km distance with a peak incident power of several Watts at the rectenna [132]. Long-distance RF-WET could also help bypass the complicated infrastructure of the electric network to power hard-to-reach locations or to provide an energy supply during emergency situations.[7]

IV. ENERGY EFFICIENT COMMUNICATION TECHNOLOGIES

Once the energy is available for operation through EP/ET mechanisms as those outlined in Sections II and III, the device(s) and/or the network(s) must ensure its efficient usage. Indeed, every technique and technology conceived for performance improvements in terms of coverage, throughput (THP), dependability, and other KPIs, can be in principle leveraged for EC reduction as well, hence, EE. That is the case of massive MIMO [133], [134], UAV-based connectivity [135], [136], [137], satellite-assisted communications [138], [139], cooperation and diversity mechanisms [140], [141], [142]. Nevertheless, there are some technologies/techniques that are natively conceived for energy-limited/efficient operation, and these constitute the scope of this section.

Notice that the primary EC sources in a device or network are i) the utilization of active circuit components such as transistors and power amplifiers, which require a power source to function, and ii) the running applications. Therefore, to enable energy-efficient and/or low-power operation, the use of active techniques and devices must be limited when possible while relying on efficient passive architectures. This comes with several challenges as active devices, although consume energy, can facilitate high-performance computing and communication, while passive devices (e.g., resistors, capacitors, inductors) are much less flexible and may dissipate energy in the form of heat as they interact with electric signals. This motivates the use of semi-passive/active architectures, which include some active techniques and

7. Please refer to https://emrod.energy/ for more details.
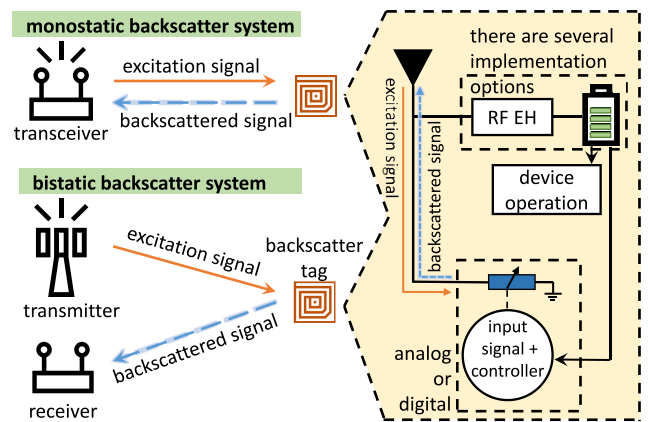
FIGURE 13. BC systems and main components.

components to efficiently support the application demands for which they are deployed, and/or tunable operational models as enabled by WuR technology. These approaches are discussed next through key technologies. Specifically, backscatter communication (BC), metasurface-aided communication, radio stripes, and WuR technologies are overviewed in Sections IV-A–IV-D, respectively. Meanwhile, the ML approach, focused on intelligently reducing the EC burden of computation/communication application tasks at the devices and networks, is discussed in Section V.[8]

A. BACKSCATTER COMMUNICATIONS

BCs are passive and involve backscatter tag(s) reflecting the signal from nearby transmitter(s) and modulating it by adjusting its amplitude, frequency, and/or phase via impedance mismatching tuning. At the receiver, the backscattered signal is processed to extract the information added by the backscatter tag. The typical components and types of BC systems are illustrated in Figure 13. Notice that the transmitter(s) may be dedicated or non-dedicated, the latter leading to the so-called ambient BC systems.

There are multiple implementation options for powering the backscatter tags, e.g., relying on EH from ambient/dedicated RF signals [146], [147], [148], solar energy, vibration, and/or thermal gradient [51], [147]. In general, the energy buffer can be as small as a capacitor of several nF, or as large as a digitally-controlled supercapacitor or rechargeable microbattery. The nature of the energy buffer depends on the application. A capacitor may suffice for ultra-low power applications, and it is not often regarded as a buffer but simply as another stage element within the circuitry. In other cases, a large and intelligently controlled energy buffer is required to enable substantial gains in RG and operating time that could not have been attained otherwise. Such energy storage topology influences also the

8. Please note that our focus in this work is on technologies (other than wireless communication standards/protocols) that we believe are most relevant for enabling low-power IoT connectivity in the near future. The interested reader can refer to other potential technologies such as molecular communication [143], printable electronics [144], and passive radar [145].

type of modulation/demodulation of the backscatter signals, which can be digital or analog as in conventional communication systems, and may be activated by a wake-up code detector [149].

Table 5 lists representative state-of-the-art BC systems along with their distinctive features and performance figures.

1) KEY BC TOPOLOGIES

As shown in Figure 13, there are two key BC topologies: monostatic and bistatic. In traditional or monostatic backscatter systems, the transmitter and receiver are integrated, e.g., in RFID, while they are separate in bistatic systems.[9] The latter implementation offers [170]:

- Temporal flexibility: The tag has additional time slots to transmit data rather than being restricted to waiting for a single reader's protocol-bound inquiry. This allows the backscatter tag (functioning as a sensor node) to transmit the sensory data immediately upon availability, which might be crucial in many sensing scenarios.
- Spatial flexibility: The position of the receivers can be optimized. Moreover, the spatial deployment of backscatter tags can be quite flexible in urban and metropolitan areas with high ambient RF APD, while in other scenarios, dedicated transmitter(s) can be strategically placed in optimal locations to balance the scalability and performance of backscatter tags.
- Technology flexibility: A variety of excitation signals can be exploited, e.g., from ambient RF sources such as TV or frequency modulation (FM) radio towers, cellular BSs, and WiFi access points, while several modulation schemes may be supported.

Due to the above, bistatic BC technology has become increasingly popular. Notice that the transmitter(s) and receiver(s) may not be independent entities in bistatic systems, but they may cooperate. For instance, the authors in [161] propose for them to share link settings and metrics to improve the BC link performance. Nevertheless, cooperation is not possible in the case of ambient BC systems.

2) USE CASES

The applications of BC technology are numerous as illustrated in Figure 14 and are rapidly increasing. The three main use cases are related to

- Invasive monitoring: BC technology may be useful in scenarios where sensors must be instrumented in an invasive manner because it can potentially eliminate the need for replacing batteries and the corresponding high cost. That is the case of applications such as i) structural monitoring, e.g., to ensure safe and reliable operation of railways, pipelines, dams, bridges, and aircraft, and ii) implantable health-care monitoring [170].

---

9. Note that the term "backscatter" may not be accurate for the bistatic architecture since the signal does not necessarily scatters back, but toward the receiver.
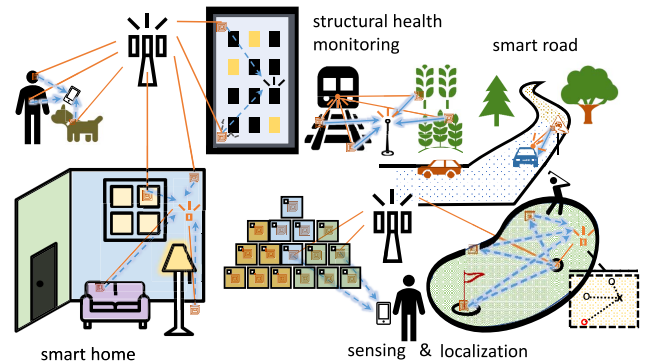

FIGURE 14. Illustration of some BC applications.

- Ubiquitous localization and sensing: The astonishing spatial diversity offered by the massive proliferation/deployment of backscatter (battery-free) devices (e.g., attached to everyday objects) can be exploited to extend networks' localization and sensing capabilities as envisioned by the 6G paradigm [171].
- Smart world: By incorporating backscatter tags into everyday objects, not only network sensing capabilities are naturally extended, but also each of these objects is digitized and becomes a source of information and/or control unit with added value.

An appropriate choice between (non-passive) radios and BC for an application depends on the operating distance, data rate requirement, and power budget [172]. There is no one-size-fits-all solution. In general, BC is more energy-efficient but may not cope with stringent requirements in terms of coverage and rate. Indeed, BLE radio is likely the best option currently to support 1 Mbps at 20-40 m, while BC fits better for applications requiring around 100 kbps at 5-10 m. However, this is also a matter of deployment topology, cost, and other KPIs such as reliability, latency, and security as required by the target application. Please refer to Section VII for further discussions on this.

3) CHALLENGES AND RESEARCH DIRECTIONS

Many state-of-the-art BC designs as those implementing LoRa and BLE backscatter in Table 5 are quite mature and deployable. However, there are still a number of challenges that require attention in the coming years to fully realize the potential of BC technology and enable its seamless integration with 5G and beyond generations, thus paving the way for the most advanced and futuristic applications. A compilation of such key challenges and associated research directions is presented below.

- Imperceptible integration to everyday objects: Fabricating BC miniaturized electronic components and integrating them into small, and diverse, form factors is challenging, and so is the imperceptible integration into everyday objects. Printed electronics may be appealing here and also the exploitation of meta-materials with favorable electrical properties (see

TABLE 5. Some representative state-of-the-art (2017-2022) BC systems.

BC System	Power ^{a}	Maximum Bit Rate		RG		Deployment	Distinctive Features
		Bit Rate	Dist.	Tx-Tg	Tg-Rx	Tx, Rx	
• LoRea [150]	70 μW	197 kbps	175 m	1 m	3.4 km	patched commodity, customized HW	Compatible with properly-tuned WiFi, Zig-Bee, and BLE radios at 2.4 GHz.
• Battery-free phone [147]	3.5 μW	N/A	N/A	15.2 m	15.2 m	customized HW	Powered by RF & light. Speech/data transmission and UE input via capacitive touch.
• LoRa backscatter [151]	9.3 μW	37.5 kbps	N/A	238 m 5 m	238 m 2.8 km	customized HW, commodity LoRa	Sensitivity of -149 dBm even under strong in(out-of-)band interference.
• FM backscatter [152]	11.1 μW	3.2 kbps	4.9 m	N/A	18.3 m	ambient FM, commodity FM	Applications: i) posters communicating to smartphones/cars, ii) smart fabric.
• Morse backscatter [153]	20 μW	250 kbps	2 m	N/A	N/A	customized HW, SDR	Morse code and on-off-keying modulation for low-complexity/power/cost.
• MOXcatter [154]	N/A	50 kbps	6 m	0.3 m	14 m	commodity WiFi	Sensing data embedded in ambient spatial-stream packets.
• BARNET [155]	3 μW	N/A	N/A	2 m	2 m	customized HW	Applications: i) identification of tagged objects, ii) 'device-free' activity recognition.
• PLoRa [156]	2.6 mW	6.25 kbps	300 m	1 m	1.1 km	commodity LoRa, SDR	Backscatter signal decoder & MAC protocol for coexisting with active LoRa nodes.
• TunnelTag [157]	57 μW	2.9 kbps	1 m	N/A	N/A	customized HW, SDR	Tunnel diode-based RF oscillator for transmission or amplifying reflection.
• RF-Mehndi [158]	N/A	N/A	N/A	0.3 m	0.3 m	customized HW	User's fingertip profiled passive RFID tag -based identifier.
• ZeroScatter [159]	850 μW	9.4 kbps	3 m	3.5 m	3.5 m	customized HW	Unmodified digital I/O pins of FPGAs and microcontrollers (MCUs) for BPSK BC.
• WiFi backscatter [160]	30.8 μW	2 Mbps	N/A	30 m	90 m	commodity WiFi	First backscatter IC communicating directly with commodity WiFi transceivers.
• BLE backscatter [161]	50 μW	N/A	N/A	1 m 2 m	42 m 37 m	commodity BLE	Closed-loop BC system exploiting frequency and spatial diversity.
• DigiScatter [162]	54.5 μW	620 kbps	7 m	7 m	7 m	customized HW	OFDMA backscatter realizing digital frequency synthesis.
• RF-powered tag [148]	840 μW	96 Mbps	N/A	N/A	N/A	customized HW	Compact tag powered by RF EH at 2.45 GHz and backscattering at 1.76 GHz.
• SyncScatter [163]	30 μW	450 kbps	32.3 m	12 m	30 m	commodity WiFi	IC achieving synchronized BC with ambient signals.
• RapidRider [164]	N/A	200 kbps	14.5 m		14 m	commodity WiFi	Exploitation of single-symbol decoding via forward and backward deinterleavers.
• IBLE [165]	N/A	8 kbps	2.5 m	1 m	18 m	commodity BLE	Reliable PHY and full BLE compatibility. IPS/GFSK modulation and FEC coding.
• Long-RG FM tag [166]	150 μW	1 kbps	19.6 m	N/A	20 m	dedicated FM carrier, SDR	Tunnel diode reflection amplifier with high gain and wide bandwidth.
• CD-ZED [51]	54 μW	N/A	N/A	7.2 km	3 m	SDR	Mobile communication based backscattering powered by indoor light.
• CAB [167]	271.2 μW	341 Mbps	0.5 m	N/A	N/A	commodity OFDM-WiFi	Content-agnostic backscatter that can demodulate both tag and ambient WiFi data.
• SubScatter [168]	613 μW	11 Mbps	2 m	N/A	20 m	customized, commodity WiFi	Subsymbol backscatter modulation for high THP.
• STScatter [169]	N/A	500 kbps	14 m	N/A	20 m	commodity WiFi, SDR	Universal space-time stream backscatter.

The two rows colored in light red correspond to mono-static BC systems, while the remaining are bistatic.

^{a} This field corresponds to the minimum power consumption (PC) of the backscattering tag.

Section IV-B) to achieve small computational materials that can communicate using backscatter.

- High-frequency operation: Most current BC systems operate at well-established microwave frequencies.

However, expanding the operation to higher frequencies (e.g., mm-wave and THz) may bring substantial benefits since more antennas per unit area can be packed, producing very directional transmit, receive,

and backscattered beams while achieving longer communication RGs. Unfortunately, ambient RF energy availability is more limited in such a high-frequency operation regime, which may prevent exploiting ambient BC. Instead, dedicated RF signals/sources are required together with novel low-complex protocols that efficiently handle the issues related to beam search, especially for flexible bistatic BC systems.[10] Notice that current high-frequency BC prototypes are monostatic, e.g., [173], [174], [175], [176], or non-flexible bistatic, e.g., [177], thus, much further research is needed to realize flexible high-frequency BC systems suited to real-life applications.

- Wide-band and frequency-agnostic designs: Wide (ultra-wide)-band BC designs (including advanced modulation schemes) are needed for supporting high data-rate applications, e.g., sensing and interaction related to augmented reality, which constitutes a challenging research direction. Also, frequency-agnostic BC systems able to operate across different protocols, locations, and applications, would be extremely appealing. In this regard, HW innovations, as well as low-power algorithms, are required to dynamically identify which frequency bands have the strongest signal.

- Security: Due to the limited power/complexity of BC systems, guaranteeing secure communications is extremely difficult. Notice that the i) accurate identification of a fake ambient BC tag, and ii) mitigation of interference generated by a tag maliciously backscattering ambient signals to a nearby reader, remain open problems in the literature. A promising direction lies in designing quantum backscattering mechanisms [178].

- Enhanced RF-EH sensitivity: RF EH, as the charging source, facilitates small form-factor and battery-free BC implementations. Unfortunately, RF-EH sensitivity is orders of magnitude worse than that of a BC receiver, which limits the connectivity RG. Improving RF-EH sensitivity, preferably up to two orders of magnitude, is a fundamental research and engineering challenge.[11] Techniques exploiting leakage power reduction, technology scaling, and sub-threshold operating using voltage scaling may be key to achieving this goal.

- BC networks: Most of BC research and prototyping is focused on the physical layer (PHY) with piecemeal evaluation. A full-layer design constitutes a challenging next step for maturing the technology and realizing scalable, integrated, and practical BC networks with the capabilities for realizing, e.g., carrier sense, network management, and polling of devices.

10. In monostatic systems, the tags can be designed to be retrodirective (i.e., capable of reflecting an incident signal toward the source direction without prior knowledge of its direction of arrival) to close a directional link with the reader, e.g., by using phase conjugate arrays, Van Atta arrays, and leaky-wave antennas [26].

11. Under free-space propagation conditions, every 6 dB improvement in sensitivity roughly doubles the operating distance [172].
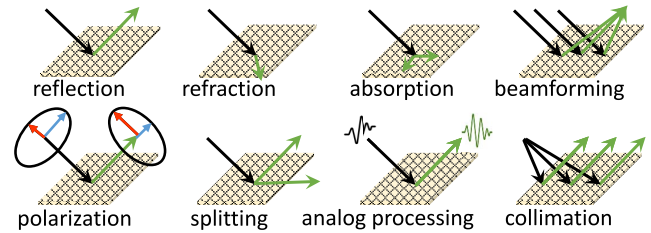

FIGURE 15. EM-based elementary metasurface functions.

B. METASURFACE-AIDED COMMUNICATIONS

Metasurfaces are surfaces composed of metamaterials with sub-wavelength thickness. The so-called metamaterials have special properties when interacting with EM radiation, and thus metasurfaces may support several functions as shown in Figure 15 [179], [180]:

- *Reflection/refraction* of the incident RF waves to a given reflecting/refracting direction;

- *Absorption*, by which the reflected/refracted signals corresponding to an incident RF wave are nulled;

- *Beamforming*, by which the incident RF waves are focused toward a given direction/location. *Collimation* is the complementary operation;

- *Polarization change* of the reflected RF waves with respect to the incident ones. For example, incident RF waves are transverse electric polarized, and reflected RF waves are transverse magnetic polarized;

- *Splitting*, by which multiple reflected or refracted RF waves are created from the incident RF waves;

- *Analog processing* includes mathematical operations at the EM level, e.g., the RF waves refracted by a metasurface can be the first-order derivative or the integral of incident RF waves.

These functions motivate the use of metasurfaces for improving wireless systems performance, e.g., for passive signal or energy relaying [181] and wireless sensing [182].

Metasurfaces are low-cost fully-passive devices with zero EC, which can be engineered to statically perform one or several of the above functions. However, more recently, the research community and industries are pursuing a more dynamic approach, where the operation of the metasurfaces is SW-controlled in real-time, thus, leading to intelligent metasurfaces. Next, we briefly discuss the main configurations of intelligent metasurfaces in the state of the art: i.e., intelligent reflective surface (IRS) and large intelligent surface (LIS), associated research challenges, and relevant research directions. A list of representative prototypes and experiments with metasurfaces in the last years is presented in Table 6 along with their main distinctive features.

1) IRS

As illustrated in Figure 16, IRS (also known as reconfigurable intelligent surface - RIS) is a metasurface composed of a large number of N subwavelength-spaced passive scattering metamaterial elements. Each scattering element can

TABLE 6. Some representative state-of-the-art (2017-2022) experiments and prototypes of reconfigurable metasurfaces.

Ref.	Elements	Controller	Control Units	Operation Freq.	Functions and Other Features
[183]	22×22	DC voltage source	PIN and varactor diodes	$11.5 - 13.5$ GHz	Dynamical beam deflection, splitting, and polarization change. $180°$ reflection phase difference.
[184]	20×20	Bias network	1 varactor diode	$3 - 4$ GHz	Discretely-tuned broadband beam reflection/ scanning.
[185]	40×40	N/A	N/A	$8.7 - 11.3$ GHz	Generating several radiation patterns. Even distribution of reflection phases from 0 to 2π.
[186]	8×8	FPGA	Binary PIN diodes	$8 - 10$ GHz	Space-time harmonic beam steering and beam shaping.
[187]	14×16	MCU	Binary relay switches	60 GHz	Beam reflection/steering. A two-stage beam-searching algorithm is implemented.
[188]	75	N/A	N/A	8.6, 10.1 GHz	Beam splitting. Two beams with different circular polarizations at each frequency.
[189]	16×16	Control board	Binary PIN diodes	5.8 GHz	Beam steering/focusing for enhancing PTE of RF-WET systems.
[190]	16×16	FPGA	PIN diodes (2-bit setups)	2.3, 28.5 GHz	Active wide-angle beam-scanning/transmission. 21.7/19.1 dBi gain achieved at 2.3/28.5 GHz.
[191]	12×12	MCU	PIN diodes	$6.1 - 6.6$ GHz	Polarization selection/reconfiguration and frequency reconfigurable polarization conversion.
[192]	24×24	FPGA	4 varactor diodes per element	5.67 to 6.15 GHz	i) spin-modulating circularly polarized waves, ii) dual-beam scanning, iii) dual-polarized shared-aperture antenna.
[193]	55×20 (0.25 m^2)	FPGA	Binary varactor diodes	$5.7 - 5.9$ GHz	Self-adaptive beamforming based on real-time feedback.
[194]	5×20	SDR reader	RFID tag	866, 870 MHz	Backscattering (virtual IRS composed of an array of commodity RF-powered RFID tags).
[195]	16×16	FPGA	N/A	$5.79 - 5.81$ GHz	Beamforming exploiting low-complexity neural beam alignment.
[196]	8×16	MCU with shift registers	Binary PIN diodes	5.3 GHz	Channel randomization to generate temporal variation and provide UE-independent entropy source.
[197]	10×10 (0.07 m^2)	N/A	Varactor diodes	$2.9 - 3.35$ GHz	Beam reflection with stable phase and amplitude responses at oblique incidences.
[198]	16×10 0.1 cm^2	MCU with shift registers	Binary PIN diodes	5.8 GHz	Beam reflection/scanning. $60°$ angle resolution for maintaining a single main lobe.
[199]	14×14	N/A	4 varactor diodes per element	$5.15 - 5.75$ GHz	Beam steering/reflection with $40°$ angular resolution.
[200]	20×20	FPGA	Binary PIN diodes	5.4 GHz	Beam steering/reflection assisted by computer vision methods leveraging a camera attached to the IRS.
[201]	10×10	MCU	3-bit RF switches & delay lines	5.3 GHz	Beam steering/reflection with high spatial resolution.

be controlled in an SW-defined manner by the so-called IRS controller, which can be embedded or separated, to properly tune the EM properties of the output signals given a set of incident RF signals impinging the scattering elements. Therefore, IRS is a promising technology to dynamically control the radio propagation environment and improve the wireless system performance in a cost/energy-effective manner. Specifically, an IRS can reflect the incident signals so they are added constructively in the desired direction to increase the signal power (so-called passive beamforming) or destructively for mitigating undesired interference, either for high data rate, dependable, secure, NLOS, or wide-coverage communications, and RF-WET (including joint communication and RF-WET).

Conventional IRSs, so-called passive IRSs, e.g., [183], [184], [185], [186], [187], [190], [191], [192], [193], [195], [196], [197], [198], [199], [200], [201], are composed of fully passive reflecting elements and the only active EC comes from the controlling HW (IRS controller and related active circuitry). This allows low-power/cost energy-efficient implementations that do not incorporate additional RF radiation into the environment, which is undoubtedly appealing in terms of sustainability.

Unfortunately, the effective coverage of passive IRS may be seriously limited since the reflected signal suffers high *product-distance path loss*, which critically constrains the signal power at the receiver(s). This can be addressed by either equipping the IRS with an increasingly massive
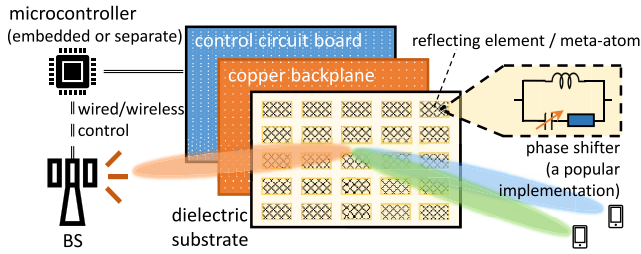
FIGURE 16. IRS architecture and assisted communication.

FIGURE 17. LIS implemented with multiple microstrips. Each microstrip is connected to a single RF chain.

number of passive elements and/or conveniently deploying the IRS closer to the transmitter/receiver to reduce the cascaded channel path loss. Unfortunately, these options are not always viable due to, e.g., deployment difficulty (limited space and unavailable site) and enormous training overhead for the CSI acquisition required to optimize the operation of the passive IRS elements in real time. This has motivated the introduction of active IRS [202], [203].

In active IRS, the reflecting elements incorporate reflection-type amplifiers, which are implemented without high-cost and power-hungry RF chains and that can simultaneously alter the signal's phase and amplitude to enhance the signal power at the receiver. This requires modestly higher EC and HW cost compared to conventional IRS.

Interestingly, the achievable signal-to-noise ratio scales as $\mathcal{O}(N)$ and $\mathcal{O}(N^2)$ when using active and conventional IRS, respectively [203]. The limited gain when using active IRSs is due to the noise power amplification introduced by the active design. Nevertheless, an active IRS implementation still provides superior rate performance (subject to a given total power budget constraint) in comparison to a conventional IRS when N is moderate thanks to the additional signal power amplification gain [203]. Table 7 illustrates the main features of conventional and active IRSs with competing relaying technologies. All in all, there is no *one-size-fits-all* solution (as usually in engineering), and the use of one technology over another depends on the specific network's characteristics, constraints, and performance requirements.

2) LIS

LIS is a metasurface equipped with RF circuits and signal processing units and composed of a virtually infinite number of elements to form a spatially continuous transceiver aperture. This structure can be used to transmit/receive communication signals across the entire surface by leveraging the hologram principle [204].[12]

For instance, an LIS may be comprised of multiple waveguides, e.g., microstrip. Each waveguide may embed a large set of radiating metamaterial elements whose frequency response can be externally and individually adjusted by

varying the local electrical properties. Each microstrip is fed by one RF chain, and the input signal is radiated by all the elements located on the microstrip, as shown in Figure 17 [131]. The figure also illustrates an example of transmitting a signal using a single microstrip with multiple elements.

3) CHALLENGES AND RESEARCH DIRECTIONS

Some key challenges and associated research directions for maturing metasurface-assisted communication technology and making it a reality in future sustainable networks are briefly discussed in the following:

- Low-cost/energy control: The use of tunable reflecting elements with discrete amplitude/phase shift levels favors cost/energy-effective implementations. However, this may significantly limit the beamforming/reflection capabilities of the metasurface, especially in the case of limited form-factor implementations. Therefore, further studies are required to unveil such underlying trade-offs. Moreover, further research is required on the metasurface controller circuitry, which interfaces with all the tunable reflecting elements and constitutes the only EC source in passive IRS implementations. Specifically, low-cost/energy-effective control mechanisms must be developed to connect and communicate with massive tunable elements, and thus agilely and jointly control their EM behaviors on demand.

- CSI acquisition: A tunable passive beamforming/reflection typically requires accurate CSI, which is challenging to acquire in practice. The two main approaches proposed in the literature for passive IRS (without any active RF chain or reflection-type amplifier), but that still require further research in terms of performance trade-offs, are [181]: i) estimate the concatenated (TX-IRS and IRS-RX) channel with some known IRS reflection patterns, or ii) exploit feedback from the TX/RX pertaining to their received signals that are reflected by the IRS (no explicit channel estimation). Meanwhile, CSI acquisition for active IRS is generally more challenging because explicit CSI of the separate TX-IRS and IRS-RX links, instead of the cascade CSI, is needed due to the amplification noise. The research addressing this issue is still incipient, e.g., cf. [203]. Notably, approaches exploiting limited CSI as those relying, e.g., on positioning information [206], may be appealing.

---

12. Alternatively, a LIS can be implemented as an IRS-aided antenna by deploying an external communication and antenna module to wirelessly feed the IRS for active communication [205], e.g., [188], [189]. Another LIS implementation can be based on discrete photonic antenna arrays integrating active optical-electrical detectors, converters, and modulators for transmission, reception, and conversion of optical or RF signals [204].
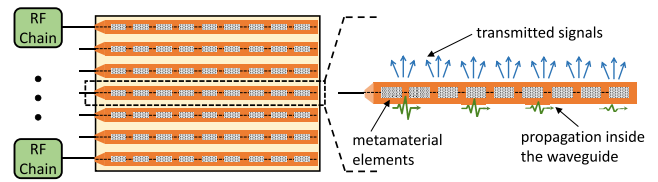
TABLE 7. Main features of IRS and competing relaying technologies.

Technology	Duplex Mode	Power Consumption	Amplification Noise	EE	HW Cost	Network Spectral Efficiency[b]	Operation Mechanism
Passive IRS	FD	Very Low	No	High (large N)	Very Low	Low	Reflect
Active IRS	FD	Low	Yes	High (small N)	Low	Moderate	Reflect & Amplify
AF Relay	HD/FD	Moderate	Yes	Low	High	Moderate-High	Amplify & Forward
DF Relay	HD/FD	High	No	Low	High	High	Decode & Forward

[b] Based on the ability to simultaneously serve multiple users and to act in other roles (e.g., as sources).
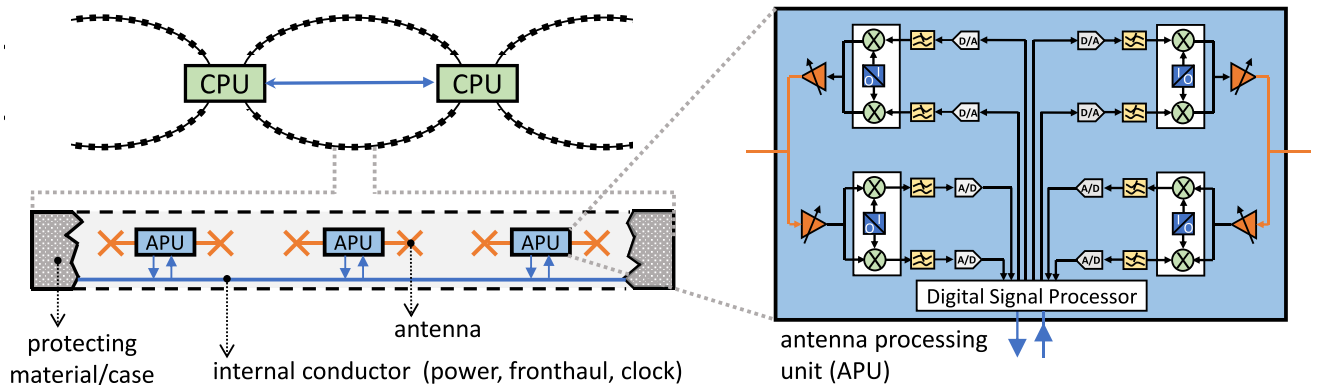

FIGURE 18. Structure of a radio stripe system. Each radio stripe can be connected to one or multiple central processing units (CPUs).

- Advanced metasurface implementations: The most commonly investigated/prototyped metasurface applications are those related to passive beamforming/reflection. However, as illustrated in Figure 15, there are several other metamaterial functions, which are so far only incorporated into non-flexible/configurable implementations. In the next years, SW-controlled engineering solutions must be developed exploiting these metamaterial functions for real-life applications. For this, accurate physics- and EM-compliant models are needed.

- Data-driven optimization: Accurate modeling/analysis and optimum design/implementation/deployment of IRSs/LISs are challenging due to the inherent complexity of such systems. This calls for efficient data-driven methods, e.g., based on deep learning (DL), reinforcement learning (RL), and transfer learning [207]. Moreover, low-complexity/energy ML mechanisms, the so-called TinyML (see Section V-C), seem appealing for incorporation into the metasurface controller circuitry for more cost/energy-effective implementations.

C. RADIO STRIPES

Massive MIMO technology addresses challenging 5G performance requirements, especially in terms of network coverage, capacity, and THP. However, the required high manufacturing and operating expenses, as well as the increased EC, make the development and deployment of truly large-scale antenna arrays extremely difficult. This motivates the research on more affordable and low-power MIMO architectures that can scale with the number of antennas more sustainably. The metasurface-aided communication architectures overviewed in Section IV-B constitute one active research front in this direction. Indeed, IRSs can be deployed to efficiently assist already-deployed MIMO networks (thus, avoiding the need to install new active HW); while LIS are basically low-power/cost massive MIMO that can replace traditional MIMO HW in many scenarios. However, the deployment of IRS/LIS alone may not fulfill all the use cases' constraints/requirements, at least in the near future, for which alternative low-power/cost MIMO architectures are still needed. That is the case of the radio stripes system discussed in the following.

Radio stripe systems enable cost/energy-effective distributed massive MIMO [208]. In such a system, with architecture depicted in Figure 18, antenna elements and circuit-mounted chips (including power amplifiers, phase shifters, filters, modulators, and A/D and D/A converters) are serially located inside the protective casing of a cable or a stripe, which also provides synchronization, data transfer, and power supply via a shared bus. Unlike traditional massive MIMO BSs, radio stripes [22], [208]: i) allow imperceptible/flexible installation in existing construction elements and alleviate the deployment permissions problem, ii) support native system resiliency to failures, and iii) facilitate low EC due to the inherent low-complex and distributed architecture functionality. Also, additional HW, including temperature/vibration sensors and microphones/speakers, can
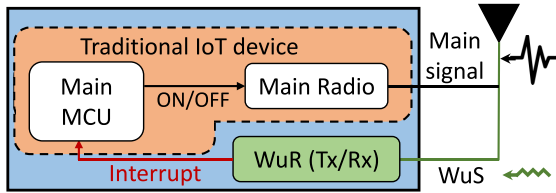
FIGURE 19. IoT device with WuR architecture for the same-band operation of WuR and main radio. Note that different architectures may or may not include all the components, such as a WuR with Tx capabilities.

FIGURE 20. Typical triggering using WuR scheme (top) and the WuS packet structure (bottom).

be deployed in the radio stripes to provide additional features/services, e.g., fire/burglar alarms, earthquake warnings, indoor positioning, and climate monitoring/control.

Due to the above, radio stripes technology is attractive for supporting energy-efficient (and sustainable) networks. Applications are numerous, e.g., to: i) facilitate high spatial multiplexing and low EC in indoor communications [209], ii) increase the coverage and end-to-end PTE of RF-WET [22], [210], and iii) support ultra-reliable low-latency communications in industrial IoT networks [211]. In general, although the last few years have witnessed significant advances in this technology (e.g., see [212], [213], [214] and references therein), more advanced/efficient resource allocation schemes, circuit implementations and prototypes, and distributed processing architectures to avoid costly signaling between the antenna elements, still pose open research challenges.

D. WuR

Duty cycling, consisting of turning off the radio component of a device, pausing its MCU, and using a timer to activate the device periodically, extends the lifetime of devices with limited battery capacity (BCAP). However, the device with the radio off cannot exchange data, which may result in excessive communication delays [215]. This motivates the adoption of WuR, which allows activating the devices' main radio on-demand when there is data to communicate. Since its EC is several orders of magnitude lower than that of the traditional main radio, e.g., about 1000 times lower [216], the WuR can be kept always on, in contrast to the duty cycling operation. Furthermore, although the EC for operating a clock is relatively low, it is still non-zero, while this EC may be eliminated entirely by employing WuRs.

Figure 19 illustrates an exemplary architecture of a WuR-capable device, which includes a low-power radio that receives and detects the WuR signal to activate the main radio for communication. The specific components and their inclusion in different architectures can vary based on the design requirements and specific use cases. A typical WuR setting is illustrated in Figure 20. The main radio of the device remains deactivated (OFF) until it is required for communication, or until a special packet known as the Wake-up signal (WuS) is received by the WuR, which generates an interrupt signal to the main MCU to switch it ON. Subsequently, the main radio can exchange data packets with the other node in
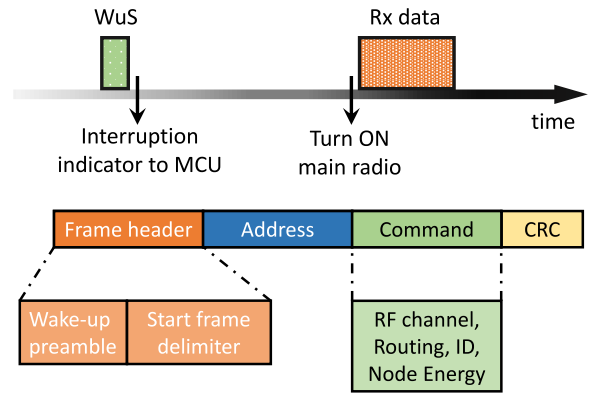
a conventional manner [217]. Figure 20 also illustrates the composition of a WuS packet. Note that the frame header consists of the wake-up preamble and start frame delimiter, a standard byte pattern agreed upon between the transmitter and the receiver. The preamble is used for synchronization whereas the start frame delimiter indicates the start of the frame that contains relevant information. Meanwhile, the address field contains the destination node identifier and the payload contains application data, commands, or extra instructions specified by the UE or application. Finally, there is an error detection frame, using cyclic redundancy check, aka CRC, to check data integrity [217].

The main benefits of using WuR are:
- EE i) avoiding unnecessary idle listening, ii) avoiding energy wasting related to start-up/power-down, and iii) combining with uplink reference signaling such that even high-speed UEs can reside in the sleep state for long periods, while not increasing the handover failure rate [218], [219].
- Short buffering delay [219].
- Synchronization assistance [218], [219].
- Low wake-up reconfiguration need for many realistic unsaturated traffic scenarios [218], [219].
- Less signaling overhead [219].

1) WUS STANDARDIZATION

Figure 21 depicts a timeline overview of the key aspects and features associated with WuS in each 3rd Generation Partnership Project (3GPP) Release. Within the 3GPP standardization process, WuS was initially introduced in Release 15 [220], [221] as a paging signal sent over the PHY downlink shared channel that "wakes up" a UE from an idle state so that it can prepare to receive data. Concerns towards enabling energy-efficient techniques resulted in a feature for 5G called low-power WuS in Release 16 [222], [223]. Next, we primarily focus on the standardization of WuR within 3GPP.

Release 16 and Release 17 incorporate improved cross-slot scheduling [226], [227]. Specifically, the network can
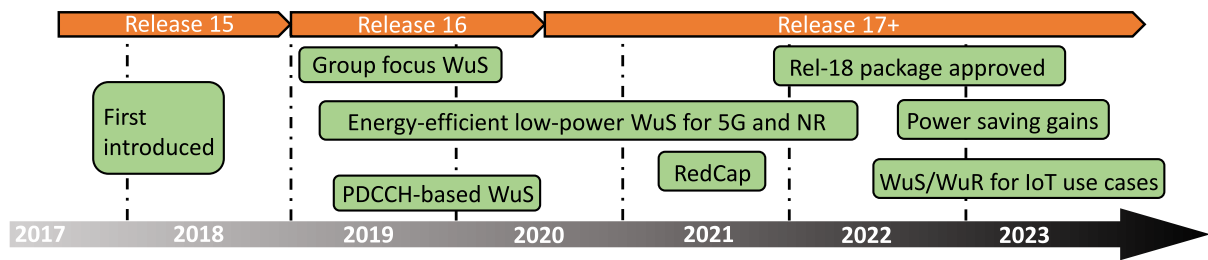
FIGURE 21. 3GPP Release evolution related to WuS.

inform a device when there is a guaranteed minimum time interval between downlink packet transmissions, leading to a significant reduction of unnecessary RF operations. In addition, group focus WuS is defined, allowing the network to wake up a configurable group of UEs (instead of all UEs) by configuring the WuR of each UE to listen for a specific pattern or sequence in the WuS that is unique to their assigned group, thereby reducing EC [226]. Another feature is that the network transmits a PHY downlink control channel-based WuS before active on-duration within discontinuous reception cycles. This allows for UEs to avoid PHY downlink control channel monitoring during on-durations within which the network is not transmitting any data to the UE, and may enable between 15% and 30% energy savings. Alternatively, WuS may allow for shorter discontinuous reception cycles, i.e., faster response time, with similar EC [226].

Release 17 specified power-saving enhancements for reduced-capability devices, referred to as RedCap, while Release 18 promises to provide further EE gains [228]. The current set of work and study items for 3GPP Release 18 radio access network (RAN) features is split into four RAN working groups. Specifically, the first group (RAN1), is responsible for the PHY specification including PHY channels and modulation, PHY multiplexing and channel coding, PHY procedures and measurements, as well as PHY-related UE capabilities. Within the 5G-Advanced Release 18 scope, low-power WuS and WuR constitute a new study item in RAN1, which aims at studying power-saving schemes that do not require existing signals to be used as WuS and includes i) evaluation methodologies for low-power WuS/WuR for power-sensitive small form-factor devices; ii) evaluation of low-power wake-up receiver architectures and WuS designs to support wake-up receivers; iii) PHY procedures and higher layer protocols to support WuS; and iv) UE EE gains compared to the existing Rel-15/16/17 UE power saving mechanisms and their coverage availability, and latency impact [229]. Finally, WUR standardization is also being carried out by IEEE, e.g., as a part of IEEE 802.11ba standard [230].

2) STATE-OF-THE-ART

WuR constitutes a promising technique for achieving IoT devices' lifespan superior to 10 years [216], [225], [231], and thus has recently received significant attention from the scientific community. For example, the authors in [231]

presented a WuR-enabled BLE device targeting an IoT scenario where such battery-powered massive IoT devices do not support direct 3GPP connections. Real-life results show that the system meets the over-10-year lifetime target while satisfying the latency requirements for 5G IoT devices. Similarly, the authors in [232] shed some light on BLE-compatible sensor devices enriched with a WUR, and their results demonstrate EE gains in low-latency applications (under 2s). In [233], the authors optimize operational parameters, determined by BS at the beginning of the session, to save energy. Meanwhile, an accurate traffic forecasting model is proposed in [218] to optimize the wake-up parameters, achieving up to 35% EC reduction. Additionally, [234] proposes a super-regenerative WuR solution to improve EE in human-body communication. Herein, WuR operates at a very low data rate, e.g., 1.25 kbps, for higher sensitivity while consuming $\sim$40 μW. Likewise, the authors in [235] show the EE from using WuR in wireless body area network applications with event-driven traffic and propose a WuR capable of receiving small control commands besides WuS. All in all, the different WuR proposals vary depending on the particular protocols used, the type of circuitry, and the application. Here, we summarize the main classification of WuR research in Table 8.

3) CHALLENGES AND TRADE-OFFS

Using a WuR brings two main impacts on the devices' performance. On the one side, there is the problem of miss-detection of the WuS, which happens when the device does not receive the page occasion scheduled within the WuS for information exchange. As a result, the device misses the chance to wake up and has to wait until the next page occasion, increasing the latency of packets/information. On the other hand, due to the inherent simplicity of the WuS, the problem of false alarms needs to be addressed. In a false alarm event, the device/WuR receives a page occasion needlessly when no information is intended to be transmitted/received. Therefore, special attention to these problems is needed when using WuR [218].

Some key challenges and research directions are:

- WuS may complicate radio resource management and device scheduling in the network due to sleep patterns, reducing potential EE gains. In this context, employing advanced ML-based scheduling algorithms considering the sleep patterns of devices may be appealing [218].

TABLE 8. WuR taxonomy.

	Classification		Ref.
WuR-based MAC protocols	**Only MAC** (access decision is done just in MAC layer)	**Cross layer** (interaction between layers to optimize MAC layer performance)	[215]
Wake-up circuit	**Duty cycled** ("duty-cycled" WuR, which is activated periodically to sense the radio channel and send WuS)	**Continues cycled** (WuR's circuit is always ON)	[224]
Communication initiator	**Transmitter initiated** (the transmitter first sends a WuS to activate the desired node and then sends data)	**Receiver initiated** (receiver sends a WuS asking for data, **bidirectional** WuS is also supported)	[215]
Addressing scheme	**Broadcast** (the WuS is received by all the neighboring nodes of the sender)	**Identity-based** (only the desired next hop node is wakening up)	[215]
Frequency band	**In-band with the main radio** (main radio has predictable dormant periods, cheaper as there is no need for a separate antenna)	**Out-of-band** (unpredictable dormant periods, it reduces interference, increases signal capacity, but may increase cost and complexity)	[217]
Power source	**Active** (battery-powered, immediate responsiveness but consumes extra power, ideal for real-time monitoring, rapid event detection, and instant data transmission)	**Fully-passive** (no internal power supply is needed, while **hybrid** combines EH with a small backup battery for responsiveness, limited operational RG, and minimal EC)	[225]

- A design where the WuR utilizes a different frequency band than the main radio increases the complexity and cost of the devices. Therefore, in-band operation and RF integrated circuit-embedded WuR implementation is desired. However, this approach complicates resource management and reduces the available spectral resources for transferring application data.
- Beamformed WuS at mm-waves and mobility management is still an open challenge since beam sweeping for WuS is required to reach a desired device. The network should be able to optimize the number of beams in a single WuS burst utilized for waking up the device.
- Applying WuR brings trade-offs between EE and other KPIs, depending on the application scenario, like latency, reliability, and robustness [219]. Therefore, more research should be directed in this direction, especially when dealing with massive low-power IoT scenarios.

V. ENERGY-EFFICIENT ML

Native ML support is essential for dealing with the increasing complexity and automation of 6G networks while improving their performance [236], [237], [238], [239]. ML techniques can help address issues such as increasing traffic demands, real-time QoS requirements, and resource allocation. However, these benefits usually come at high computational and memory requirements. Therefore, energy-efficient ML algorithms are paramount for network sustainability [240].

Several ML features affect EC. In general, larger models, i.e., with more parameters, require more energy to be trained and exploited. Also, training time (TRNT) and inference time (INFT) directly affect the EC at the training and inference phases, respectively. Therefore, the trade-offs between power cost and performance reward require special attention.

As shown in Figure 22, ML approaches are broadly classified as supervised learning (SL), unsupervised learning (UL),
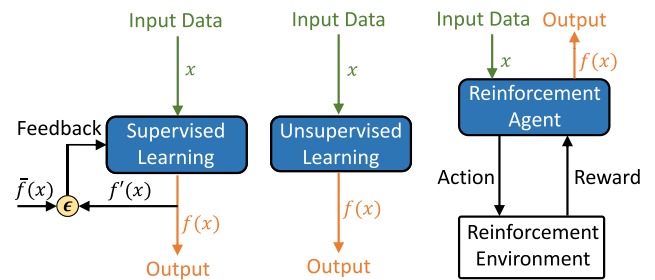
FIGURE 22. High-level representation of SL, UL, and RL working principles.

and RL. In SL, the model is trained on a labeled dataset that contains the correct outcome for the corresponding input. SL can be used for tasks such as predicting energy usage based on historical data or identifying energy-efficient products based on specific features [241]. Meanwhile, UL does not require labeled data and thus can be used, e.g., to identify patterns in energy usage data, detect anomalies that may indicate energy waste, or reduce the dimensionality of the data [242]. In RL, effective solutions are learned over time given constraints imposed by the inputs and without attempting to find hidden categories or structures. RL is suitable for solving problems with multiple optimal solutions [243], such as optimizing energy usage in buildings or predicting optimal times to charge electric vehicles. Notice that RL algorithms can be computationally intensive, but they have the potential to improve EE over time by learning from experiences and making adjustments accordingly [244].

ML in wireless communications can be approached from three distinct perspectives: i) the network side, ii) the edge, and iii) the device side, as illustrated in Figure 23. Note that network-based computing treats mobile devices as data collectors sending data to cloud servers for processing. The drawback of this scenario is the introduced overhead and potentially severe latency [245]. Also, ML algorithms can be complex, data-hungry, and computationally costly on the

FIGURE 23. ML at the network, at the edge, and at the device side.

network side. Contrary, ML at the device side must be simpler (i.e., of lower power/cost), while a balance/trade-off betwee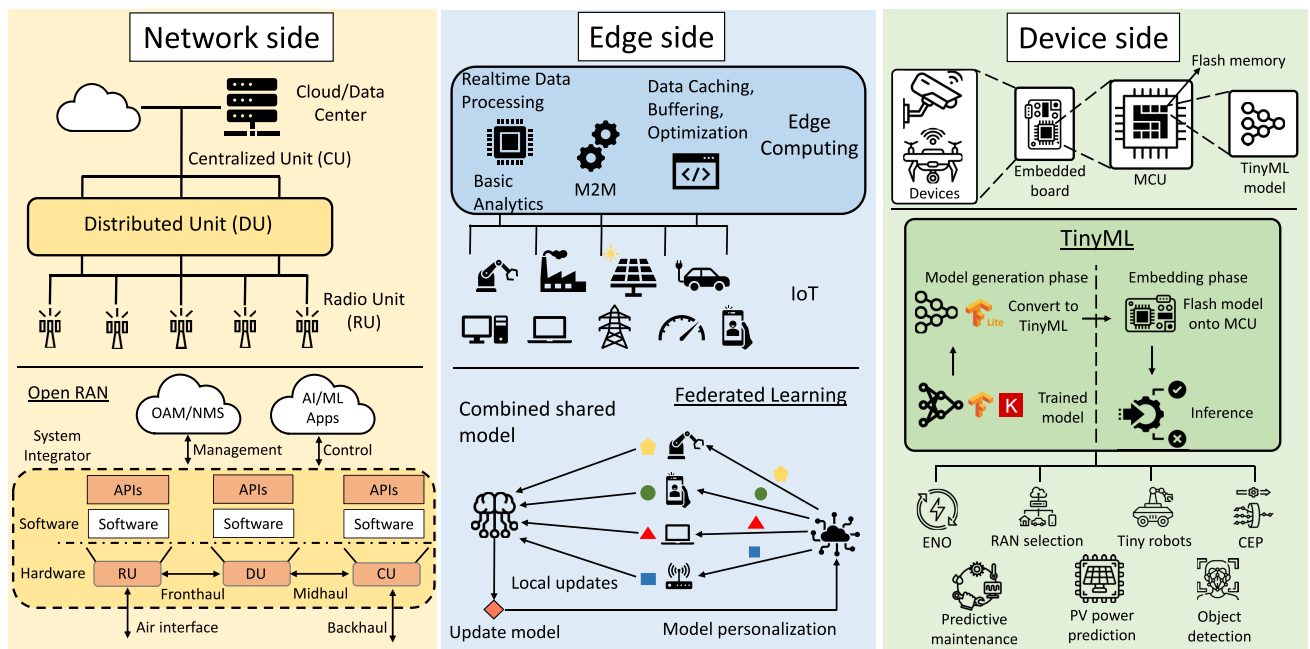n these two extremes is achievable by implementing ML at the edge. Indeed, ML at the edge takes advantage of local processing and data storage capabilities while communicating with the network and devices. Specifically, pre-trained models are offloaded from the cloud to individual devices, such that they can make inferences locally. This allows for more efficient and effective use of ML in wireless communications. However, it can only support tasks that require lightweight computations [245].

Next, we discuss each of the aforementioned approaches, Indeed, our purpose here is to explore the application of ML algorithms in specific local aspects of IoT connectivity, such as computation task optimization and offloading decision-making, which provide EE enhancements.

A. ML AT THE NETWORK SIDE
Currently, RANs account for 73% of the total EC in modern cellular networks [246]. Therefore, intelligent resource management is imperative to maximize EE, and thus minimize EC. Specifically, the current trend is to replace rule-based heuristics with optimal parameters configured through the knowledge acquired by data-driven approaches [246]. In this regard, ML models play a crucial role in enabling intelligent networks to characterize their environment, predict system changes in real-time, and react accordingly [240].

1) ML ALGORITHMS
Choosing the most appropriate ML algorithm is critical for solving EE problems on the network side. The algorithm choice should be based on the specific problem, network architecture, and available data. A comprehensive/holistic

approach that considers all aspects of the network, including HW, SW, and communication protocols, is needed to design energy-efficient solutions [247], [248], [249]. Moreover, the specific EE problem, the architecture and the characteristics of the network, the available data, and the strengths and weaknesses of each ML algorithm must be considered [247], [248], [249]. For instance, decision trees are suitable for routing optimization, while support vector machines are effective for network anomaly detection and prediction. Genetic and clustering algorithms are well-suited for optimizing the placement of network and edge resources [250], [251], [252], while clustering algorithms are also appealing for grouping network devices, e.g., based on their EC patterns [252].

DL is one of the most popular ML algorithms at the network side [242]. Specifically, deep SL (DSL) and deep RL (DRL) are commonly used on the network side due to the availability of labeled data and the effectiveness and flexibility of the models. Both can be used to optimize EC by adjusting the network resources based on the network traffic load. By combining feature extraction with prediction, DL models classify, predict, and make accurate decisions more effectively than traditional ML algorithms [242]. This is because DL architectures such as multi-layer perceptrons, convolutional neural networks (NN) [253], and transformers [254] can estimate complex mappings between input and labels in the training data [253], all while efficiently utilizing HW-based accelerators such as graphical processing units (GPU) for training and inference [255]. Other benefits include the distribution of processing, avoiding redundant capacity in hotspots, and the efficient marshaling of big data, generated in-network or at user devices.

Notably, DL has its own drawbacks, e.g., i) DL requires large amounts of training data, whose curation and labeling may be costly and face privacy concerns; ii) DL algorithms are largely black boxes with low interpretability and explainability; and iii) DL may require dedicated ML accelerators for efficient operation [246], [256]. All in all, the DL benefits may be outweighed by the costs in many cases, calling for careful and informed designs/implementations.

2) USE CASES AND APPLICATIONS

In 6G networks, ubiquitous intelligence is key for providing efficient and personalized services. However, this poses a challenge in terms of data management and EC [257]. Indeed, EC can increase considerably unless energy-efficient approaches are used [258]. Thus, there is a pressing need for developing lightweight, flexible, and adaptive solutions that effectively tame the dynamics of the environment and minimize EC in practical IoT networks [246]. This has motivated the adoption of DRL for solving a variety of wireless communication problems [259], making it a valuable tool for data-driven optimization in new-generation systems [246].

ML forecasting network environment can help DRL algorithms converge faster to proper operational policies, resulting in faster adaptation to changing conditions and greater EE in network operations [246]. Indeed, they can trigger well-informed decisions to address integration issues with IoT devices. For instance, ML models can analyze historical data from IoT devices and predict their traffic patterns and energy requirements. This information can then be used to optimize resource allocation, dynamically adjust power management settings, and improve EE. In [260], ML is used to estimate empirical path loss and shadowing. These estimations guide transmission power adjustments, leading to significant energy savings of up to 43% compared to conventional protocols. Moreover, DRL is exploited in [261] to design an adaptive LoRa strategy for improved reliability in industrial applications, while [262] uses a combination of decentralized and centralized RL for allocating spreading factor and transmission power to the devices, showing significant improvement in both network level good-put and EC, especially for large and highly congested networks. Additionally, ML models can help identify IoT behavior patterns, such as peak usage times and idle periods of devices, enabling the implementation of energy-saving strategies like sleep modes and adaptive transmission power control. Ultimately, ML-driven insights can contribute to achieving a coordinated balance between network performance and EE in IoT environments.

Notably, novel DRL schemes have been proposed to manage advanced sleep modes in BSs. Specifically, in [263], the sleeping level length is set by the BS in a sequential manner. When the cell becomes idle, the BS departs from the deepest level of sleep and gradually switches to higher activity levels. At each stage, the BS decides the number of slots for the current sleep mode. Most relevant components must thus only be active (consuming energy) when handling actual data [246].

In this scheme, the BS is the agent, and the network is the environment, including the traffic load, available energy, and the state of other network devices. The BS takes actions (i.e., sets the sleeping level length) based on its observations of the environment (i.e., traffic load and energy budget), and the RL algorithm provides feedback (i.e., rewards or penalties) based on the EC and network performance. The RL algorithm aims to maximize the reward by finding the optimal sleeping level and duration for the BS while maintaining the required network performance.

On the other hand, demand forecasting, i.e., predicting how much time and resources will be spent on applications, is a key problem in data center management. Notice that good forecasting techniques can lead to minimizing EC by scheduling jobs efficiently [264]. Network-level data usually exhibit significant spatiotemporal variations, which can be utilized for network diagnosis and management, UE mobility analysis, and public transportation planning [245]. In this context, DL has the potential to improve EE in a variety of settings by improving demand forecasting, optimizing resource allocation, and identifying patterns and trends in data that can be used to reduce EC.

Establishing a data collection path model is another solution for minimizing EC. Specifically, data collection can avoid visiting needless nodes and collecting unreliable/correlated data, resulting in outperforming traditional data collection methods in both energy and delay. In that direction, proactive caching can also contribute to EC minimization using forecasted lookup patterns to jointly optimize computation offloading policies and caching decisions [265]. These approaches may minimize radio interface usage, which is a significant EC source in IoT devices.

3) OPEN-RAN (O-RAN) AND VIRTUALIZATION AS KEY TECHNOLOGICAL ENABLERS

Open interfaces are appealing for the operators to swiftly introduce novel services and tailor the network to their own requirements. Motivated by this, RAN is evolving towards the O-RAN concept, which focuses on openness and intelligence [266]. O-RAN brings new business opportunities and encourages local 5G innovations. O-RAN aims at decoupling the RAN components from their underlying SW and HW components, enabling operators to cover more users in a cost-effective, secure, and energy-efficient manner [267].

In O-RAN, the RAN is disaggregated into three main building blocks: i) radio unit (RU), ii) distributed unit (DU), and iii) centralized unit (CU) [266] as shown in Figure 23. Also, the O-RAN ALLIANCE has defined different interfaces within the RAN including those for i) fronthaul between RU and DU and ii) midhaul between DU and CU. Another feature of O-RAN is the RAN intelligent controller, which adds programmability to the RAN and the ability to introduce new services and features [266], [267], [268].

The open interfaces/protocols of O-RAN enable a seamless integration of ML models to optimize radio resource management, interference mitigation, and network planning.

In addition, O-RAN supports real-time data processing and analysis, which is essential for ML algorithms that require fast and accurate decision-making. This enables intelligent and adaptive management of the RAN, resulting in better network performance, improved UE experience, and lower operating costs. These all may be crucial to extend the battery lifetime (BL) of IoT devices [268], [269], and support energy-sustainable IoT ecosystems in general.

Notice that network functions virtualization (NFV) enables network functions to be implemented as SW applications running on virtualized infrastructure. This improves EE in several ways, such as reducing the required number of physical network devices, dynamically scaling network functions based on demand to avoid over-provisioning, and optimizing energy usage across the network through centralized management and orchestration [270]. Furthermore, NFV allows for the implementation of EE features like sleep modes and power management in network devices, leading to even further EC reduction while maintaining network performance. In general, O-RAN and NFV are two complementary technologies that can work together to improve the EE and flexibility of wireless networks.

O-RAN initiative is still in its infancy, with a lot of work in progress. Hence, it is important that future research activities specifically focus on practical and real-world trials with respect to the virtualized RAN concepts [271], which heavily rely on self-organization and other ML-based approaches.

B. ML AT THE EDGE

ML at the edge (popularly known as edge ML [236] or edge intelligence [272]) refers to the training and use of ML models across the computing continuum: on UEs, edge nodes, and cloud servers, rather than only in cloud-based centralized setups. his is a rapidly evolving field with prominent innovation opportunities and impact [273], [274].

This topic is often studied from two viewpoints. First, *ML on edge* refers to adapting ML methods for the distributed edge environment, while *ML for edge* is the use of ML methods for the benefit of the edge environment [236], [275]. As an example of ML on edge, distributed learning and inference allow the efficient distribution of ML computations across the computing continuum. This allows optimizing the overall processing time, resulting in lower EC of in/near-sensor devices [240]. Moreover, ML on edge can reduce latency, allow localized filtering of unwanted data, and increase system uptime as data is locally processed. On the other hand, edge ML can introduce various benefits, including the application of a predictive approach in troubleshooting. With edge ML, real-time data can be analyzed by ML models to proactively identify potential issues or anomalies before they escalate into critical problems. This can provide EE gains and minimize the downtime of IoT systems. Moreover, operational efficiency and EE can be promoted in IoT environments by reducing unnecessary offloads and tasks and optimizing resource utilization [276], [277].

Shortcomings that need to be addressed include: i) complexity due to coordination issues related to IoT constraints like processing power, memory, and delay in real-time applications; ii) heterogeneity, opportunism, and geographical distribution of computing resources; iii) fluctuating or intermittent connectivity; iv) security and privacy; v) standardization-related concerns like interoperability of IoT with ML integration; and vi) accuracy and latency issues in real-time applications [238], [272].

1) ML ALGORITHMS

ML can be used for multiple purposes in the context of edge computing. For example, edge computing can filter data, with only relevant data getting transmitted between the user devices, the edge nodes, and the cloud. This results in substantial savings in terms of bandwidth and cost of data transmission [23]. Moreover, advanced ML techniques can be utilized, for example, to optimize computation tasks, make offloading decisions on a wireless device, and identify the best scheduling solutions for working and sleeping time, thus lowering EC and enhancing EE [257].

There are several ML approaches that can be applied at the edge depending on the application requirements and constraints. These include distributed learning, referring to a family of methods such as federated learning (FL) for distributing the learning and training data/process across various nodes [278], and transfer learning, a method for transferring knowledge between ML models in different domains [23]. For example, FL and other variants of distributed learning have emerged as possible solutions for solving complex operational decisions at the edge side [257] (Figure 23). Although undoubtedly beneficial, FL faces some challenges [257]:

- *Expensive communication and synchronization.* FL communication overhead may be limited by reducing the number of i) communication rounds and ii) gradients in each communication round.
- *Security/privacy/robustness issues.* FL must: i) provide protection against malicious attacks, ii) support privacy, iii) tolerate heterogeneous HW, and iv) support robust aggregation algorithms.
- *Model size (MDS).* A large FL MDS might be unsuitable for an IoT device, especially given stringent real-time constraints. Thus, efficient training and inference are necessary for massive and heterogeneous networks.

2) USE CASES AND APPLICATIONS

IoT places significant demands on three main areas: transmission, storage, and computation. Indeed, IoT devices generate a large volume of data and have limited storage and computational capabilities, thus edge-based storage and computing are appealing. Such an edge integration effectively accelerates data uploads/computation, reduces response and device active time, and improves EE [279].

Edge computing is a rapidly evolving field, and the trend is towards moving cloud functions to the network edges [280].

Edge computing benefits the advancement and implementation of 5G and beyond networks with applications such as augmented/virtual reality [281], low-complexity IoT [279], Internet of Vehicles [282], and video stream analysis [283]. Delay-sensitive augmented/virtual reality applications can be migrated to edge servers to guarantee a high-quality user experience with a timely response. For low-complexity IoT tasks, edge servers can reduce the HW complexity and increase device lifespan by processing tasks that were originally done locally. Additionally, edge cloud networks close to vehicles can improve transportation safety, reduce traffic congestion, and provide value-added services. By utilizing the capabilities of MEC networks, video playback can be expedited, and user experience can be enhanced. Moreover, smart meters can be used to detect electricity consumption and gather data into a central controller to facilitate real-time power control, leading to improved systematic EE.

The spatial and temporal correlation of IoT traffic data is exploited in dual prediction and data compression techniques. Specifically, dual prediction techniques use the correlation between the current and previous data to predict future data values. On the other hand, data compression techniques use the correlation between the different sensor nodes to compress the data before transmission. Both techniques reduce the amount of data that needs to be transmitted and hence the EC and bandwidth requirements [10].

At the edge, communication-efficient FL and efficient training aims at reducing the required communication overhead of the training process while maintaining or improving the model's performance. Federated parallelization is an efficient training FL technique for parallelizing the training process across multiple devices and thus accelerating the training. Efficient training can also be conducted via distillation which consists of training a small model that mimics the behavior of the original model; therefore, significantly reducing the amount of required training data [257].

3) MEC - A KEY TECHNOLOGICAL ENABLER

MEC, or multi-access edge computing, is an architecture for mobile networks. MEC was introduced to address the latency issue during mobile cloud computing offloading, pushing computing and storage resources to the edge with the aim of bringing those resources as well as applications and services near the end-users. MEC is characterized by two key features, namely, low latency and high workload capacity, stemming from proximity to users and their devices [284]. However, MEC has unique design considerations, such as complex wireless environments and MEC servers' inherently limited computational capabilities [10].

Computation offloading is a significant MEC feature as it may prolong the lifespan of IoT devices by delegating computation tasks to edge devices as long as the communication overhead remains reasonable [252]. Specifically, the amount of energy that a mobile node can save by offloading an application depends on the number of computation instructions and communication data. If the computation instructions
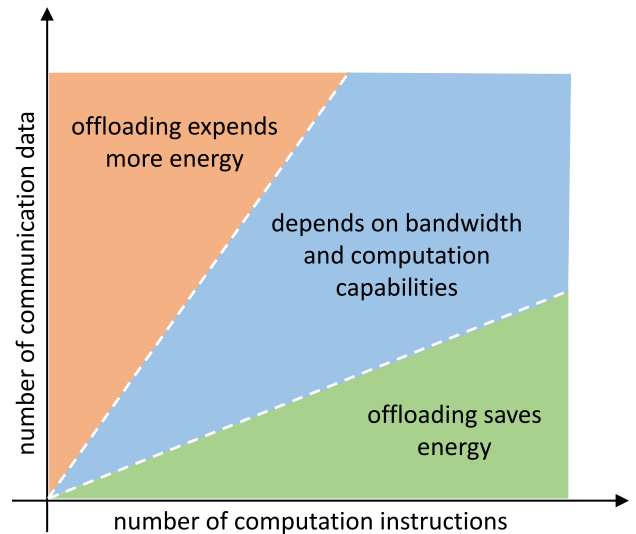

FIGURE 24. EE when using computation offloading [285].

are much larger than the communication data, it is more energy-efficient to offload the computation-intensive application to the server. However, if communication is expensive, it is better to carry out the application at the mobile node itself. The condition also depends on the bandwidth available for communication, where a large bandwidth can save communication time and improve EE [252]. Figure 24 illustrates when offloading can save energy [285].

There are various energy-efficient offloading methods: i) computation-based methods, which involve partitioning the offloading application program and offloading computation-intensive parts; ii) communication-based methods, which involve reducing the amount of communication by aggregating or compressing data [285]; and iii) hybrid/joint optimizations methods, which involve collaboratively executing tasks on the mobile node and the cloud to minimize EC while ensuring the total execution of tasks [286]. The study in [286] demonstrated that computation offloading can reduce EC and increase battery life up to 50% for practical applications.

Notice that establishing a MEC infrastructure, especially server locations, is the first step toward constructing a MEC system. For this, the system planners and administrators must consider the deployment cost, demand for computation, and availability of renewable energy sources for EP in the case of green MEC [250], [251], [284], [287]. Indeed, green MEC is an emerging technology that combines the benefits of MEC with energy-efficient computing and EP from renewable sources to create a more sustainable and eco-friendly approach to mobile computing [287]. The need for green MEC arises from the growing awareness of the environmental impact of mobile computing and the increasing demand for sustainable solutions. As mobile networks and devices become more ubiquitous and essential to our daily lives, their EC and carbon footprint also increase. Green MEC may reduce this impact by optimizing the use

of resources and energy-efficient computing techniques. One method for designing green MEC systems is dynamic right-sizing, where the EC of a MEC server is dependent on its utilization, and a server should consume energy proportional to its workload [288].[13] To achieve energy-efficient servers, the processing speeds of lightly-loaded edge servers can be reduced. Another method is geographical load balancing (GLB) [287], which utilizes spatial variations in workload patterns, temperatures, and electricity costs to direct the flow of workload between different data centers. These methods can help to reduce the EC of MEC systems and provide a carbon-neutral energy supply, leading to more sustainable and environmentally friendly network operations [287].

C. ML AT THE DEVICE: TINYML

TinyML focuses on computationally efficient and HW-constrained ML algorithms for low-power IoT devices [289]. TinyML-equipped IoT devices can energy-efficiently execute tasks and make autonomous decisions without continuously relying on cloud/network services. This reduces network traffic and latency in decision-making and increases privacy. Due to this and its on-device data processing capability, TinyML may assist with intermittent connectivity/computing.

1) USE CASES AND APPLICATIONS

In the following, several use cases and applications motivating the deployment of ML on the device side are discussed.

Energy Management is a promising solution to enable i) energy-neutral operation (ENO) in always-on IoT devices utilizing EH [290] and ii) efficient allocation of PV energy from a single-panel or off-grid system to multiple tasks [291]. In this regard, TinyML models can forecast future EH values to help devise a proactive ENO strategy for IoT devices. Besides, when the number of EH samples becomes insufficient to elaborate a forecast on the incoming ambient energy, energy management strategies can still benefit from TinyML by using the current battery state and previous EH measurements [290]. Again, PV power prediction is necessary for proper energy management/distribution among tasks. The study in [291] employs TinyML to perform PV power prediction and suggests that it can also serve as an indicator for measuring the effects of aging on the power-generating capacity of solar panels.

In addition, TinyML may jointly ensure EE and the QoS of the data transmitted by an IoT device with multiple radio access technologies (RATs) by enabling a proper RAT selection. This can be implemented by considering situational characteristics, such as the available energy on the device and the size/urgency level of data to be transmitted [292]. Also, TinyML may support complex event processing (CEP),

which aims to identify complex event patterns in real-time data streams from multiple sources using predefined logic rules.[14] Note that large IoT deployments pose a challenge for CEP in performing sequence matching over raw data due to unexpected events or outliers not covered by predefined rules, motivating the use of ML-based CEP. Nowadays, the data privacy and latency issues faced by centralized IoT are empowering the idea of performing ML-based on-device CEP in IoT. Interestingly, the authors in [293] designed a framework that puts together TinyML and CEP for machine safety monitoring in a distributed IoT network.

Object detection running in EH IoT devices constitutes another application opportunity for TinyML. Indeed, data privacy and latency requirements in various object detection applications, such as intelligent video surveillance and number-plate recognition, are encouraging the idea of on-device processing in IoT, thus calling for efficient TinyML implementations. Due to the dynamic nature of the EH environment, on-device processing must be aware of the available energy on the device, and also of the time, energy that must be collected, and the deadline to complete a task. Interestingly, TinyML is used in [25] to detect a person over a battery-less IoT device. This study also suggests the use of TinyML for applications where battery-less devices require a long lifetime and are difficult to reach.

TinyML allows performing predictive maintenance procedures, e.g., for detecting and preemptively solving the impending failures that a system might face, over an MCU-based sensor device rather than on the cloud. For instance, authors in [297] implemented TinyML to perform anomalous behavior detection tasks over the sound recordings of the ToyADMOS dataset [298], where spectrograms (images generated from audio) are used as input. It is shown that optimizing the audio sampling rate used to form a spectrogram can lead to a decrement in the spectrogram dimensions, which further leads to a reduced INFT, required memory, and EC. In addition, [299] introduces TinyML for detecting various kinds of faults in a solenoid valve, which is an electro-mechanical actuator. Here, the transient response of the drive current of the solenoid valve is used as input to the TinyML model. This encapsulates information about the solenoid valve's electro-mechanical action, from which predictions about the present working condition can be made.

Finally, TinyML is appealing for realizing tiny robots. Note that legged robots use imitation learning [300] and RL to learn their walking gaits. However, such procedures are not functional for tiny robots, which are low-cost resource-constrained robots that can potentially be used in search & rescue operations, military reconnaissance, space robotics, and routine equipment monitoring [301]. To address this, authors in [302] proposed exploiting TinyML techniques, such as graph freezing and float16 quantization, to shrink the

---

13. In a typical MEC system, the EC of a server is often fixed, regardless of its utilization or workload. This means that even if the server is underutilized, it still consumes a significant amount of energy, leading to wastage and higher energy costs. Refer to the EPC metric in Section VI-D1.

14. For instance, CEP can be used in a manufacturing plant to detect a sequence of events that occur before a machine breaks down by monitoring, e.g., changes in pressure, temperature, and vibration, and thus triggering an alert for maintenance.
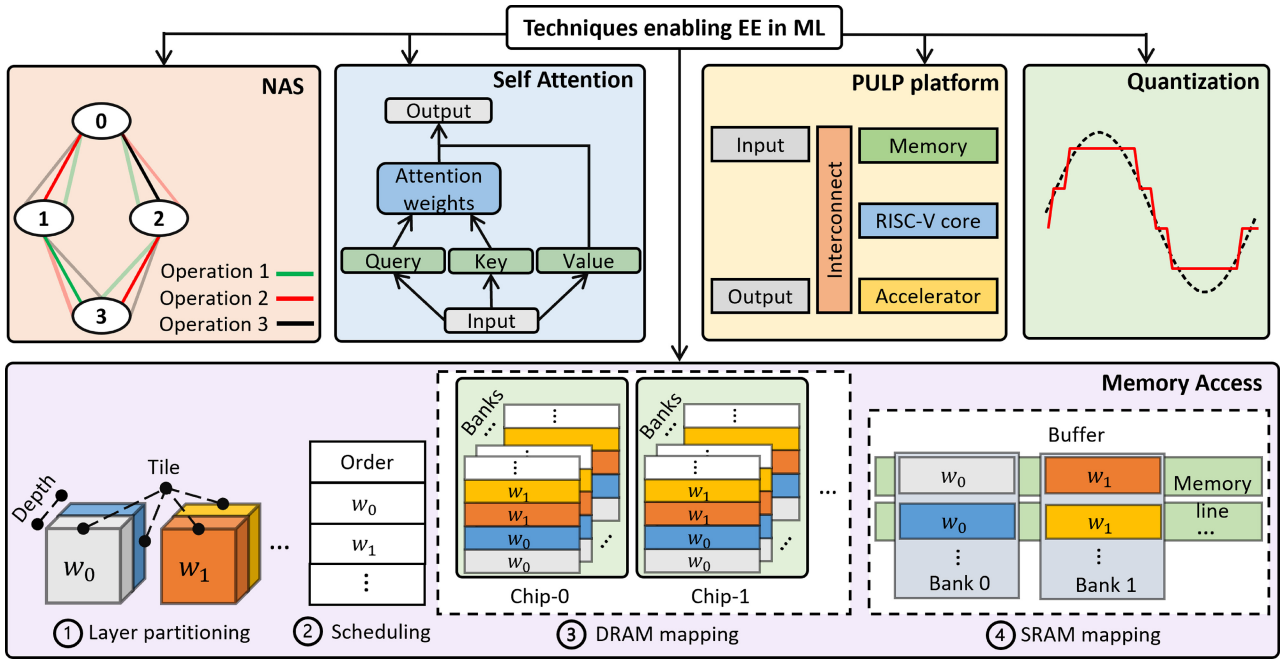
FIGURE 25. Techniques enabling EE in ML [294], [295], [296].

trained walking gaits of the NN MDS by 8×. Graph freezing made all NN variables constant and float16 quantization changed the floating point weights from 32-bit to 16-bit.

2) ENABLING TECHNIQUES

Next, we discuss the techniques to support energy-efficient ML models. They are illustrated in Figure 25, while representative state-of-the-art works using them are listed in Table 9 together with key features.

- *Neural architecture search (NAS)* aims to find the best NN architecture that can fit in the available MCU resources. For instance, the authors in [303] proposed MCUNet, a framework to optimally handle/utilize the MCU resources. MCUNet consists of TinyNAS, a low-complexity NAS, and TinyEngine, an inference library that reduces the required runtime memory. The latter utilizes a memory scheduling method to reduce the SRAM required by an ML model during its inference, thus reducing the EC per inference. Meanwhile, a drop in the required SRAM frees up the MCU memory, thus allowing TinyNAS to select a larger ML model to achieve higher accuracy. As shown in Table 9, MCUNet achieves higher accuracy compared to MobileNetV2 [304] and ResNet-18 [305], while utilizing relatively less SRAM and Flash. All in all, MCUNet may provide energy-cum-memory efficient NN architectures that can be implemented on an MCU.
- *Parallel ultra-low power (PULP)* is an IoT processor architecture providing SW-level acceleration for TinyML. PULP adopts both data and thread-level parallelism to deliver a steady performance irrespective of the operating voltages and mW-level computations/inference. As shown in Table 9, PULP

uses 3.5× less energy than ARM Cortex-M4 processor and also exhibits a significantly lower INFT. For instance, the authors in [306] adopted PULP-based MCU to facilitate the parallel run in non-neural ML algorithms, while a fast NN (FANN)-on-MCU framework supporting both ARM Cortex-M series MCUs and PULP-based MCUs is used in [307]. The latter framework can be used with both fixed-point and floating-point NN models.
- *Model compression* allows the creation of models that are compact enough to fit on limited-capability devices and efficient enough to run with limited power. Specifically, model compression is a collection of techniques aimed at reducing the ML MDS, thereby enhancing its efficiency in terms of memory usage and computational requirements [308]. Several model compression techniques are commonly employed in TinyML: i) quantization, which reduces the precision of the numbers used in the model, significantly decreasing MDS and computational requirements [309]. For example, coarse quantization drastically reduces the precision of an NN model's parameters to less than 8 bits, making the model suitable for an MCU. Moreover, the quantized NN model enjoys less EC and faster computation than its full-precision counterpart. However, these advantages come at the cost of lower NN accuracy [310]; ii) pruning, which eliminates parts of the NN that contribute little to the output, such as weights that are close to zero [311]; iii) knowledge distillation, which trains a smaller model (the student) to mimic the behavior of a larger model (the teacher). The smaller model is then used in place of the larger model, resulting in significant computational savings [312]; and iv) weight sharing,

TABLE 9. Representative state-of-the-art works on energy-efficient TinyML (2020-2022).

Ref.	Application	EE strategy	Additional information
[290]	ENO	tinyMAN	EC = 27.75 μJ per inference, memory footprint (SRAM + Flash) = 91 KB
[314]	Face recognition	Quantization	MNIST dataset. Chainer ML framework (prediction latency = 0.005 ms, training time = 3.54 s) and Keras/Tensorflow (prediction latency = 5.3 ms, training time = 243 s)
[294], [315]	Smart health-care, image classification, object detection	ROMANet	Compared to the baseline scheme SmartShuttle [316], the number of DRAM accesses is 12%, 36% and 45% smaller for the AlexNet, VGG-16, and MobileNet, respectively. Moreover, the DRAM access energy is 12%, 36%, and 46% smaller, while the numbers are 12%, 34%, and 45% smaller in terms of DRAM operations
[303]	Image classification, object detection, keyword spotting in an audio	MCUNet, linear quantization	ImageNet dataset. MCUNet (4-bit Quantization): SRAM = 0.49 MB, Flash usage = 1.9 MB, accuracy = 70.7%. MobileNetV2 (8-bit Quantization): SRAM = 1.7 MB, Flash usage = 2.5 MB, accuracy = 69.8%, ResNet-18 (8-bit Quantization): SRAM = 0.9 MB, Flash usage = 11.2 MB, accuracy = 69.8%
[307]	Hand gesture recognition, fall detection, human activity classification	PULP	Applied to a NN with 103800 multiply-add operations. FANN-on-MCU with ARM Cortex-M4 processor: EC = 183 μJ per inference, INFT = 17.6 ms. FANN-on-MCU with PULP processor: EC = 50 μJ per inference, INFT $\leq$ 1 ms
[317]	Object recognition	AttendNets	ImageNet$_{50}$ dataset. MobileNetV2: accuracy = 68.7%, number of parameters = 2290 K, number of multiply-add operations = 299.7 M. AttendNets: accuracy = 71.7%, number of parameters = 782 K, number of multiply-add operations = 191.3 M
[299]	Fault detection	PreMa	PC = 140 mW per inference, INFT = 76.2 ms

which involves using the same weights for multiple neurons, reducing the total number of weights that need to be stored [313]. By applying these and other techniques, it is possible to create small, efficient models that can run on low-power devices, thereby expanding the ML applications.

- *Self-attention* allows each input in a sequence to weigh the importance of all other inputs, thereby enabling each element to "attend to" or "focus on" all other elements in the sequence [254]. This enables high levels of parallelization and increased efficiency, and it is a significant departure from traditional recurrent NNs or long short-term memory networks, which typically focus on preceding words in a sequence. Note that the self-attention mechanism is a core component of the Transformer architecture [254], which is revolutionizing natural language processing. In the context of TinyML, the benefits of self-attention and Transformer architectures can be manifold. Firstly, the ability to process sequence data in a non-temporal manner allows for more efficient computation, which is crucial for low-power edge devices [318]. Secondly, the parallelizability of Transformer architectures makes them well-suited to the resource-constrained environments typical of TinyML applications. By processing all elements of a sequence simultaneously rather than sequentially, Transformers can deliver faster INFT, which is crucial for real-time applications on edge devices [319]. However, the original Transformer models are often too large and computationally intensive for TinyML applications. Therefore, further research on model compression techniques and efficient Transformer variants, such as TinyBERT [320] and DistilBERT [321] is needed. For example, AttendNets, a TinyML model, combines self-attention with a machine-driven design exploration, resulting in a compact deep NN with low-precision parameters. As shown in Table 9, AttendNets achieves better performance in terms of accuracy, EC, and memory consumption than MobileNet-V2 for pictorial recognition with ImageNet-50 dataset [322].

- *Memory access techniques* aim to reduce the average DRAM access energy in an inference. Note that HW accelerators, consisting of DRAM (off-chip part), and SRAM and compute engine (on-chip parts), are required to implement complex NNs, such as CNN, on a device. The portion of the NN layer that an accelerator can process at a single time instance depends on the data storage capacity of SRAM. Also, the same input data could be used in multiple NN operations, which leads to multiple DRAM accesses for the same data. Notably, the energy required for DRAM access is relatively higher than that for a NN operation, which means that a significant amount of energy can be saved by exploiting memory access techniques. Indeed, authors in [315] recommend using ROMANet [294], a memory access technique to cut down the average energy-per-DRAM-access, on-chip buffer access energy, and the number of DRAM accesses. Specifically, ROMANet partitions an NN layer and then schedules the processing of the portions to minimize the number of DRAM accesses. Based on the data storage capacity of SRAM, data is partitioned into various blocks. Then, ROMANet maps them to the available DRAM and SRAM resources to minimize the row buffer conflicts and maximize the bank-level parallelism, respectively. The advantages of using ROMANet with various NN architectures, such

as AlexNet [323], VGG-16 [324], and MobileNet [325], are clearly visible in Table 9.

The reader interested in TinyML can refer to the python package hls4ml [326], which allows designing ML algorithms for low-power FPGA or ASIC devices. Through hls4ml, it is possible to optimize the HW implementation of TinyML models by leveraging the parallelism and pipelining capabilities of FPGA. This can lead to more efficient use of HW resources and reduced EC.

3) FUTURE DIRECTIONS

Bio-inspired optimization [327] may brush aside unnecessary ML computations, providing potential benefits for TinyML. Furthermore, alternate NN models, such as spiking NN [328] and analog NN [329], should be investigated for TinyML design. Also, new computing models for TinyML systems are needed, while the existing computing models, such as in-memory computing, require further research in the context of TinyML systems. Note that MobileNet is a baseline deep NN model for edge computing, however, no baseline TinyML model is currently available for the end devices. Indeed, a baseline TinyML model serving as a reference point for future TinyML models is appealing. Finally, a ferroelectric RAM [330] based MCU is a viable option for tinyML applications where some data must be retained on the device along with the tinyML code. Ferroelectric RAM offers non-volatile storage and allows flexible memory partitioning for tinyML code and data storage. Also, it allows writing a single data word to memory in ~ 100 ns, whereas flash memory demands several ms to perform the same operation. However, these non-volatile RAMs are currently bulky and have a significantly limited number of write cycles, so further research and technological advances are still needed to make them really appealing.

VI. ENERGY-RELATED PERFORMANCE METRICS

The potential of a technique/technology, especially when compared to a competitor, can only be assessed via relevant metrics quantifying performance. Herein, we discuss several performance metrics related to energy that serve this purpose for the discussed IoT techniques and technologies at a component/device and/or system/network level. The key features of the discussed performance metrics in terms of related energy processes, application level (at the component/device or system/network level), and relevance/applications are summarized in Table 10. Meanwhile, Table 11 lists relevant performance targets (i.e., KPIs) for the current and next generation of wireless systems. The numeric values are indicative and were extracted from the vast literature consulted for this work, together with data extrapolation and trend analysis in some cases.

A. ENERGY CONVERSION AND TRANSFER METRICS

Every energy conversion/transfer process introduces losses and other potential non-linearities affecting device/network EE. Some related metrics are discussed below.

1) COMPONENT EE (CEE)

CEE refers to the ratio of the output to the total energy at the input of a certain electronic component. CEE considers losses due to heat, friction, and other inefficiencies, and describes the effectiveness of electronic components such as power supplies, power amplifiers, motors, filter circuits, phase shifters, etc. The ideal CEE value is 100% or 1.

2) EH INPUT/OUTPUT RELATIONSHIP

The main parameters impacting the performance of EH circuits are sensitivity, saturation, and CE. Sensitivity refers to the minimum magnitude or change in the input signal required for the EH circuit to produce a usable electric signal. Saturation, on the other hand, occurs when the circuit reaches maximum output and thus the harvested power is independent of any input power increase. Both sensitivity and saturation metrics correspond to absolute values of power or energy, while the EH CE is defined as the ratio of the output to the total input power. Notably, CE is a non-linear function of the input power, and EH datasheets commonly include input-output power transfer curves under different settings and emphasize the maximum achievable CE (hereinafter referred to as max CE). Notice that in WET systems, one can exploit the fact that CE depends on the operating conditions, such as the operating frequency, incident energy, and distance to the source, to drive the EH circuits to their maximum CE.

3) APD OF EH TRANSDUCERS

This metric refers to the ratio of the EH transducer's peak output power to its size under specific operating conditions. It is given in W per unit of area/volume and characterizes the miniaturization level achievable by an EH transducer.

4) PTE

This metric, given as a percentage or a dimensionless quantity, refers to the ratio of the power captured by EH receiver(s) to the power consumed by the corresponding transmitter(s) and captures the joint energy CE of the transmitter(s), the medium, and the EH circuit(s). Several factors can affect PTE, including the distance between the ET and EH receiver and the specific WET technology.

5) EH COVERAGE

This comprises a set of metrics characterizing the EH effectiveness in a given IoT network. For example,

- Average harvested energy (AHE) characterizes the average energy harvested by (either a specific or random) device in the network in a given time period. It can be given in J, or W if time is normalized.
- Energy outage probability (EOP) denotes the probability that the energy harvested by a device in a certain period falls below an energy threshold ξ. Such a threshold may indicate the energy required for the device operation or the execution of a relevant task. EOP is the complement of the energy coverage probability.

TABLE 10. Performance metrics and related key features.

Metric	Energy Process			Level		Relevance/Applications
	EP	ET	EE	Comp.	Syst.	
Conversion & Transfer						
CEE			X	X		EE assessment of power supplies/amplifiers, motors, filters, phase shifters
sensitivity^c	X	X		X		KPI of EH circuits specifying minimum input power for operation
saturation	X	X		X		KPI of EH circuits specifying maximum input power for operation
max CE	X	X	X	X		KPI of EH circuits specifying the maximum attainable CE
APD	X	X	X	X		KPI of EH transducers useful for determining miniaturization levels
PTE		X	X		X	EE assessment of a WET technology/system
AHE	X	X			X	Assessment of the EH capabilities of device(s) in a network/deployment
EOP	X	X			X	Operational availability of battery-less EH device(s) in a network/deployment
MDHE	X	X			X	Fine-grained theoretical performance analysis of battery-less EH networks
Storage & Consumption						
BCAP	X			X		Assessment of the devices' lifetime in a network/deployment
ED	X			X		Form-factor assessment for given devices' energy requirements
lifetime	X	X	X	X	X	Operational assessment and deployment/maintenance planning
EC			X	X	X	Determination of the energy demands. Critical for energy-limited systems
FLOPS			X	X	X	EC KPI for digital signal processors
MIPS			X	X	X	EC KPI for computer systems
NHE	X	X	X	X		Assessment of energy availability after losses from EH-related protocols
RPS			X	X	X	Assessment/comparison of techniques in terms of triggered EC reduction
PKD			X	X	X	Assessment of maximum energy demands. Critical for energy-limited systems
GECS	X		X	X	X	Assessment of the penetration of green energy
LCOE	X		X		X	Cost assessment for energy generation. Critical for feasibility studies
GRS	X	X			X	Reliability/stability assessment of energy generation/transfer systems
EER						
-data transfer			X		X	Typical KPI for wireless communication networks with/without QoS constraints
-RG,CA,CV			X	X	X	Energy-efficient coverage assessment of a wireless communication technology
-bandwidth			X	X		EE assessment of the utilization of the bandwidth by a technology/device
-AroI (-BoI)			X	X		Energy-efficient coverage (bandwidth utilization) assessment of an IRS
-MLA			X	X	X	EE assessment of the accuracy of an ML model
-LA			X	X	X	EE assessment of the accuracy of a positioning/localization system
-WUA			X	X	X	EE assessment of a WuR system responsiveness
Others						
EPC,EPI, IPR,LDR			X	X	X	Evaluation of system's EE according to the load. Useful for proper operation planning and resource allocation/scheduling
WET exposure level		X			X	Assessment of the environmental disturbances of deploying a WET technology
WuT			X	X		Assessment of a WuR responsiveness and corresponding EC
INFT			X	X	X	Assessment of the INFT (being proportional to EC) of an ML model
MDS,SRAM			X	X		KPI for the deployability of an ML model in a device/platform

^c Please note that sensitivity is not solely related to EH but also applies to radios in general, including WuRs. Specifically in WuR, sensitivity is taken into account in WUA, since the sensitivity of the WuR directly contributes to achieving higher WUA values.

- Meta distribution of the harvested energy (MDHE) is the distribution of the EOP conditioned on certain time/spatial-varying network parameters, thus, it is a finer-grained performance metric compared to the typical EOP. For instance, within the stochastic geometry framework, the authors in [21], [331] express MDHE as the distribution of the EOP conditioned on the locations of the RF ETs. In this context, MDHE can be interpreted as the fraction of IoT devices in each realization of the point process (i.e., each possible network realization) that have harvested energy above ξ with probability at least ε.

Notice that EOP and MDHE are relevant metrics for networks with battery-less IoT devices. Meanwhile, AHE can suffice in scenarios where devices are equipped with

TABLE 11. Some performance metrics and associated target values for current and next generation of wireless systems.

KPI	Technology	Current Values	Target Values, 2030	Enabling techniques
max CE, APD	PV	20%, 175 mW/cm^2	50%	Multijunction cells/CPV
	TEG	10%, 20 mW/cm^2	30%, 70 mW/cm^2	Highly efficient thermoelectric materials
	MFC	NE, 5 mW/cm^3	NE, 500 mW/m^3	Electrode engineering
	Piezoelectric	NE, < 1 mW/cm^3	NE, 5 W/cm^3	Structural optimization
	Triboelectric	NE, 50 mW/cm^2	NE, 300 mW/cm^2	Structural optimization
	Electrostatic	NE, 20 mW/cm^3	NE, 200 mW/cm^3	Structural optimization
	Vibration EM	NE, < 1 mW/cm^3	NE, 3 mW/cm^3	Structural optimization
	Wind	$35 - 50\%$, NE	60%, NE	Blade geometry optimization
	RF-EH	$30 - 70\%$, NE	90%, NE	Adaptive receiver design
PTE	RF-WET	$< 50\%$	70%	Joint end-to-end efficiency optimization
	Inductive WET	$80 - 95\%$	99%	Joint end-to-end efficiency optimization
	Capacity WET	$70 - 90\%$	99%	Joint end-to-end efficiency optimization
	Laser WET	$< 10\%$	30%	Joint end-to-end efficiency optimization
	Acoustic WET	$< 5\%$	50%	Joint end-to-end efficiency optimization
AHE, EOP	Ambient RF-EH	$0.01 - 1$ μW, 10^{-1}	≥ 0.1 μW, 10^{-3}	Dynamic (online) hybrid combing
	RF-WET	≥ 1 μW, 10^{-3}	≥ 50 μW, 10^{-5}	Low-power massive MIMO and IRS
ED, BL	Lithium-ion	$100 - 260$ Wh/kg, $300 - 500$ cycles	500 Wh/kg, $-$	High-capacity electrode materials
	Supercapacitor	10 Wh/kg, $10^5 - 10^6$ cycles	300 Wh/kg, $-$	Novel carbon-based composite materials
	Solid-state battery	300 Wh/kg, 5000 cycles	500 Wh/kg, 30$\times$ current value	High-capacity electrode materials
EC	BC	$10 - 1000$ μW	≤ 10 μW	Meta-materials and printed electronics
	Passive IRS	$0.1 - 0.5$ mW/element	~ 10 μW/element	Low-resolution reflecting elements & novel meta-materials
	Active IRS	≥ 0.5 mW/element	~ 0.1 mW/element	Advanced negative resistance components, e.g., tunnel diodes
	WuR	$1 - 50$ μW	< 1 μW	Optimization of WuS detection Edge traffic prediction for smart wake up
	TinyML	$0.1 - 50$ μJ/inference	$0.05 - 5$ μJ/inference	PULP-based MCU & coarse quantization
THP-EER	BC	$0.1 - 1$ nJ/bit	1 pJ/bit	Passive frequency tuning and low-power amplifiers using, e.g., tunnel diodes
	LIS	$0.01 - 1$ μJ/bit	< 10 nJ/bit	Effective deployment & low-power HW
RG-EER	BC	100 μW/m	10 μW/m	Passive frequency tuning and low-power amplifiers using, e.g., tunnel diodes
	Radio stripes	$10 - 100$ μW/m/element	$1 - 50$ μW/element	Effective/dynamic deployment
MLA-EER	ML at network	$90 - 99\%$	$> 99.99\%$	Specialized HW & high-quality datasets
	ML at edge	$70 - 99\%$	$> 99.9\%$	Specialized infrastructure & ensembling
	TinyML	$65 - 98\%$	$> 99\%$	Advanced model search space and NAS
INFT, MDS	ML at network	$10 - 30$ ms, < 250 MB	< 0.1 ms, < 1 MB	Model compression & optimization
	ML at edge	$10 - 100$ ms, < 15 MB	< 1 ms, < 1 MB	Knowledge distillation
	TinyML	$10 - 1000$ ms, < 2 MB	< 10 ms, < 50 KB	Coarse quantization and NAS
WuT	WuR	$10 - 200$ ms	$\sim 1 - 24$ ms	Fusion of WuR and main radio Shorter transmission time intervals

batteries, especially when AHE is measured per device. All in all, these metrics measure the extent to which a region or network has access to EH services and/or can reliably maintain an energy supply over time. Therefore, they can be used to assess the level of energy availability/ubiquity.

B. ENERGY STORAGE AND CONSUMPTION METRICS
1) BCAP AND ENERGY DENSITY (ED)

BCAP, given in ampere-hour (Ah), refers to the amount of potential energy (usually chemical energy) that batteries

can store. Over time, batteries' electrodes degrade due to continuous chemical reactions causing capacity fades and unexpected voltage drops. Notice that even when batteries remain disconnected from an external circuit, the internal chemical reactions will cause a self-discharge thus reducing the amount of energy stored over time.

The ED of a battery measures the amount of power a battery can store per unit of volume (Wh/m^3) or weight (Wh/kg). High ED batteries can provide longer battery life while maintaining a compact and lightweight form factor. The factors that mainly impact batteries' ED are their chemistry and internal cell design.

2) LIFETIME

In general terms, the lifetime of a device/component is a measure of its durability and refers to the period of time over which the device/component is expected to remain functional and perform its intended tasks without requiring significant repairs or maintenance. When the device/component is a battery, we refer to BL, which is influenced by many factors such as the chemistry and design of the battery, the conditions under which it is used and stored, the level of usage and discharge, and the charging and maintenance practices over time. The BL may be given by the number of charging cycles the battery can endure before its capacity decreases significantly, or by the time the battery can hold a charge before it needs to be recharged.

Finally, there is also the so-called network lifetime, which refers to the time period a network functions as intended. This metric is obviously related to the lifetime of the devices, but the relationship depends on the specific network's applications and performance requirements.

3) EC

The EC of a component/device, specified in J, depends on the specific operation modes, i.e., active, idle, and sleep states, and the time spent in each. As briefly discussed in Section IV-D, duty cycling takes care of properly scheduling these states to reduce EC subject to QoS requirements. Notably, tasks such as computation and communication, mostly executed in active modes, consume different amounts of energy. For instance, transmitting typically consumes more energy than receiving data wirelessly, and the EC from computational tasks increases with the computation complexity. Obviously, the EC of a system/network is given by the sum of the EC of its components/devices.

The EC scales linearly with the floating-point operations per second (FLOPS) and the million instructions per second (MIPS) in digital signal processors and computer systems, respectively [332]. The scaling factor depends on the amount of computational work, the number and type of functional units, the clock frequency, and the complexity of the instruction set architecture. Therefore, FLOPS and MIPS units are usually more useful than J units for EC

performance comparisons in these systems.[15] Nevertheless, when assessing network EC performance figures, one may inevitably rely on J units and averages over different states, workloads, components/devices, etc.

Other three key metrics related to EC are:
- Net harvested energy (NHE), which represents the amount of energy that is available for use by an EH device after considering losses, i.e., due to energy conversion and storage, and EC related to the EH protocol, e.g., for CSI acquisition in RF-WET systems [333].
- Relative power saving (RPS), which quantifies the EC reduction driven by a certain approach compared to a specific benchmark [218].
- Peak demand (PKD), which measures the highest level of energy demand during a given period. It can be used to assess the capacity of a renewable energy system to meet maximum demand and identify opportunities for demand management and energy storage [334].

4) GREEN EC SHARE (GECS)

The exploitation of renewable sources is fundamental to support IoT sustainability and thus will continue expanding in the next few years. Consequently, quantifying their relative contribution to the total energy budget of IoT devices/networks/systems is increasingly relevant. For this, the GECS metric, which quantifies the portion of the energy consumed from renewable/green sources relative to the total EC, is undoubtedly attractive. Through GECS, one can get a clear understanding of the extent to which a given IoT device/network/system relies on green energy sources. GECS can even evolve to quantify the energy contribution from specific renewable sources, thus, supporting more granular insights. Notice that this metric can be used as a benchmark to set targets for increasing the green energy share(s) of a solution, as part of a broader sustainability strategy.

5) LEVELIZED COST OF ELECTRICITY (LCOE)

This metric evaluates the economic feasibility of electricity generation systems. It is defined as the ratio of the estimated electricity cost and the estimated power plant's electricity generation during its lifetime, thus given in monetary units per kiloWatt-hour (e.g., $/kWh). In the context of renewable energy, LCOE estimation is heavily determined by the ambient energy availability over the system lifetime. In fact, the number, size, and complexity of the EH transducer/circuits and the energy storage vary depending on the geographic location. Besides, the uncertainty in the amount of electricity generated by the system may cause unexpected expenses

---

15. However, comparing different systems in terms of FLOPS/MIPS may be sometimes misleading for EC assessment. For instance, there are processors with architectures specialized for specific tasks. That is the case of security accelerators, whose designs are optimized for encryption/decryption operations. Due to their specialized architecture, these processors require significantly less energy compared to conventional general-purpose processors when performing the required computations.

due to outages or energy trading with the grid to cope with sporadic high electricity demands. Fortunately, energy trading can also be a source of revenue if the network sells its surplus energy (cf. Section VII-A3). One important consideration is the market dynamics that may occur over the anticipated lifespan of EH systems, which is typically around 30 years. This timeframe allows for the potential increase in renewable energy market penetration, as well as the development of technological breakthroughs. These advancements could lead to reduced LCOE due to lower costs for spare parts and more efficient electricity generation.

Finally, LCOE can be extrapolated in WET-enabled networks to estimate the ratio between the electricity cost to the harvested energy. In such cases, the system PTE, the deployment of the ETs, and CE of the EH circuits may be the main impacting factors when estimating LCOE.

6) GRID RELIABILITY/STABILITY (GRS)

GRS captures the ability of the (micro)grid to maintain a stable energy supply, despite fluctuations in demand and supply, especially when considering contributions from renewable energy sources. It can be given in percentage and used to assess the resilience of the grid and identify opportunities for improving grid management and infrastructure.

C. EER METRICS

An alternative (and popular) way to define EE is the ratio between the achievable QoS performance and the corresponding EC/PC that is required, or the ratio between the EC/PC and the corresponding achievable QoS performance. We define this category of EE metrics as EER and notice that they differ from one another in the type of QoS performance metric that is used.

Resource allocation based on optimized EER metrics has been a popular approach in wireless communication engineering, e.g., [133], [134], [141]. However, this may not always be the most effective way to balance performance and EC in all scenarios, unless the EER metrics already include some guarantee for QoS support and/or maximum EC. If not, minimizing EC while maintaining QoS requirements, e.g., [135], [137], [140], [210], [335], or maximizing performance given energy constraints, e.g., [131], [202], [205], [290], [336], [337], may be more important.

Although algorithmically/analytically optimizing EER metrics may not always make sense due to intrinsic operational constraints, they are certainly useful for comparison and drawing valuable performance insights. Next, we briefly discuss relevant QoS performance metrics and associated EER units.

1) DATA TRANSFER METRICS

This is a set of highly related measures used to evaluate data transmission performance over wireless networks, e.g.,

- spectral efficiency, in bps/Hz, measures the amount of data that can be transmitted per unit of time and frequency spectrum;

- THP (capacity), in bps, measures the amount of data that can be transmitted over a wireless link or network (under ideal conditions) within a given period of time;
- $\epsilon-$capacity, in bps, constitutes the best upper-bound for the attainable THP (capacity) supporting an outage probability that does not exceed ϵ;
- goodput, in bps, measures the useful data rate (considering overhead and error correction) delivered to the end-user/application;
- effective capacity, given in bits per channel use (bpcu), constitutes the highest arrival rate that can be served by a network under a particular latency constraint, thus capturing PHY and link layer characteristics.

The corresponding EER metrics are often given in bits/Hz/J, bits/J (or J/bit), or bpcu/J.

2) RANGE (RG) OR COVERAGE AREA (CA)

RG refers to the maximum distance over which a specific device, or a generic technology device, can transmit/receive signals effectively, according to, e.g., target QoS guarantees. Similarly, CA refers to the geographical area within which wireless connections can be established and maintained with certain QoS guarantees.[16] The corresponding EER metrics are often given in W/m or W/m^2, i.e., characterizing the required amount of power per distance/area unit.

Related to CA, but specifically for IRS-assisted networks, the authors in [338] proposed the so-called area of influence (referred to as AroI) metric. AroI comprises the area of significantly improved wireless connectivity triggered by the IRS(s) when optimizing for the whole area under consideration instead of a single nominal receive position. Notably, the spatial resolution of IRSs in horizontal and vertical axes depends on the specific element array, thus, AroI specified in m^2 might be insufficient. Similarly, in non-terrestrial networks, including those composed at least partially of UAVs, high-altitude platforms, and satellites, the network becomes inevitably three-dimensional. In such scenarios, the CA may evolve to coverage volume (CV) and the corresponding EER to be measured in W/m^3.

3) BANDWIDTH

This metric corresponds to the frequency region in which a technology or device operates, e.g., harvests sufficient energy in the case of RF-EH networks. Note that for some communication systems, especially those with frequency hopping (e.g., BLE or LoRa), there might be two different bandwidth notations: i) signal bandwidth referring to the bandwidth of one channel and ii) system bandwidth referring to the bandwidth of all channels used by the system. Another special case of interest is in IRS-assisted networks, where there is the so-called bandwidth of influence (BoI), which indicates

---

16. Sometimes, RG/CA and associated QoS guarantees, e.g., expressed via data transfer metrics, appear directly intertwined in the form of new metrics such as area spectral efficiency and area THP/capacity.

the frequency region in which, any wave hitting the IRS, will be reflected [338]. The corresponding EER metric is given in Hz/W.

4) ACCURACY

This metric somewhat characterizes the proximity of a measured/estimated value to the true value. Both relative units, such as percentages and relative errors (e.g., probability of error, miss-detection (PMD), false-alarm (PFA)), and absolute units, such as mean absolute error (MAE), root mean square error (RMSE), and standard deviation (STD), can be used depending on the scenario/application. Next, three specific examples are briefly discussed.

- ML accuracy (MLA) is a crucial KPI for any ML model, especially TinyML, which is subject to tight HW and SW-related constraints. In general, the definition of learning accuracy depends on the task for which the ML model has been employed. For example, for anomalous behavior detection [339] and object detection [25] tasks, MLA may be the percentage of correct predictions an ML model makes. Meanwhile, for tasks such as PV power prediction [291], learning accuracy can either be defined by the RMSE or MAE among the predictions and actual observations.
  Notice that the EER metric corresponding to an ML model may be expressed as the ratio between the relative/absolute accuracy that is achievable per EC at each inference step. Specifically, the corresponding EER metric can be expressed in %/J when using relative accuracy metrics. Finally, recall that FLOPs can be also used as a measure of energy as discussed in Section VII-B3, thus, resulting in %/FLOP EER units.
- Localization accuracy (LA) is the precision with which a system can estimate the position of an object or feature in a given environment. LA can be given in angular or distance units, which may correspond to RMSE, MAE, or STD statistics. For instance, indoor localization systems supported by future 6G wireless communication systems may operate with sub-meter STD LA [340]. The corresponding EER metric may be given in rad/W or m/W.
- Wake-up accuracy (WUA) refers to the ability of a WuR to reliably detect WuS and avoid false alarms. WUA is typically characterized by two other metrics: PFA and PMD [218]. PFA refers to the probability that the WuR wakes up erroneously in the absence of WuS (false positive, FP). On the other hand, PMD refers to the probability that the WuR fails to wake up in the presence of WuS (false negative, FN). Specifically, PFA is given by FP/(FP+TN) while PMD is FN/(TP+FN), where TP (true positive) is the number of times the WuR correctly detects WuS, and TN (true negative) is the number of times the WuR correctly determines that WuS is not present. Moreover, WUA is calculated as (TP + TN)/(TP + TN + FP + FN). Both PFA and

PMD depend on various factors such as the type of WuS used, the power level of the signal, the distance between the transmitter and receiver, and the RF conditions of the environment. The accuracy of WuR can be improved by optimizing these parameters and using more sophisticated WuS detection algorithms [219].

D. OTHER METRICS
1) ENERGY PROPORTIONALITY COEFFICIENT (EPC)

This metric represents the PC of a device/system as a function of the offered load. In general, the observed PC increases non-linearly with the load, and the PC in an idle state is often non-negligible, e.g., network switches consume up to 85% of their peak PC when idle [341]. EPC is defined in $[-1, 1]$, where EPC = 1 (-1) means that each increase in load leads to an equal increase (decrease) in EC, while EPC = 0 describes the case when the system EC is constant and does not depend on the load.

Three other metrics related to EPC are [341]:

- Energy proportionality index (EPI), which captures the difference between the measured power and the ideal power, i.e., the power that the device should consume if it is fully energy proportional. EPI is expressed in the region between idle and peak PC only. EPI = 0 (1) indicates that the EC is agnostic (fully energy proportional) to the workload.
- Idle-to-peak power ratio (IPR), which measures the ratio between the idle and the peak PC. IPR values tending to zero indicate energy-proportional designs.
- Linear deviation ratio (LDR), which captures the deviation of the observed PC from the fully proportional case, i.e., a straight line connecting idle and peak PC values. LDR = 0 corresponds to a linear system.

2) WET EXPOSURE LEVEL

This metric describes how different WET energy-carrying signals disturb the surrounding environment.

A common metric to characterize RF transmissions is the power spectral density (in W/Hz), which describes the power distribution of a signal across different frequency components. In the spatial domain, one can resort to the effective isotropic radiated power (in W) which is the hypothetical power that an isotropic antenna must radiate to yield the same signal strength as the actual RF transmission in the direction of the antenna's maximum gain. Besides, the EM field (EMF) radiated by RF sources may cause disturbances in nearby equipment operating in the same frequency band. For such a case, international organizations have resorted to EM compatibility, a boolean metric that takes on "pass" or "fail" and describes the ability of electronic equipment to successfully operate in a certain EM environment without being affected by (or affecting) other devices.

International EMF exposure limits are available, e.g., ICNIRP guidelines [342], which are set with large margins to protect against both short- and long-term health effects. Limits are provided for whole-body and local exposure scenarios while considering both general public and occupational environments (cf. [343] for their impact on the RF-WET performance). The exposure limit values relate to basic physical quantities that are known to induce adverse health effects. For instance, the specific absorption rate (SAR, in W/kg), specific absorption (SA, in kJ/kg), absorbed power density (in W/m^2), and absorbed ED (in kJ/m^2) are basic physical quantities measuring the absorption rate of EM energy in the human body. Other reference quantities that are more easily evaluated, e.g., incident power density (in W/m^2), incident ED (kJ/m^2), electric field strength (in V/m), and magnetic field strength (in A/m) are also provided. These have been derived from the basic restrictions to provide a more practical way to demonstrate compliance. Reference levels provide an equivalent degree of protection as the basic restrictions, and thus an exposure is taken to be compliant with the guidelines if it is shown to be below either the relevant basic restrictions or relevant reference levels. There are also EMF exposure assessment standards available for wireless communication and WET technologies.

The exposure to ultrasound signals is commonly evaluated by measuring the intensity of the pressure wave in dB, scaled to a frequency sensitivity response curve. However, the exposure duration is also relevant, as prolonged exposure can cause hearing impairments in animals whose hearing response falls within the operating frequency. In humans, the impact of ultrasound signals is primarily characterized by two indexes: the thermal index and the mechanical index. The thermal index is a unitless metric that measures the ratio between the acoustic power penetrating the skin and the amount of power required to raise the body temperature by one degree Celsius. On the other hand, the mechanical index indicates the ability of the acoustic signal to create tissue mechanical stress and damage.

3) WAKE-UP TIME
Wake-up or start-up time refers to the time it takes for an IoT device to wake up from a low-power sleep state and become fully operational, thus it is given in time units. This is an important factor in determining EC and responsiveness of a network implementing duty cycling and/or WuR. Typically, wake-up time depends on several factors such as the type of device, the complexity of the wake-up process, and the power management scheme used. For instance, some devices may require a longer wake-up time to initialize, calibrate, or preheat sensors or to establish a wireless connection. Similarly, wake-up time can be longer if the device is in a deep sleep mode that requires more time to restore the device's state. To optimize EC and responsiveness, it is important to minimize the wake-up time. This can be achieved by using low-power HW components, optimizing the SW for fast wake-up time, and selecting appropriate power management schemes [344].

4) INFT
INFT refers to the time it takes for an ML model to make a prediction (i.e, inference) on a new data sample. INFT can be affected by the complexity of the model, the size of the input data, the HW configuration, and the SW implementation. This metric is relevant for scenarios demanding real-time inference and/or low EC (since EC is proportional to active time) such as power management and sleep mode handling.

5) MDS
This metric refers to the amount of memory required to store the model's parameters and configuration, thus it is given in bytes (kB, MB, etc.). A small MDS makes the ML model easier to deploy on resource-constrained devices and/or can lead to fast INFT since the number of computation parameters during the inference phase is limited [345]. However, reducing the MDS can also have a negative impact on model performance, since smaller models may have limited capacity to capture complex patterns in the data.

6) PEAK SRAM FOR INFERENCE
SRAM is an onboard memory space that accepts both read and write operations. Thus, an ML model's mutable parameters during its runtime are stored in SRAM. Notably, the peak memory required by a TinyML model during its inference, also called peak SRAM, becomes noteworthy in MCUs because of the limitations on onboard available memory. The standard units to express peak SRAM are kB and MB. Also, the peak SRAM depends on the memory scheduling procedure carried out by an inference library, while NAS decides MDS. The available SRAM on an MCU sets an upper bound on peak SRAM, while flash memory on an MCU constrains MDS [303].

VII. INTEGRATION, PROTOCOLS, AND DEPLOYMENTS
Energy-sustainable IoT ecosystems require the seamless integration of the EP, ET, and EE-related technologies overviewed in this paper. This must be supported by proper multi/cross-layer protocol designs, energy-aware/efficient connectivity solutions, advanced simulation/testing tools, and validation procedures. This section delves into these critical aspects, namely integration, protocols, and real-world deployments as illustrated in Figure 26, highlighting their pivotal role in advancing sustainability KPIs.

A. TECHNOLOGY INTEGRATION
Herein, we discuss the integration of the surveyed technologies with IoT ecosystems while indicating the unique challenges and opportunities that arise.

1) HYBRID EH
The performance of ambient EH technologies is severely limited by the uncontrollability and unpredictability of ambient sources. This can be addressed by integrating multiple transducers into a single device. As shown in Figure 27, a hybrid
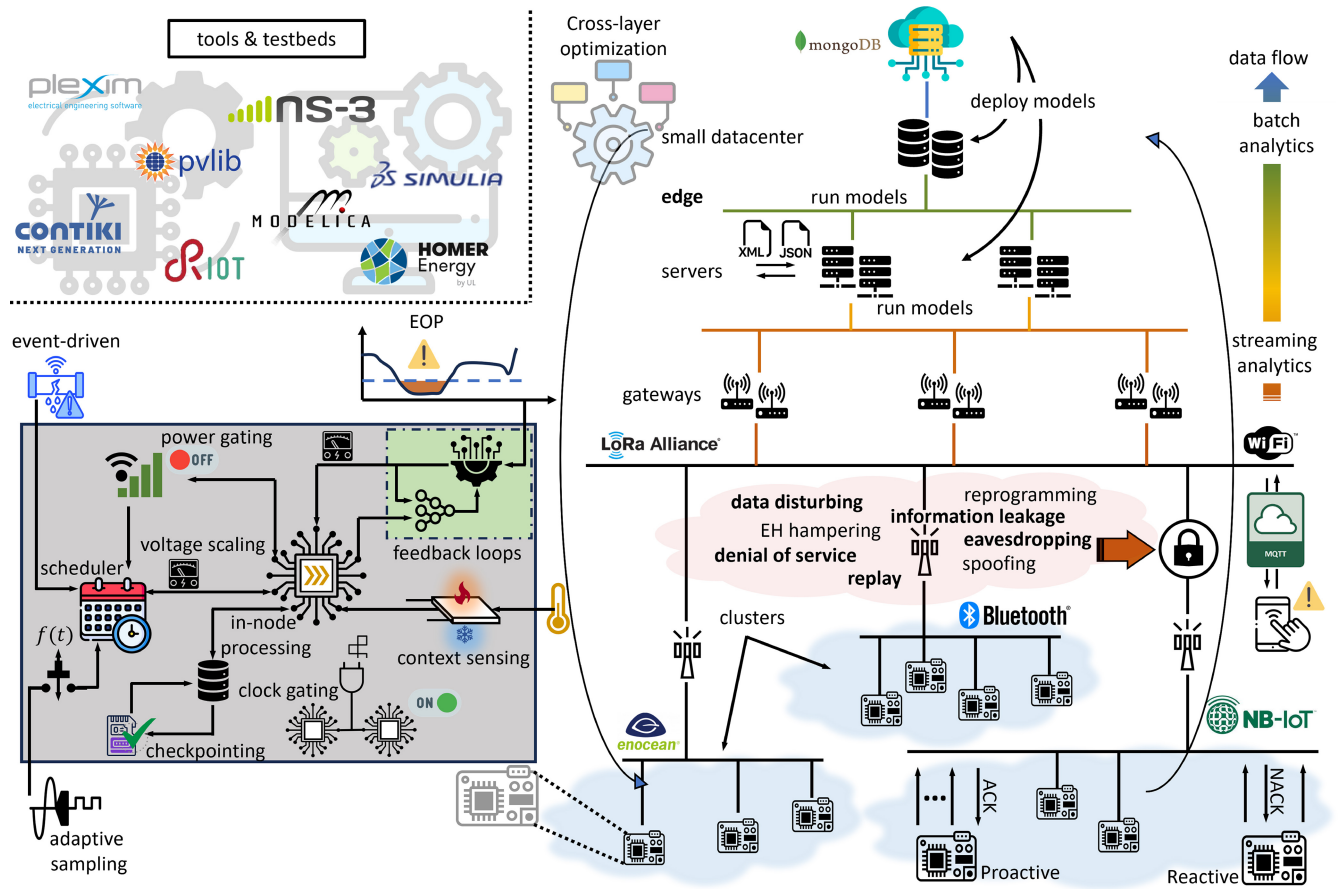
FIGURE 26. Integration, protocols, and deployments for energy-sustainable IoT. Herein, we also illustrate lightweight data-interchange formats, e.g., .xml and .json, and the database management tool MongoDB, which can be used for ML model storage and distribution, configuration files, and measurement formatting.

energy harvester can harness energy from either multiple or a single energy source. In the former case, the system can benefit from a higher AHE if all the sources are simultaneously available or can reduce the EOP of the energy supply as it is more likely that at least one of the sources is available [346]. As an example, Figure 27a illustrates a device harvesting energy from the wind and sunlight simultaneously. Observe that AHE is greater during sunny hours while harnessing energy from the wind compensates when sunlight is unavailable. In the case of single-source hybrid EH, the system incorporates additional transducers to harness the wasted energy from another EH process. For instance, Figure 27b shows that the excessive heat created in CPVs for light EH can constitute an additional EH source by using a heat-based transducer [347]. Besides, multiple transducers designed for harnessing energy from the same source type can be combined to create a more versatile EH system. For instance, Figure 27c shows that one can increase the AHE, enable dynamic matching to the frequency of the source, or allow the transducer to resonate at multiple frequencies by combining different types of VEH transducers [76]. Finally, EH from ambient or dedicated sources can amicably coexist, e.g., dedicated ET may be turned off when ambient energy

is sufficient for the device operation, thus saving energy at the energy transmitter [348].

Packing multiple transducers into a single device to realize hybrid EH is not exempt from challenges. First, it increases the manufacturing costs and compromises the aesthetics and form factor of the device. Hence, the trade-offs between extended device lifetime with the added costs require careful analysis. Second, it increases the HW's complexity in dealing with transducers' varying characteristics and optimal operating conditions. Indeed, the power management circuits must combine different output voltage levels and match the internal impedance of multiple transducers [349] to maximize the NHE. Finally, the environmental conditions can affect each transducer differently, which may affect the overall performance and lifetime of a hybrid EH system.

2) ENERGY MANAGEMENT AND FORECASTING

Harvesting energy from ambient sources and its judicious usage are equally essential to achieving IoT ENO. Figure 28 shows the energy management techniques discussed next.

Digital circuits' EC comprises dynamic and static PC components [350]. The former results from the switching activity when load capacitances are charged and discharged
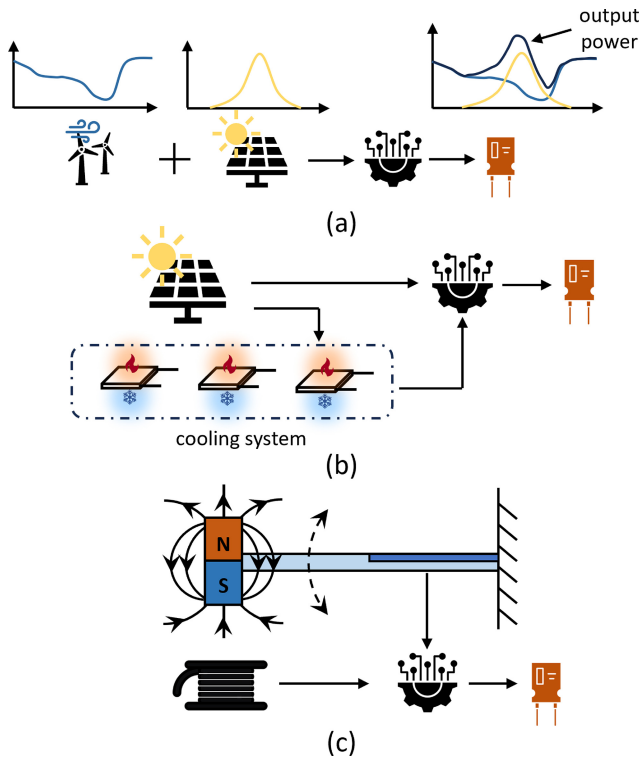
FIGURE 27. Hybrid EH systems from (a) multiple sources using PVs and wind turbines, (b) a single source using a CPV and heat-based energy harvesters for heat waste recovery in the cooling system, and (c) a single source using different VEH transducers.

FIGURE 28. Energy management techniques for digital circuits.

and short-circuit currents flow during state transitions. Meanwhile, the static PC is caused by the leakage currents throughout reverse-biased semiconductors' junctions and sub-threshold conduction of inactive transistors. Notice that EC increases with the supply voltage following a square and linear law in the dynamic and static component cases, respectively. Hence, an effective technique to reduce the overall EC consists of dynamically scaling the supply voltage and the clock frequency to just meet the application demands [351]. However, reducing the supply voltage/clock frequency leads to slower circuits, motivating modern digital circuits to be designed using multi-voltage architectures where each sub-system is operated at a different voltage/clock frequency. Such a design increases the complexity of the power distribution network and requires voltage level shifters to guarantee the correct transfer of logic levels between different power domains [351]. Finally, the supply/clock signal on inactive sub-systems in digital circuits may be cut off to prevent leakage using the so-called power/clock gating techniques. Under such a paradigm, conditional logic operates the circuits that provide power and clock to the sub-systems, pushing them into a deep sleep state whenever inactive. This technique was used in [352] to design a heterogeneous integrated circuit composed of an always-on ultra-low-power MCU for power management and reduced wake-up time and a general-purpose MCU, with higher computational capabilities, for communication, processing, and sensing. Notice that the additional logic
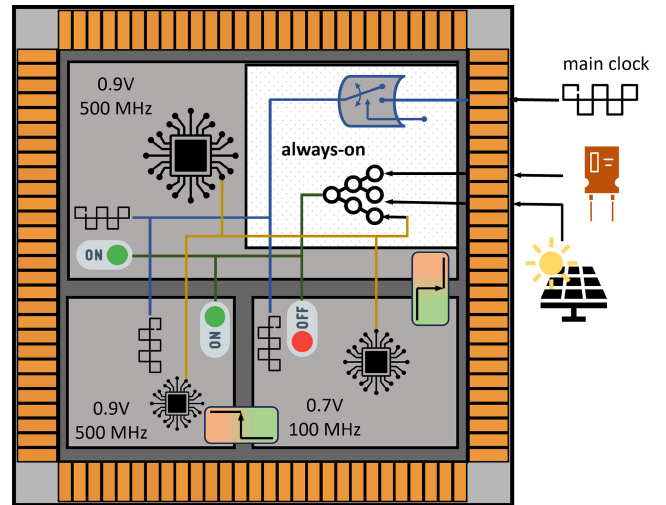
required for implementing power and clock gating techniques increases circuit complexity, cost, clock signal delays, and EC [353].

Combining the aforementioned approaches with proactive energy management techniques is key to boosting the reliability and lifetime of EH-enabled applications [354]. As a first step, one can enable SW-reconfigurable power management circuits that cater to varying operating circuit conditions while adaptively pushing the system into μW PC levels when energy is scarce [355]. This enables running ML algorithms to aid in the complex process of optimizing the previously discussed power management techniques [351]. Indeed, ML methods may help predict the availability and demand for energy through historical data extrapolation and thus facilitate preemptive scheduling maximizing EE.

In general, real-time data aggregation, advanced predictions, and adjustable algorithms are needed to ensure accurate predictions and efficient energy allocation [356]. For instance, at a device level, an RL-based TinyML algorithm using past samples from the EH process and the current battery level may judiciously allocate energy for future operations [290]. The main limitation of the approach in [290] is that the TinyML algorithm is trained offline which ignores that unpredictable changes in EH patterns can lead to improper energy management strategies. A potential solution to address this limitation could involve integrating retraining procedures at the edge or cloud level.

Leveraging predictive analytics and ML algorithms for proactive energy management shows promise for efficient IoT energy management [354]. Depending on the model's complexity, robustness, and delay considerations (cf. Section VI-C4, VI-D4, and VI-D5), implementation may be more feasible to occur on-device or on the edge/network side. These strategies together allow for the optimization of the EC of IoT devices in accordance with their real-time needs, leading to longer BL and contributing to the sustainable operation of IoT ecosystems.

3) ENERGY TRADING & MICROGRIDS

The exploitation of renewable sources to power self-sufficient energy transmitters becomes an appealing solution on the road toward sustainable networks [357]. However, although energy management and forecasting techniques can help match energy generation and demand, this is not always possible, calling for energy trading mechanisms.

The above vision aligns with the emergence of microgrids, which are becoming increasingly popular to provide reliable and sustainable electricity in remote areas. A microgrid is a local energy network that comprises renewable electricity generators (cf. Section II), energy storage, fuel-based generators, and controllers/inverters. Notice that fuel-based generators can be used as a backup source of electricity when renewable sources cannot meet the energy demand alone while the controllers/inverters ensure that energy flows efficiently between the various microgrid components.

Since energy trading with the main grid serves as a backup for renewable-powered WET systems [358], minimizing/maximizing LCOE/GECS while achieving a reliable network operation is necessary. This is not only more environmentally friendly but also reduces energy production and distribution costs. Moreover, direct energy trading among energy transmitters can also be possible using wired and wireless links that bypass the hierarchy of the power grid distribution network. In the case of wireless energy trading, static energy transmitters can exchange energy using, for instance, RF-WET and laser power beaming, to transfer energy over large distances. Moving/flying energy transmitters can benefit from these but also from near-field WET technologies, such as inductive coupling and acoustic WET, which provide high PTE, to trade energy with static transmitters. The idea here is that those transmitters cannot be equipped with large solar panels or wind turbines, so it is more beneficial for them to trade with other ground infrastructures with superior ambient EH capabilities [16], [359]. Finally, energy transmitters can trade their surplus energy using the grid distribution network [360]. This scenario becomes useful in the case when there is no direct connection between the parties, e.g., when the seller and the buyer transmitters are operated by different companies or deployed so distant from each other that a direct connection becomes infeasible.

Scalable algorithms are needed to support the increasing number of trading energy transmitters. Also, the latter may be managed by different operators, posing a challenge in terms of privacy and security. This motivates the use of DLT, which can provide transparency to the transactions and protect the ledger against possible forging by energy transmitters/receivers behaving maliciously [361]. In other situations, the operators may not want to disclose their generation capacity and demand, hence potentially limiting the information exchange during the trading process [362].

4) ENERGY-AWARE SENSING AND COMPUTING

Although accurate energy availability forecasting, efficient energy management, and energy trading techniques (cf. Sections VII-A2 and VII-A3) can help prevent energy outages and task execution interruptions, they may not avoid them completely. These adverse events may still occur in those cases where the devices must uninterruptedly monitor/sense the environment, report, and/or execute high-complex tasks given limited energy availability. Sensing, computing, and communication protocols must consume extremely low energy and be resilient against these situations. Herein, we focus on the former two, while communication technologies and protocols are addressed in Section VII-B.

In terms of sensing, the concept of "context sensing from EH patterns" is gaining traction. This is based on the observation that EH patterns often reflect the context in which energy is being harvested. For instance, kinetic-powered wearable IoTs can identify and tally the user's steps as the energy harvester produces distinctive peaks in its signal with each footstrike [363], while a thermoelectric energy harvester can detect changes in surface temperature from the variations in the EH signal [364]. Note that by substituting dedicated sensors with a context detection algorithm based on the EH patterns, the device EC might reduce significantly.[17] The two basic approaches for sensing from EH signals are [365]: i) analyzing the patterns of the instantaneous power generated by the EH transducer, which allows the detection of a rich set of contexts at the expense of frequent sampling of the fluctuating power values; and ii) analyzing the amount of the total energy accumulated in the storage over a specific period of time, which consumes less energy by sampling the stored energy only once in a while at the expense of more coarse-grained context sensing. Very simple TinyML (cf. Section V-C) algorithms, e.g., for peak detection, moving average computation, etc, may be used for these tasks.

In terms of computation, efficient "checkpointing" techniques are necessary to periodically save volatile states in non-volatile memory so that the program execution can restart from a known state if the power supply fails. Notice that all variables and registers stored in the volatile memory during the program execution are completely lost in an energy-outage situation. The main challenges encountered by checkpointing techniques for EH IoT devices and main state-of-the-art approaches are [365]:

- Non-negligible EC: the energy overhead of checkpointing increases with the number of variables to be copied/stored/reloaded. Therefore research on the area focuses on identifying how often to implement checkpointing and specific stages/locations for it. One appealing approach lies in inserting potential checkpoints to

---

17. Readers may refer to [365] for further details on possible applications, e.g., human activity recognition, transportation mode detection, estimation of calorie expenditure, gait recognition, hotword detection, HVAC airflow monitoring, and acoustic communication.

the program at compile-time and activating checkpointing processes at run-time. Observe that checkpointing after a function call, instead of during the call, is desirable since it requires copying less volatile variables to non-volatile memory. However, having too separated checkpoints may lead to "Sysyphean loops", which happen when the energy harvested since the previous power cut-off is repeatedly not enough to reach the next checkpoint. This motivates the use of a watchdog timer to be triggered when the program re-executes from a checkpoint or a task boundary, e.g., to launch a checkpointing process when the timer interrupt occurs to mandatorily split previously uncompleted code group into two or multiple smaller groups [366].

- State inconsistency: complete state preservation is guaranteed when variables are exclusively stored in either volatile or non-volatile memory, not split between both. All variables in volatile memory are saved in non-volatile memory at a checkpoint, allowing consistent state recovery after power loss. Yet, most systems employ a small volatile memory for frequently used variables, placing the rest in non-volatile memory for efficiency. During rollbacks, volatile memory variables revert, while non-volatile memory variables retain new values, leading to potential data inconsistencies. This can be addressed by decomposing a longrunning program into a sequence of short and atomic tasks [367], [368], and leveraging the concept of "idempotency".[18]

- Timing inconsistency: power failures can cause timestamp discrepancies, especially in IoT devices interfaced with multiple sensors. Notice that battery-powered devices use a real-time clock to prevent this, but EH IoT devices lack consistent power, leading to inconsistent timestamps in reported sensor data. This motivates the tracking of the time elapsed between power failures, e.g., using remanence decay of SRAM and capacitor for timekeeping [369], [370]. Note that time tracking can also be done by receiving relevant information from the network (e.g., cellular or satellite) after recovering, but this might lead to more EC.

The use of non-volatile RAMs combining features of RAM (fast access to any address) and flash (being non-volatile) may be appealing to avoid the above challenges (cf. Section V-C3).

Finally, efficient sensing and computation mechanisms for ultra-low-power EH devices can leverage i) event-driven processing, e.g., triggering computation tasks only when specific sensor data thresholds or conditions are met, reducing the need for continuous polling; ii) in-node pre-processing, e.g., essential features extraction from sensor data before transmission; iii) adaptive sampling, e.g., dynamically adjusting sampling rates based on the current operating conditions

or the importance of the sensed data, conserving energy during periods of lower activity; iv) task schedulers that prioritize sensing and computation tasks according to their urgency and importance; v) feedback loops, which may allow adjusting computation and sensing parameters based on changing environmental conditions or system performance, ensuring optimal resource utilization; and vi) HW/SW co-design to exploit the strengths of both domains for efficient computation and sensing.

5) SECURITY AND PRIVACY IN ENERGY-SUSTAINABLE IOT

IoT ecosystems comprise low-capability devices that are responsible for collecting data, pre-processing it, and/or controlling certain processes, e.g., sensors and actuators, but also edge and network devices and cloud servers, which execute complex processing tasks to support user applications and services. IoT protocols and data management/analytics must optimize and harmonize the operation of all these heterogeneous system components, and do so while supporting security and privacy, which is inherently challenging.

For low-capability IoT devices, the main security attacks are related to [371], [372]:

- eavesdropping, which consists of intercepting and storing messages interchanged by the authorized devices, and thus supposes a data leakage and privacy threat;

- denial of service, including mainly jamming and flooding. A typical example of a jamming attack is the transmission of interfering signals that affect the performance of the communication channel. Meanwhile, a flooding attack may consist of the transmission of bogus messages, whose processing consumes energy and may ultimately deplete the IoT battery, e.g., BLE and LoRa devices may be flooding-sensitive [373]. In fact, a WuR device may experience the so-called *denial of sleep*, by which it is maintained awake by a malicious system continuously sending it wake-up requests.

- side-channel information leakage, where the malicious node leverages magnetic leaks, EM, EC, and/or timing information PHY features to obtain private data or secret keys;

- spoofing, which consists of impersonating trusted entities or devices to deceive the system, thus often leading to unauthorized access or data manipulation;

- replay, which consists of retransmitting previously captured legitimate data to gain unauthorized privileges or manipulate system behavior;

- data disturbing and reprogramming, which consists of partially/totally modifying information and/or program data on a sensor, e.g., security parameters, sensing conditions, reporting destinations, by direct access to the node, or simply by feeding it with malicious data. Intermittent computing procedures such as checkpointing are typically used to support the operation of EH devices (cf. Section VII-A4), but may, unfortunately, facilitate these security breaches, e.g., a malicious

---

18. Specifically, an idempotent task can be executed multiple times without producing different results. To ensure idempotency, the task should not contain any idempotency violation, which is a write instruction to a non-volatile memory that was first accessed by a read instruction.

adversary could read the data on the checkpoint, checkpoint tamper after system restoration, or execute a checkpoint replay attack;

- EH hampering, which consists of hampering the amount of ambient energy that the IoT device harvests, e.g., to force the device to lower its security level or to delay the transmission of highly sensitive data.

Notice also that EH-IoT devices in certain applications may go off for long periods and then back on, thus hiding possible attacks that tamper with the device.

Considering that performing cryptographic operations and processing/transmitting large data overhead, including signatures, keys, padding, drains valuable energy resources, extremely lightweight security and privacy-preserving mechanisms must be carefully designed [374] and operate at lower layers mostly. For instance, they could rely on adaptable security levels and waiting queues for highly confidential data such that the device transmits data when it has the energy to encrypt it [372], and TinyML-assisted approaches [315]. In the case of *denial of sleep* attacks, a simple approach is to generate/update the wake-up address of each WuR node in a pseudo-random fashion based on key material known only by authorized peers [375].

Unfortunately, even lightweight mechanisms may not be incorporated into such IoT devices in many cases due to their extremely low energy availability and complexity, and they must completely rely on network/edge protocol stack security to secure their data. The IoT network/edge protocol stack must ensure compatibility and security across a wide variety, and potentially massive number, of IoT devices, platforms, and communication technologies. Moreover, keeping IoT devices up-to-date with the latest security patches and firmware updates may be challenging, especially in remote or inaccessible locations. These issues call for efficient encryption, key management, authentication, and authorization mechanisms. Also, IoT ecosystems may be divided into segments or zones, each with its own security policies and access controls, to limit the impact of a security breach and help contain potential threats.

ML is a fundamental tool to support IoT security [374], e.g., by helping detect anomalies and patterns indicative of security breaches or attacks. Due to its tendency to be resource-intensive, ML approaches are mainly leveraged on the network side, although edge and device sides are increasingly exploiting them as well (cf. Section V). Note that ML may also help a malicious system to become smarter. For instance, a malicious system may correlate a certain IoT device activity with the surrounding environment conditions, and use such information to predict the type of sensed data or energy availability. Therefore, every system action and potential information that may be derived from it, and observed/sensed, must be carefully considered.

Finally, ML and DLT (e.g., blockchain) integration can drastically improve EE and privacy/security of transactions in 5G and beyond networks [376], [377], [378]. Specifically, DLT allows value transactions between parties through decentralized trust but suffers from large communication overhead and difficulty in handling massive two-way connections, which may be mitigated by ML approaches. Indeed, ML can enhance DLT security and efficiency by providing intelligent data analysis and prediction capabilities [257]. A relevant application scenario for ML and DLT integration within the context of self-sustainable and secure IoT ecosystems is that of energy trading and microgrids (cf. Section VII-A3).

6) TOOLS AND TESTBEDS

Tools and testbeds provide controlled environments for realistic experimentation, enabling the identification/solution of potential energy inefficiencies, protocol bottlenecks, and integration issues. By simulating and analyzing the performance of IoT devices and systems, potential IoT designs/ strategies can be validated prior to real-world deployment.

Key tools for EH circuit/HW modeling and design include LTspice, PSIM, Simplis, MathWorks Simscape Electrical, Altium Designer, and KiCad. Meanwhile, the following tools are undoubtedly appealing for device/system-level simulation and forecasting purposes:

- PLECS, which is a simulation platform that includes support for modeling EH systems. It allows the modeling and simulation of various EH sources and circuits in both continuous-time and discrete-time domains.
- OpenModelica [379], which is an open-source modeling and simulation environment supporting the development of models for various physical systems, including EH devices. It allows the modeling and analysis of the behavior of energy harvesters in complex systems.
- WRF, which is a widely used atmospheric simulation system that provides detailed weather forecasts, especially for solar radiation, wind speed, and other weather parameters that impact energy generation from solar panels and wind turbines.
- SIMULIA (part of Dassault Systèmes), which is a suite of simulation tools that includes capabilities for studying multiphysics systems, including EH applications. It can be used to simulate the mechanical and electrical aspects of EH devices.
- PVLIB Python [380], which is an open-source library providing functions to estimate/forecast solar irradiance, module temperature, and energy output from photovoltaic systems.
- EH Module in NS-3 [381], which allows the simulation of the EH capabilities of IoT devices and the study of their interaction with communication protocols.
- HOMER Energy, which allows forecasting energy generation based on historical data and system

TABLE 12. Features and KPIs of IoT communication technologies[d].

Technology	Freq. band	Topology	Sensitivity	Data Rate	Range	Average PC	Security
RFID (EPC-global Gen2)	sub-GHz	P2P	−85 dBm	40−640 kbps	>10 m	1−10 μW	AES-128, Grain-128A
BLE 5	2.4 GHz	P2P, mesh	−97 dBm	0.125−2 Mbps	<200 m	20 mW	AES-CCM, P256 ECDH, HMAC-SHA-256
WiFi HaLow	sub-GHz	star, relays	−100 dBm	0.04−4 Mbps	<1 km	10 − 100 mW	WPA3, AES-CCMP
Z-Wave	sub-GHz	mesh	−110 dBm	9.6−100 kbps	<100 m	1 mW	Security 2 (S2), AES-128
Thread	2.4 GHz	mesh	−105 dBm	250 kbps	30 m	0.5 mW	AES-128,IP-based security, QR-aided authentication
ZigBee Pro	2.4 GHz, sub-GHz	mesh	−100 dBm	250 kbps	<100 m	0.1−0.5 mW	CCM, IEEE 802.15.4-based
EnOcean	sub-GHz	P2P, mesh, star	−98 dBm	125 kbps	10 − 30 m	0.05 − 50 mW	rolling code, AES-128
NFC	HF	P2P	−10 dBm	424 kbps	10 cm	5 mW	AES-128, NDEF
LoRa	sub-GHz	star-of-stars	−148 dBm	0.3−50 kbps	>10 km	150 mW	AES-128 in CCM
SigFox	sub-GHz	star	−142 dBm	1−100 bps (uplink)	30−50 km (rural), 3−10 km (urban)	100 mW	application/manufacturer-dependent
NB-IoT	licensed LTE bands	star	−141 dBm	204.8 kbps (uplink)	10−35 km	500 mW	3GPP security
LTE-M	licensed LTE bands	star	−108 dBm	4 Mbps (downlink), 7 Mbps (uplink)	<15 km	> 500 mW	3GPP security

[d]The data in this table come from an exhaustive search over numerous technical specifications and are intended for qualitative, rather than quantitive, comparisons. Moreover, a thorough technology comparison may require considering other indicators that are not captured in this table.

specifications and provides microgrid optimization capabilities.

These tools and many others allow for investigating integration capabilities, expected performance, and valid protocols of EH-based devices and systems.

Although the above tools also provide testbed capabilities to some extent, they are often limited, thus calling for more advanced platforms for modeling and testing complex ecosystems. For instance, researchers may resort to Contiki or RIOT, open-source operating systems specifically designed for IoT devices, to develop and test IoT applications with a focus on energy optimization. Indeed, Contiki provides various built-in protocols, including CoAP and MQTT,[19] and supports energy-efficient communication and management features. Moreover, COOJA, which is a network simulator specifically designed for Contiki, may be used to simulate large-scale IoT networks and evaluate the EC of IoT devices under different conditions. Finally, tools such as OpenLCA and SimaPro for life cycle assessment are becoming appealing to quantify the environmental impact of IoT products/services while considering the entire lifetime.

19. CoAP is defined in RFC 7252 as a UDP-based transport protocol, a limited HTTP, developed for devices with limited memory, storage and computing power, limited battery power, and low bandwidth. Meanwhile, MQTT is a lightweight, publish-subscribe, IoT network protocol for message queue/message queuing service.

B. PROTOCOLS, CROSS-LAYER DESIGNS, AND SCALABILITY
1) IOT CONNECTIVITY LANDSCAPE

The IoT connectivity landscape is broad, with both short- and long-RG technology solutions as illustrated in Table 12. Also, connectivity protocols like MQTT, CoAP, and AMQP can enable seamless communication among heterogeneous IoT devices. In general, energy-sustainable IoT connectivity solutions must i) support EH, WET, and/or energy trading, ii) achieve high EE and spectral efficiency, iii) guarantee long BL, iv) be scalable and support high connection density, and v) require low maintenance and intervention. Current IoT connectivity solutions are far from achieving these goals, but firm steps are currently being taken in these directions.

Short-RG IoT connectivity technologies feature protocols with a sub-km communication RG, e.g., RFID, Bluetooth, WiFi, Z-Wave, Thread, ZigBee, EnOcean, and NFC, and target use cases such as audio streaming, well-being monitoring, and home/industry automation. RFID is a passive BC technology [12], [158], [173], [194], which often directly incorporates EH and WuR techniques, while Bluetooth and WiFi have dedicated protocols for low-power devices, namely i) BLE [232], [382], which incorporates WuR, standard duty cycling, and an energy-efficient adaptive frequency hopping mechanism, and ii) IEEE 802.11ah (WiFi HaLow) [383], which implements smart power management

techniques such as distance-based power control. Z-Wave is designed specifically for smart home devices and applications, utilizing a mesh network topology such that devices can act as repeaters and may choose the most energy-efficient communication routes within the mesh network [384], [385]. Thread also utilizes a mesh network topology, although supporting IPv6, which enables direct IP connectivity for devices without requiring translation or additional overhead [384]. Finally, standards like EnOcean and ZigBee Green Power have defined a radio interface to specifically operate battery-less EH devices (even using RF-EH/WET in the case of EnOcean) [384], while the NFC alliance is currently focusing efforts on developing a standard for EP at short distances over the existing NFC radio interface.

Meanwhile, long-RG IoT connectivity technologies feature protocols with a communication RG from several to tens km, e.g., LoRa, ZigFox, NB-IoT, and LTE-M, and suit use cases such as smart agriculture, asset tracking, and environmental monitoring. Notably, LoRa devices are capable of dynamically re-configuring their transceivers, e.g., the spreading factor and bandwidth, so that transmissions can achieve longer distances with a given power budget [151], [385], [386]. SigFox and NB-IoT devices communicate over narrow bandwidth channels to reduce the received noise power and simplify the receiver's front-end design [227], [385]. NB-IoT and LTE-M implement extended discontinuous reception that allows the devices to enter a sleep mode for long durations while periodically waking up to check for incoming data, thus reducing the waking-up frequency and conserving energy [222], [227]. A key difference is that LTE-M supports higher bandwidth (1.4 MHz) and thus cannot be deployed in guard bands, like NB-IoT. Finally, note that contrary to LoRa/Sigfox, both NB-IoT and LTE-M support IP from UE while the transceivers often implement TCP/IP, MQTT, CoAP, and even FTP and HTTP stacks.

2) ENERGY-AWARE PROTOCOLS

Energy-aware protocols are crucial to meet/improve energy-sustainability KPIs (cf. Section VI). The main challenges are associated with the diverse composition of IoT devices, variable network conditions, and specific application requirements. Therefore, IoT protocols must be adaptive and manage energy resources based on availability and demand patterns [246]. For instance, data transmissions may be triggered avoiding ultimately compromising present and future system states, which inevitably requires energy-awareness.

At the MAC and network layer, Low-Energy Adaptive Clustering Hierarchy, Time-Slotted Channel Hopping, and Routing Protocol for Low Power and Lossy Networks, designed to adapt to the data paths that are most frequently traveled, are incipient examples of energy-aware protocols. Meanwhile, content-based protocols focused on retrieving content to minimize expenditure of energy [387] are appealing at higher layers.

Incorporating intelligence into energy-aware protocols may promote efficient resource allocation, real-time power management, and EE increase [246]. Additionally, ML facilitates the recognition of common device behavior patterns, aiding the implementation of energy-saving strategies like sleep modes and adaptive power control. Ultimately, this alliance of power-driven insight can combine the efficiency of network power and the EE of the network, and promote the integration of energy-aware protocols into the IoT.

The grant-free spectrum access scenario, where the IoT devices attempt to access the transmission medium directly, without prior control signaling, thus saving energy but increasing the collision probability, constitutes an ML application opportunity. Indeed, integrating ML here may allow IoT devices to acquire optimal transmission strategies progressively, thereby minimizing collision occurrences and promoting network scalability. Numerous ML-based grant-free spectrum access mechanisms have already surfaced in existing literature, e.g., [388], [389], [390] discussing distributed RL-based spectrum access mechanisms. Specifically, the time-slot selection task is modeled as a Markov game in [388] and equilibrium points within this game are learned through a reward-inaction RL algorithm. In the case of [389], the spectrum access mechanism leverages the information about the inner states of the device to determine time-slot selection probabilities. Here, the RL algorithm approximates the intricate mapping between the inner states and slot selection probabilities. Meanwhile, the authors in [390] portray the IoT devices as RL agents trained over time, while developing coordination among their spectrum access mechanisms, leading to successful message transmissions. Still, further work is needed to consider the energy availability issues and thus design energy-aware MAC protocols.

Finally, note that these protocols are not only limited by the IoT devices' characteristics but also the network side's and the variety of emerging technologies such as distributed MIMO, IRS, and LIS. Such technologies may not only have advantages but also intrinsic limitations regarding resource allocation and interoperability, which must be considered.

3) CROSS-LAYER OPTIMIZATION

An IoT ecosystem mainly includes the following layers [391]: i) HW/sensing/actuation, ii) local processing and storage, iii) communication, iv) cloud computing (and storage), and v) application. Each of them has its own sublayers and typical EE-promoting techniques. For instance, sleep/wake-up, self-organization, and approximate computing in TinyML [315] are HW/sensing layer techniques, while EH and intermittent computing-related techniques, including NAS, network pruning, quantization, and ROMANet in TinyML [315], correspond to the local processing and storage layer. TinyML MDS is decided by NAS, while TinyML peak SRAM during inference can be reduced by optimizing the memory scheduling using ROMANet. The communication layer comprises the (sub-)layers of the communication
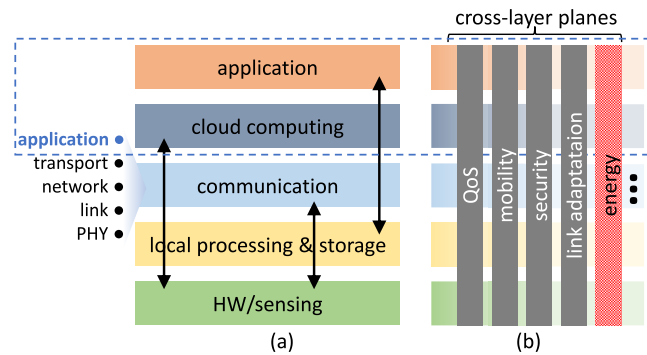
FIGURE 29. Cross-layer designs: (a) direct layer interactions and (b) layer interactions via coordination planes.

protocol stack, including scheduling, access, and routing protocols. Meanwhile, centralized optimization, virtual machines, and caching correspond to the cloud computing layer, and data aggregation, predictive analytics, and user-centric customization techniques pertain to the application layer. By considering the interdependencies and interactions between layers and optimizing accordingly, the system performance may largely improve, increasing lifetime, reducing costs, enhancing QoS in dynamic environments, and improving sustainability.

The traditional cross-layer perspective is illustrated in Figure 29a and lies in creating new interfaces establishing direct paths between non-adjacent layers (or sub-layers) for parameter sharing. Many of the energy-aware protocols from Section VII-B2 exploit this approach. However, this can lead to tightly coupled layers as changes in one layer may require corresponding adjustments in multiple other layers, thus compromising modularity and making the system complex and difficult to manage. Also, excessive information sharing between distant layers might raise security and privacy concerns. This may be mitigated by introducing coordination planes, which are cross-section views of the protocol stack on which interlayer coordination algorithms can be applied. This was proposed in [392] within the communication protocol stack environment/layer while identifying four key coordination planes, namely security, QoS, mobility, wireless link adaptation, and energy. This framework must evolve to incorporate energy and sustainability aspects over the layers and sub-layers as illustrated in Figure 29b.

All in all, holistic approaches are needed for orchestrating the wide spectrum of technical approaches discussed throughout the paper across the IoT ecosystem layers.

4) TOPOLOGY, DATA, AND LOAD MANAGEMENT

Topology, data, and load management play crucial roles in achieving scalable energy-sustainable IoT networks. Indeed, by optimizing the physical layout and data processing tasks of the network, EC, EH/EP capabilities, and connectivity performance can be improved. For instance, clustering nodes into groups or hierarchies can alleviate scalability issues and minimize EC [393], while an appropriate

and dynamic deployment of RF PBs can improve the EH statistics of RF-EH IoT networks and/or lower the deployment/maintenance costs of the PBs network [126], [357]. Also, the expansion of MEC (cf. Section V-C) and distributed data storage capabilities can help reduce latency, communication, and energy costs, supporting scalability and EE. Energy-aware self-organizing networks and routing protocols (cf. Section VII-B2) are appealing as well. In terms of data management, techniques such as i) data fusion, integrating data collected from multiple sensors while filtering out redundant or irrelevant information; ii) resource-constrained data storage models, optimizing storage and computing resources; iii) distributed and replicated storage, reducing reliance on centralized data centers; and iv) energy-aware compression techniques reducing the size of data being transmitted; may be leveraged depending on the scenario characteristics and performance requirements [394], [395]. Note that data management is intrinsically related to load management as well, which comprises i) load balancing techniques, distributing tasks and data processing load evenly across devices; ii) edge and fog computing offload workloads to devices and fog nodes, reducing the need for data transmission to remote clouds [396]; iii) dynamic resource allocation, adjusting network topology and parameters based on real-time conditions. Regarding the latter, ML techniques (cf. Section V) can be exploited to analyze data patterns and predict workload demands for efficient resource allocation.

Finally, DLTs provide a tamper-proof and transparent ledger that is crucial for data fusion and storage [377], [397]. Also, the corresponding consensus mechanisms optimize resource allocation and facilitate load balancing, thus promoting EE. This synergy between IoT strategies and DLT integration promises enhanced data security, efficient ET and trading (cf. Section VII-A3), and optimized resource utilization, advancing energy-sustainability goals.

C. REAL-WORLD CASE STUDIES AND APPLICATIONS

Herein, we delve into real-life implementations that pertain to the surveyed technologies on the road to achieving energy-sustainable IoT. In particular, we focus on those relying on the discussed EP and ET technologies to supply energy. Table 13 summarizes our discussion.

Light-based EH is a popular technology for commercial EH-powered solutions due to its high availability/harvestability compared to other solutions. Inspired by this, SODAQ[20] developed Track Extreme, a device for asset tracking. Every time the Tracker Extreme detects the asset moves, it obtains its location based on WiFi and GPS signals which then can report over BLE, NB-IoT, or LTE-M radio solutions. Similarly, Xeelas[21] developed Series S2, a light-based EH device with applications in industrial monitoring and asset tracking that can communicate

---

20. Please refer to https://sodaq.com/ for more details.
21. Please refer to https://www.xeelas.nl/?lang=en for more details.

TABLE 13. Real-world IoT implementations comprising EP/ET.

Solution	EP/ET	Communication	Intelligence	Use Cases
Track Extreme	light	BLE, NB-IoT, LTE-M. multi-RAT	N/A	asset tracking
enerSENSE	indoor light	LoRa, NFC. Multi-RAT	edge ML for data processing, visualization, and control	industry 4.0, smart buildings
Series S2	light	2G, LTE-M, 2.4 GHz. Multi-RAT	N/A	industrial monitoring, asset tracking
Jack	light	BLE	cloud ML for damage evaluation and location	fleet management
EnOcean	vibration, heat, light	EnOcean®, Bluetooth, Zigbee. WuR support	N/A	smart spaces, EE, smart homes
ONiO.zero	vibration, heat, RF-EH, light	BLE. WuR support	N/A	ultra-low PC applications
IoT Pixels	RF-EH	Bluetooth	cloud ML for mapping measurements into the physical world	smart healthcare, supply chain
MIMO tag [398]	RF-EH	MIMO and backscattering	N/A	digital twins, RFID tags
Infinity	heat, light	BLE	cloud ML	machine health monitoring
AirCord™	laser-ET	WiFi	cloud ML for power allocation	industrial IoT, healthcare, gaming, retail
Cota® Real Wireless Power™	RF-ET	N/A	cloud ML for collection and analysis of device data	building automation, smart homes, lighting control, and industrial applications
Emrod	RF-ET	N/A	N/A	energy distribution, space-based energy infrastructure, connecting renewables to the grid

over 2G, LTE-M, and 2.4 GHz. In the transportation sector, AGC Automotive[22] has developed the Jack device for fleet management operations. The Jack utilizes vibration sensors to detect impacts on the vehicles' windshields while the measurements are uploaded to the ML-powered cloud application via a BLE gateway for damage evaluation and reporting. Accounting for the duration of indoor human activities, enerthing® has developed a thin-film PV specifically optimized for indoor EH. The solution is utilized for powering the enerSENSORS devices for Industry 4.0 and smart building applications. The enerSENSORS utilize a LoRa radio to upload the measurements to a cloud-based service that runs ML algorithms to aid decision-making processes based on the received measurements. The enerSENSORS also support over-the-air configurations via an NFC radio.[23]

Undoubtedly, RF-EH stands over its technological counterparts in achieving small form factor implementations due to its relative simplicity. This technology is exploited by Wiliot[24] for the IoT Pixels, which is a credit card-sized smart label intended for smart healthcare and supply chain management applications. The device communicates over Bluetooth with a cloud service to report temperature measurements and location information. The cloud service transforms the raw sensory information into the corresponding physical

domains using ML algorithms. Meanwhile, researchers at the Georgia Institute of Technology developed a multi-antenna solution in [398] for a card-sized device to boost the output power of RF-EH. Their prototype implements a low-power MIMO solution using Rotman lenses. Rotman lenses achieve beam scanning by adjusting the time delays created in the microstrips network that feeds the antenna elements; hence, eliminating the need for phase shifters.

EnOcean[25] designs devices capable of harvesting energy from vibrations, light, and heat, and communicating over Bluetooth, Zigbee, and EnOcean®. EnOcean's solutions also provide support for adding an external WuR circuitry as an energy-saving mechanism. The company targets smart spaces (occupation maximization), EE, and smart homes. Similarly, Sensemore[26] developed Infinity, a self-sustained device that can harvest energy from heat and light. The Infinity device communicates over BLE to an ML-powered cloud service, which can evaluate the condition of industrial machinery based on sensory data. Aiming at reducing cost and form factor, ONiO® developed ONiO.zero, a tiny MCU that integrates power management, memory, BLE transceiver, and support for WuR circuitries. Currently, ONiO.zero targets batteryless applications, such as a marketed remote control, which can be sustained by EH from vibrations, multiband RF-EH, light, and thermal energy.[27]

22. Please refer to https://www.smart-jack.com/ for more details about their Jack project.

23. Please refer to https://www.enerthing.com/en/products/ for more details about enerthing's portfolio.

24. Please refer to https://www.wiliot.com/ for more details.

25. Please refer to https://www.enocean.com/ for more details.

26. Please refer to https://sensemore.io/ for more details.

27. Please refer to https://www.onio.com/ for more details.

TABLE 14. Summary of the main challenges and associated research directions of the overviewed technologies.

Tech.	Challenges	Research Directions
Light/	Energy-intensive manufacturing	Exploitation of new materials such as perovskite and organic PVs
Laser	Efficient operation	Multi-junction PVs, CPV
	LOS dependency	UAVs, supporting reflectors, hybrid RF and light-EH/laser-WET
Heat	Limited temperature variations	Transducer arrays, large surface area, materials engineering
	High thermal stress/strain	Exploitation of new materials for the contacts
MFC	Limited lifespan	Genetically engineered bacteria for improved activity and lifetime
	Poor cost-effectiveness	Cheap, yet durable, materials and stacked or multilayer MFCs
Vibration	Reduced resonant frequency	Frequency up-conversion and low-stiffness materials
	Frequency matching	Passive frequency tuning and bandwidth broadening schemes
Flow	Footprint and threat to wildlife	Compact designs, contained/vertical blades, bladeless generators
	Operation in variable wind conditions	Wind prediction, real-time optimization of cut-in/cut-off speed
RF-EH/	Very low ambient EH	Multi-antenna, metasurface-aided, and self-reconfigurable receivers
RF-WET	Low end-to-end efficiency	CSI-free beamforming, robotic WET, holistic optimization, distributed MIMO
Inductive	Sensitive to coils' misalignment	Efficient coils design and deployment
	Large form-factor	Meta-materials-aided design for EM field focusing
Capacitive	Requires high/variable reactive compensation	Active variable reactance rectifier
	Limiting fringing fields	Phased multi-plate charging
Acoustic	Limited charging distance/efficiency	Phased acoustic arrays, optimal deployment of acoustic transmitters
BC	Imperceptible integration to everyday objects	Exploitation of meta-materials
	High-frequency operation	Low-complexity protocols for beam search/exploitation
	Wide-band designs	Advanced, yet affordable, modulation schemes
	Frequency-agnostic designs	HW and algorithms for proper frequency tuning
	Security	Quantum backscattering mechanisms
	Enhanced RF-EH sensitivity	Advanced leakage power reduction and technology and voltage scaling
IRS/LIS	Low-cost/energy control	Trade-offs studies and low-cost/energy-effective control mechanisms
	CSI acquisition, especially for active IRS	Low-complexity/overhead training protocols
	Advanced metasurface implementations	SW-controlled techniques exploiting diverse metamaterial functions
	Data-driven optimization	Low-complexity/energy ML mechanisms, including TinyML
Radio	Costly signaling between antenna elements	Efficient distributed processing architectures
stripes	Advanced HW and protocols	Novel circuits/prototypes and optimized resource allocation
WuR	Radio resource management and scheduling	Characterization of EE, latency, reliability, and robustness trade-offs
	HW complexity and cost	Inband and RFIC-embedded WuR implementation
	Beamformed WuS at mm-waves	Beamforming design for WuS, how many beams in a single WuS burst?
O-RAN	Lack of results for envisioned goals	Practical and real-world trials on the virtualized RAN
MEC	Complexity due to coordination issues	Distributed coordination frameworks
	Heterogeneity and standardization	Edge-native frameworks to provide a unified and standardized approach to managing devices and data
	Security and privacy concerns	Secure communication protocols and data encryption techniques
	Accuracy & latency in real-time applications	Data pre-processing, caching, and filtering, QoS management
	Designing a green MEC	Dynamic right-sizing, GLB, and renewable energy-powered MEC systems
FL	Expensive communication & synchronism	Reducing communication rounds and vanishing gradients in each round
	Security/Privacy issues	Protection against malicious attacks and privacy-enhancing
	Robustness issues	Heterogeneous HW support and robust aggregation algorithms
	MDS	Efficient training and inference for massive heterogeneous networks
TinyML	Limited processing power of devices	Novel HW to reduce the computational burden of ML algorithm
	Noise sensitivity of in-memory computing	Error-correcting codes & advanced, yet affordable, signal-processing

Efforts to develop commercial ET applications are also substantial. For example, Wi-Charge[28] developed the

28. Please refer to https://www.wi-charge.com/ for more details.

laser-based AirCord™ technology, which aims at meeting the energy demands of industrial IoT, healthcare, gaming, and retail applications. Wi-Charge transmitters are hybrid nodes capable of transmitting energy and communicating

TABLE 15. Summary of key (sub-)system-level technical challenges and associated research directions.

Challenge	Research Directions	Tools
Quantifying the energy-sustainability grade of IoT ecosystems: IoT ecosystems are composed of heterogeneous subsystems and devices with complex dynamics/interactions. Also, energy sustainability comprises economic, societal, and environmental aspects, which must be captured and is especially challenging since both short and long-term perspectives must be considered.	A unifying quantitative metric for assessing the energy-sustainability grade of an IoT ecosystem may be unfeasible. Instead, key metrics separately for economic, societal, and environmental aspects could be used and they should be readily understandable by the general public, e.g., labels. Their expected evolution over time should be provided. The definition of such key quantities and measurement and forecasting procedures requires intensive interdisciplinary research. It should be up to the specific stakeholders, regulators, administrations, etc to establish comparisons and targets.	Energy-related performance metrics, including those in Section VI. Big-data collection and analysis, e.g., via ML. Life cycle assessment SW, e.g., SimaPro and OpenLCA. Economic and societal modeling tools.
Intelligent energy management: EC must be minimized without compromising QoS, calling for intelligent energy management mechanisms, which must cope with EH/EC variability; dynamic network topologies (since devices may frequently connect and disconnect); and real-time decision-making requirements.	Developing context-aware low-energy EH/EC predictive approaches (at network, edge, and device), tailored energy-efficient HW designs, and multi-EH transducers with efficient HW/SW-level integration. Developing energy-aware routing protocols and efficient self-organizing and self-healing network architectures. Characterizing cloud- vs edge- vs local-based decision-making trade-offs and proposing novel related protocols.	EH forecasting/simulation tools (cf. Section VII). Statistical and optimization tools, including ML. HW simulation/design tools, IoT-tailored operating systems & network simulators (cf. Section VII). Power profilers.
Network extreme-heterogeneity and scalability support: How to efficiently support the operation of the increasingly massive and heterogeneous (in terms of characteristics, supported protocols, and performance requirements) IoT devices?	Developing protocols that enable seamless interoperability among heterogeneous IoT devices (e.g., by strengthening MQTT, CoAP, and AMQP protocols). Developing mechanisms to distribute tasks/data processing among the devices (e.g., based on their energy levels, computational capabilities, and proximity to each other). Developing O-RAN functionalities that facilitate serving such extremely massive and heterogeneous IoT networks.	Protocol analyzers and stacks. IoT middleware platforms. IoT-tailored operating systems and network simulators (cf. Section VII). Game theory. Statistical and optimization tools, including ML (especially FL).
Distributed energy trading: Energy trading in microgrids is increasingly gaining traction. The current infrastructure-based model limits the flexibility and scalability of the technology, motivating distributed implementations. The latter are challenged by a greater coordination and interoperability complexity, difficulty in guaranteeing grid stability, reduced resilience, and data management and security issues.	Developing control algorithms that can handle many distributed energy resources and heterogeneous ET mechanisms (including WET), manage energy trading autonomously at the microgrid level, and scale efficiently. Designing multi-agent systems that can negotiate energy transactions, optimize energy flows, and adapt to dynamic supply/demand changes. DLT is promising for the latter, but further research is needed to reduce the large communication overhead and support massive two-way connections and EE. Developing and standardizing practical WET protocols.	Game theory. DLT tools. Statistical and optimization tools, including ML (e.g., RL, FL, and evolutionary algorithms). Graph theory and network analysis tools. EH forecasting/simulations tools (cf. Section VII). Quantum computing.
Low-power signal processing and task execution: Signal processing and task execution procedures directly affect the EC of IoT devices but are inevitably necessary for supporting sensing, communication, management, and interoperability capabilities and specific high-level applications. How to minimize their energy footprint without QoS degradation?	Integrating EH forecasting capabilities on computation offloading decision mechanisms and MEC systems. Developing novel techniques for optimizing TinyML platforms, e.g., based on bio-inspired optimization, spiking and analog NN, and in-memory computing. Developing baseline TinyML models and efficient (HW-aided) intermittent computing mechanisms. Developing intelligent energy management mechanisms (cf. 2nd row). Developing efficient context sensing techniques from EH patterns. Developing low-overhead protocols (cf. 3rd row).	EH forecasting/simulation tools (cf. Section VII). Game and control theory. Markov decision processes. MCU, cloud computing, and MEC (e.g., OpenFog and EdgeX) platforms. IoT middleware. ML (especially clustering, pattern recognition, and evolutionary algorithms).
Cross-layer design and holistic integration: Jointly optimizing different layers, technologies, and subsystems is crucial for realizing energy-sustainable IoT ecosystems. However, cross-layer designs and holistic integration require a deep understanding of the layers' interactions and trade-offs to preserve security and privacy, prevent unforeseen interactions, enable real-time adaptation, and mediate over conflicting objectives.	Characterizing system layers' interactions and potential overheads from implementing cross-layer optimizations. Developing adaptive and self-organizing mechanisms that allow IoT devices to adjust their behavior across layers based on changing conditions. Designing communication protocols that explicitly consider interactions and trade-offs between different layers. Exploring methods for achieving real-time adaptation across layers, including rapid decision-making mechanisms and efficient feedback loops to enable timely adjustments. Developing metrics and evaluation frameworks that capture the energy-sustainability benefits of cross-layer designs.	IoT-tailored operating systems & network simulators (cf. Section VII). Queueing and control theory. Protocol and network analysis tools. Statistical and optimization tools, including ML (especially RL and NN). MEC platforms (e.g., OpenFog and EdgeX). DLT tools.

with the devices via WiFi. The transmitters communicate through a cloud-based ML controller that allocates power to the different devices based on their power requirements, battery level, power availability, and customer-set priority. Similarly, Ossia® is working on an RF-WET/EH MIMO system. Ossia's Cota® Real Wireless Power[TM] technology

also relies on an ML-powered cloud service for user charging management.[29] Finally, a company named Emrod is aiming to solve some of the challenges faced in the current power distribution grid network via an RF-WET solution. Emrod envisions a system for transferring energy over long distances by utilizing a network of relays in the path between the transmitter and the receiver to guarantee LOS radio link. Each node is equipped with a massive antenna array to enhance coverage. Currently, the company is also targeting space-based energy applications that redirect the solar energy harvested in space to power earth applications via RF-WET and aims at using this technology as a means for connecting hard-to-reach renewable energy generators to the main power grid.

VIII. CONCLUSION & OUTLOOK

In this work, we provided valuable insights into energy-sustainable IoT and claimed it could only be supported by the harmonious coexistence of EP, ET, and EE processes. These processes refer to charging exploiting green energy sources, the intentional movement of energy from one device/system to another, and the ability of a device/system/process to perform its intended function with minimal energy. We overviewed the main technologies corresponding to these processes together with their use cases, recent advances, challenges, and research directions. Specifically, EH technologies based on light, heat, MFCs, vibration, flow, and RF were discussed within the EP processes. In the case of ET, the focus was on RF, inductive/capacitive coupling, laser, and acoustic-based technologies. Meanwhile, in terms of EE support, we focused on technologies enabling low-power communication, namely BC, metasurface-aided communication, radio stripes, and WuR, and also ML approaches aiming to reduce the EC burden of application tasks at the device/edge/network side. In general, the appropriateness of a specific technology/technique depends on the characteristics and performance requirements of the target application, and increasingly on their sustainability support level. This may be assessed by considering proper performance metrics. Indeed, we discussed relevant performance metrics to assess energy-sustainability potential and listed some relevant target values for specific technologies in the next generation of wireless systems. Moreover, we discussed protocol, integration, and implementation issues at the technology and system level.

Table 14 compiles a summary of the main challenges and corresponding research directions of all the discussed technologies, while Table 15 does the same but from a system-level perspective instead of technology-wise. Observe that although the focus of this work was on IoT energy-sustainability aspects during the operation phase, a truly self-sustainable ecosystem must consider sustainability aspects along the entire product lifecycle, i.e., planning, manufacturing, deployment, operation (including the target use case),

29. Please refer to https://www.ossia.com/ for more details.

maintenance, and disposal. The natural progression of our work is to expand in this direction.

Finally, policy and regulatory frameworks play a pivotal role in shaping the adoption and deployment of truly sustainable IoT solutions. Indeed, ensuring effective government policies, incentivizing sustainable manufacturing practices, and fostering international cooperation are essential. This is often challenging since policy and regulatory factors i) vary significantly across different regions, ii) are subject to frequent changes, and iii) can be context-specific and dependent on the specific application of IoT technologies. Additionally, the establishment of robust data privacy laws and standardized regulations must strike a balance between responsible IoT deployment and safeguarding user information. Going forward, it is imperative to address these policy and regulatory challenges and explore their intersection with technical advancements to promote the development of self-sustainable IoT ecosystems along the entire product lifecycle.

REFERENCES

[1] M. Munasinghe, *Environmental Economics and Sustainable Development*, vol. 3. New York, NY, USA: World Bank, 1993.

[2] A. A. Martins, T. M. Mata, C. A. Costa, and S. K. Sikdar, "Framework for sustainability metrics," *Ind. Eng. Chem. Res.*, vol. 46, no. 10, pp. 2962–2973, 2007.

[3] P.-M. Boulanger, "Sustainable development indicators: A scientific challenge, a democratic issue," *Surveys Perspect. Integr. Environ. Soc.*, vol. 1, no. 1, p. 9, 2008.

[4] F.-D. Vivien, "Sustainable development: An overview of economic proposals," *Surveys Perspectives Integr. Environ. Soc.*, vol. 1, no. 2, p. 16, 2008.

[5] W. R. Blackburn, *The Sustainability Handbook: The Complete Management Guide to Achieving Social, Economic and Environmental Responsibility*. London, U.K.: Routledge, 2012.

[6] C. Ratiu and B. B. Anderson, "The identity crisis of sustainable development," *World J. Sci. Technol. Sustain. Develop.*, vol. 11, no. 1, pp. 4–15, 2014.

[7] M. Mura, M. Longo, P. Micheli, and D. Bolzani, "The evolution of sustainability measurement research," *Int. J. Manag. Rev.*, vol. 20, no. 3, pp. 661–695, 2018.

[8] B. Corona, L. Shen, D. Reike, J. R. Carreón, and E. Worrell, "Towards sustainable development through the circular economy—A review and critical assessment on current circularity metrics," *Resource Conservation Recycling*, vol. 151, Dec. 2019, Art. no. 104498.

[9] J. Walzberg, G. Lonca, R. J. Hanes, A. L. Eberle, A. Carpenter, and G. A. Heath, "Do we need a new sustainability assessment method for the circular economy? A critical literature review," *Front. Sustain.*, vol. 1, Jan. 2021, Art. no. 620047.

[10] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, 4th Quart., 2017.

[11] Y. Meng, Y. Yang, H. Chung, P.-H. Lee, and C. Shao, "Enhancing sustainability and energy efficiency in smart factories: A review," *Sustainability*, vol. 10, no. 12, p. 4779, 2018.

[12] S. H. Alsamhi, O. Ma, M. S. Ansari, and Q. Meng, "Greening Internet of Things for greener and smarter cities: A survey and future prospects," *Telecommun. Syst.*, vol. 72, pp. 609–632, Jul. 2019.

[13] S. Nižetić et al., "Internet of Things (IoT): Opportunities, issues and challenges towards a smart and sustainable future," *J. Clean. Prod.*, vol. 274, Nov. 2020, Art. no. 122877.

[14] N. H. Motlagh, M. Mohammadrezaei, J. Hunt, and B. Zakeri, "Internet of Things (IoT) and the energy sector," *Energies*, vol. 13, no. 2, p. 494, 2020.

[15] S. Hu, X. Chen, W. Ni, X. Wang, and E. Hossain, "Modeling and analysis of energy harvesting and smart grid-powered wireless communication networks: A contemporary survey," *IEEE Trans. Green Commun. Netw.*, vol. 4, no. 2, pp. 461–496, Jun. 2020.

[16] O. Cetinkaya, D. Balsamo, and G. V. Merrett, "Internet of MIMO Things: UAV-assisted wireless-powered networks for future smart cities," *IEEE Internet Things Mag.*, vol. 3, no. 1, pp. 8–13, Mar. 2020.

[17] M. Matinmikko-Blue et al. "White paper on 6G drivers and the UN SDGs." 2020. [Online]. Available: http://jultika.oulu.fi/files/isbn9789526226699.pdf

[18] N. Mahmood et al., "White paper on critical and massive machine type communication towards 6G [white paper]," 2020. [Online]. Available: https://arxiv.org/abs/2004.14146

[19] N. H. Mahmood, H. Alves, O. L. A. López, M. Shehab, D. P. M. Osorio, and M. Latva-Aho, "Six key features of machine type communication in 6G," in *Proc. 2nd 6G Wireless Summit (6G SUMMIT)*, 2020, pp. 1–5.

[20] N. H. Mahmood et al., "Machine type communications: Key drivers and enablers towards the 6G era," *EURASIP J. Wireless Commun. Netw.*, vol. 2021, no. 1, pp. 1–25, 2021.

[21] H. Alves and O. L. A. López, *Wireless RF Energy Transfer in the Massive IoT Era: Towards Sustainable Zero-Energy Networks.* Hoboken, NJ, USA: Wiley, 2022.

[22] O. L. A. López, H. Alves, R. D. Souza, S. Montejo-Sánchez, E. M. G. Fernández, and M. Latva-Aho, "Massive wireless energy transfer: Enabling sustainable IoT toward 6G era," *IEEE Internet Things J.*, vol. 8, no. 11, pp. 8816–8835, Jun. 2021.

[23] M. A. Albreem, A. M. Sheikh, M. H. Alsharif, M. Jusoh, and M. N. Mohd Yasin, "Green Internet of Things (GIoT): Applications, practices, awareness, and challenges," *IEEE Access*, vol. 9, pp. 38833–38858, 2021.

[24] P. Bradu et al., "Recent advances in green technology and industrial revolution 4.0 for a sustainable future," *Environ. Sci. Pollution Res.*, to be published.

[25] A. Sabovic, M. Aernouts, D. Subotic, J. Fontaine, E. De Poorter, and J. Famaey, "Towards energy-aware tinyML on battery-less IoT devices," *Internet Things*, vol. 22, Jul. 2023, Art. no. 100736.

[26] H. Rahmani et al., "Next-generation IoT devices: Sustainable eco-friendly manufacturing, energy harvesting, and wireless connectivity," *IEEE J. Microw.*, vol. 3, no. 1, pp. 237–255, Jan. 2023.

[27] R. Venkatesan and A. Tamilvanan, "A sustainable agricultural system using IoT," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, 2017, pp. 763–767.

[28] A. Maroli, V. S. Narwane, and B. B. Gardas, "Applications of IoT for achieving sustainability in agricultural sector: A comprehensive review," *J. Environ. Manag.*, vol. 298, Nov. 2021, Art. no. 113488.

[29] N. Khan, R. L. Ray, G. R. Sargani, M. Ihtisham, M. Khayyam, and S. Ismail, "Current progress and future prospects of agriculture technology: Gateway to sustainable agriculture," *Sustainability*, vol. 13, no. 9, p. 4883, 2021.

[30] E. M. Grua, M. De Sanctis, I. Malavolta, M. Hoogendoorn, and P. Lago, "Social sustainability in the e-health domain via personalized and self-adaptive mobile apps," in *Software Sustainability*. Cham, Switzerland: Springer Int., 2021, pp. 301–328.

[31] J. Zhao, X. Zheng, R. Dong, and G. Shao, "The planning, construction, and management toward sustainable cities in China needs the environmental Internet of Things," *Int. J. Sustain. Develop. World Ecol.*, vol. 20, no. 3, pp. 195–198, 2013.

[32] L. Belli et al., "IoT-enabled smart sustainable cities: Challenges and approaches," *Smart Cities*, vol. 3, no. 3, pp. 1039–1071, 2020.

[33] A. K. Jha, A. Ghimire, S. Thapa, A. M. Jha, and R. Raj, "A review of AI for urban planning: Towards building sustainable smart cities," in *Proc. 6th Int. Conf. Inventive Comput. Technol. (ICICT)*, 2021, pp. 937–944.

[34] S. Narendra, P. Pradeep, and M. V. Ramesh, "An Internet of Things (IoT) based sustainable water management," in *Proc. IEEE Global Humanitarian Technol. Conf. (GHTC)*, 2017, pp. 1–6.

[35] E. Aivazidou et al., "Smart technologies for sustainable water management: An urban analysis," *Sustainability*, vol. 13, no. 24, 2021, Art. no. 13940.

[36] S. Butt, M. Ahmadi, and M. Razavi, "Green computing: Sustainable design and technologies," in *Proc. 5th Int. Conf. Innov. Technol. Intell. Syst. Ind. Appl. (CITISIA)*, 2020, pp. 1–7.

[37] K. M. Law, A. W. Ip, B. B. Gupta, and S. Geng, *Managing IoT and Mobile Technologies With Innovation, Trust, and Sustainable Computing.* Boca Raton, FL, USA: CRC Press, 2021.

[38] M. Jasiulewicz-Kaczmarek and A. Gola, "Maintenance 4.0 technologies for sustainable manufacturing—An overview," *IFAC PapersOnLine*, vol. 52, no. 10, pp. 91–96, 2019.

[39] S. Keivanpour and D. A. Kadi, "Internet of Things enabled real-time sustainable end-of-life product recovery," *IFAC PapersOnLine*, vol. 52, no. 13, pp. 796–801, 2019.

[40] M. Jasiulewicz-Kaczmarek, S. Legutko, and P. Kluk, "Maintenance 4.0 technologies—New opportunities for sustainability driven maintenance," *Manage. Prod. Eng. Rev.*, vol. 52, no. 10, pp. 91–96, 2020.

[41] K. S. Kumar, T. A. Kumar, S. Sundaresan, and V. K. Kumar, "Green IoT for sustainable growth and energy management in smart cities," in *Handbook of Green Engineering Technologies for Sustainable Smart Cities*. Boca Raton, FL, USA: CRC Press, 2021, pp. 155–172.

[42] P. Vrignat, F. Kratz, and M. Avila, "Sustainable manufacturing, maintenance policies, prognostics and health management: A literature review," *Rel. Eng. Syst. Safety*, vol. 218, Feb. 2022, Art. no. 108140.

[43] M. M. Razip et al., "The development of sustainable IoT E-waste management guideline for households," *Chemosphere*, vol. 303, Sep. 2022, Art. no. 134767.

[44] *Data Centres and Data Transmission Networks*, Int. Energy Agency, Paris, France, 2022.

[45] C. Freitag, M. Berners-Lee, K. Widdicks, B. Knowles, G. S. Blair, and A. Friday, "The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations," *Patterns*, vol. 2, no. 9, 2021, Art. no. 100340.

[46] S. Wankhede, S. Sachan, and V. Kumar, *Wireless Power Transmission Market by Technology, Type, and Application: Global Opportunity Analysis and Industry Forecast, 2021-2030*. London, U.K.: Allied Market Res., 2021.

[47] F. Sioshansi, *Energy Efficiency: Towards the End of Demand Growth*. Waltham, MA, USA: Academic, 2013.

[48] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, 2021.

[49] T. Sanislav, G. D. Mois, S. Zeadally, and S. C. Folea, "Energy harvesting techniques for internet of things (IoT)," *IEEE Access*, vol. 9, pp. 39530–39549, 2021.

[50] A. Shore, J. Roller, J. Bergeson, and B. H. Hamadani, "Indoor light energy harvesting for battery-powered sensors using small photovoltaic modules," *Energy Sci. Eng.*, vol. 9, no. 11, pp. 2036–2043, 2021.

[51] D.-T. Phan-Huy, D. Barthel, P. Ratajczak, R. Fara, M. d. Renzo, and J. de Rosny, "Ambient backscatter communications in mobile networks: Crowd-detectable zero-energy-devices," *IEEE J. Radio Freq. Identification*, vol. 6, pp. 660–670, 2022.

[52] Y. Liu, D. Li, B. Du, L. Shu, and G. Han, "Rethinking sustainable sensing in agricultural Internet of Things: From power supply perspective," *IEEE Wireless Commun.*, vol. 29, no. 4, pp. 102–109, Aug. 2022.

[53] M. Yamaguchi, F. Dimroth, J. F. Geisz, and N. J. Ekins-Daukes, "Multi-junction solar cells paving the way for super high-efficiency," *J. Appl. Phys.*, vol. 129, no. 24, 2021, Art. no. 240901.

[54] N. Torabi et al., "Progress and challenges in perovskite photovoltaics from single-to multi-junction cells," *Mater. Today Energy*, vol. 12, pp. 70–94, Jun. 2019.

[55] M. Azadinia, M. Ameri, R. T. Ghahrizjani, and M. Fathollahi, "Maximizing the performance of single and multijunction MA and lead-free perovskite solar cell," *Mater. Today Energy*, vol. 20, Jun. 2021, Art. no. 100647.

[56] N. Vaidya and O. Solgaard, "Immersion graded index optics: Theory, design, and prototypes," *Microsyst. Nanoeng.*, vol. 8, no. 1, p. 69, 2022.

[57] P. Soonsawad, K. E. Jeon, and J. She, "Improved energy harvesting with one-time adjusted solar panel for BLE beacon," in *Proc. IEEE 93rd Veh. Technol. Conf. (VTC-Spring)*, 2021, pp. 1–5.

[58] M. Alves, A. Pérez-Rodríguez, P. J. Dale, C. Domínguez, and S. Sadewasser, "Thin-film micro-concentrator solar cells," *J. Phys. Energy*, vol. 2, no. 1, 2019, Art. no. 12001.

[59] D. Sato et al., "Stretchable micro-scale concentrator photovoltaic module with 15% efficiency for three-dimensional curved surfaces," *Commun. Mater.*, vol. 2, no. 1, p. 7, 2021.

[60] X. Li, P. Li, Z. Wu, D. Luo, H.-Y. Yu, and Z.-H. Lu, "Review and perspective of materials for flexible solar cells," *Mater. Rep. Energy*, vol. 1, no. 1, 2021, Art. no. 100001.

[61] A. Perera, M. Katz, R. Godaliyadda, J. Häkkinen, and E. Strömmer, "Light-based Internet of Things: Implementation of an optically connected energy-autonomous node," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2021, pp. 1–7.

[62] G. Wang, M. A. Adil, J. Zhang, and Z. Wei, "Large-area organic solar cells: Material requirements, modular designs, and printing methods," *Adv. Mater.*, vol. 31, no. 45, 2019, Art. no. 1805089.

[63] A. J. Williams, M. F. Torquato, I. M. Cameron, A. A. Fahmy, and J. Sienz, "Survey of energy harvesting technologies for wireless sensor networks," *IEEE Access*, vol. 9, pp. 77493–77510, 2021.

[64] S. Zhu et al., "Review on wearable thermoelectric generators: From devices to applications," *Energies*, vol. 15, no. 9, p. 3375, 2022.

[65] N. Wang, J.-N. Zhang, H. Ni, H.-Z. Jia, and C. Ding, "Improved MPPT system based on FTSMC for thermoelectric generator array under dynamic temperature and impedance," *IEEE Trans. Ind. Electron.*, vol. 69, no. 10, pp. 10715–10723, Oct. 2022.

[66] R. A. Kishore, "Harvesting thermal energy with ferroelectric materials," in *Ferroelectric Materials for Energy Harvesting and Storage*. Amsterdam, The Netherlands: Elsevier, 2021, pp. 85–106.

[67] R. Rodriguez, M. Preindl, J. S. Cotton, and A. Emadi, "Review and trends of thermoelectric generator heat recovery in automotive applications," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 5366–5378, Jun. 2019.

[68] A. Nozariasbmarz et al., "Review of wearable thermoelectric energy harvesting: From body temperature to electronic systems," *Appl. Energy*, vol. 258, Jan. 2020, Art. no. 114069.

[69] M. D. Albert, K. O. Bennett, C. A. Adams, and J. G. Gluyas, "Waste heat mapping: A U.K. study," *Renew. Sustain. Energy Rev.*, vol. 160, May 2022, Art. no. 112230.

[70] G. Palanisamy, H.-Y. Jung, T. Sadhasivam, M. D. Kurkuri, S. C. Kim, and S.-H. Roh, "A comprehensive review on microbial fuel cell technologies: Processes, utilization, and advanced developments in electrodes and membranes," *J. Cleaner Prod.*, vol. 221, pp. 598–621, Jun. 2019.

[71] S. Maddalwar, K. K. Nayak, M. Kumar, and L. Singh, "Plant microbial fuel cell: Opportunities, challenges, and prospects," *Bioresource Technol.*, vol. 341, Oct. 2021, Art. no. 125772.

[72] M. Ramya and P. S. Kumar, "A review on recent advancements in bioenergy production using microbial fuel cells," *Chemosphere*, vol. 288, no. 2, 2022, Art. no. 132512.

[73] A. A. Yaqoob, M. N. M. Ibrahim, and S. Rodríguez-Couto, "Development and modification of materials to build cost-effective anodes for microbial fuel cells (MFCs): An overview," *Biochem. Eng. J.*, vol. 164, Dec. 2020, Art. no. 107779.

[74] A. A. Yaqoob, M. N. M. Ibrahim, and C. Guerrero-Barajas, "Modern trend of anodes in microbial fuel cells (MFCs): An overview," *Environ. Technol. Innov.*, vol. 23, Aug. 2021, Art. no. 101579.

[75] M. Shirvanimoghaddam et al., "Towards a green and self-powered Internet of Things using piezoelectric energy harvesting," *IEEE Access*, vol. 7, pp. 94533–94556, 2019.

[76] S. Khalid, I. Raouf, A. Khan, N. Kim, and H. S. Kim, "A review of human-powered energy harvesting for smart electronics: Recent progress and challenges," *Int. J. Precis. Eng. Manuf. Green Technol.*, vol. 6, no. 4, pp. 821–851, 2019.

[77] M. Li, A. Luo, W. Luo, and F. Wang, "Recent progress on mechanical optimization of MEMS electret-based electrostatic vibration energy harvesters," *J. Microelectromech. Syst.*, vol. 31, no. 5, pp. 726–740, 2022.

[78] Y. Li, K. Tao, B. George, and Z. Tan, "Harvesting vibration energy: Technologies and challenges," *IEEE Ind. Electron. Mag.*, vol. 15, no. 1, pp. 30–39, Mar. 2021.

[79] Y. Li, C. Zhou, X. Wang, J. Wang, D. Qiao, and K. Tao, "A vibration energy harvester with targeted frequency-tuning capability," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022.

[80] K. Wang, W. Liu, Y. Tang, J. Pei, S. Kang, and Z. Wu, "Widening the bandwidth of vibration energy harvester by automatically tracking the resonant frequency with magnetic sliders," *Sustain. Energy Technol. Assess.*, vol. 58, Aug. 2023, Art. no. 103368.

[81] A. Morel, D. Gibus, G. Pillonnet, and A. Badel, "Opportunities for electrically-based frequency tuning of piezoelectric vibration energy harvesters," in *Proc. IEEE 20th Int. Conf. Micro Nanotechnol. Power Gener. Energy Convers. Appl. (PowerMEMS)*, 2021, pp. 236–239.

[82] Y. Feng, Z. Zhou, H. Luo, R. Wang, Y. Han, and Y. Xiong, "Frequency-tunable resonant hybrid vibration energy harvester using a piezoelectric cantilever with electret-based electrostatic coupling," *IET Nanodielectrics*, vol. 6, no. 2, pp. 46–56, 2023.

[83] N. Kumari and M. Rakotondrabe, "Design and development of a lead-freepiezoelectric energy harvester for wideband, low frequency, and low amplitude vibrations," *Micromachines*, vol. 12, no. 12, p. 1537, 2021.

[84] X. Li, G. Hu, Z. Guo, J. Wang, Y. Yang, and J. Liang, "Frequency up-conversion for vibration energy harvesting: A review," *Symmetry*, vol. 14, no. 3, p. 631, 2022.

[85] N. Jackson, "Tuning and widening the bandwidth of vibration energy harvesters using a ferrofluid embedded mass," *Microsyst. Technol.*, vol. 26, no. 6, pp. 2043–2051, 2020.

[86] H. Zhao, X. Wei, Y. Zhong, and P. Wang, "A direction self-tuning two-dimensional piezoelectric vibration energy harvester," *Sensors*, vol. 20, no. 1, p. 77, 2019.

[87] G. Shi et al., "A piezo-electromagnetic coupling multi-directional vibration energy harvester based on frequency up-conversion technique," *Micromachines*, vol. 11, no. 1, p. 80, 2020.

[88] Z. Zhao, D. Wang, T. Wang, W. Shen, H. Liu, and M. Chen, "A review: Approaches for aerodynamic performance improvement of lift-type vertical axis wind turbine," *Sustain. Energy Technol. Assess.*, vol. 49, Feb. 2022, Art. no. 101789.

[89] M. Ebrahimpour, R. Shafaghat, R. Alamian, and M. Safdari Shadloo, "Numerical investigation of the savonius vertical axis wind turbine and evaluation of the effect of the overlap parameter in both horizontal and vertical directions on its performance," *Symmetry*, vol. 11, no. 6, p. 821, 2019.

[90] D. Y. Villarreal, *VIV Resonant Wind Generators*, Vortex Bladeless SL, Madrid, Spain, 2018.

[91] O. L. A. López, B. Clerckx, and M. Latva-Aho, "Dynamic RF combining for multi-antenna ambient energy harvesting," *IEEE Wireless Commun. Lett.*, vol. 11, no. 3, pp. 493–497, Mar. 2022.

[92] M. Song et al., "Up-and-coming physical concepts of wireless power transfer," 2021, *arXiv:2107.14086*.

[93] J. Zhou, P. Zhang, J. Han, L. Li, and Y. Huang, "Metamaterials and metasurfaces for wireless power transfer and energy harvesting," *Proc. IEEE*, vol. 110, no. 1, pp. 31–55, Jan. 2022.

[94] W. Wang, Q. Zhang, H. Lin, M. Liu, X. Liang, and Q. Liu, "Wireless energy transmission channel modeling in resonant beam charging for IoT devices," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3976–3986, Apr. 2019.

[95] K. Jin and W. Zhou, "Wireless laser power transmission: A review of recent progress," *IEEE Trans. Power Electron.*, vol. 34, no. 4, pp. 3842–3859, Apr. 2019.

[96] M. Qaraqe, M. Usman, A. Serbes, I. S. Ansari, and M.-S. Alouini, "Power hotspots in space: Powering CubeSats via inter-satellite optical wireless power transfer," *IEEE Internet Things Mag.*, vol. 5, no. 3, pp. 180–185, Sep. 2022.

[97] M.-A. Lahmeri, M. A. Kishk, and M.-S. Alouini, "Charging techniques for UAV-assisted data collection: Is laser power beaming the answer?" *IEEE Commun. Mag.*, vol. 60, no. 5, pp. 50–56, May 2022.

[98] P. Jaffe, "Spooky power at a distance: Researchers have beamed substantial amounts of energy at distances over 1 kilometer," *IEEE Spectr.*, vol. 59, no. 6, pp. 36–41, Jun. 2022.

[99] Q. Liu, M. Xiong, M. Liu, Q. Jiang, W. Fang, and Y. Bai, "Charging a smartphone over the air: The resonant beam charging method," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 13876–13885, Aug. 2022.

[100] M. Xiong, Q. Liu, G. Wang, G. B. Giannakis, and C. Huang, "Resonant beam communications: Principles and designs," *IEEE Commun. Mag.*, vol. 57, no. 10, pp. 34–39, Oct. 2019.

[101] M. Xiong, M. Liu, Q. Zhang, Q. Liu, J. Wu, and P. Xia, "TDMA in adaptive resonant beam charging for IoT devices," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 867–877, Feb. 2019.

[102] M. Liu, S. Xia, M. Xiong, M. Xu, Q. Liu, and H. Deng, "NLOS transmission analysis for mobile SLIPT using resonant beam," *IEEE Trans. Wireless Commun.*, early access, May 24, 2023, doi: 10.1109/TWC.2023.3277593.

[103] A. Allam, K. G. Sabra, and A. Erturk, "Aspect ratio-dependent dynamics of piezoelectric transducers in wireless acoustic power transfer," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 5, pp. 984–996, May 2020.

[104] V. F.-G. Tseng et al., "Self-detachable through-metal acoustic wireless power transfer," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 69, no. 7, pp. 2381–2389, Jul. 2022.

[105] V. F.-G. Tseng, S. S. Bedair, and N. Lazarus, "Phased array focusing for acoustic wireless power transfer," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 65, no. 1, pp. 39–49, Jan. 2018.

[106] K. Kim, S. G. Jang, H. G. Lim, H. H. Kim, and S.-M. Park, "Acoustic power transfer using self-focused transducers for miniaturized implantable neurostimulators," *IEEE Access*, vol. 9, pp. 153850–153862, 2021.

[107] V. F.-G. Tseng, T. Kiebala, D. Bruno, B. Novick, N. Lazarus, and S. S. Bedair, "Wireless power transmission using acoustic-to-inductive relayed transfer," in *Proc. 21st Int. Conf. Micro Nanotechnol. Power Gener. Energy Convers. Appl. (PowerMEMS)*, 2022, pp. 98–101.

[108] X. Tang, M. Sameer, and S. Mandal, "Acoustic wireless power and data telemetry for structural health monitoring," in *Proc. IEEE SENSORS*, 2018, pp. 1–4.

[109] V. F.-G. Tseng, S. S. Bedair, and N. Lazarus, "Acoustic wireless power transfer with receiver array for enhanced performance," in *Proc. IEEE Wireless Power Transfer Conf. (WPTC)*, 2017, pp. 1–4.

[110] Z. Kashani, S. J. Ilham, and M. Kiani, "Design and optimization of ultrasonic links with phased arrays for wireless power transmission to biomedical implants," *IEEE Trans. Biomed. Circuits Syst.*, vol. 16, no. 1, pp. 64–78, Feb. 2022.

[111] V. F.-G. Tseng, S. S. Bedair, J. J. Radice, T. E. Drummond, and N. Lazarus, "Ultrasonic lamb waves for wireless power transfer," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 3, pp. 664–670, Mar. 2020.

[112] M. L. Wang, A. Singhvi, G. Nyikayaramba, B. Murmann, and A. Arbabian, "Adaptive beamforming for wireless powering of a network of ultrasonic implants," in *Proc. IEEE Int. Ultrasonics Symp. (IUS)*, 2022, pp. 1–4.

[113] H.-J. Kim, H. Hirayama, S. Kim, K. J. Han, R. Zhang, and J.-W. Choi, "Review of near-field wireless power and communication for biomedical applications," *IEEE Access*, vol. 5, pp. 21264–21285, 2017.

[114] M. R. V. Moghadam and R. Zhang, "Multiuser wireless power transfer via magnetic resonant coupling: Performance analysis, charging control, and power region characterization," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 1, pp. 72–83, Mar. 2016.

[115] M. R. V. Moghadam and R. Zhang, "Node placement and distributed magnetic beamforming optimization for wireless power transfer," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 2, pp. 264–279, Jun. 2018.

[116] Y. Zhang, D. Chen, and T. Jiang, "Robust beamforming design for magnetic MIMO wireless power transfer systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 5066–5077, 2021.

[117] Y. Liu, T. Wu, and M. Fu, "Interleaved capacitive coupler for wireless power transfer," *IEEE Trans. Power Electron.*, vol. 36, no. 12, pp. 13526–13535, Dec. 2021.

[118] S. Sinha, A. Kumar, B. Regensburger, and K. K. Afridi, "Design of high-efficiency matching networks for capacitive wireless power transfer systems," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 10, no. 1, pp. 104–127, Feb. 2022.

[119] Y. Hou, S. Sinha, and K. Afridi, "Tunable multistage matching network for capacitive wireless power transfer system," in *Proc. IEEE Wireless Power Transf. Conf. (WPTC)*, 2021, pp. 1–4.

[120] M. Z. Erel, K. C. Bayindir, M. T. Aydemir, S. K. Chaudhary, and J. M. Guerrero, "A comprehensive review on wireless capacitive power transfer technology: Fundamentals and applications," *IEEE Access*, vol. 10, pp. 3116–3143, 2022.

[121] C. Cai, X. Liu, S. Wu, X. Chen, W. Chai, and S. Yang, "A misalignment tolerance and lightweight wireless charging system via reconfigurable capacitive coupling for unmanned aerial vehicle applications," *IEEE Trans. Power Electron.*, vol. 38, no. 1, pp. 22–26, Jan. 2023.

[122] C. Buyle, B. Cox, L. V. der Perre, and L. De Strycker, "A multi-band solution for interacting with energy-neutral devices," in *Proc. 55th Asilomar Conf. Signals Syst. Comput.*, 2021, pp. 329–333.

[123] O. L. A. López, H. Alves, S. Montejo-Sánchez, R. D. Souza, and M. Latva-Aho, "CSI-free rotary antenna beamforming for massive RF wireless energy transfer," *IEEE Internet Things J.*, vol. 9, no. 10, pp. 7375–7387, May 2022.

[124] O. L. A. López, F. A. Monteiro, H. Alves, R. Zhang, and M. Latva-Aho, "A low-complexity beamforming design for multiuser wireless energy transfer," *IEEE Wireless Commun. Lett.*, vol. 10, no. 1, pp. 58–62, Jan. 2021.

[125] S. Shen, J. Kim, and B. Clerckx, "Closed-loop wireless power transfer with adaptive waveform and beamforming: Design, prototype, and experiment," *IEEE J. Microw.*, vol. 3, no. 1, pp. 29–42, Jan. 2023.

[126] O. M. Rosabal, O. L. A. López, H. Alves, R. D. Souza, and S. Montejo-Sánchez, "Massive wireless energy transfer with multiple power beacons for very large Internet of Things," in *Proc. Int. Workshop Very Large Internet Things (VLIoT)*, 2021, pp. 1–5.

[127] O. M. Rosabal, O. L. A. López, H. Alves, S. Montejo-Sánchez, and M. Latva-Aho, "On the optimal deployment of power beacons for massive wireless energy transfer," *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10531–10542, Jul. 2021.

[128] H. Dai et al., "Area charging for wireless rechargeable sensors," in *Proc. 29th Int. Conf. Comput. Commun. Netw. (ICCCN)*, 2020, pp. 1–9.

[129] X. Yuan, X. Jiang, Y. Hu, and A. Schmeink, "Joint analog beamforming and trajectory planning for energy-efficient UAV-enabled nonlinear wireless power transfer," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 10, pp. 2914–2929, Oct. 2022.

[130] H. Zhang, N. Shlezinger, F. Guidi, D. Dardari, M. F. Imani, and Y. C. Eldar, "Near-field wireless power transfer for 6G internet of everything mobile networks: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 60, no. 3, pp. 12–18, Mar. 2022.

[131] H. Zhang, N. Shlezinger, F. Guidi, D. Dardari, M. F. Imani, and Y. C. Eldar, "Beam focusing for near-field multiuser MIMO communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7476–7490, Sep. 2022.

[132] C. T. Rodenbeck et al., "Microwave and millimeter wave power beaming," *IEEE J. Microw.*, vol. 1, no. 1, pp. 229–259, Jan. 2021.

[133] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Designing multi-user MIMO for energy efficiency: When is massive MIMO the answer?" in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2014, pp. 242–247.

[134] K. N. R. S. V. Prasad, E. Hossain, and V. K. Bhargava, "Energy efficiency in massive MIMO-based 5G networks: Opportunities and challenges," *IEEE Wireless Commun.*, vol. 24, no. 3, pp. 86–94, Jun. 2017.

[135] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Mobile Internet of Things: Can UAVs provide an energy-efficient mobile architecture?" in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2016, pp. 1–6.

[136] X. Lin, G. Su, B. Chen, H. Wang, and M. Dai, "Striking a balance between system throughput and energy efficiency for UAV-IoT systems," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10519–10533, Dec. 2019.

[137] O. Martínez Rosabal, O. L. A. López, D. E. Pérez, M. Shehab, H. Hilleshein, and H. Alves, "Minimization of the worst case average energy consumption in UAV-assisted IoT networks," *IEEE Internet Things J.*, vol. 9, no. 17, pp. 15827–15838, Sep. 2022.

[138] A. Kapovits et al., "Satellite communications integration with terrestrial networks," *China Commun.*, vol. 15, no. 8, pp. 22–38, 2018.

[139] F. A. Tondo, V. D. P. Souto, O. L. Alcaraz López, S. Montejo-Sánchez, S. Céspedes, and R. D. Souza, "Optimal traffic load allocation for aloha-based IoT LEO constellations," *IEEE Sensors J.*, vol. 23, no. 3, pp. 3270–3282, Feb. 2023.

[140] C. You and K. Huang, "Exploiting non-causal CPU-state information for energy-efficient mobile cooperative computing," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4104–4117, Jun. 2018.

[141] O. L. Alcaraz López, H. Alves, and M. Latva-aho, "Joint power control and rate allocation enabling ultra-reliability and energy efficiency in SIMO wireless networks," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5768–5782, Aug. 2019.

[142] O. L. Alcaraz López, N. H. Mahmood, and H. Alves, "Enabling URLLC for low-cost IoT devices via diversity combining schemes," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2020, pp. 1–6.

[143] I. F. Akyildiz, M. Pierobon, and S. Balasubramaniam, "Moving forward with molecular communication: From theory to human health applications [point of view]," *Proc. IEEE*, vol. 107, no. 5, pp. 858–865, May 2019.

[144] J. Wiklund et al., "A review on printed electronics: Fabrication methods, inks, substrates, applications and environmental impacts," *J. Manuf. Mater. Process.*, vol. 5, no. 3, p. 89, 2021.

[145] H. D. Griffiths and C. J. Baker, *An Introduction to Passive Radar*. London, U.K.: Artech House, 2022.

[146] K. Lin et al., "Throughput optimization in backscatter-assisted wireless-powered underground sensor networks for smart agriculture," *Internet Things*, vol. 20, Nov. 2022, Art. no. 100637.

[147] V. Talla, B. Kellogg, S. Gollakota, and J. R. Smith, "Battery-free cellphone," *ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 1, no. 2, pp. 1–20, 2017.

[148] R. Reed, F. L. Pour, and D. S. Ha, "An energy efficient RF backscatter modulator for IoT applications," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2021, pp. 1–5.

[149] T. Soyata, L. Copeland, and W. Heinzelman, "RF energy harvesting for embedded systems: A survey of tradeoffs and methodology," *IEEE Circuits Syst. Mag.*, vol. 16, no. 1, pp. 22–57, Aug. 2016.

[150] A. Varshney, O. Harms, C. Pérez-Penichet, C. Rohner, F. Hermans, and T. Voigt, "LoRea: A backscatter architecture that achieves a long communication range," in *Proc. 15th ACM Conf. Embedded Netw. Sensor Syst.*, 2017, pp. 1–14.

[151] V. Talla, M. Hessar, B. Kellogg, A. Najafi, J. R. Smith, and S. Gollakota, "LoRa backscatter: Enabling the vision of ubiquitous connectivity," *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 1–24, 2017.

[152] A. Wang, V. Iyer, V. Talla, J. R. Smith, and S. Gollakota, "FM backscatter: Enabling connected cities and smart fabrics," in *Proc. 14th USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, 2017, pp. 243–258.

[153] S. N. Daskalakis, G. Goussetis, S. D. Assimonis, M. M. Tentzeris, and A. Georgiadis, "A uW backscatter-morse-leaf sensor for low-power agricultural wireless sensor networks," *IEEE Sensors J.*, vol. 18, no. 19, pp. 7889–7898, Oct. 2018.

[154] J. Zhao, W. Gong, and J. Liu, "Spatial stream backscatter using commodity WiFi," in *Proc. 16th Annu. Int. Conf. Mobile Syst. Appl. Services (MobiSys)*, 2018, pp. 191–203.

[155] J. Ryoo, Y. Karimi, A. Athalye, M. Stanaćević, S. R. Das, and P. Djurić, "BARNET: Towards activity recognition using passive backscattering tag-to-tag network," in *Proc. 16th Annu. Int. Conf. Mobile Syst. Appl. Services*, 2018, pp. 414–427.

[156] Y. Peng et al., "PLoRa: A passive long-range data network from ambient LoRa transmissions," in *Proc. Conf. ACM Special Interest Group Data Commun. (SIGCOMM)*, 2018, pp. 147–160.

[157] A. Varshney, A. Soleiman, and T. Voigt, "TunnelScatter: Low power communication for sensor tags using tunnel diodes," in *Proc. 25th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2019, pp. 1–17.

[158] C. Zhao et al., "RF-Mehndi: A fingertip profiled RF identifier," in *Proc. IEEE Conf. Comput. Commun.*, 2019, pp. 1513–1521.

[159] A. Dadkhah, J. Rosenthal, and M. S. Reynolds, "ZeroScatter: Zero-added-component backscatter communication using existing digital i/O pins," in *Proc. IEEE Topical Conf. Wireless Sensors Sensor Netw. (WiSNet)*, 2019, pp. 1–3.

[160] P.-H. P. Wang, C. Zhang, H. Yang, M. Dunna, D. Bharadia, and P. P. Mercier, "A low-power backscatter modulation system communicating across tens of Meters with standards-compliant Wi-Fi transceivers," *IEEE J. Solid-State Circuits*, vol. 55, no. 11, pp. 2959–2969, Nov. 2020.

[161] M. Katanbaf, V. Jain, and J. R. Smith, "Relacks: Reliable backscatter communication in indoor environments," *ACM Interact. Mobile Wearable Ubiquitous Technol.*, vol. 4, no. 2, pp. 1–24, 2020.

[162] F. Zhu, Y. Feng, Q. Li, X. Tian, and X. Wang, "DigiScatter: Efficiently prototyping large-scale OFDMA backscatter networks," in *Proc. 18th Int. Conf. Mobile Syst. Appl. Services (MobiSys)*, 2020, pp. 42–53.

[163] M. Dunna, M. Meng, P.-H. Wang, C. Zhang, P. Mercier, and D. Bharadia, "SyncScatter: Enabling WiFi like synchronization and range for WiFi backscatter communication," in *Proc. 18th USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, 2021, pp. 923–937.

[164] Q. Wang, S. Chen, J. Zhao, and W. Gong, "RapidRider: Efficient WiFi backscatter with uncontrolled ambient signals," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2021, pp. 1–10.

[165] M. Zhang, S. Chen, J. Zhao, and W. Gong, "Commodity-level BLE backscatter," in *Proc. 19th Annu. Int. Conf. Mobile Syst. Appl. Services (MobiSys)*, 2021, pp. 402–414.

[166] J. Hu, L. Zhong, T. Ma, Z. Ding, and Z. Xu, "Long-range FM backscatter tag with tunnel diode," *IEEE Microw. Wireless Compon. Lett.*, vol. 32, no. 1, pp. 92–95, Jan. 2022.

[167] Y. Yang, L. Yuan, J. Zhao, and W. Gong, "Content-agnostic backscatter from thin air," in *Proc. 20th Annu. Int. Conf. Mobile Syst. Appl. Services (MobiSys)*, 2022, pp. 343–356.

[168] L. Yuan and W. Gong, "SubScatter: Sub-symbol WiFi backscatter for high throughput," in *Proc. IEEE 30th Int. Conf. Netw. Protocols (ICNP)*, 2022, pp. 1–11.

[169] Y. Yang and W. Gong, "Universal space–time stream backscatter with ambient WiFi," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, 2022, pp. 101–110.

[170] C. Xu, L. Yang, and P. Zhang, "Practical backscatter communication systems for battery-free Internet of Things: A tutorial and survey of recent research," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 16–27, Sep. 2018.

[171] C. de Lima et al., "6G white paper on localization and sensing." 2020. [Online]. Available: https://arxiv.org/abs/2006.01779

[172] V. Talla, J. Smith, and S. Gollakota, "Advances and open problems in backscatter networking," *Mobile Comput. Commun.*, vol. 24, no. 4, pp. 32–38, 2021.

[173] P. F. Freidl, M. E. Gadringer, D. Amschl, and W. Bcosch, "mm-Wave RFID for IoT applications," in *Proc. Integr. Nonlinear Microw. Millimetre-wave Circuits Workshop (INMMiC)*, 2017, pp. 1–3.

[174] A. Guerra, F. Guidi, D. Dardari, A. Clemente, and R. D'Errico, "A millimeter-wave indoor backscattering channel model for environment mapping," *IEEE Trans. Antennas Propag.*, vol. 65, no. 9, pp. 4935–4940, Sep. 2017.

[175] D. Matos, M. D. da Cruz Jordão, R. Correia, and N. B. Carvalho, "Millimeter-wave BiCMOS backscatter modulator for 5 G-IoT applications," *IEEE Microw. Wireless Compon. Lett.*, vol. 31, no. 2, pp. 173–176, Feb. 2021.

[176] J. Simonjan, B. D. Unluturk, and I. F. Akyildiz, "In-body bio-nanosensor localization for anomaly detection via inertial positioning and THz backscattering communication," *IEEE Trans. Nanobiosci.*, vol. 21, no. 2, pp. 216–225, Apr. 2022.

[177] S. Adibelli, P. Juyal, M. Prvulovic, and A. Zajic, "THz bistatic backscatter side-channel sensing at a distance," *IEEE Trans. Antennas Propag.*, vol. 70, no. 2, pp. 1440–1450, Feb. 2022.

[178] R. Di Candia, R. Jäntti, R. Duan, J. Lietzen, H. Khalifa, and K. Ruttik, "Quantum backscatter communication: A new paradigm," in *Proc. 15th Int. Symp. Wireless Commun. Syst. (ISWCS)*, 2018, pp. 1–6.

[179] M. Di Renzo et al., "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.

[180] W. Long, R. Chen, M. Moretti, W. Zhang, and J. Li, "A promising technology for 6G wireless networks: Intelligent reflecting surface," *J. Commun. Inf. Netw.*, vol. 6, no. 1, pp. 1–16, 2021.

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

[182] A. M. Elbir, K. V. Mishra, M. R. B. Shankar, and S. Chatzinotas, "The rise of intelligent reflecting surfaces in integrated sensing and communications paradigms," *IEEE Netw.*, early access, Dec. 26, 2022, doi: 10.1109/MNET.128.2200446.

[183] C. Huang, C. Zhang, J. Yang, B. Sun, B. Zhao, and X. Luo, "Reconfigurable metasurface for Multifunctional control of electromagnetic waves," *Adv. Opt. Mater.*, vol. 5, no. 22, 2017, Art. no. 1700485.

[184] S. Tian, H. Liu, and L. Li, "Design of 1-bit digital reconfigurable reflective metasurface for beam-scanning," *Appl. Sci.*, vol. 7, no. 9, p. 882, 2017.

[185] Y. Zhou et al., "Design of phase gradient coding metasurfaces for broadband wave modulating," *Sci. Rep.*, vol. 8, no. 1, p. 8672, 2018.

[186] L. Zhang et al., "Space–time-coding digital metasurfaces," *Nat. Commun.*, vol. 9, no. 1, p. 4334, 2018.

[187] X. Tan, Z. Sun, D. Koutsonikolas, and J. M. Jornet, "Enabling indoor mobile millimeter-wave networks based on smart reflect-arrays," in *Proc. IEEE Conf. Comput. Commun.*, 2018, pp. 270–278.

[188] Q. Luo et al., "Multibeam dual-circularly polarized reflectarray for connected and autonomous vehicles," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3574–3585, Apr. 2019.

[189] N. M. Tran, M. M. Amri, J. H. Park, S. I. Hwang, D. I. Kim, and K. W. Choi, "A novel coding metasurface for wireless power transfer applications," *Energies*, vol. 12, no. 23, p. 4488, 2019.

[190] L. Dai et al., "Reconfigurable intelligent surface-based wireless communications: Antenna design, prototyping, and experimental results," *IEEE Access*, vol. 8, pp. 45913–45923, 2020.

[191] G. Y. Liu et al., "Frequency-domain and spatial-domain reconfigurable metasurface," *ACS Appl. Mater. Interfaces*, vol. 12, no. 20, pp. 23554–23564, 2020.

[192] X. G. Zhang et al., "Polarization-controlled dual-programmable metasurfaces," *Adv. Sci.*, vol. 7, no. 11, 2020, Art. no. 1903382.

[193] X. Pei et al., "RIS-aided wireless communications: Prototyping, adaptive beamforming, and indoor/outdoor field trials," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8627–8640, Dec. 2021.

[194] I. Vardakis, G. Kotridis, S. Peppas, K. Skyvalakis, G. Vougioukas, and A. Bletsas, "Intelligently wireless batteryless RF-powered reconfigurable surface," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2021, pp. 1–6.

[195] C. Xiao, H. Huo, W. Xu, H. Sun, and C. Shu, "Reconfigurable intelligent surface-aided indoor communication with neural beam alignment," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, 2021, pp. 546–550.

[196] P. Staat, H. Elders-Boll, M. Heinrichs, R. Kronberger, C. Zenger, and C. Paar, "Intelligent reflecting surface-assisted wireless key generation for low-entropy environments," in *Proc. IEEE 32nd Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, 2021, pp. 745–751.

[197] J. C. Liang et al., "An angle-insensitive 3-bit reconfigurable intelligent surface," *IEEE Trans. Antennas Propag.*, vol. 70, no. 10, pp. 8798–8808, Oct. 2021.

[198] G. C. Trichopoulos et al., "Design and evaluation of reconfigurable intelligent surfaces in real-world environment," *IEEE Open J. Commun. Soc.*, vol. 3, pp. 462–474, 2022.

[199] R. Fara, P. Ratajczak, D.-T. Phan-Huy, A. Ourir, M. Di Renzo, and J. de Rosny, "A prototype of reconfigurable intelligent surface with continuous control of the reflection phase," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 70–77, Feb. 2022.

[200] M. Ouyang, F. Gao, Y. Wang, S. Zhang, P. Li, and J. Ren, "Computer vision-aided reconfigurable intelligent surface-based beam tracking: Prototyping and experimental results," *IEEE Trans. Wireless Commun.*, early access, Apr. 11, 2023, doi: '10.1109/TWC.2023.3264752.

[201] M. Rossanese, P. Mursia, A. Garcia-Saavedra, V. Sciancalepore, A. Asadi, and X. Costa-Perez, "Designing, building, and characterizing RF switch-based reconfigurable intelligent surfaces," in *Proc. 16th ACM Workshop Wireless Netw. Testbeds Exp. Eval. Characterization (WiNTECH)*, 2022, pp. 69–76.

[202] R. Long, Y.-C. Liang, Y. Pei, and E. G. Larsson, "Active reconfigurable intelligent surface-aided wireless communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 4962–4975, Aug. 2021.

[203] Z. Kang, C. You, and R. Zhang, "Active-IRS-aided wireless communication: Fundamentals, designs and open issues," 2023, *arXiv:2301.04311*.

[204] C. Huang et al., "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 118–125, Oct. 2020.

[205] V. Jamali, A. M. Tulino, G. Fischer, R. R. Müller, and R. Schober, "Intelligent surface-aided transmitter architectures for millimeter-wave ultra massive MIMO systems," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 144–167, 2021.

[206] O. L. A. López, N. H. Mahmood, H. Alves, C. M. Lima, and M. Latva-aho, "Ultra-low latency, low energy, and massiveness in the 6G era via efficient CSIT-limited scheme," *IEEE Commun. Mag.*, vol. 58, no. 11, pp. 56–61, Nov. 2020.

[207] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?" *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 7331–7376, Oct. 2019.

[208] G. Interdonato, E. Björnson, H. Quoc Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, pp. 1–13, Mar. 2019.

[209] U. K. Ganesan, E. Björnson, and E. G. Larsson, "RadioWeaves for extreme spatial multiplexing in indoor environments," in *Proc. 54th Asilomar Conf. Signals Syst. Comput.*, 2020, pp. 1007–1011.

[210] O. L. A. López, D. Kumar, R. D. Souza, P. Popovski, A. Tölli, and M. Latva-Aho, "Massive MIMO with radio stripes for indoor wireless energy transfer," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7088–7104, Sep. 2022.

[211] E. N. Tominaga, O. L. A. López, R. D. Souza, and H. Alves, "Statistical analysis of received signal strength in industrial IoT distributed massive MIMO systems," in *Proc. Joint Eur. Conf. Netw. Commun. 6G Summit (EuCNC/6G Summit)*, 2022, pp. 100–105.

[212] H. He, X. Yu, J. Zhang, S. Song, and K. B. Letaief, "Cell-free massive MIMO for 6G wireless communication networks," *J. Commun. Inf. Netw.*, vol. 6, no. 4, pp. 321–335, 2021.

[213] Z. H. Shaik, E. Björnson, and E. G. Larsson, "MMSE-optimal sequential processing for cell-free massive MIMO with radio stripes," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7775–7789, Nov. 2021.

[214] Z. H. Shaik, E. Björnson, and E. G. Larsson, "Distributed computation of a posteriori bit likelihood ratios in cell-free massive MIMO," in *Proc. 29th Eur. Signal Process. Conf. (EUSIPCO)*, 2021, pp. 935–939.

[215] M. Ghribi and A. Meddeb, "Survey and taxonomy of MAC, routing and cross layer protocols using wake-up radio," *J. Netw. Comput. Appl.*, vol. 149, Jan. 2020, Art. no. 102465.

[216] J. Oller, I. Demirkol, J. Casademont, J. Paradells, G. U. Gamm, and L. Reindl, "Has time come to switch from duty-cycled MAC protocols to wake-up radio for wireless sensor networks?" *IEEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 674–687, Apr. 2016.

[217] R. Piyare, A. L. Murphy, C. Kiraly, P. Tosato, and D. Brunelli, "Ultra low power wake-up radios: A hardware and networking survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2117–2157, 4th Quart., 2017.

[218] D. E. Ruíz-Guirola, C. A. Rodríguez-López, S. Montejo-Sánchez, R. D. Souza, O. L. A. López, and H. Alves, "Energy-efficient wake-up signalling for machine-type devices based on traffic-aware long short-term memory prediction," *IEEE Internet Things J.*, vol. 9, no. 21, pp. 21620–21631, Nov. 2022.

[219] S. Rostami, P. Kela, K. Leppanen, and M. Valkama, "Wake-up radio-based 5G mobile access: Methods, benefits, and challenges," *IEEE Commun. Mag.*, vol. 58, no. 7, pp. 14–20, Jul. 2020.

[220] *NR; NR & NG-RAN Overall Description; Stage 2*, 3GPP Standard TS 38.300, 2019.

[221] *Section 10.2.6B; Narrowband Wake Up Signal*, 3GPP Standard TS 36.211, 2018.

[222] *RP-192875, Additional MTC Enhancements for LTE (Work-Item Description)*, 3GPP, Sophia Antipolis, France, 2019.

[223] "Study on UE power saving in NR," 3GPP, Sophia Antipolis, France, Rep. TR 38.840, 2019.

[224] F. Z. Djiroun and D. Djenouri, "MAC protocols with wake-up radio for wireless sensor networks: A review," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 587–618, 1st Quart., 2016.

[225] J. Rinne, J. Keskinen, P. R. Berger, D. Lupo, and M. Valkama, "Viability bounds of M2M communication using energy-harvesting and passive wake-up radio," *IEEE Access*, vol. 5, pp. 27868–27878, 2017.

[226] *White Paper: 3GPP Releases 16 & 17 and Beyond*, 5G Americas, Bellevue, WA, USA, Jan. 2021.

[227] *RP-201306, Rel-17 Enhancements for NB-IoT and LTE-MTC (Work-Item Description). New SID: Study on UE Power Saving in NR, Document Rec. RP-181463*, 3GPP, Sophia Antipolis, France, Jun. 2018.

[228] X. Lin, "An overview of 5G advanced evolution in 3GPP release 18," *IEEE Commun. Stand. Mag.*, vol. 6, no. 3, pp. 77–83, Sep. 2022.

[229] *RP-213645, 3GPP TSG RAN Meeting #94e. Study on Low-Power Wake-Up Signal and Receiver for NR Study Item RAN 1, 2, 4. 3GPP RAN Release-18 Package Approved At the RAN Plenary Meeting*, 3GPP, Sophia Antipolis, France, Dec. 2021.

[230] S. Shellhammer, A. Asterjadhi, and Y. Sun. "Wake-up radio concept." 2023. [Online]. Available: https://oa.upm.es/12207/1/INVE_MEM_2011_97102.pdf

[231] A. Froytlog et al., "Ultra-low power wake-up radio for 5G IoT," *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 111–117, Mar. 2019.

[232] K. Mikhaylov and H. Karvonen, "Wake-up radio enabled BLE wearables: Empirical and analytical evaluation of energy efficiency," in *Proc. IEEE 14th Int. Symp. Med. Inf. Commun. Technol. (ISMICT)*, 2020, pp. 1–5.

[233] S. Rostami, H. D. Trinh, S. Lagen, M. Costa, M. Valkama, and P. Dini, "Wake-up scheduling for energy-efficient mobile devices," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 6020–6036, Sep. 2020.

[234] J. Petäjäjärvi, K. Mikhaylov, R. Vuohtoniemi, H. Karvonen, and J. Iinatti, "On the human body communications: Wake-up receiver design and channel characterization," *EURASIP J. Wireless Commun. Netw.*, vol. 2016, pp. 1–17, Aug. 2016.

[235] J. Petäjäjärvi, H. Karvonen, K. Mikhaylov, A. Pärssinen, M. Hämäläinen, and J. Iinatti, "WBAN energy efficiency and dependability improvement utilizing wake-up receiver," *IEICE Trans. Commun.*, vol. 98, no. 4, pp. 535–542, 2015.

[236] L. Lovén et al., "EdgeAI: A vision for distributed, edge-native artificial intelligence in future 6G networks," in *Proc. 1st 6G Wireless Summit*, 2019, pp. 1–2.

[237] I. Ahmad et al., "Machine learning meets communication networks: Current trends and future challenges," *IEEE Access*, vol. 8, pp. 223418–223460, 2020.

[238] E. Peltonen et al., "6G white paper on edge intelligence," 2020, *arXiv:2004.14850*.

[239] O. L. A. López et al., "Statistical tools and methodologies for URLLC—A tutorial," 2022, *arXiv:2212.03292*.

[240] M. Shafique et al., "An overview of next-generation architectures for machine learning: Roadmap, opportunities and challenges in the IoT era," in *Proc. Design Autom. Test Europe Conf. Exhibit. (DATE)*, 2018, pp. 827–832.

[241] M. E. Morocho-Cayamcela, H. Lee, and W. Lim, "Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions," *IEEE Access*, vol. 7, pp. 137184–137206, 2019.

[242] H. Fourati, R. Maaloul, and L. Chaari, "A survey of 5G network systems: Challenges and machine learning approaches," *Int. J. Mach. Learn. Cybern.*, vol. 12, pp. 385–431, Feb. 2021.

[243] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A survey of machine learning techniques applied to self-organizing cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2392–2431, 3rd Quart., 2017.

[244] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 30, 2016, pp. 2094–2100.

[245] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2224–2287, 3rd Quart., 2019.

[246] D. López-Pérez et al., "A survey on 5G radio access network energy efficiency: Massive MIMO, lean carrier design, sleep modes, and machine learning," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 653–697, 1st Quart., 2022.

[247] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.

[248] H. Kokkonen et al., "Autonomy and intelligence in the computing continuum: Challenges, enablers, and future directions for orchestration," 2022, *arXiv:2205.01423*.

[249] D. E. Ruíz-Guirola, O. L. A. López, S. Montejo-Sánchez, R. D. Souza, and M. Bennis, "Performance analysis of ML-based MTC traffic pattern predictors," *IEEE Wireless Commun. Lett.*, vol. 12, no. 7, pp. 1144–1148, Jul. 2023.

[250] T. Lähderanta et al., "Edge computing server placement with capacitated location allocation," *J. Parallel Distrib. Comput.*, vol. 153, pp. 130–149, Jul. 2021.

[251] L. Ruha et al., "Capacitated spatial clustering with multiple constraints and attributes," 2020, *arXiv:2010.06333*.

[252] X. Cao, L. Liu, Y. Cheng, and X. Shen, "Towards energy-efficient wireless networking in the big data era: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 303–332, 1st Quart., 2018.

[253] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[254] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 5998–6008.

[255] L. F. Maimó, F. J. G. Clemente, M. G. Pérez, and G. M. Pérez, "On the performance of a deep learning-based anomaly detection system for 5G networks," in *Proc. IEEE SmartWorld Ubiquitous Intell. Comput. Adv. Trusted Comput. Scalable Comput. Commun. Cloud Big Data Comput. Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, 2017, pp. 1–8.

[256] A. Reuther, P. Michaleas, M. Jones, V. Gadepally, S. Samsi, and J. Kepner, "AI accelerator survey and trends," in *Proc. IEEE High Perform. Extreme Comput. Conf. (HPEC)*, 2021, pp. 1–9.

[257] Y. Liu, X. Yuan, Z. Xiong, J. Kang, X. Wang, and D. Niyato, "Federated learning for 6G communications: Challenges, methods, and future directions," *China Commun.*, vol. 17, no. 9, pp. 105–118, 2020.

[258] A. Abrol and R. K. Jha, "Power optimization in 5G networks: A step towards GrEEn communication," *IEEE Access*, vol. 4, pp. 1355–1374, 2016.

[259] Y. Guo, F. Fang, D. Cai, and Z. Ding, "Energy-efficient design for a NOMA assisted STAR-RIS network with deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 72, no. 4, pp. 5424–5428, Apr. 2023.

[260] M. González-Palacio, D. Tobón-Vallejo, L. M. Sepúlveda-Cano, S. Rúa, and L. B. Le, "Machine-learning-based combined path loss and shadowing model in LoRaWAN for energy efficiency enhancement," *IEEE Internet Things J.*, vol. 10, no. 12, pp. 10725–10739, Jun. 2023.

[261] T. Fedullo, A. Morato, F. Tramarin, P. Bellagente, P. Ferrari, and E. Sisinni, "Adaptive LoRaWAN transmission exploiting reinforcement learning: The industrial case," in *Proc. IEEE Int. Workshop Metrol. Ind. 4.0 IoT (MetroInd4.0 IoT)*, 2021, pp. 671–676.

[262] S. U. Minhaj et al., "Intelligent resource allocation in LoRaWAN using machine learning techniques," *IEEE Access*, vol. 11, pp. 10092–10106, 2023.

[263] F. E. Salem, T. Chahed, Z. Altman, and A. Gati, "Traffic-aware advanced sleep modes management in 5G networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2019, pp. 1–6.

[264] M. Demirci, "A survey of machine learning applications for energy-efficient resource management in cloud computing environments," in *Proc. IEEE 14th Int. Conf. Mach. Learn. Appl. (ICMLA)*, 2015, pp. 1185–1190.

[265] W. Shafik, M. Matinkhah, and M. N. Sanda, "Network resource management drives machine learning: A survey and future research direction," *J. Commun. Technol. Electron. Comput. Sci.*, vol. 2020, pp. 1–15, Jun. 2020.

[266] T. Theodoropoulos et al., "Cloud-based XR services: A survey on relevant challenges and enabling technologies," *J. Netw. Netw. Appl.*, vol. 2, no. 1, pp. 1–22, 2022.

[267] "Update: Open RAN explained." Accessed: Dec. 1, 2023. [Online]. Available: https://www.nokia.com/about-us/newsroom/articles/open-ran-explained/

[268] M. Karbalaee Motalleb, V. Shah-Mansouri, S. Parsaeefard, and O. L. Alcaraz López, "Resource allocation in an open RAN system using network slicing," *IEEE Trans. Netw. Service Manag.*, vol. 20, no. 1, pp. 471–485, Mar. 2023.

[269] S. Niknam et al., "Intelligent O-RAN for beyond 5G and 6G wireless networks," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2022, pp. 215–220.

[270] B. Yi, X. Wang, K. Li, S. K. Das, and M. Huang, "A comprehensive survey of network function virtualization," *Comput. Netw.*, vol. 133, pp. 212–262, Mar. 2018.

[271] L. Gavrilovska, V. Rakovic, and D. Denkovski, "From cloud RAN to open RAN," *Wireless Pers. Commun.*, vol. 113, no. 3, pp. 1523–1539, 2020.

[272] D. Xu et al., "Edge intelligence: Empowering intelligence to the edge of network," *Proc. IEEE*, vol. 109, no. 11, pp. 1778–1837, Nov. 2021.

[273] J. Park, S. Samarakoon, M. Bennis, and M. Debbah, "Wireless network intelligence at the edge," *Proc. IEEE*, vol. 107, no. 11, pp. 2204–2239, Nov. 2019.

[274] E. Peltonen et al., "The many faces of edge intelligence," *IEEE Access*, vol. 10, pp. 104769–104782, 2022.

[275] S. Deng, H. Zhao, W. Fang, J. Yin, S. Dustdar, and A. Y. Zomaya, "Edge intelligence: The confluence of edge computing and artificial intelligence," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7457–7469, Aug. 2020.

[276] L. Lovén, E. Peltonen, E. Harjula, and S. Pirttikangas, "Weathering the reallocation storm: Large-scale analysis of edge server workload," in *Proc. Joint Eur. Conf. Netw. Commun. 6G Summit (EuCNC/6G Summit)*, 2021, pp. 317–322.

[277] L. Lovén, E. Peltonen, L. Ruha, E. Harjula, and S. Pirttikangas, "A dark and stormy night: Reallocation storms in edge computing," *EURASIP J. Wireless Commun. Netw.*, vol. 2022, no. 1, p. 86, 2022.

[278] P. Kairouz et al., "Advances and open problems in federated learning," *Found. Trends Mach. Learn.*, vol. 14, nos. 1–2, pp. 1–210, 2021.

[279] W. Yu et al., "A survey on the edge computing for the Internet of Things," *IEEE Access*, vol. 6, pp. 6900–6919, 2017.

[280] M. Maray and J. Shuja, "Computation offloading in mobile cloud computing and mobile edge computing: Survey, taxonomy, and open issues," *Mobile Inf. Syst.*, vol. 2022, Jun. 2022, Art. no. 1121822.

[281] X. Qiao, P. Ren, S. Dustdar, and J. Chen, "A new era for Web AR with mobile edge computing," *IEEE Internet Comput.*, vol. 22, no. 4, pp. 46–55, Jul./Aug. 2018.

[282] X. Zhu, Y. Luo, A. Liu, M. Z. A. Bhuiyan, and S. Zhang, "Multiagent deep reinforcement learning for vehicular computation offloading in IoT," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 9763–9773, Jun. 2021.

[283] C. Feng, P. Han, X. Zhang, B. Yang, Y. Liu, and L. Guo, "Computation offloading in mobile edge computing networks: A survey," *J. Netw. Comput. Appl.*, vol. 202, Jun. 2022, Art. no. 103366.

[284] K. Gasmi, S. Dilek, S. Tosun, and S. Ozdemir, "A survey on computation offloading and service placement in fog computing-based IoT," *J. Supercomput.*, vol. 78, no. 2, pp. 1983–2014, 2022.

[285] K. Kumar and Y.-H. Lu, "Cloud computing for mobile users: Can offloading computation save energy?" *Computer*, vol. 43, no. 4, pp. 51–56, 2010.

[286] R. P. Mathur and M. Sharma, "A machine learning approach for offload decision making in mobile cloud computing," in *Proc. 7th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, vol. 1, 2021, pp. 1148–1154.

[287] X. Sun and N. Ansari, "Green cloudlet network: A distributed green mobile cloud network," *IEEE Netw.*, vol. 31, no. 1, pp. 64–70, Jan./Feb. 2017.

[288] L. A. Barroso and U. Hölzle, "The case for energy-proportional computing," *Computer*, vol. 40, no. 12, pp. 33–37, 2007.

[289] R. Kallimani, K. Pai, P. Raghuwanshi, S. Iyer, and O. L. López, "TinyML: Tools, applications, challenges, and future research directions," *Multimedia Tools Appl.*, to be published.

[290] T. Basaklar, Y. Tuncel, and U. Y. Ogras, "TinyMAN: Lightweight energy manager using reinforcement learning for energy harvesting wearable IoT devices," 2022, *arXiv:2202.09297*.

[291] G. Gruosso and G. S. Gajani, "Comparison of machine learning algorithms for performance evaluation of photovoltaic energy forecasting and management in the TinyML framework," *IEEE Access*, vol. 10, pp. 121010–121020, 2022.

[292] R. Sanchez-Iborra and A. F. Skarmeta, "TinyML-enabled frugal smart objects: Challenges and opportunities," *IEEE Circuits Syst. Mag.*, vol. 20, no. 3, pp. 4–18, Aug. 2020.

[293] H. Ren, D. Anicic, and T. A. Runkler, "The synergy of complex event processing and tiny machine learning in Industrial IoT," in *Proc. 15th ACM Int. Conf. Distrib. Event Syst.*, 2021, pp. 126–135.

[294] R. V. W. Putra, M. A. Hanif, and M. Shafique, "ROMANet: Fine-grained reuse-driven off-chip memory access management and data organization for deep neural network accelerators," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 29, no. 4, pp. 702–715, Apr. 2021.

[295] "Transformer and self-attention mechanism." Accessed: Mar. 23, 2023. [Online]. Available: https://www.youtube.com/watch?v=5T38-2J5CcY

[296] "PULP platform." Accessed: Mar. 25, 2023. [Online]. Available: https://pulp-platform.org/

[297] E. Njor, J. Madsen, and X. Fafoutis, "A primer for tinyML predictive maintenance: Input and model optimisation," in *Proc. Artif. Intell. Appl. Innov.*, 2022, pp. 67–78.

[298] Y. Koizumi, S. Saito, H. Uematsu, N. Harada, and K. Imoto, "ToyADMOS: A dataset of miniature-machine operating sounds for anomalous sound detection," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, 2019, pp. 313–317.

[299] P. BN, H. Yelchuri, V. Shastry, and T. Prabhakar, "PreMa: predictive maintenance of solenoid valve in real-time at embedded edge-level," 2022, *arXiv:2211.12326*.

[300] X. B. Peng, E. Coumans, T. Zhang, T.-W. Lee, J. Tan, and S. Levine, "Learning agile robotic locomotion skills by imitating animals," 2020, *arXiv:2004.00784*.

[301] S. M. Neuman et al., "Tiny robot learning: Challenges and directions for machine learning in resource-constrained robots," in *Proc. IEEE 4th Int. Conf. Artif. Intell. Circuits Syst. (AICAS)*, 2022, pp. 296–299.

[302] J. Jabbour et al. "Closing the Sim-to-real gap for ultra-low-cost, resource-constrained, quadruped robot platforms." 2022. [Online]. Available: https://sim2real.github.io/assets/papers/2022/jabbour.pdf

[303] J. Lin et al., "MCUNet: Tiny deep learning on IoT devices," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 11711–11722.

[304] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetv2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 4510–4520.

[305] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[306] E. Tabanelli, G. Tagliavini, and L. Benini, "DNN is not all you need: Parallelizing non-neural ML algorithms on ultra-low-power IoT processors," *ACM Trans. Embedded Comput. Syst.*, vol. 22, no. 3, pp. 1–33, 2023.

[307] X. Wang, M. Magno, L. Cavigelli, and L. Benini, "FANN-on-MCU: An open-source toolkit for energy-efficient neural network inference at the edge of the Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4403–4417, May 2020.

[308] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "A survey of model compression and acceleration for deep neural networks," 2017, *arXiv:1710.09282*.

[309] B. Jacob et al., "Quantization and training of neural networks for efficient integer-arithmetic-only inference," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2704–2713.

[310] Q. Lu and B. Murmann, "Improving the energy efficiency and robustness of tinyML computer vision using log-gradient input images," 2022, *arXiv:2203.02571*.

[311] S. Han, J. Pool, J. Tran, and W. Dally, "Learning both weights and connections for efficient neural network," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1135–1143.

[312] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*.

[313] W. Chen, J. Wilson, S. Tyree, K. Weinberger, and Y. Chen, "Compressing neural networks with the hashing trick," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2285–2294.

[314] R. Dogaru and I. Dogaru, "Fast training of light binary convolutional neural networks using chainer and cupy," in *Proc. 12th Int. Conf. Electron. Comput. Artif. Intell. (ECAI)*, 2020, pp. 1–4.

[315] M. Shafique, A. Marchisio, R. V. W. Putra, and M. A. Hanif, "Towards energy-efficient and secure edge AI: A cross-layer framework," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, 2021, pp. 1–9.

[316] J. Li et al., "SmartShuttle: Optimizing off-chip memory accesses for deep learning accelerators," in *Proc. Design Autom. Test Europe Conf. Exhibit. (DATE)*, 2018, pp. 343–348.

[317] A. Wong, M. Famouri, and M. J. Shafiee, "AttendNets: Tiny deep image recognition neural networks for the edge via visual attention condensers," 2020, *arXiv:2009.14385*.

[318] H. Zhang, Y. N. Dauphin, and T. Ma, "Fixup initialization: Residual learning without normalization," 2019, *arXiv:1901.09321*.

[319] A. Howard et al., "Searching for MobileNetv3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1314–1324.

[320] I. Turc, M.-W. Chang, K. Lee, and K. Toutanova, "Well-read students learn better: On the importance of pre-training compact models," 2019, *arXiv:1908.08962*.

[321] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," 2019, *arXiv:1910.01108*.

[322] B. Fang, X. Zeng, and M. Zhang, "NestDNN: Resource-aware multitenant on-device deep learning for continuous mobile vision," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2018, pp. 115–127.

[323] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[324] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[325] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[326] F. Fahim et al., "HLS4ML: An open-source codesign workflow to empower scientific low-power machine learning devices," 2021, *arXiv:2103.05579*.

[327] P. Gupta, D. K. Saini, P. Rawat, and K. Zia, *Bio-Inspired Optimization in Fog and Edge Computing Environments: Principles, Algorithms, and Systems*. Boca Raton, FL, USA: CRC Press, 2023.

[328] E. Donati, M. Payvand, N. Risi, R. Krause, and G. Indiveri, "Discrimination of EMG signals using a neuromorphic implementation of a spiking neural network," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 5, pp. 795–803, Oct. 2019.

[329] A. Shafiee et al., "ISAAC: A convolutional neural network accelerator with *in-situ* analog arithmetic in crossbars," in *Proc. ACM/IEEE 43rd Annu. Int. Symp. Comput. Archit. (ISCA)*, 2016, pp. 14–26.

[330] V. Rzehak, *Low-Power FRAM Microcontrollers and Their Applications*, Texas Instrum., Dallas, TX, USA, 2019.

[331] N. Deng and M. Haenggi, "The energy and rate meta distributions in wirelessly powered D2D networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 2, pp. 269–282, Feb. 2019.

[332] T. Chen, H. Kim, and Y. Yang, "Energy efficiency metrics for green wireless communications," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, 2010, pp. 1–6.

[333] Y. Zeng and R. Zhang, "Optimized training design for wireless energy transfer," *IEEE Trans. Commun.*, vol. 63, no. 2, pp. 536–550, Feb. 2015.

[334] F. Stern and J. Spencer, "Chapter 10: Peak demand and time-differentiated energy savings cross-cutting protocol," Nat. Renew. Energy Lab., Golden, CO, USA, Rep. NREL/SR-7A40-68566, 2017. [Online]. Available: http://www.nrel.gov/docs/fy17osti/68566.pdf

[335] Y. Wang, M. Sheng, X. Wang, L. Wang, and J. Li, "Mobile-edge computing: Partial computation offloading using dynamic voltage scaling," *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4268–4282, Oct. 2016.

[336] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy harvesting sensor networks," *ACM Trans. Embedded Comput. Syst.*, vol. 6, no. 4, p. 32, Sep. 2007.

[337] G. Bhat, J. Park, and U. Y. Ogras, "Near-optimal energy allocation for self-powered wearable systems," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Design (ICCAD)*, 2017, pp. 368–375.

[338] M. Crozzoli et al., *Deliverable D2.4 Metrics and KPIs for RISE Wireless Systems Analysis: Final Results*, document H2020-ICT-52/RISE-6G/D2.4, RISE 6G, Paris, France, 2022.

[339] X. Zhao, L. Zhang, Y. Cao, K. Jin, and Y. Hou, "Anomaly detection approach in industrial control systems based on measurement data," *Information*, vol. 13, no. 10, pp. 1–43, 2022.

[340] O. L. A. López, D. Kumar, and A. Tölli, "Polarization diversity-enabled LOS/NLOS identification via carrier phase measurements," *IEEE Trans. Commun.*, vol. 71, no. 3, pp. 1–13, Mar. 2020.

[341] C. Fiandrino, D. Kliazovich, P. Bouvry, and A. Y. Zomaya, "Performance and energy efficiency metrics for communication systems of cloud computing data Centers," *IEEE Trans. Cloud Comput.*, vol. 5, no. 4, pp. 738–750, Oct.–Dec. 2017.

[342] International Commission on Non-Ionizing Radiation Protection (ICNIRP), "Guidelines for limiting exposure to electromagnetic fields (100 kHz to 300 GHz)," *Health Phys.*, vol. 118, no. 5, pp. 483–524, 2020.

[343] O. L. A. López et al., "High-power and safe RF wireless charging: Cautious deployment and operation," *IEEE Wireless Commun.*, early access.

[344] I. Ketata, S. Ouerghemmi, A. Fakhfakh, and F. Derbel, "Design and implementation of low noise amplifier operating at 868 MHz for duty cycled wake-up receiver front-end," *Electronics*, vol. 11, no. 19, p. 3235, 2022.

[345] A. Kumar, S. Goyal, and M. Varma, "Resource-efficient machine learning in 2 KB RAM for the Internet of Things," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 1935–1944.

[346] S. C. Chandrarathna, S.-Y. Moon, and J.-W. Lee, "A power management system for hybrid energy harvesting from multiple multitype sources and ultrawide range source tracking," *IEEE Trans. Power Electron.*, vol. 38, no. 4, pp. 4859–4875, Apr. 2023.

[347] M. N. Ibrahim, H. Rezk, M. Al-Dahifallah, and P. Sergeant, "Hybrid photovoltaic-thermoelectric generator powered synchronous reluctance motor for pumping applications," *IEEE Access*, vol. 7, pp. 146979–146988, 2019.

[348] Q. Wu, G. Zhang, D. W. K. Ng, W. Chen, and R. Schober, "Generalized wireless-powered communications: When to activate wireless power transfer?" *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8243–8248, Aug. 2019.

[349] A. Abuellil, J. J. Estrada-López, A. Bommireddipalli, A. Costilla-Reyes, Z. Zeng, and E. Sánchez-Sinencio, "Multiple-input harvesting power management unit with enhanced boosting scheme for IoT applications," *IEEE Trans. Ind. Electron.*, vol. 67, no. 5, pp. 3662–3672, May 2020.

[350] H. Ali et al., "A survey on system level energy optimisation for MPSoCs in IoT and consumer electronics," *Comput. Sci. Rev.*, vol. 41, Aug. 2021, Art. no. 100416.

[351] S. Hou et al., "Fine-grained online energy management of edge data Centers using per-core power gating and dynamic voltage and frequency scaling," *IEEE Trans. Sustain. Comput.*, vol. 8, no. 3, pp. 522–536, Sep. 2023.

[352] M. Weißbrich, H. Blume, and G. Payá-Vayá, "A silicon-proof controller system for flexible ultra-low-power energy harvesting platforms," in *Proc. 11th Int. Conf. Mod. Circuits Syst. Technol. (MOCAST)*, 2022, pp. 1–6.

[353] G. Akgün, M. Ali, and D. Göhringer, "Power-aware computing systems on FPGAs: A survey," in *Proc. 31st Int. Conf. Field Program. Logic Appl. (FPL)*, 2021, pp. 45–51.

[354] T. Han, K. Muhammad, T. Hussain, J. Lloret, and S. W. Baik, "An efficient deep learning framework for intelligent energy management in IoT networks," *IEEE Internet Things J.*, vol. 8, no. 5, pp. 3170–3179, May 2021.

[355] P. Mayer, M. Magno, and L. Benini, "Smart power unit mW-to-nW power management and control for self-sustainable IoT devices," *IEEE Trans. Power Electron.*, vol. 36, no. 5, pp. 5700–5710, May 2021.

[356] H. Zhou, Q. Liu, K. Yan, and Y. Du, "Deep learning enhanced solar energy forecasting with AI-driven IoT," *Wireless Commun. Mobile Comput.*, vol. 2021, pp. 1–11, Jun. 2021.

[357] O. M. Rosabal, O. L. A. López, H. Alves, and M. Latva-Aho, "Sustainable radio frequency wireless energy transfer for massive Internet of Things," 2023, *arXiv:2306.15051*.

[358] X. Chen, W. Ni, T. Chen, I. B. Collings, X. Wang, and G. B. Giannakis, "Real-time energy trading and future planning for fifth generation wireless communications," *IEEE Wireless Commun.*, vol. 24, no. 4, pp. 24–30, Aug. 2017.

[359] D. H. Nguyen, "Electric vehicle—Wireless charging-discharging lane decentralized peer-to-peer energy trading," *IEEE Access*, vol. 8, pp. 179616–179625, 2020.

[360] H. Kim, J. Lee, S. Bahrami, and V. W. S. Wong, "Direct energy trading of microgrids in distribution energy market," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 639–651, Jan. 2020.

[361] Z. Li, J. Kang, R. Yu, D. Ye, Q. Deng, and Y. Zhang, "Consortium blockchain for secure energy trading in Industrial Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 14, no. 8, pp. 3690–3700, Aug. 2018.

[362] Y. Xia, Q. Xu, Y. Huang, Y. Liu, and F. Li, "Preserving privacy in nested peer-to-peer energy trading in networked microgrids considering incomplete rationality," *IEEE Trans. Smart Grid*, vol. 14, no. 1, pp. 606–622, Jan. 2023.

[363] H. Kalantarian and M. Sarrafzadeh, "Pedometers without batteries: An energy harvesting shoe," *IEEE Sensors J.*, vol. 16, no. 23, pp. 8314–8321, Dec. 2016.

[364] B. Campbell, B. Ghena, and P. Dutta, "Energy-harvesting thermoelectric sensing for unobtrusive water and appliance metering," in *Proc. 2nd Int. Workshop Energy Neutral Sens. Syst.*, 2014, pp. 7–12.

[365] D. Ma, G. Lan, M. Hassan, W. Hu, and S. K. Das, "Sensing, computing, and communications for energy harvesting IoTs: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 1222–1250, 2nd Quart., 2020.

[366] M. Hicks, "CLANK: Architectural support for intermittent computation," *ACM SIGARCH Comput. Architect. News*, vol. 45, no. 2, pp. 228–240, 2017.

[367] K. Maeng, A. Colin, and B. Lucia, "ALPACA: Intermittent execution without checkpoints," in *Proc. ACM Program. Lang.*, vol. 1, 2017, pp. 1–30.

[368] A. Sabovic, A. K. Sultania, C. Delgado, L. D. Roeck, and J. Famaey, "An energy-aware task scheduler for energy-harvesting batteryless IoT devices," *IEEE Internet Things J.*, vol. 9, no. 22, pp. 23097–23114, Nov. 2022.

[369] A. Rahmati, M. Salajegheh, D. Holcomb, J. Sorber, W. P. Burleson, and K. Fu, "TARDIS: Time and remanence decay in SRAM to implement secure protocols on embedded devices without clocks," in *Proc. 21st USENIX Security Symp. (USENIX Security)*, 2012, pp. 221–236.

[370] J. Hester et al., "Persistent clocks for batteryless sensing devices," *ACM Trans. Embedded Comput. Syst. (TECS)*, vol. 15, no. 4, pp. 1–28, 2016.

[371] H. Elahi, K. Munir, M. Eugeni, S. Atek, and P. Gaudenzi, "Energy harvesting towards self-powered IoT devices," *Energies*, vol. 13, no. 21, p. 5528, 2020.

[372] A. Di Mauro, D. Papini, and N. Dragoni, "Security challenges for energy-harvesting wireless sensor networks," in *Proc. 2nd Int. Conf. Pervasive Embedded Comput. Commun. Syst. (PECCS)*, 2012, pp. 422–425.

[373] K. Mikhaylov, R. Fujdiak, A. Pouttu, V. Miroslav, L. Malina, and P. Mlynek, "Energy attack in LoRaWAN: Experimental validation," in *Proc. 14th Int. Conf. Availability Rel. Security*, 2019, pp. 1–6.

[374] J. Cook, S. U. Rehman, and M. A. Khan, "Security and privacy for low power IoT devices on 5G and beyond networks: Challenges and future directions," *IEEE Access*, vol. 11, pp. 39295–39317, 2023.

[375] A. T. Capossele, V. Cervo, C. Petrioli, and D. Spenza, "Counteracting denial-of-sleep attacks in wake-up-radio-based sensing systems," in *Proc. 13th Annu. IEEE Int. Conf. Sens. Commun. Netw. (SECON)*, 2016, pp. 1–9.

[376] T. Nguyen, N. Tran, L. Loven, J. Partala, M.-T. Kechadi, and S. Pirttikangas, "Privacy-aware blockchain innovation for 6G: Challenges and opportunities," in *Proc. 2nd 6G Wireless Summit (6G SUMMIT)*, 2020, pp. 1–5.

[377] T. Nguyen, L. Lovén, J. Partala, and S. Pirttikangas, "The intersection of blockchain and 6G technologies," in *6G Mobile Wireless Networks*. Cham, Switzerland: Springer, 2021, pp. 393–417.

[378] A. Patil, S. Iyer, O. L. Lopez, R. J. Pandya, A. Kalla, and R. Kallimani, "A comprehensive survey on spectrum sharing techniques for 5G/B5G intelligent wireless networks: Opportunities, challenges and future research directions," 2022, *arXiv:2211.08956*.

[379] P. Fritzson et al., "The OpenModelica integrated environment for modeling, simulation, and model-based development," *Model. Identification Control*, vol. 41, no. 4, pp. 241–295, 2020.

[380] W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, "PVLIB python: A python package for modeling solar energy systems," *J. Open Source Softw.*, vol. 3, no. 29, p. 884, 2018.

[381] C. Tapparello, H. Ayatollahi, and W. Heinzelman, "Energy harvesting framework for network simulator 3 (ns-3)," in *Proc. 2nd Int. Workshop Energy Neutral Sens. Syst.*, 2014, pp. 37–42.

[382] M. Afaneh, *Intro to Bluetooth Low Energy*. New York, NY, USA: Novel Bits, 2018.

[383] L. Tian, S. Santi, A. Seferagić, J. Lan, and J. Famaey, "Wi-Fi HaLow for the Internet of Things: An up-to-date survey on IEEE 802.11 ah research," *J. Netw. Comput. Appl.*, vol. 182, May 2021, Art. no. 103036.

[384] S. Marksteiner, V. J. E. Jimenez, H. Valiant, and H. Zeiner, "An overview of wireless IoT protocol security in the smart home domain," in *Proc. Internet Things Bus. Models Users Netw.*, 2017, pp. 1–8.

[385] J. C. Cano, V. Berrios, B. Garcia, and C. K. Toh, "Evolution of IoT: An industry perspective," *IEEE Internet Things Mag.*, vol. 1, no. 2, pp. 12–17, Dec. 2018.

[386] A. Hoeller, R. D. Souza, O. L. Alcaraz López, H. Alves, M. de Noronha Neto, and G. Brante, "Analysis and performance optimization of LoRa networks with time and antenna diversity," *IEEE Access*, vol. 6, pp. 32820–32829, 2018.

[387] S. Gangwar, Yogita, I. B. Prasad, S. S. Yadav, V. Pal, and N. Kumar, "GTFR: A game theory based fuzzy routing protocol for WSNs," *IEEE Sensors J.*, early access, Feb. 28, 2023, doi: 10.1109/JSEN.2023.3248226.

[388] A. Rech and S. Tomasin, "Coordinated random access for Industrial IoT with correlated traffic by reinforcement-learning," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2021, pp. 1–6.

[389] Z. Jiang, S. Zhou, and Z. Niu, "Distributed policy learning based random access for diversified QoS requirements," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[390] S. u. Haque, S. Chandak, F. Chiariotti, D. Gündüz, and P. Popovski, "Learning to speak on behalf of a group: Medium access control for sending a shared message," *IEEE Commun. Lett.*, vol. 26, no. 8, pp. 1843–1847, Aug. 2022.

[391] K. Kumar, S. Kumar, O. Kaiwartya, Y. Cao, J. Lloret, and N. Aslam, "Cross-layer energy optimization for IoT environments: Technical advances and opportunities," *Energies*, vol. 10, no. 12, p. 2073, 2017.

[392] G. Carneiro, J. Ruela, and M. Ricardo, "Cross-layer design in 4G wireless terminals," *IEEE Wireless Commun.*, vol. 11, no. 2, pp. 7–13, Apr. s2004.

[393] B. Zhu, E. Bedeer, H. H. Nguyen, R. Barton, and J. Henry, "Improved soft-k-means clustering algorithm for balancing energy consumption in wireless sensor networks," *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4868–4881, Mar. 2021.

[394] B. Diène, J. J. Rodrigues, O. Diallo, E. H. M. Ndoye, and V. V. Korotaev, "Data management techniques for Internet of Things," *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106564. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S088832701930785X

[395] R. Krishnamurthi, A. Kumar, D. Gopinathan, A. Nayyar, and B. Qureshi, "An overview of IoT sensor data processing, fusion, and analysis techniques," *Sensors*, vol. 20, no. 21, p. 6076, 2020.

[396] J. Wan, B. Chen, S. Wang, M. Xia, D. Li, and C. Liu, "Fog computing for energy-aware load balancing and scheduling in smart factory," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4548–4556, Oct. 2018.

[397] S. Sun, X. Zheng, J. Villalba-Díez, and J. Ordieres-Meré, "Data handling in industry 4.0: Interoperability based on distributed ledger technology," *Sensors*, vol. 20, no. 11, p. 3046, 2020.

[398] A. Eid, J. G. D. Hester, and E. Tentzeris, "High gain and large beamwidth Rotman-lens-based and mm-Wave backscattering and energy harvesting systems and associated methods," U.S. Patent App. 18 008 244, Sep. 2023.

ONEL L. A. LÓPEZ (Member, IEEE) received the B.Sc. degree (First-Class Hons.) from the Central University of Las Villas, Cuba, in 2013, the M.Sc. degree from the Federal University of Paraná, Brazil, in 2017, and the D.Sc. degree (with Distinction) in electrical engineering from the University of Oulu, Finland, in 2020, where he currently holds an Assistant Professorship (tenure track) of Sustainable Wireless Communications Engineering with the Centre for Wireless Communications. He is the co-author of the book entitled "*Wireless RF Energy Transfer in the Massive IoT Era: Towards Sustainable Zero-Energy Networks*", (Wiley, December 2021). His research interests include wireless communications, signal processing, sustainable IoT, and wireless RF energy transfer. He is a collaborator to the 2016 Research Award given by the Cuban Academy of Sciences, a co-recipient of the 2019 IEEE EuCNC Best Student Paper Award, and the recipient of the 2020 Best Doctoral Thesis Award granted by Finland TEK and TFiF in 2021.

OSMEL M. ROSABAL (Graduate Student Member, IEEE) received the B.Sc. degree (Hons.) in electrical engineering from the Central University of Las Villas, Cuba, in 2013, and the M.Sc. degree (Hons.) in electrical engineering from the University of Oulu, Finland, in 2021. He is currently pursuing the Ph.D. degree in wireless communications engineering with a focus on sustainable wireless energy transfer systems, Centre for Wireless Communications, University of Oulu, Finland. From 2013 to 2018, he served as a specialist in telematics with Cuban Telecommunications Company (ETECSA).

KONSTANTIN MIKHAYLOV (Senior Member, IEEE) is an Assistant Professor for Convergent IoT Communications with the Centre for Wireless Communications, University of Oulu, Oulu, Finland. He has authored or coauthored more than one hundred research papers on wireless connectivity for IoT, IoT system and device design, and applications. His research focuses on radio access and beyond-access technologies and protocols for massive and dependable IoT and the matters related to the design and practical use of IoT systems.

DAVID E. RUIZ-GUIROLA (Graduate Student Member, IEEE) received the B.Sc. (First-Class Hons) and M.Sc. (with Distinction) degrees in telecommunications and electronic engineering from the Central University of Las Villas (UCLV), Santa Clara, Cuba, in 2018 and 2019, respectively. He is currently pursuing the Ph.D. degree with the University of Oulu, Finland. He served as an Associate Professor with UCLV from 2018 to 2021. He joined the Centre for Wireless Communications, University of Oulu in 2021. His research interests include sustainable IoT, energy harvesting, wireless RF energy transfer, machine-type communications, machine learning, and traffic prediction.

LAURI LOVÉN (Senior Member, IEEE) received the D.Sc. degree from the University of Oulu, Finland, in 2021, where he is a Postdoctoral Researcher in the 6G Flagship Research Program with the Center for Ubiquitous Computing, He has worked with the Distributed Systems Group, TU Wien in 2022. He has coauthored two patents and ca. 40 research articles in international peer-reviewed journals, conferences, and workshops. His current research concentrates on edge intelligence, and in particular on the orchestration of resources as well as distributed learning and decision-making in the computing continuum.

PRASOON RAGHUWANSHI (Student Member, IEEE) received the B.Tech. and M.Tech. degrees in electronics and communication engineering from the National Institute of Technology Hamirpur, Hamirpur, India, in 2020. He is currently pursuing the D.Sc. degree in communications engineering with the University of Oulu, Oulu, Finland, where he joined the Centre for Wireless Communications in 2022. His research interests include random access protocols for IoT networks, machine-type communications, deep reinforcement learning, and TinyML.

SRIDHAR IYER (Senior Member, IEEE) received the M.S. degree from New Mexico State University, USA, in 2008, and the Ph.D. degree from Delhi University, India, in 2017. He currently serves as a Professor with the Department of CSE (AI), KLE Technological University, Dr. M. S. Sheshgiri Belagavi Campus, India. His current research focus includes semantic communications and spectrum enhancement techniques for intelligent wireless systems. He received the Young Scientist Award from the DST/SERB, Government of India in 2013 and the Young Researcher Award from the Institute of Scholars in 2021. He is the Recipient of the 'Protsahan Award' from IEEE ComSoc, Bangalore Chapter as a recognition to his contributions towards paper published/tutorial offered in recognized conferences/journals, from January 2020 to September 2021 and from October 2021 to October 2022.