Context-Aware Intelligence in Resource-Constrained IoT Nodes: Opportunities and Challenges

Baibhab Chatterjee
Purdue University

Shreyas Sen
Purdue University

Ningyuan Cao and Arijit Raychowdhury
Georgia Institute of Technology

Editor's note:
This article provides an academic perspective of the problem, starting with a survey of recent advances in intelligent sensing, computation, communication, and energy management for resource-constrained IoT sensor nodes and leading to a future outlook and needs.
—Shreyas Sen, Purdue University

THE PERSONAL, HEALTHCARE, AND CONSUMER ELECTRONIC industries have experienced rapid advancements in the past few decades due to aggressive technology scaling and low-power, low-cost implementation of sensor electronics built on mobile computing/communication platforms having small form factors. This has resulted in a pervasive growth of connected devices, leading to what is known today as the Internet of Things (IoT), as shown in Figure 1a. Increased fidelity and higher bandwidths are expected to result in 50 billion connected devices, generating 30+ exabytes of data per month by 2020, which would extend beyond 500 billion devices by 2030 [1].

Digital Object Identifier 10.1109/MDAT.2019.2899334
Date of publication: 13 February 2019; date of current version: 11 April 2019.

Background and motivation

The dynamic nature of the IoT devices, coupled with their stringent resource constraints (small-size and portability requirements, leading to smaller energy backup, less computing and memory resources in harsh environments, varying channel conditions, asymmetric data rates in uplink and downlink, etc.) and the availability of multiple communication modalities [wired, proximity, low-energy Bluetooth (BTLE), ANT, LoRa, ZigBee, Human Body Communication (HBC), MedRadio, and millimeter-wave (mm-wave), to name a few] necessitate proper selection of communication architecture based on the application and corresponding resource constraints. In addition, the power cost of communication (≈1 nJ/bit in standard wireless networks [2]) may warrant intelligent allocation between local and remote computing resources, which would require context-aware operation corresponding to different scenarios, leading to minimum energy consumption for a certain amount of information

Copublished by the IEEE CEDA, IEEE CASS, IEEE SSCS, and TTTC

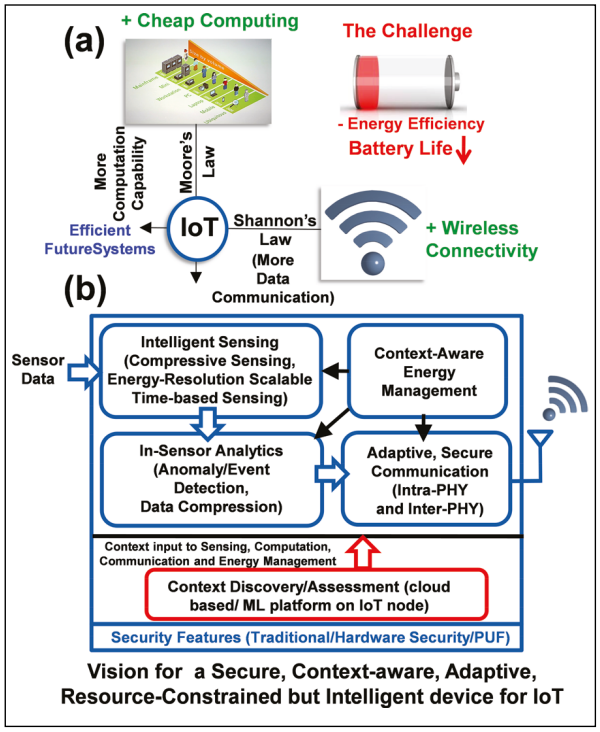

Figure 1. (a) IoT at the juncture of Moore's law (more computation enabled by technology scaling) and Shannon's law (more data rate enabled by modern wireless standards), consuming a prohibitively high amount of energy [2]. (b) Our vision for an RC-IoT device for optimum energy efficiency.

transfer, thereby improving the lifespan of the network. Also, energy-resolution scalable sensing technology [preferably in the compressed domain (CD)] enables consumption of sensing energy when required and present-day context-agnostic IoT systems are typically overdesigned to take care of all possible contexts/scenarios, which trade off fidelity with power consumption and, hence, degrade the energy efficiency. Cloud computing is usually employed in a larger system that enables data analytics, remote device monitoring, visualization, and client delivery [3]. This requires the IoT nodes to upload the digitized data from the sensors to the gateway/cloud. The cloud then performs data analytics and notifies specific management systems (energy, memory, and real-time OS) to take suitable actions. However, implementing the entire computation framework in the cloud would mean a higher communication payload at the edge device, which in turn leads to higher communication power. Also, the closed loop (from the sensor to cloud and back to the sensor) latency might be prohibitively large for certain cases (e.g., tactile internet, autonomous driving, and medical emergencies). The situation demands truly intelligent devices that are aware of the operating conditions and contexts and can dynamically adapt itself for optimal energy efficiency and performance by switching among different modes (computation-heavy, communication-heavy, high-security, low-power, etc.). This is shown in Figure 1b in the form of our vision for a secure, context-aware, adaptive, resource-constrained yet intelligent IoT device, represented as a combination of multiple sensing, computation, and communication modalities with different power and performance. Parts of the context information can be generated in the cloud (for latency-relaxed applications), whereas the latency-limited context assessment needs to happen in the sensor node itself, using smart learning algorithms.

Challenges in asymmetric IoT networks

Before delving into the implementation details of contextual, adaptive machine intelligence, let us discuss the specific challenges for a generic scenario of an IoT ecosystem that contains a multitude of heterogeneous connected devices (Figure 2). These IoT devices include resource-constrained and resource-rich nodes, gateways, and cloud data centers. The focus of this article is on the resource-constrained leaf nodes that are defined in [4] as the ones that do not have the hardware and software capabilities to support the Transmission Control Protocol (TCP)/IP protocol suite.

1) Finite Resources: In view of the IoT ecosystem, a resource can either be physical (such as memory, computation power, energy, and network bandwidth) or virtual (software procedures to perform data compression, outlier detection, etc.). An IoT device may lack one or more of these aspects because of size limitation and specific applications.

2) Heterogeneity: An IoT subsystem, such as a smart home, can have a significant amount of heterogeneity with respect to hardware and software [4]. Various degrees of resource constraints may coexist in the same ecosystem, which makes context-awareness a challenging task since it now becomes a function of application, device location, available computation/memory resources, channel conditions, and communication modalities. For example, smartphones with relatively high computation power can support advanced learning and data compression algorithms, while a small temperature sensor must resort to elementary learning and data processing methods [5].

3) Security: IoT systems envision automatic discovery and support for a new target device without human intervention [6], [7], which immediately raises concerns on the security and privacy aspects of the network. With the limited resources and latency constraints, proper authentication and/or authorization mechanisms become all the more challenging, as the devices with fewer resources often tend to sacrifice security for lower energy consumption.

4) Context and context-based adaptability: To optimize the performance of the individual nodes, specific contexts/modes need to be defined as discussed in the previous section, with a proper switching arrangement between modes for adaptability. The definition of context, as will be discussed later, is highly application dependent, and therefore the implementation of context-based adaptability would be different for every application and either needs to be decided beforehand by the designer or learned on the fly by the system.

5) Scalability and reconfigurability: In addition, the IoT ecosystem should be capable of handling a variable number of nodes due to the mobility and dynamic properties of the devices, and the hardware and software implementations should be scalable to a large population of devices. It is important to note that the previously described challenges also create asymmetry among the nodes in the network, as there could be a need for communication between two devices with unequal resources and capabilities. Indeed, IoT has a communication bottleneck in the uplink, as typical IoT applications (smart sensing, wearable devices, healthcare, etc.) involve uploading the collected data from multiple sensors to a single base station [5]. This asymmetry can be optimally leveraged with a high-level goal to reduce the energy consumption of the overall system, as will be explained in the following sections.

Common terminologies used throughout the article

1) IoT: Small-scale developments of internet-connected devices were materialized as early as 1982, when researchers at Carnegie Mellon University deployed a Coke vending machine with an online inventory [8]. Mark Weiser's famous 1991 paper on ubiquitous computing [9] envisioned the concept of a large scale implementation, and the term *Internet of Things* was coined by Ashton in a presentation at Proctor and Gamble in 1999 [10]. According to the International Telecommunication Union, IoT is a vision that ensures "from anytime, anyplace connectivity for anyone—we will now have connectivity for anything."

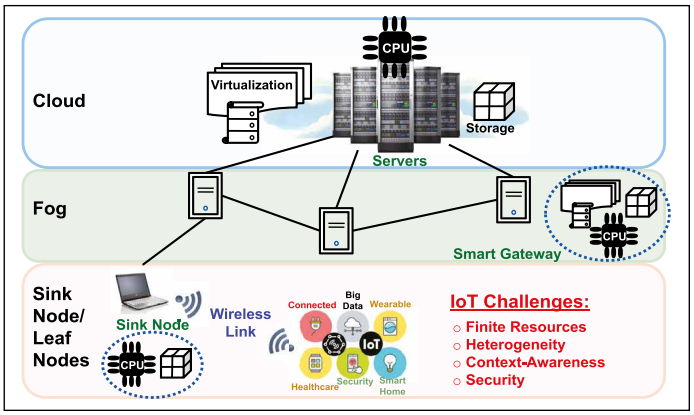

Figure 2. IoT ecosystem and its specific challenges [4].

2) Machine intelligence: Machine intelligence is usually associated with Machine learning (ML), which is defined in [11] as "the adoption of computational methods for improving machine performance by detecting and describing consistencies and patterns in training data." In view of the resource-constrained IoT (RC-IoT) nodes, however, intelligence or edge intelligence refers to the process of context discovery and assessment, which is imperative in the realization of context-aware, adaptive techniques and strategies (hardware/algorithmic/learning-based) for sensing, computing, and communication in the constrained environment.

3) Resource: Adopting the generic, all-encompassing definition [12], a resource is defined as "any object which can be allocated within a system." For IoT systems, the most important resources are memory (for storage), energy (for battery lifetime), compute capability (for computation), and network bandwidth (for communication). Depending on the available memory, RC-IoT devices are categorized into Class-0, Class-1, and Class-2 devices as shown in Table 1 [13], with Class-0 devices having the most stringent constraints.

4) Context and context awareness: The notion of context-aware computing was first introduced by Schilit and Theimer [14]. Although many definitions exist for context and context awareness, the one provided by Abowd et al. [15] is widely accepted as a concrete definition of context based on the five Ws (who, what, where, when, and why). As has been argued

Table 1. Available resources for RC-IoT Devices [13].

RC-IoT device	RAM(data Size)	ROM(code size)	Example
Class-0 (C0)	$\ll$ 10KB	$\ll$ 100KB	Single biosensor
Class-1 (C1)	$\approx$ 10KB	$\approx$ 100KB	Multi-sensor node
Class-2 (C2)	$\approx$ 50KB	$\approx$ 250KB	Local hub with TCP/IP

in [16], all previous definitions [17]–[20] of context and context-awareness suffered from the specificity of the example applications that they were referred to and could not be used to define the new context. References [15] and [16] defined context as follows:

> Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.

In light of the above definition of context, context-awareness is defined as follows: "A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task" [15]. A context is usually represented by a model and its attributes, and it is often described based on the application scenario. Further details on this can be found in [6] and [21].

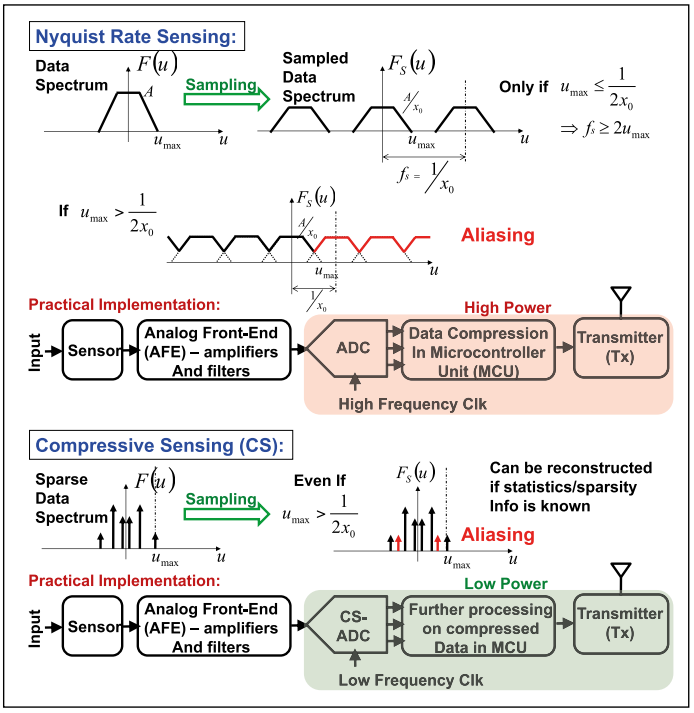

Figure 3. Nyquist rate sampling/sensing versus compressive sensing.

5) Quality of context: Quality of context (QoC) is related to how well the context model and its attributes are extracted from raw sensor data. QoC is defined using a combination of parameters such as validity, precision, and update rate of the context information, which is processed out of the raw data from a sensor. Reference [6] presents a detailed survey on QoC.

Building toward the concept of a context-aware, adaptive RC-IoT system, we shall discuss intelligent hardware techniques for sensing (compressive sensing, time-based sensing), computation (edge analytics/in-sensor analytics in the form of anomaly detection and data compression), communication [Intraphysical layer (Intra-PHY) and Inter-PHY adaptation, along with two recent sub-10pJ/b communication modalities, i.e., proximity communication and HBC], and energy management (dynamically reconfigurable LDO, switched-mode LDO, and intermittent powering) in this article. We shall also present two examples of cross-layer adaptive systems that employ more than one approach discussed in this article for optimum power-efficiency. Finally, after describing various security considerations and learning techniques for RC-IoT devices, we present our view of the current state-of-the-art and where it needs to be in the near future, based on the learnings and research in the domain of secure, context-aware, adaptive Rc-IoT nodes over the last two decades.

Intelligent sensing

Compressed-domain signal acquisition

Compressed-domain sensing/compressive sensing (CS) [22], [23] is a mathematical tool in signal processing that defies the Shannon–Nyquist sampling theorem by sampling a sparse signal at a rate lower than the Nyquist paradigm and still being able to reconstruct the signal with negligible errorrate (Figure 3). Since its inception, CS has found multiple applications including image processing [24], medical imaging [25], RADAR technology [26], in-sensor analytics [27], gesture recognition [28], [29], and healthcare [30]. CS algorithms assume that the signal to be sampled has a sparse representation, and it was shown that sparse signals with randomly [from independent and identically distributed (i.i.d) Gaussian distribution] undersampled data can be recovered with a low error by formulating it as an optimization problem. Hence, the advantage of CS is twofold: 1) CS allows a lower sampling rate that

reduces the power consumption in the analog-to-digital converter (ADC) and clock generation circuitry, and 2) compression creates a smaller amount of data with rich information-content that reduces the burden on the subsequent processing and communication modules. Since many of the naturally occurring signals such as sound, visual image, or seismic data can be represented in the sparse form [31], it is possible to leverage the superior energy efficiency of CS in an IoT scenario. Two comprehensive reviews on CS can be found in [31] and [32].

To mathematically represent CS more clearly, let us assume that the orthogonal basis $\{\psi_i\}_{i=1}^{n}$ span the n-dimensional real space $\mathbb{R}^n$. Then, any signal $x \in \mathbb{R}^n$ can be represented by matrix multiplication of the matrix ψ with the elements of a sparse vector $S = [S_1, S_2, S_3, ..., S_n]^T \in \mathbb{R}^n$ such that $x = \sum_{i=1}^{n} \psi_i S_i$. If the vector S has only $k \ll n$ nonzero entries, then the signal x is said to be k-sparse, and ψ is called the sparsifying/representation matrix for x. For CS, the $n \times 1$ input signal x is pre-multiplied by an $m \times n$ sensing matrix Φ to get an $m \times 1$ compressed signal y, where $m < n$ and the ratio (n/m) is termed as the compression factor. This is represented by the following equation and is shown in Figure 4

$$y = \Phi x = \Phi \psi S. \quad (1)$$

If the coherence (correlation) $\mu(\Phi,\psi) = \sqrt{n} \max |\Phi_j, \psi_i|$ (where $1 \le i \le n$ and $1 \le j \le m$ with Φ_j being the jth row of Φ) is low, it can be proved that fewer samples are required to reconstruct the signal [23].

Compressed domain processing and computational data converters

Multiple CS algorithms for IoT applications have been developed in the last decade. Reference [34] showed a matrix-multiplying ADC (MM-ADC) in 130-nm CMOS technology and demonstrated two applications: 1) electrocardiogram (ECG)-based cardiac arrhythmia detection (9.7× energy savings as compared to traditional ADC followed by arrhythmia detection) and 2) image-pixel-based gender detection (23× energy savings as compared to traditional ADC followed by gender detection). Feature extraction and classification were combined in a single measurement matrix (Φ in CS theory) for lightweight applications as shown in the work. Our earlier work [33] demonstrated a light-powered smart camera with CD gesture detection. To enable always on and self-powered operation on IoT devices,

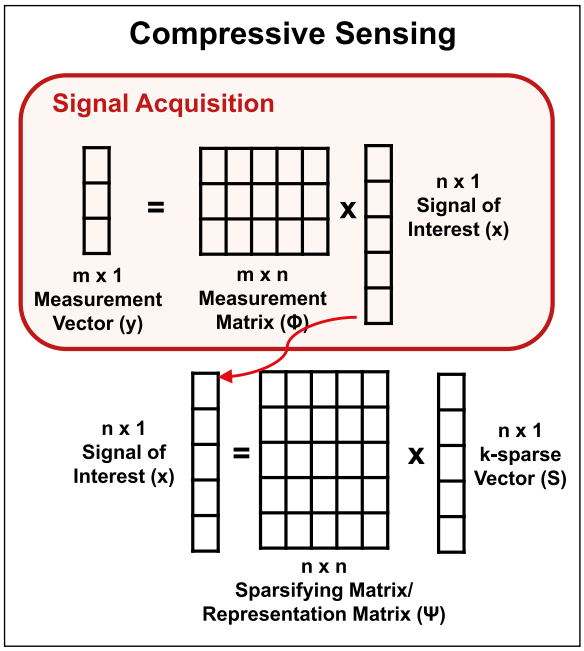

Figure 4. CS: creation of an *m* × 1 measurement vector from an *n* × 1 signal of interest (m < n).

Amaravati et al. [33] exploit CD data processing, which allows trigger detection with significantly lower power and computational requirements. This

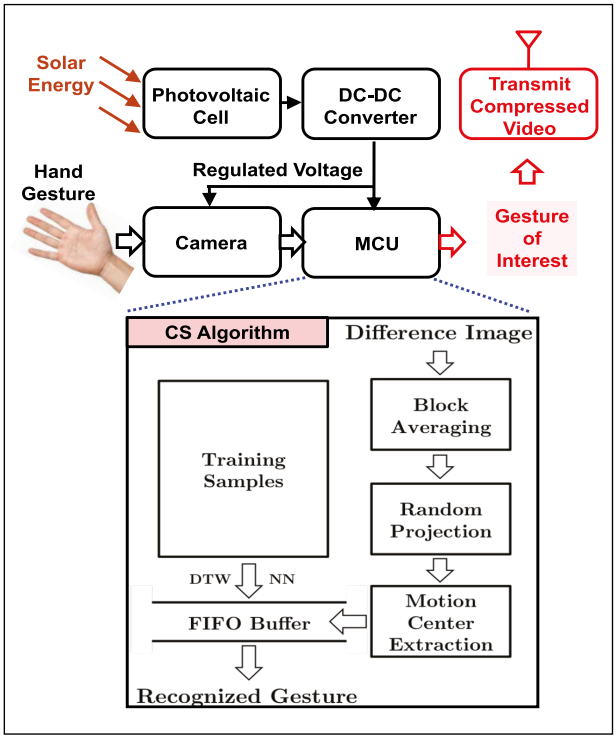

Figure 5. Block diagram of the CS algorithm for gesture detection [33].

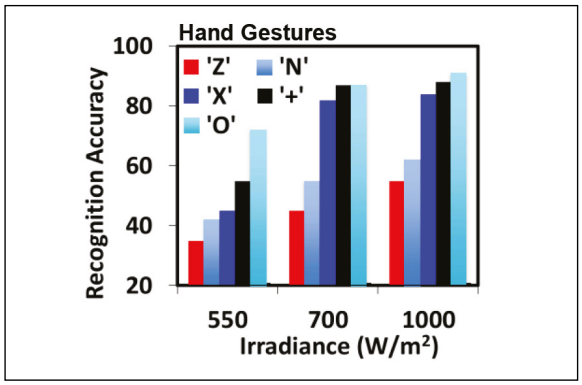

Figure 6. Accuracy of the detection of hand gestures "Z," "X," "O," "N," and "+" as a function of the irradiance level of the light-powered smart camera with CS [33].

is in contrast to existing algorithms that work directly in the pixel domain. Given the objective of the camera front end (FE), the computation complexity can be largely reduced (768×, as demonstrated in [33]) from existing algorithms that are targeted for continuous gesture recognition [35], [36].

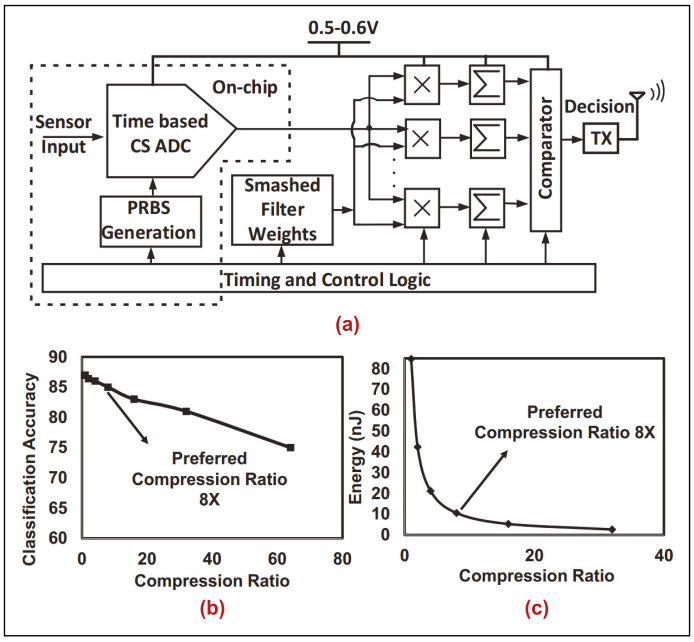

Figure 7. (a) Arrhythmia detection using time-based CS ADC with embedded classification and INL-aware training [30]. (b) Classification accuracy versus compression ratio. (c) Energy efficiency (nJ/classification) versus compression ratio, showing that a compression ratio of 8× achieves ≈90% lower energy with an accuracy hit of <5%.

In this system, the gesture motion is captured by a sequence of difference images between consecutive frames. Each difference image passes through two layers of compression to reduce its resolution and to be transferred to the CD. In the first layer, the resolution is reduced by dividing the whole image into several blocks and taking the average of each block. In the second layer, coded combinations of these block-averaged pixels are extracted. It then estimates the center of the motion directly from these compressed measurements. These motion centers are passed to a classifier for gesture recognition. Figure 5 shows the block diagram of the proposed system, while Figure 6 presents the accuracy of detecting different hand gestures as a function of the irradiance levels in the environment where the camera operates. In the work presented in [33], the sparse compression algorithm was performed in the microcontroller unit instead of the ADC. In [27], the authors have shown an ASIC implementation in 130-nm technology that utilizes CS DAC and MM-ADC together to achieve only 165-nJ/frame classification.

Figure 7 shows the results from an arrhythia detection ASIC [30] with a time-based CS ADC. A total of 160 parallel processing units were employed on-chip, and an accuracy of 84% was achieved with only 10.5-nJ energy per classification for a compression ratio of 8×. One key idea here is the introduction of computational ADCs, where analog input signals are not only digitized but also computed upon during acquisition. In particular, computational ADCs provide linear transformations of the signal in a single stage, thus improving the system energy-efficiency.

In a more recent work [37], a submicrowatt CS hardware is presented in 65-nm CMOS technology with online self-adaptivity for incoming signals with varying sparsity. Initial efforts of self-adaptivity were earlier demonstrated in [38] using an asynchronous ADC with an adjustable sampling rate and in [39] using temporal decimation and wavelet shrinkage. Both of these techniques were utilized with specific incoming signals. On the other hand, Roose et al. [37] offer a more general technique that exploits the online sensory data statistics for dynamic reconfiguration (in terms of the compression algorithm, compression harshness, and sampling frequency).

Sensing using time/frequency

Many of the naturally occurring signals in IoT are slowly varying, such as temperature, humidity,

and vibration. Most of the energy content in these signals is contained within extremely low frequencies. The resolution and dynamic range (DR) requirements for these applications, however, can be small (e.g, temperature and humidity), large (e.g, vibration), or variable based on environment (e.g, radiation). Voltage-mode and current-mode ADC designs in these scenarios become limited by the ambient noise, supply rails, and power consumption. Time-based ADCs, on the other hand, can utilize the availability of time (since signals are of very low frequency) in an energy-resolution scalable manner as shown in [40]. For high-resolution requirements, the signal to be sensed is converted into an equivalent frequency (using a resistive sensor and a ring oscillator-based resistance to frequency converter) and is simply observed (using a counter) for a longer amount of time for a change in the average frequency. For low-resolution requirements, the frequency is observed for a shorter amount of and then can be turned off (through duty cycling) for saving energy. Figure 8 shows the working principle for the time-based ADC for detecting the difference between a frequency f and its slightly modified version f_2. The minimum amount of time for which we need to observe/count the frequencies to detect the difference is $1/|f_1 - f_2|$. Hence, for a smaller $|f_1 - f_2|$ (high-resolution requirement), the time to enable the counter needs to be higher.

Even though this method ensures energy-resolution scalability within a range, the resolution cannot be made infinitely high by waiting for a longer time. The ambient noise statistics, process, voltage, and temperature (PVT) variation, and jitter accumulation in the ring oscillator would limit the achievable resolution, out of which jitter accumulation is shown to be the dominant factor in [40] in a controlled environment for radiation measurement. This is demonstrated in Figure 9, where it is shown that the scaled quantization error in measuring a fixed frequency within a predefined amount of time goes down with the time of measurement. However, the accumulated jitter from the ring oscillator goes up with the total time of measurement. If the slope of the linear plot of accumulated jitter versus measurement time is k, then the achievable resolution is shown to be limited to $\log_2(1/k)$ bits.

The system in [40] achieves 18-bit resolution with 861-nW power consumption (one reading per second) and 12-bit resolution with 9.04-nW power

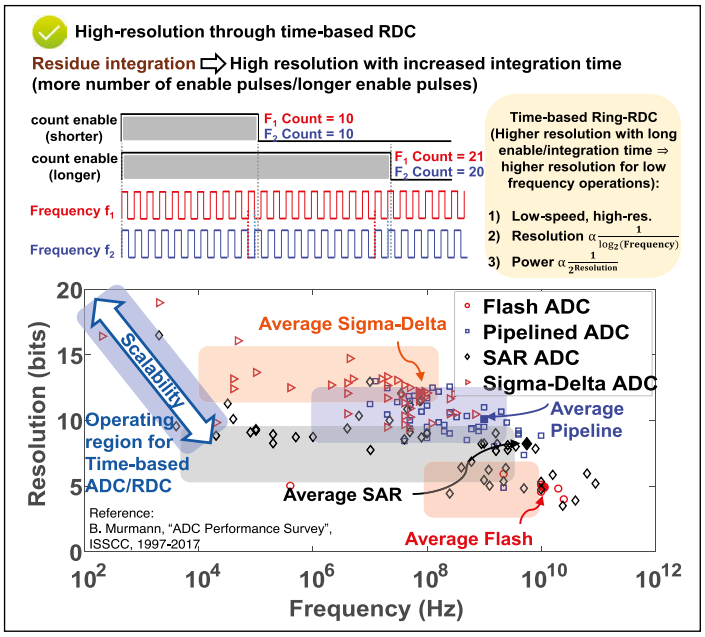

Figure 8. Working principle of time-based ADC: higher resolution with more integration time and frequency/energy-resolution scalability as compared to traditional ADCs [40].

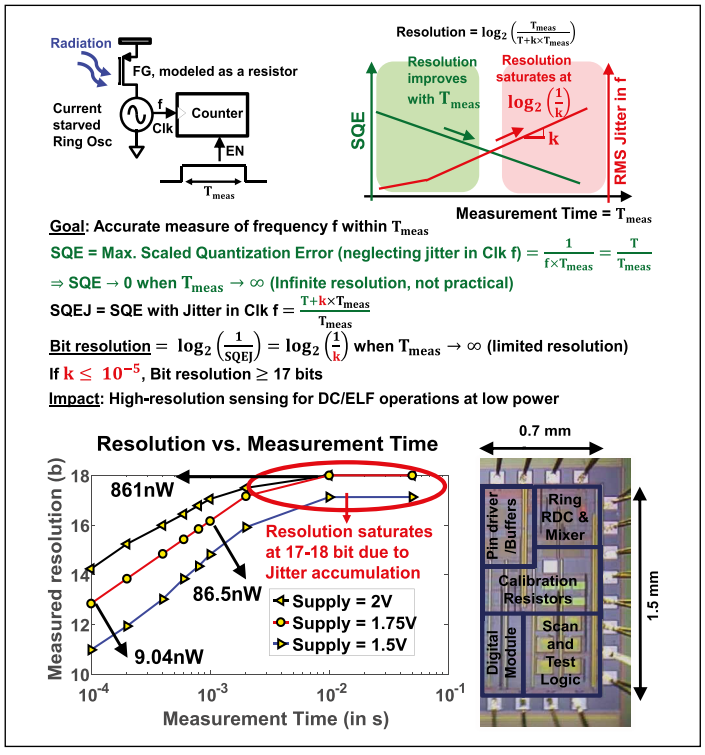

Figure 9. Application of time-based ADC in radiation sensing [40] using a resistive floating gate sensor, a three-stage differential ring oscillator, and counters. 18-bit resolution is achieved with 861-nW power, utilizing the tradeoffs among measurement time, bit resolution, and accumulated jitter.

consumption (one reading per second). The resolution can be improved by phase noise reduction techniques for the ring oscillator at the cost of a higher power.

Collaborative sensing

Collaborative wireless sensor networks [41]–[43] can sense an analog signal over a large-area test bed (e.g., soil nitrate sensing for smart agricultural application) utilizing collaborative efforts among the sensor nodes and their communication with each other/with the cloud, which brings us to the next part of the article—the tradeoffs and power optimization among in-sensor computation, short-range communication, and long-range communication.

Intelligent computing platforms

As the number of distributed sensors and IoT end-nodes are increasing, the total amount of data transfer to the backend cloud servers are becoming prohibitively large, resulting in network congestion and high energy consumption during data transmission at the sensor node [44]. This motivates the need for in-sensor data analytics that would perform context-aware data acquisition with compression, followed by transmit if necessary.

Need for intelligent computing

The computation and communication energies (E_{comp} and E_{comm}, respectively) in a system can be written as

$$E_{\text{comp}} = (E_{\text{comp}}/bit) \times \text{Number of bits switched}$$

$$E_{\text{comm}} = (E_{\text{comm}}/bit) \times \text{Number of bits sent.} \quad (2)$$

(E_{comp}/bit) will be dominated by the dynamic power for frequencies above the leakage-dominant region, as shown in [45] and [46], and, hence, can be approximated by $(E_{\text{comp}}/bit) = CV^2$, which scales with technology. If an ideal technology had allowed zero device capacitances, then (E_{comp}/bit) would be very close to the theoretical limit posed by the Landauer principle [47] as given in the following:

$$(E_{\text{comp}}/bit)_{\text{th\_min}} = k_B T \times \ln 2, \quad (3)$$

where k_B is Boltzmann's constant and T is the ambient temperature. For room temperatures, $(E_{\text{comp}}/\text{bit})_{\text{th\_min}}$ is calculated to be about 2.9×10^{-21} J. For a standard 45-nm CMOS technology node, the bit switching energy was simulated to be ≈1fJ for this analysis. However, even if a fictitious technology could potentially offer zero capacitances, a zero-power receiver (Rx), and 100% efficiency for the transmitter (Tx), (E_{comm}/bit) would still be limited by the free-space path loss (PL_{FS}) of the physical channel, which is given by Frii's equation [48], [49] and shown in the following:

$$PL_{FS} = G_{Tx} \cdot G_{Rx} \left(\frac{\lambda}{4\pi d}\right)^m, \quad (4)$$

where G_{Tx} and G_{Rx} represent the gains of the transmitting and receiving antennas, respectively, λ is the wavelength, d is the distance between the Tx and Rx, and m is a parameter (typically between two and three) that represents the fading margin. For $d = 10$ m and a typical ANT protocol operating at 915 MHz, the most optimistic PL_{Fs} ($m = 2$, $G_{Tx} = 2$ dB, $G_{Rx} = 2$ dB) turns out to be about 48 dB, which means, with a state-of-the-art Rx sensitivity of –100 dBm [50], the Tx needs to transmit a minimum of –52 dBm. This translates to a power consumption of 6.3 nW (theoretical limit—assuming no capacitance and 100% efficiency) and an energy/bit of 105 fJ/b (for a maximum data rate of 60 kbps for ANT), which is 10^7 times higher than the theoretical minimum energy/bit for computation, as given by the Landauer principle.

From the foregoing analysis, the theoretical minimum energy/bit for communication is given by the physical limits of the channel, and can be written as

$$(E_{\text{comm}}/\text{bit})_{\text{th\_min}} = \frac{Rx_{\text{sen}}}{\left(G_{Tx} \cdot G_{Rx}\left(\frac{\lambda}{4\pi d}\right)^n\right) \times \eta . DR}, \quad (5)$$

where $Rx_{\text{sen}} = k_B T_{50\Omega} \times NF \times SNR \times k \times DR$ is the Rx sensitivity as a function of DR [51], η is Tx efficiency, and DR is the data rate supported.

Figure 10 shows the comparison of E_{comm} and E_{comp} for the same number of bits transmitted, or switched. The state-of-the-art wireless transceivers [52] consume ≈10^4 times more energy as compared to computational bit switching in 45- and 65-nm nodes. This bottleneck analysis directly signifies that some amount of intelligent computation at the sensor node (in-sensor analytics) would help in bringing down the total energy by enabling selective data transmission, which will reduce E_{comm} at the cost of additional E_{comp}.

In-sensor analytics as a form of edge intelligence

Based on the communication and computation energy tradeoffs and the amount of resources

available at the RC-IoT device, partial or complete processing of the sensor data (e.g., anomaly detection and data compression for sensor readout, and object localization and segmentation for video surveillance) can take place in the leaf node itself. In this section, we discuss the two most common ISA techniques for RC-IoT devices, namely, anomaly/outlier detection and data compression. The anomaly detection methods can enable selective (and immediate) data transmission when an anomaly occurs in an otherwise normal sensor readout. As a healthcare example, selective ECG data transmission with arrhythmia (anomaly) detection would ensure immediate notification with minimum communication cost. Data compression, on the other hand, would ensure that the maximum amount of information between transmissions can be stored in a small amount of on-sensor memory.

1) Anomaly/Outlier Detection: According to Barnett and Lewis [69], "an outlier is an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data." Figure 11 shows an example of anomaly in a sensor readout and explains the three classes of anomaly that are common in IoT devices and wireless sensor networks [70]. It is to be noted that the primary difference between outlier detection and event detection is the fact that an outlier is detected by comparing the readings from the sensor with each other and without any prior semantics that define the trigger conditions of an anomaly. On the other hand, trigger conditions for event detection are usually defined a priori, and the sensor readouts are compared with that trigger condition to detect an event.

Outlier detection algorithms utilize *spatio-temporal correlations* among the data points from the same node and/or neighboring nodes to distinguish between normal operation and anomalies [70]. Table 2 shows some of the most common anomaly detection techniques for wireless sensor networks and IoT. These methods include both learning (supervised and unsupervised)-based techniques and algorithmic (statistics-based) techniques and offer various orders of resource requirements and accuracy. Some of the most recent works include a hybrid statistical method from Twitter [71], which has low latency and high accuracy but needs more computational resources. Simplistic techniques such as mean- and average-based statistical analysis [53], on the other hand, can be implemented easily on the RC-IoT device itself for optimum computation-communication tradeoff.

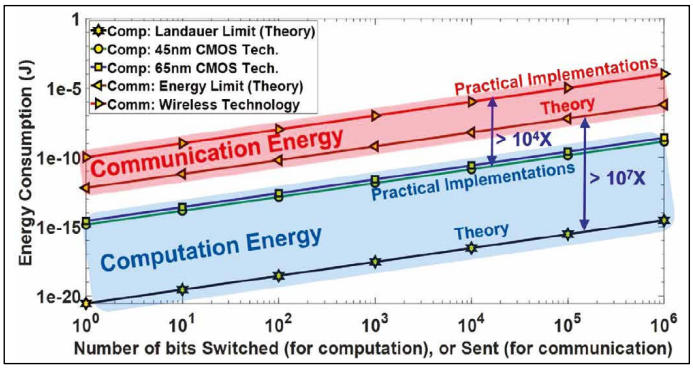

Figure 10. Comparison of communication and computation energies (both theoretical and from standard implementations [52]) that show that communication energy is ≈104 times more than computation energy (with same number of bits). Leakage power is ignored in the analysis.

2) Data Compression: As shown in the "Intelligent Computing Platforms" section, compressive-sensing techniques can result in significant energy savings in the ADC, on-sensor processor, and communication modules. It must be noted that CS-ADC is still an emerging technology and has not yet become an integral part of commercially available embedded frameworks. In-sensor data compression techniques on the IoT processor, however, have also shown energy benefits by bringing down the communication power. Some of the earliest reported works on the tradeoff between the raw data communication and the compressed data communication are from MIT's Computer Science and Artificial Intelligence Laboratory [72] and from CMU's Odyssey Project [73]. Reference [73] used application-aware adaptation that trades off data quality with resource consumption

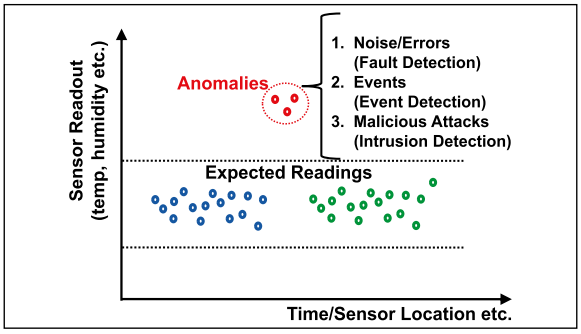

Figure 11. Example of anomaly in sensor readout.

Table 2. Outlier detection techniques for wireless sensor networks/IoT.

Technique	Example	Salient features	Drawbacks
Statistics-based	1) Gaussian Parametric Estimation [53], 2) Non-Gaussian Parametric Estimation [54], 3) Kernel Density Estimation (non-parametric) [55], [56], and 4) histogram-based method (nonparametric) [57]	1) Spatiotemporal correlation for Gaussian nonanomalous data, fixed thresholds for anomaly detection, 2) anomalies are treated as SaS-distributed impulsive events, 3) no a priori PDF is assumed. Kernel density functions approximate the PDF, and 4) works on histograms and not on raw data (inherent compression—reduced communication cost)	Simplistic and can suffer from low accuracy
Nearest-Neighbor-based	Euclidean distance [58] and dynamic time warping methods [59]	Simple implementations for both univariate and multivariate data	Resource-extensive for multivariate data
Clustering-based	Creates clusters based on raw data and detects outliers that do not fall into any cluster [60]	Can be employed to take care of incremental processing	Resource-extensive for multivariate data, suffer from the choice of an appropriate cluster width
Classification-based	1) SVM approach [61], [62], 2) Bayesian Network approach [63], [64], 3) long short term memory (LSTM) [65]/hierarchical temporal memory (HTM) [66] approach	1) Maximally separated classes (one/two class approach to reduce complexity, 2) uses Bayesian Intuitions to predict anomalies, and 3) uses LSTM/HTM for time-series data pattern of unknown length	Computationally intensive
Spectral-decomposition-based	PCA-based approach [67], [68]	Added advantage of dimensionality reduction/data compression	Selecting suitable principal components is computation-heavy

with the help of an embedded OS. Reference [72] experimentally showed that the ratio of energy required to transmit 1 bit is ≈480–1270 times higher than that of a 32-bit addition under varying channel conditions. This means that a compression algorithm that is able to remove more than 1 bit from a string of data would have energy benefits if the algorithm is equivalent to (or less than) 480 addition instructions. The standard compression algorithms explored in [72] [such as bzip2/Burrows–Wheeler transform (BWT), Lempel–Ziv–Welch (LZW), Lempel–Ziv–Oberhumer (LZO), and prediction by partial matching (PPMd)] are much smaller than 480 additions, which means that any of these algorithms would be beneficial. However, the key limitation in an IoT implementation comes from the runtime memory requirement for these algorithms, which is in tens of kilobytes for LZO to hundreds of kilobytes for BWT. This readily makes these algorithms infeasible for C0, C1, and C2 RC-IoT devices (referring to Table 1). More lightweight compression techniques, such as miniLZO and sensor LZW with mini cache (S-LZW-MC) [74], require only 8.192 and 3.250 Kbytes memory, respectively, and can be used in C2 and some C1 devices. Other important techniques for data compression in IoT devices include 1) principal component analysis (PCA) ([75]–[77], which use lightweight PCA for dimensionality reduction and data compression), 2) coding by ordering ([78], where the data from one node is shown to be encoded by the order at which other nodes in the same hierarchy communicate with the parent node), 3) burst mode/pipelined techniques ([79], where data are stored, packetized, and transmitted in burst mode to remove redundancies and number of transmitter switch on/off), 4) frame difference-based compression ([33] and [80] that store differences in consecutive frames for video compression), and 5) distributed data compression [81] using conditional entropy encoding with correlated data between two nodes that perform spatial data compression through short-range communication between the sensor nodes. For optimum resource utilization, this short-range communication can be a low-power communication scheme, such as MedRadio or HBC (for body area networks within a few meters), which consumes hundreds of microwatts, or ANT/BTLE, which consumes a few milliwatts to ≈10 mW when ON, as will be shown in the next section. After spatial compression is done, we envision that the node with the highest amount of battery life

Table 3. Comparison of state-of-the-art wireless techniques for IoT nodes [82].

	Proximity comm. [83]	HBC [84], [85]	NFC	ZigBee	BT/BTLE	ANT	WiFi	LoRa WAN
Distance	1 mm	2–5 m through human body	10 cm	10–100 m	10–100 m	10–20 m	30–50 m	≈1 km
Data rate (bps)	8–32 G	10's of M	20–400 k	20–200 k	0.8–2.1 M	60 k	300 M (802.11 g), 7 G (802.11ac/11d)	200 k
Energy efficiency (J/b)	4 p	6.3 p	1 n	5 n	15 n	10 n	5 n	1 μ
Security	High	High	Medium	Low	Low	Medium	Medium/High	Low/Medium

(or the node that is closest to the Rx) would take the responsibility of sending the compressed data to larger distances, possibly through a high-power communication protocol such as LoRa WAN.

Intelligent communication

Continuous device scaling over the last few decades have resulted in cheap computation through Moore's law, and the ability to support higher data bandwidths has created cheap wireless communication paradigms through Shannon's law. However, the progress in battery technology has been relatively slower, making the available energy one of the most sought after resources in modern IoT systems, thereby motivating the research needs toward low-energy sensing, computation, and communication. As supported by the analysis presented in [86] and in the "Intelligent Computing Platforms" section, the energy cost per bit for communication is 10^3–10^4 times higher than the energy cost of computation for raw data bits. In the vision of the truly intelligent IoT nodes presented in this article, most optimum energy efficiencies are expected from the communication subsystems based on the specific operating conditions/context (such as communication distance, channel conditions, latency, quality of service requirements, data rate, battery conditions, and process variation) when turned on. Table 3 shows the state-of-the-art communication modalities available for IoT devices, which range from 4-pJ/b proximity communication for ≈1-mm distance to 1-μJ/b long-range (LoRaWAN) communication to ≈1 km. We readily notice the possibility of optimizing the communication framework within a modality and among different modalities, hereinafter called Intra-PHY and *Inter-PHY communication* as explained in [2]. The concept of Intra-PHY and Inter-PHY communication is presented in Figure 12a, where the switching of PHY is shown to occur based on communication distance (as an example of context), while the adaptation within a PHY is performed for optimum energy efficiency based on the operating conditions.

Intra-PHY channel-adaptive radios

For Intra-PHY adaptation, the energy-performance tunability knobs are dynamically optimized without changing the PHY. Traditional techniques of scaling the energy consumption over varying channels involve adaptive modulation and coding [87], which increases the order of modulation (from QPSK to 16-QAM to 64-QAM) as the channel quality becomes better and corresponding error vectors become more and more manageable. Although this increases the spectral efficiency of overall transmission, the power consumption of the radio frequency (RF) FE effectively remains constant. As shown in [88], 70%–90% of the overall power in a low-power transceiver (Tx+Rx) system is consumed in the Rx FE/Tx power amplifier (PA) and LO generation subsystems, and, hence, significantly more energy efficiencies can be obtained by dynamically scaling the FE power and performance according to the application. Most of the research efforts in building channel-adaptive designs are concentrated toward the Tx PA and employ techniques such as digital predistortion, Tx power control, envelope tracking, polar implementation, and dynamic companding with PA bias control [89], [90]. Rx circuit-level adaptation techniques include automatic gain control and field-programmable low-noise amplifiers (LNA) with power-linearity tradeoff [91]. Some of the recent advancements include an adaptive DR and BW Rx [92] that use a

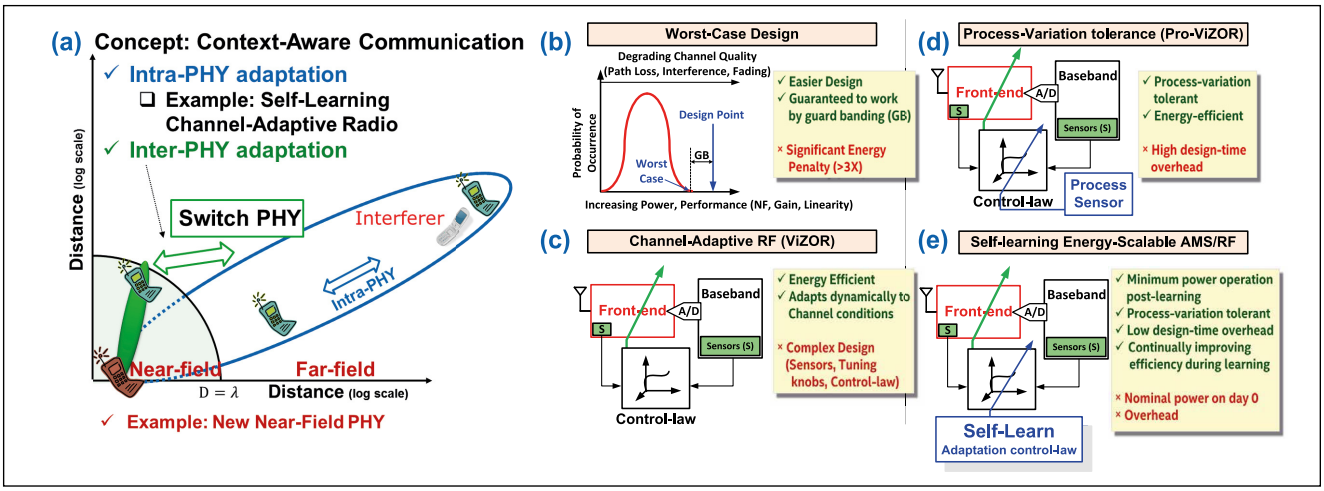

Figure 12. Vision for adaptive communication in IoT [2]. (a) Context-aware communication PHY, which can adapt to its surroundings to perform more efficiently with experience by self-learning the optimum operating points. Adaptation can be intra-PHY or inter-PHY based on context, indicating the need for incorporating incorporating multiple adaptive PHYs per device. (b) Today's worst case design philosophy. Circuits/systems are generally designed to handle the worst case conditions plus a guard band. This leads to significant loss in energy efficiency. (c) Dynamically channel-adaptive radio (ViZOR). (d) Process-variation tolerant ViZOR (Pro-ViZOR). (e) Self-learning energy-scalable wireless systems.

programmable gain amplifier (PGA) and an adaptive intermediate-frequency filter. Discrete-time spectrum sensing was utilized in [93] to modify the modes of an Rx filter to achieve adaptive interference rejection. An interference-aware adaptive ADC was shown to adapt itself to a low-power mode in absence of any blocker using a simple built-in spectrum analyzer [94]. A channel adaptive ADC and a successive-approximation-register-based time-to-digital converter (TDC) was shown for a 28-Gbps wireline system in [95]. These implementations have shown benefits in the standalone adaptive subsystems (such as the LNA, PGA, or TDC). However, it must be noted that, unlike a Tx where most of the power is consumed in the PA, the Rx power consumption is more distributed among different blocks; hence, the entire Rx should be considered as a unit for power-performance tradeoff analysis. It was shown in [96] and [97] that the best-case energy-savings in an Rx FE can be obtained by distributing the instantaneous performance-slack optimally across different building blocks in the Rx. A precharacterized control law (defined during design) was employed to achieve multidimensional adaptation of multiple Rx components with virtually zero-margin (ViZOR) Rx operation. Figure 12c shows the operation of ViZOR using design-time tuning knobs and sensors in the Rx FE to dynamically optimize power and performance. If the tuning knobs are designed in an orthogonal manner (i.e., operation of one knob will modify only one specification out of linearity, gain, and NF of the FE), the controller was shown to achieve ≈3× better energy savings for best-case channel conditions [98], [99] and can be optimized for either maximum data-rate (data-priority) or minimum energy (energy-priority) for any channel [100], [101]. However, it was also reported that the adaptation control law, which was fixed during design time, cannot work optimally under manufacturing process variations. References [90] and [102] solve this problem by detecting the process corner of the device under consideration using built-in process sensors and updating the control law accordingly during postmanufacturing tuning. This technique (i.e., Pro-ViZOR) is shown in Figure 12d and requires high design-time effort to cover the entire process-corner space for the power-performance adaptation [103].

Intra-PHY adaptation: Self-learning radios

The high design-time complexity of Pro-ViZOR was significantly reduced by employing self-learning wireless systems (Figure 12e) that gradually learn the adaptation control law when the device is in idle condition [104]. Figure 13 shows how the learning

algorithm populates the power-performance channel space during an intermediate time instant and at the final time instant when learning is complete. Figure 14 presents the average power consumption of such a self-learning channel-adaptive wireless system over multiple days to include various channel conditions. The initial overhead is due to the need for controlled on-line experiments during real-time operation that gathers useful data points for learning the control law. It is shown that this system becomes increasingly energy-efficient with experience [105]–[107]. When the power consumption saturates with the learning, the overhead of controlled experiments is removed (day 29 in Figure 14).

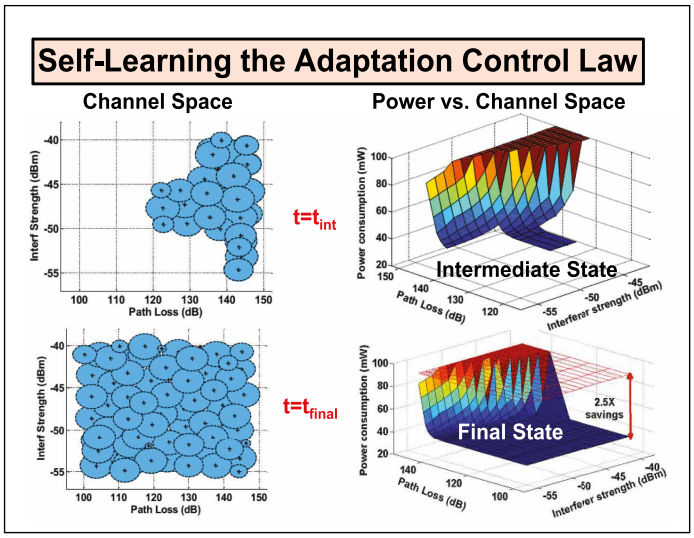

Figure 13. Self-learning the adaptation control law: learning of channel space and power profile at intermediate and final time instances [2].

Communicating with ultralow-energy (<10 pJ/b) PHYs

Today's wireless technology is limited by the high channel losses (≈60–80 dB in standard operating conditions) that increase the power consumption in the Tx to compensate for this channel loss. In addition, narrowband wireless techniques employed in standard implementations involve frequency upconversion (Tx) and downconversion (Rx), which increase the power overhead to enable smaller antennas and multiplexing.

Due to these reasons, traditional wireless techniques such as near-field communication (NFC), Zigbee, BTLE, ANT, and Wi-Fi can only achieve a best-case energy efficiency of ≈1 nJ/b, while low-power wireless body-area networks (WBANs) and MedRadio implementations usually operate at hundreds of pJ/b for short-distance (1–5 m) communication [2]. However, recent progress in wireline-like broadband techniques can enable sub-10-pJ/b communication over low-loss channels (mm-scale device proximity communication, or meter-scale data transfer through the human body), as shown with two examples in this section. Wireline-like techniques eliminate the need for antennas as well as modulation, thus lowering the power consumption dramatically. However, communicating with multiple devices would now require time division multiplexing instead of frequency division multiplexing, thereby increasing the latency if proper scheduling techniques are not employed.

1) mm-scale proximity communication: Figure 15 demonstrates the modality for mm-scale multi-Gbps proximity communication [83], [108] at ≈4 pJ/b. Proximity communication is implemented by employing metal plates (couplers) in both communicating devices, which are brought in close proximity (≈1 mm) to establish connectivity through capacitive coupling. Because of the antennaless capacitive terminations in both the devices, the channel behaves like a simple capacitive divider with low loss and maximally flat frequency response, thus enabling broadband signaling. One key challenge in this mode of communication arises from the self-resonance frequency (SRF) of the inductive vias with the distributed parasitic capacitance of the coupler plates. The SRF creates a peak in the channel frequency response, thus disturbing its otherwise flat

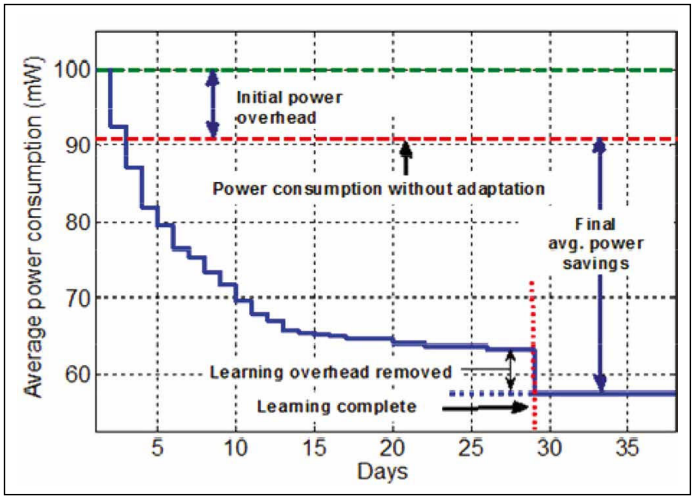

Figure 14. System behavior and power savings for self-learning radio.

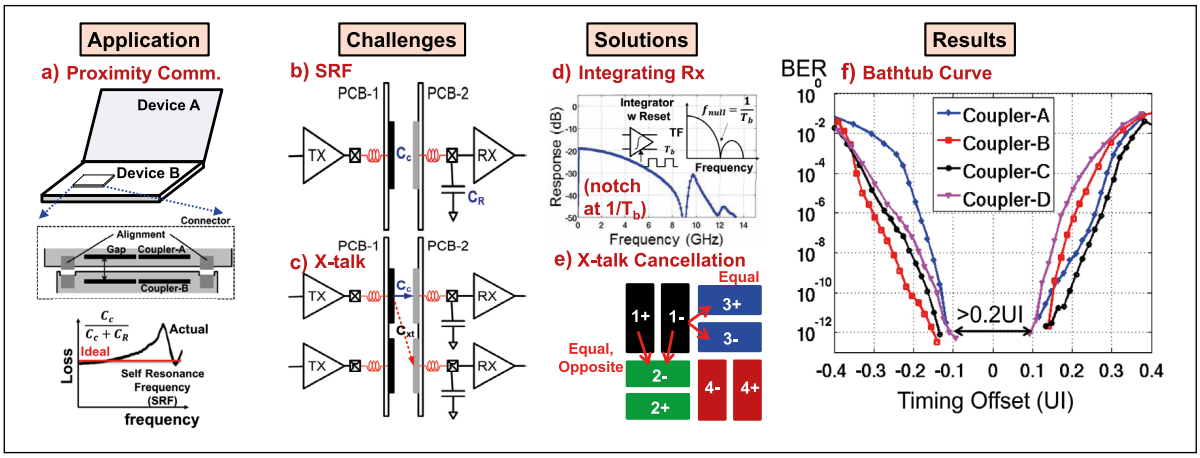

Figure 15. Emerging PHY example 1. (a) mm-scale proximity communication [83]. (b)–(c) Specific challenges (SRF of the interface and crosstalk). (d)–(e) Their solutions [integrating dual data-rate (DDR) Rx that creates a notch at the SRF to mitigate ringing and alternating rectangular differential couplers to mitigate crosstalk]. (f) Measurement results showing BER of <10^{-12}.

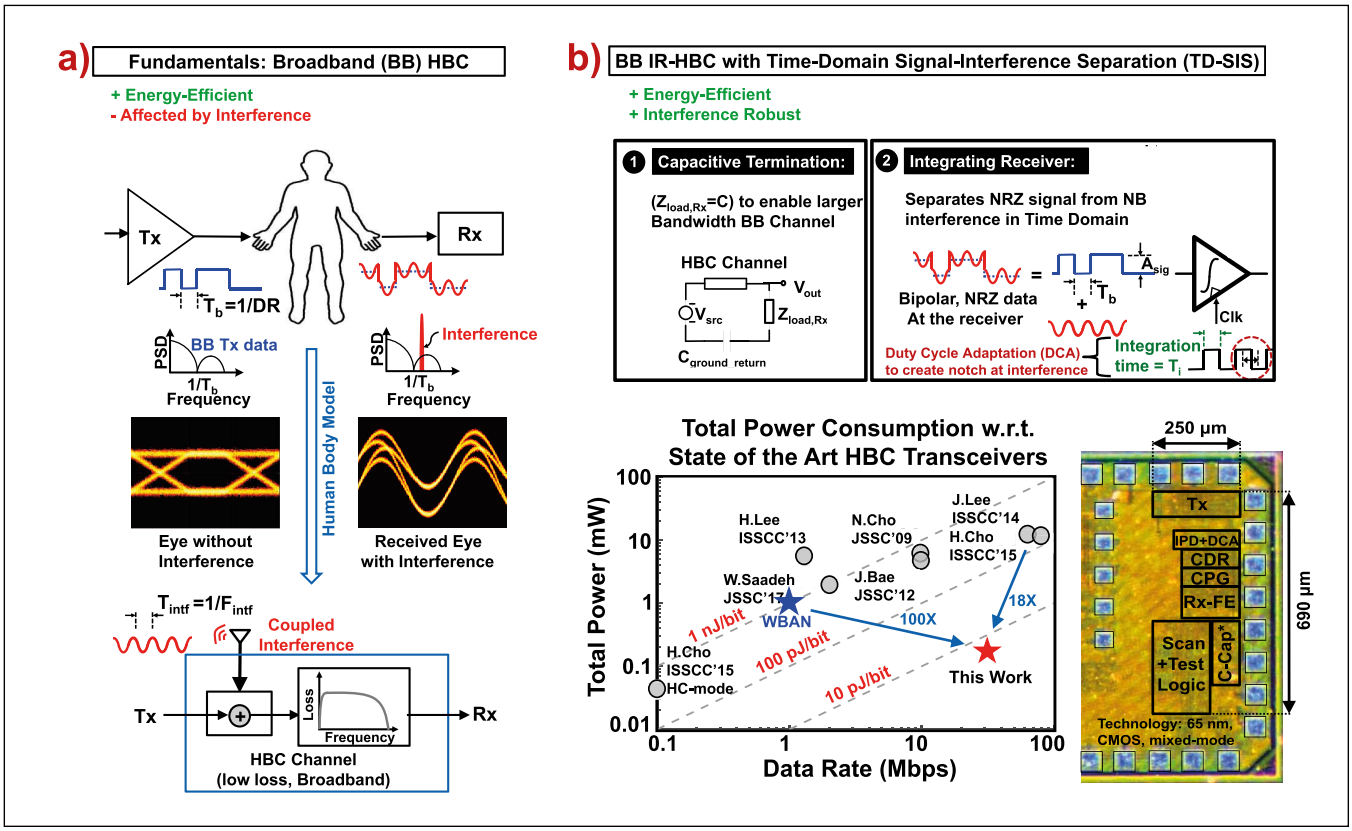

Figure 16. Emerging PHY example 2. (a) Broadband HBC is affected by interference from the environment. (b) IR-HBC using time-domain signal-interference separation [84], enabled by 1) capacitive termination (offers larger frequency range for broadband application) and 2) integrating DDR Rx for interference rejection using signal-interference separation and duty-cycle adaptation. In comparison with state-of-the-art HBC transceivers, broadband IR-HBC achieved 18× better energy efficiency (6.3 pJ/b), which is ≈100× better than traditional WBAN.

nature. An integrating receiver (Rx) with a tunable notch placed at the SRF can solve this problem as illustrated in [83]. The other key challenge in this technique is the crosstalk between parallel channels, which is solved using alternating rectangular differential couplers that employ inherent passive crosstalk cancelation. As shown in Figure 15e, the crosstalk from 1+ and 1– to 2– is equal and opposite, and hence, cancels each other, while the crosstalk from 1– to 3+ and 3– are equal and therefore cancel each other differentially. Using these two techniques, Thakkar et al. [83] successfully demonstrate 32-Gbps data transfer with bit error rate (BER) $<10^{-12}$ using four parallel channels up to 0.8-mm distance and 4-pJ/b energy efficiency (which is ≈100× lower than contemporary mm-wave gigabits-per-second implementations [109], [110]).

2) Interference-Robust Human Body Communication (IR HBC): Many future healthcare [111], human–computer interaction [112], [113], and neuroscience applications rely on the Internet of Body (IoB), to connect wearable and implantable devices on, in, and around the human body, which are typically interconnected though WBAN, consuming upward of 1 nJ/b. Using the human body itself as a low-loss broadband communication medium [114]–[116], energy efficiencies [84], [117] similar to the proximity communication, or wireline input–output (IO) [118]–[120] achieve high *physical security* [85]. Capacitive termination along with voltage-mode signaling allows broadband communication in which low loss and absence of upconversion and downconversion give rise to the extreme energy efficiencies. The key challenge in broadband HBC comes from the antenna effect in the human body that picks up unwanted interferences that corrupt the signal. An interference detection and rejection loop using an adaptive notch [121], [122] at the integrating Rx has enabled the lowest energy (6.3 pJ/b for 30-Mbps data transfer through the body, which is ≈100× lower than traditional WBAN), as well as the most interference robust (can tolerate –30-dB signal-to-interference ratio) HBC transceiver built to date [84], as shown in Figure 16.

Inter-PHY adaptation: Communication with context switching

Like humans, a truly intelligent RC-IoT node needs knowledge (context-awareness) and adaptation according to the situation (reconfigurability). Inter-PHY context-aware adaptation is most effective when multiple PHYs with different orders of energy

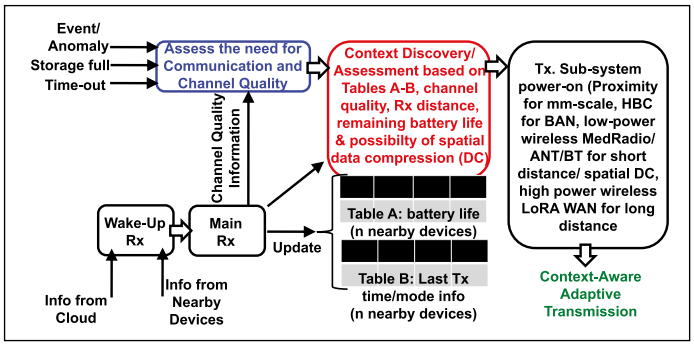

Figure 17. Vision for context-aware adaptive PHY. The IoT node needs to store minimal information on nearby devices (last transmit time, mode information, and battery life) along with its own remaining battery life. In case of an event/anomaly detection, if the sensor storage is full (even with data compression) or there is a transmit timeout, the node would then asses the context and turn the corresponding transmit subsystem on. If a change in context requires change in Tx modality, it will be taken care of by the context discovery/assessment block, which can employ a structured algorithm/learning framework.

efficiencies, channel loss, data rates, and distance support are incorporated in the same transceiver. Figure 17 presents the vision for a context-aware adaptive PHY that involves the following:

- assessment of the need for communication based on event/anomaly detection (in-sensor analytics), memory (storage) buffer information, and channel quality information,
- context discovery and assessment based on battery life of current and nearby devices (helps to understand which device has the most resources for long-range high-power communication, if required),
- last transmit time and modality of current and nearby devices (helps to understand the spatial statistics of the data and the sensors), Rx distance and location (e.g., whether both the Tx and Rx devices are on the human body),
- the possiblity of spatial data compression based on the information from nearby devices (if successful, this will require long-range data communication for only one node among a cluster of sensors), along with any other information from the cloud.

Equipped with all the knowledge, the RC-IoT device can now adapt itself to the context and transmit

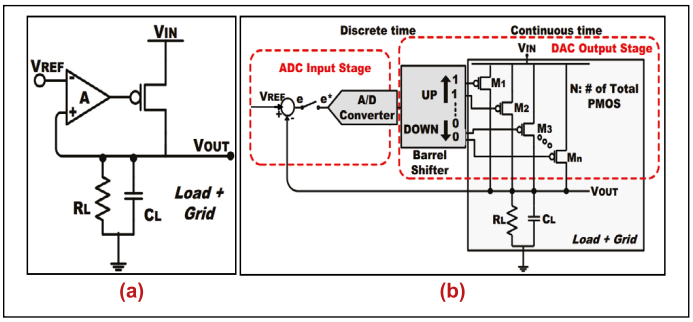

Figure 18. (a) Analog and (b) digital LDO [123].

using the available modes (proximity communication for mm-scale, IR-HBC for body-connected devices, low-power wireless/MedRadio for short distances within 5 m, ANT/BTLE for distances up to tens of meters, and high-power LoRa WAN for distances >100 m).

Intelligent energy management

By exploiting aggressive technology scaling and low-power design techniques, IoT nodes are continuously trying to reduce the overall energy consumption. Context-aware and adaptive future IoT devices would require different power supplies for different modes of operation, thereby necessitating an adaptive energy management unit to handle different scenarios. For example, in-sensor analytics (which are performed in a digital on-chip/on-board processor) can utilize minimal supply voltages for low-power, near-threshold (≈0.35 V for 14-nm or lower nodes) operation, while the high-power RF Tx for long-distance communication may need a higher supply voltage (≈1 V) for high-power delivery. Short-range communication, on the other hand, may require a different supply voltage that falls between the minimum and maximum values. Some of these modules may be connected to the same power delivery network (PDN) (e.g., the entire communication module with multiple modalities can be supported by only one on-chip voltage regulator), thereby offering a variable load to the power management unit (PMU). The PMU may even be supported by a dynamic source if opportunistic energy harvesting is used to supplement the onboard battery [123]. Hence, the PDN in the context-aware, adaptive RC-IoT scenario should be able to support 1) a wide DR, 2) high-power conversion efficiency throughout the range, and 3) a platform and interfacing circuitry for optimum power transfer to the load with minimum losses. LDOs have been traditionally used in CMOS ASICs to provide ripple-free constant voltages, the analog and digital implementations of which are shown in Figure 18.

The power conversion efficiency (η) of an LDO is given by

$$\eta = \frac{V_{\mathrm{OUT,LDO}}}{V_{\mathrm{IN,LDO}}} \times \frac{I_{\mathrm{Load}}}{I_{\mathrm{Load}} + I_{\mathrm{Control}}}, \tag{6}$$

where $V_{\mathrm{OUT,LDO}}$ and $V_{\mathrm{IN,LDO}}$ are the output and input voltages from the LDO, I_{Load} is the load current drawn from the LDO, and I_{Control} is the controller current consumption. It has been shown in [124] that a digital LDO can offer fast switching at low controller currents, apart from being synthesizable and process/voltage scalable as compared to its analog counterpart.

Dynamically reconfigurable power conversion LDO

To meet the requirements of large DR and high efficiency, which effectively remains constant over the DR of operation, a reconfigurable digital LDO with sampling rate adaptation was demonstrated in [124] and [125], using an IBM 130-nm CMOS technology. The design comprises of a 128-bit barrel shifter, controlling 128 identical P-MOSFETs that provide line and load regulation at the output side (Figure 19a). A clocked, sense-amplifier-based comparator compares the regulated voltage with a reference voltage and, depending on the result of the comparison, increases or decreases the number of P-MOSFETs supplying current to the load. It has been explained in [125] that a linearized control loop model for the LDO has two open loop poles—one at DC (from the integrator) and the other at a frequency determined by the ratio of the load's pole frequency to the sampling frequency of the controller. As the load current dynamically changes in different IoT scenarios, the pole from the load and, hence, the second pole for the LDO would also change in a baseline design without adaptation, leading to overdamped behavior in heavy-load condition, and oscillatory behavior in light-load condition, which reduces the current efficiency drastically. However, if the sampling frequency of the controller is also modified according to the load current (this information is obtained from how many P-MOSFETs are on in the LDO), it has been shown that an effectively constant current efficiency can be maintained. This technique for adaptation has two significant system-level advantages: 1) the closed-loop system poles are bounded within a certain range, leading to stable and consistent system behavior over a large DR, and 2) as the sampling frequency is lowered for light-load

condition, the digital controller's power also scales, leading to an improved current efficiency (Figure 19c). The overall design achieves >80% peak-current efficiency over 0.45–1.14 V, with 0.1–4.60-mA load current range.

Switched mode control (SMC) for LDOs

To achieve 1) fast transient response across a wide current and voltage range, 2) rapid droop mitigation, and 3) dynamic switching by exploiting decoupling between small signal (SS) gain and large-signal (LS) transient behavior, Nasir et al. [126] demonstrated an SMC-based hybrid LDO (analog LDO with SS control and digital LDO with large signal control) as shown in Figure 20. SMC combines the advantages of both analog LDO (high gain, low droop, high-power supply rejection) with digital LDO (fast LS operation and adaptivity as shown in Figure 19c) and achieves >80% peak current efficiency over 0.5–1.1 V, with 1–12-mA load current and only 6-ns droop recovery time.

Energy management for intermittently powered devices

Since an IoT device can employ intermittent sensing, computation, and communication, which is supported from small-energy sources (or from harvested energy in extremely resource-constrained scenarios), energy management considerations for intermittent operation become extremely important, and lightweight software procedures for control flow, optimal checkpointing, concurrence, and data consistency [127], [128] need to be developed. This along with improved techniques of high-dynamic-range, adaptive PDNs is believed to be one of the major research directions for context-aware RC-IoT devices.

Intelligent cross-layer adaptive systems

System-level IoT designs can incorporate more than one approach discussed previously to optimally enhance machine intelligence and achieve performance improvement/energy reduction, as shown in [129] and [130].

Cao et al. [129] proposed a camera-based wireless sensor node with a self-optimizing end-to-end computation and communication design, targeted for surveillance applications. The demonstrated system supports multiple feature-extraction and classification algorithms, tunable processing depth (PD), and PA gain. Minimum-energy operating point is

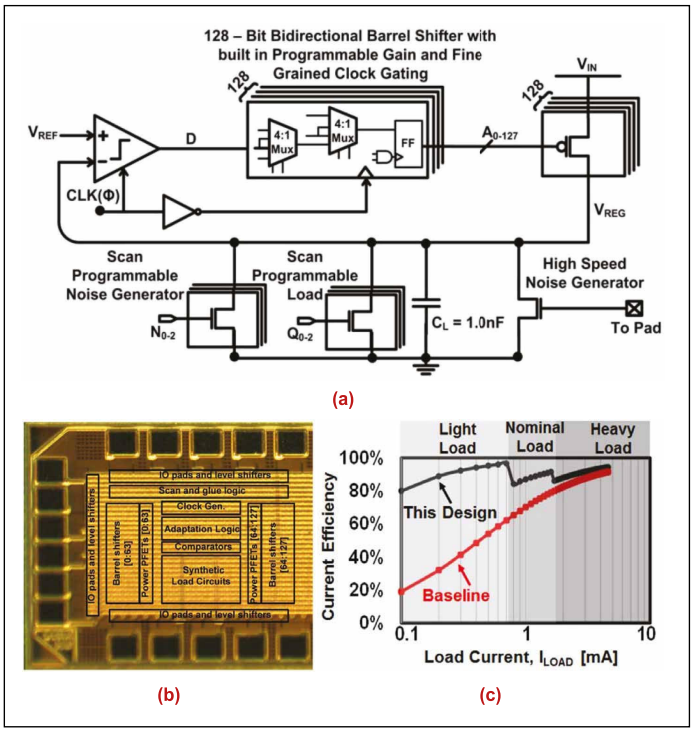

Figure 19. (a) Digital LDO with autonomous adaptation of sampling CLK that offers a wide DR [124]. (b) Chip micrograph. (c) Current efficiencies with and without adaptation.

dynamically and intelligently chosen depending on information content, accuracy targets, and wireless channel conditions. The computing platform,

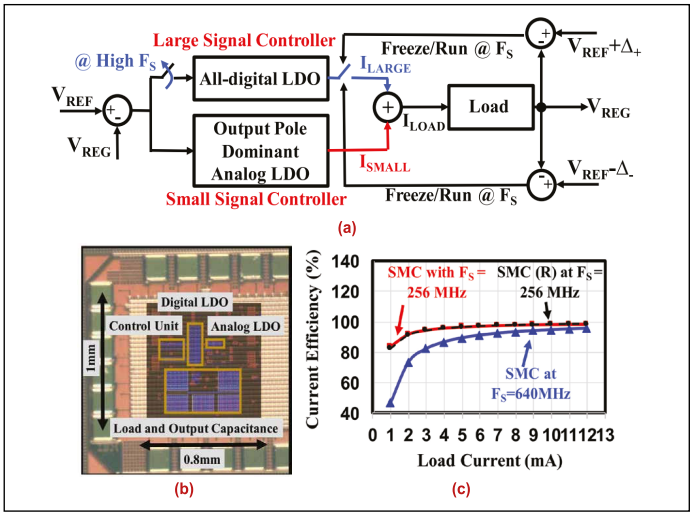

Figure 20. (a) Hybrid LDO with SMC for wide DR [126]. (b) Chip micrograph. (c) Current efficiencies with and without adaptation of sampling frequency (F_S). SMC(R) is an SMC mode with reset for faster droop recovery.

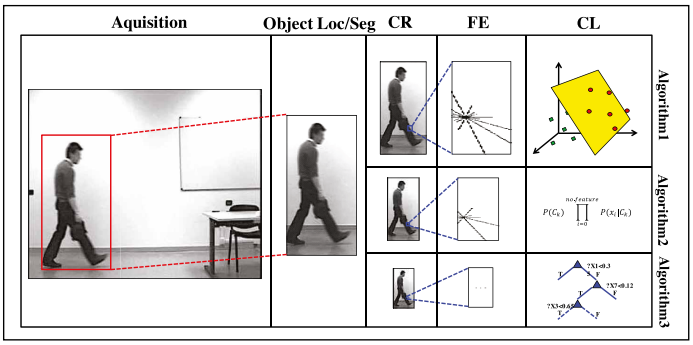

Figure 21. Algorithm demonstration with a real video frame [129].

sensor module, and transmission blocks are ADI ADSP-BF707 image processor, OV7670 camera sensor, and USRP B200 SDR software-defined radio, respectively, and achieve a 4.3× reduction in energy consumption compared to a baseline design.

The sensing+computation process is demonstrated in Figure 21. After image data acquisition, the consecutive frame differences are computed as a preprocessing step. Low-power preprocessing not only locates and segments the potential human objects but also enables adaptive *in situ* processing depending on the information content defined as a number of active pixels of frame difference. The platform demonstrates three different algorithms to provide accuracy/energy scalability and each algorithm is further divided into three PDs—i.e., compression, feature extraction, and classification—to support computation and communication tradeoff. On the transmission side, the platform applies adaptive radio whose PA gain is dynamically adjusted, guaranteeing minimum required BER with respect to time-varying wireless channel condition (pathloss in this case). It is intuitive to understand that as the PD increases, the energy cost to compute increases, but the data volume required to transmit decreases, thus reducing the energy cost to communicate. As the channel condition changes (from clean to noisy channel), the minimum energy point also changes. For a clean channel, a lower PD is preferred (as the energy to communicate is low), whereas with increasing pathloss a higher PD is preferred. The end-to-end self-optimization, which dynamically adapts the PD depending on the channel condition to always track the point of minimum total energy, is measured and demonstrated in Figure 22.

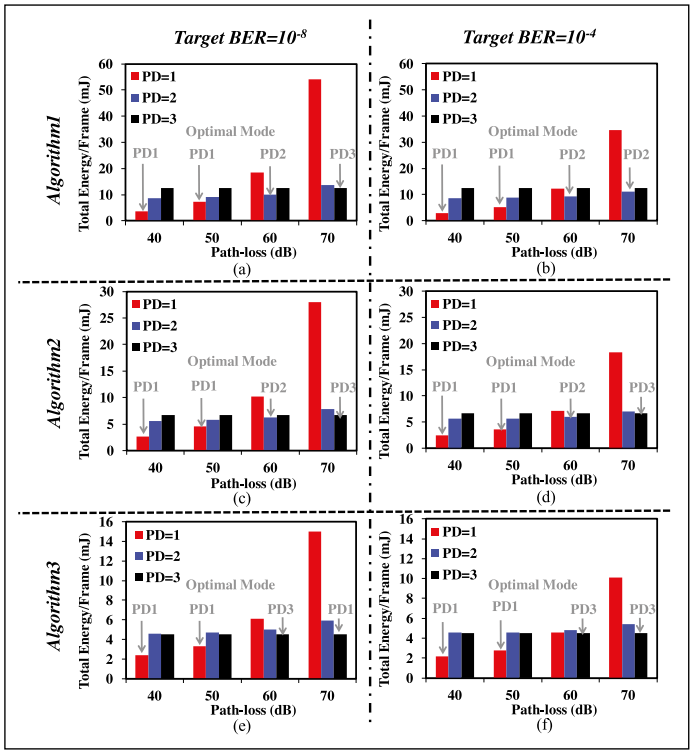

Figure 22. Measured total energy (computation+communication) per frame for different PD with increasing pathloss [129]. Experimental results are demonstrated for the three algorithms described here and two BER targets. When pathloss is high, the general trend is that optimal mode moves to more FE-embedded processing.

Platform proposed in [130] is a collaborative intelligent heating, ventilation, and air-conditioning system (HVAC) occupancy detection solution via data fusion between optical (OP) and infrared (IR) camera-based sensor nodes together in a smart wireless sensor network. Figure 23 demonstrates that data fusion has enabled accurate human detection compared with baseline designs, especially in severe lighting/heating conditions, and the consequent low miss rate (5×) in turn reduces sampling rate, resulting in expanded sensor lifetime (3×) while maintaining the required detection latency.

Collaborative intelligence is achieved at the sensor node as well as among the sensor nodes at the backend server, which is located at the HVAC and controls the HVAC. When detection accuracy is fixed after the employment of the sensor network, minimizing the latency of occupancy detection depends on reducing the sample interval (i.e., the number of OP and IR images captured per second). However, a high sampling rate will lead to severe sensor energy expenditure and limited sensor lifetime. It is also noted that

the occupancy of a particular region in a building is dependent on its neighboring regions. For example, consider a typical floor-plan of a building with three rooms, A, B, and C. The occupancy of room A is dependent on room B if a door between A and B is available and people can walk from B to A, as shown in Figure 24a and vice versa. This motivates the proposed dynamic HVAC control strategy, targeting minimized latency of occupancy detection based on a collaborative scheme among neighboring HVAC sections.

Consider a network of sensors deployed as shown in Figure 24a. The sensor node at B estimates the presence of an occupant. If an occupant is detected, then it further tracks the occupant via difference of frames and estimation of the direction of motion. The direction of motion is sent to the backend, which resolves the potential adjoining HVAC areas that can be subsequently occupied. In this example, an occupant moving from B toward A will allow the backend to send an alert to the sensor node at A. Now, this sensor node increases its sampling rate to reduce the latency of detection. The effective sampling interval, T_{eff}, is reduced as shown in Figure 24b.

References [129] and [130] aim at adaptively minimizing energy expenditure of IoT devices in a time-varying environment (wireless condition, object moving direction, etc.) while maintaining decent performance (accuracy, BER, detection latency, etc.) through distributed control on the fly or centralized control at the backend.

Security considerations for RC-IoT devices

From an implementation point of view, the IoT architecture is usually divided into 3, 4, 5 or 7 layers as shown in [131]–[135]. References [132]–[134] demonstrated the three-layer architecture as shown in Figure 25. The details of the three layers and their security concerns are presented in Table 4. These security concerns involve data confidentiality, integrity, and availability (commonly known as the CIA triad [136]), which are related to privacy, correctness, and authentication, respectively. Constrained IoT devices (most notably, CO devices such as small biosensors) have limited resources and can, hence, support only a subset of the intended security features. This makes these devices extremely prone to privacy attacks [136]–[138].

In addition to these three layers, today's IoT devices employ a fourth layer called support layer,

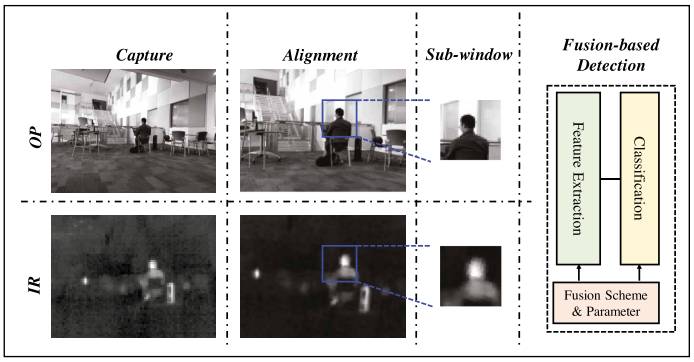

Figure 23. Demonstration of the algorithm presented in [130].

in between the network layer and application layer, which is dedicated toward security features and performs authentication using preshared secrets, keys, and passwords. However, this layer can also suffer from DoS attacks and malicious insider attack, as illustrated in [137], [155], and [156]. Moreover, the big-data problem in IoT (network exhaustion due to inundation of data) has resulted in modem IoT architects to move to a five-layer architecture with added security and data-processing capabilities [3], [157]–[163].

CISCO currently defines a seven-layer IoT structure as shown in [135]. In this discussion, we shall limit ourselves to only the perception layer, and hence, a detailed description of the advanced layer models (4, 5, and 7 layers) is out of the scope of this article.

Traditional techniques against perception layer attacks

A large number of security breaches in RC-IoT occur in the perception layer which is most vulnerable to privacy attacks due to its resource constraints. The most common security measures against attacks on the perception layer are listed in

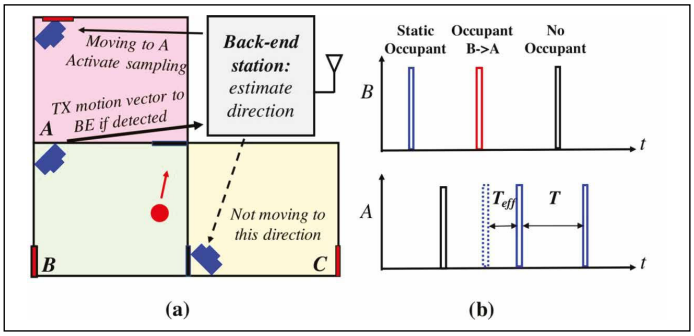

Figure 24. Illustrative representation of (a) simple sensor network with interdependence and (b) demonstration of event-driven sampling [130].

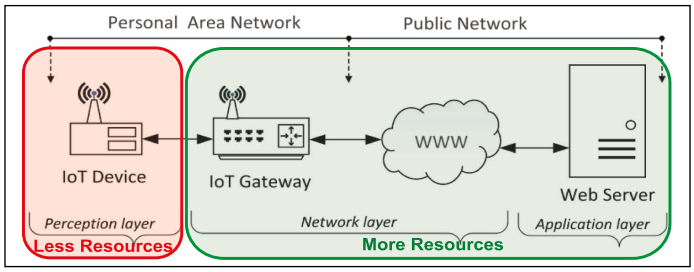

Figure 25. Basic three-layer architecture of IoT ecosystem. The perception layer devices are usually resource constrained and is more prone to attacks due to elementary security features [131].

Table 5. A complete description of each of these techniques is out of the scope of this article. Further details on each can be found in [137] and the corresponding references.

Hardware security

Encryption engines are at center stage for achieving IoT security. The computationally secure 256-bit Advanced Encryption Standard (AES) is traditionally believed to provide data confidentiality through encryption as the mathematical complexity of the recovery algorithm becomes 2^{256}. AES-256 with 14 rounds is used today in banking, military, and government applications, and hence, there has been significant research efforts to break (as well as to enhance) the standard. Related subkey-based recovery attacks have shown to significantly decrease the complexity of key recovery for reduced-round AES (2^{22} for five rounds [164], 2^{39} for nine rounds, and 2^{45} for ten rounds [165]). However, recovery complexity for practical applications with 14 rounds is still significantly high with traditional mathematical attacks. However, nontraditional techniques such as electromagnetic side channel attacks (EM-SCA) or power SCA have proven to reduce the recovery complexity to a mere 2^{13}, breakable within 50 s as shown in [166]. These emerging SCA techniques exploit information leaks from the physical cryptohardware in the form of EM emanation (for EM-SCA) and supply-line fluctuations (for power SCA), and pose a much bigger threat to the current standards than computational attacks, motivating the need for hardware techniques to suppress the side-channel leaks in the physical system.

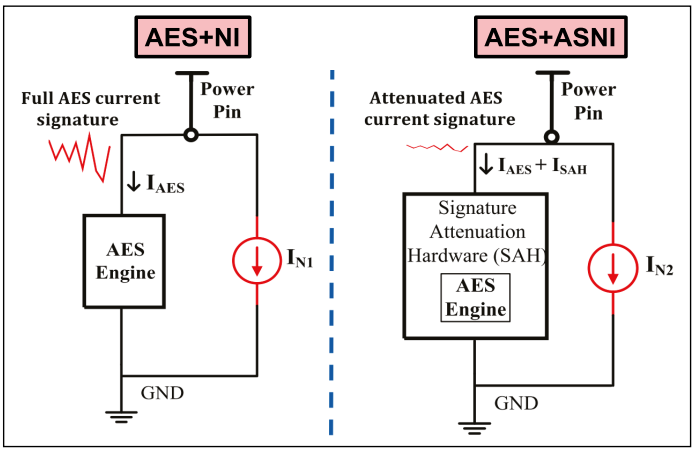

Figure 26. Noise injection and ASNI as countermeasures for power SCA.

1) Power SCA and its countermeasures: Correlation power analysis (CPA) [167] has shown to be an efficient technique for power SCA as it reduces the search space of AES-128/192/256 to just $2^8 = 256$ for each key byte (hence, the overall complexity becomes 2^{13} for 256-bit, i.e., 2^5 byte keys). Traditional power SCA countermeasures try to reduce the SNR of the leaked information through power balancing or gate-level masking but incur significant area, power, and performance overhead [168]. Attenuated Signature Noise Injection (ASNI) was utilized in [168] and [169] (Figure 26), which obfuscates the AES power traces through parallel noise injection and performs signature attenuation through a signature-attenuating hardware implemented using an on-chip shunt LDO. This method attenuates the critical AES signature in the supply current by >200× with 60% additional overhead in area and 68% overhead in power. More recently, Kar et al. [170] and Singh et al. [171] demonstrated the signature attenuation technique using an integrated voltage regulator (IVR) and loop randomizing/random fast voltage dithering techniques with only ≈5% overhead in power and area.

2) EM SCA and its countermeasures: Except for hardware masking [172], the amount of protective approaches against correlation EM analysis (CEMA [173])-based EM-SCA has remained relatively scarce. In [174], a ground-up approach was presented to find the specific source of EM emanation within the metal stack of an ASIC built using Intel's 32-nm technology, and was generalized using other popular technologies such as Taiwan Semiconductor Manufacturing Company (TSMC) 65-nm. It was shown that EM emanations from metals lower than layer eight are barely distinguishable using a commercially available EM probe, and hence, it was proposed (STELLAR) that a cryptographic core, with power supply routed

Table 4. Details and security considerations of the three-layer IoT architecture [137].

Layer	Purpose	Security threats	Remarks
Perception/sensor	Collecting information from sensors/ devices	Eavesdropping [137], Node Capture [139], Add Malicious Node [137], Replay Attack [140], Timing Attack [141]	Most attacks are on data confidentiality and integrity [137], [138]
Network/transmission	Connects devices to each other and to higher layer through wired/wireless media	Denial of Service (DoS) [137], Man-in-the-Middle Attack [142], Storage Attack [137], Exploit Attack [137]	Most attacks are on data integrity [137]
Application	Has the responsibility to extend sensor-specific services to applications/clients	Cross-site Scripting [137], Malicious Code Attack [137]	Most attacks are on data availability [137]

Table 5. Security measures against perception layer attacks.

Security measure	Details	Advantages	Drawbacks
HMAC [143], [144]	Hash Functions along with Encryption Algorithms (SHA, MD5, CBC etc) are used	Employed to maintain data integrity	Key-hacking is possible through invasive/ semi-invasive/software/side-channel attacks
Public Key Infrastructure (PKI) protocols [145], [146]	Base station communicates with the devices to get the public key while the private keys are stored separately	More secure than passwords —Malicious user needs both the secret private key and a passphrase to pose any threat	1) Key hacking: The private key needs to be protected and 2) not very scalable
Open Authentication (OAuth/OAuth 2.0) [147]–[149]	Client-server-based system where server has the list of authorized clients. Everyone can request for access, but server grants access tokens only to authorized clients	Access is granted in a secure way	1) Vulnerable to cross-cite-recovery-forgery (CSRF) and 2) implementation becomes cumbersome as the network grows since the user needs to authenticate each device
Mutual authentication [150], [151]	Client-server-based system where Client creates a request and an HMAC-SHA signature, and sends both the request and signature to server. The server retrieves the HMAC-SHA signature using a secret access key and verifies the signature with client's signature	Both client and server certificates are verified	Requires a PKI with high cost of initial deployment
Lightweight cryptography [131], [152]	Cryptographic Keys are used to convert messages	Plain text to cipher text by using symmetric, asymmetric keys and hash functions	Hard to implement for Class-0 devices with stringent resource constraints
Embedded security framework [153], [154]	Provides secure secondary storage, runtime environment and secure memory management	Provides a complete security package	Extremely resource-intensive

entirely in lower level metals locally, and equipped with signature suppression techniques like ASNI, before reaching higher-level metal routing, would be resistant against both EM and Power SCA.

3) Machine learning SCA—X-DeepSCA and possible countermeasures: Recently, ML SCA attacks have been shown as a big threat as it can uncover the secret key within a few traces using previously learned models. Das et al. [175] demonstrate cross-device deep-learning-based SCA (X-DeepSCA) using training data that contains augmented power traces from multiple devices with AES-128. With ≈200k traces and proper choice of hyperparameters, it was shown that X-DeepSCA attacks can recover keys with 99.9% accuracy from different target devices with ≈10× lower minimum number of traces as compared to traditional CPA. This increases the threat surface of SCAs significantly and puts further emphasis on SCA countermeasures such as IVR, ASNI, and STELLAR.

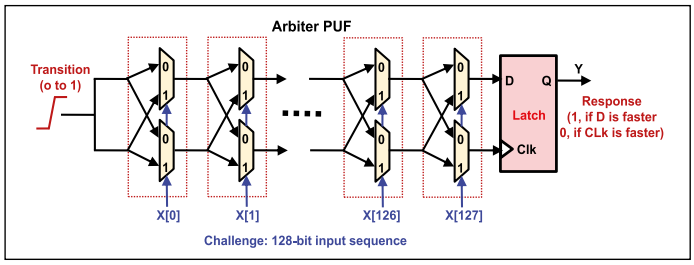

Figure 27. 128-bit Arbiter PUF [176]. Assuming same layout length, the circuit creates two delay paths for each input X, and produces the 1-bit output Y based on which path is faster.

PUF-based techniques

Physical unclonable functions (PUFs) have emerged as a promising augmentation (sometimes even as an alternative to key/token-based cryptography), which leverage manufacturing process variations to generate a unique and device-specific identity for a physical system [176][178]. PUF implementations are simpler in terms of hardware and do not need to store the secret key that is used to employ complex cryptographic algorithms.

1) Digital PUFs: Traditional digital PUFs employ simple circuitry such as ring oscillators [179], Arbiters [180], SRAMs [181]–[183], and dynamic RAMs [184] for PUF implementation, which consume much less power and area than key-based cryptographic implementations wherein the secret key is stored in a battery-backed SRAM or in a nonvolatile memory/electrically erasable programmable read-only memories (EEPROMs), which are all expensive resources for a constrained devices such as C0 or C1. Moreover, any invasive tampering mechanism usually changes the PUF's output, thereby letting the user know about the attack. These advantages, coupled with low resource requirements, makes PUFs a suitable choice for IoT environments. As an example, an Arbiter PUF is shown in Figure 27, wherein a 128-bit challenge (input) produces a 1-bit response (output) according to the respective path delays in the data and clock paths due to the random manufacturing variations [176], [180], and hence, can be utilized for device authentication. Even though it was shown later in [185] that the randomness of the output can be modeled with reasonably low complexity, an improved design with XORed outputs from multiple Arbiter PUFs demonstrated high tolerance against modeling attacks [185], [186].

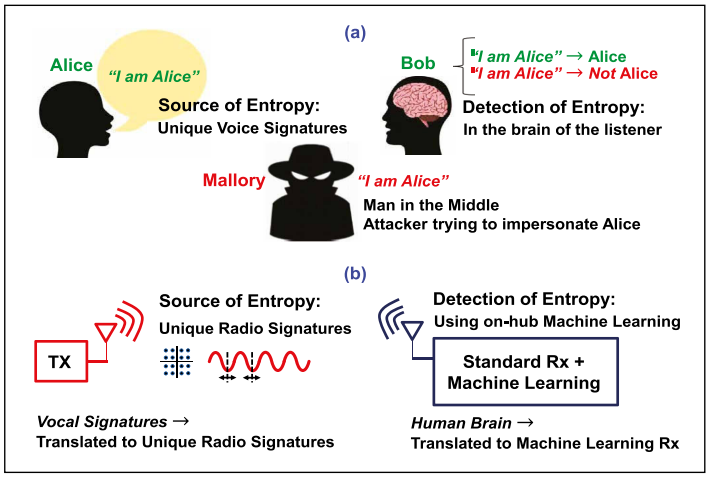

Figure 28. Principle of RF-PUF [187]. (a) Authentication in human voice communication: Bob (the receiver) can identify Alice (the transmitter) based on the unique voice signatures, and not based on the contents of what Alice speaks. Mallory (the impersonator) can also be identified (as not Alice), since his unique voice signatures would be different from Alice. (b) Analogous system that utilizes an RF-PUF framework for secure radio communication.

2) RF-PUF: Traditional PUF designs discussed above still require a minimal amount of additional hardware at the transmitter side of the RC-IoT device. Chatterjee et al. [187] proposed a new kind of PUF for RC-IoT scenario, which exploits the effects of inherent analog and RF process variations at the transmitter (Tx) side by detecting them with an *in-situ* ML hardware at the resource-rich receiver (Rx). This method embraces the already existing nonidealities at the Tx, which are usually discarded in a traditional communication scenario and, hence, do not require any additional hardware for PUF generation. The method is inspired by the inherent authentication in human voice communication as shown in Figure 28, with unique human voice being replaced by unique Tx signatures, and human brain replaced by a neural network at the Rx. The holistic system-level view for RF-PUF implementation is shown in Figure 29, while the number of unique transmitters that can be identified with varying channel conditions and Rx signatures is shown in Figure 30. It has been shown with the simulation results that up to 8000 RC-IoT devices can be uniquely identified with 99% accuracy. Proof-of-concept hardware evaluations were also demonstrated. Since this method does not require any additional hardware at the Tx, the framework can be

utilized as an extremely useful security feature for RC-IoT devices for a small-to-medium-scale smart system.

Learning frameworks for RC-IoT devices

With the above background in energy-constrained sensing, computing, communication, energy management, and security, let us look into the ML models that have been developed and can effectively combine different modalities based on context. Since this is a developing field, there is no single learning framework that overcomes all the challenges in IoT, and hence, it is important to characterize the most promising learning frameworks based on the applications [5]. The current literature on learning in an IoT framework can broadly be classified into three categories: ML, sequential learning (SL), and reinforcement learning (RL).

Machine learning (ML)

ML techniques usually build regression-based models on labeled or unlabeled data (for supervised and unsupervised learning, respectively). ML techniques are computationally complex and require an extensive training data set for acceptable performance, both of which require expensive resources [188]. Hence, instead of executing the ML algorithms in the RC-IoT device, many of the implementations resort to a centralized cloud-based processing unit for ML [3], [188], [189]. However, this means that the sensor data have to be communicated to the cloud for further processing and, hence, pose a burden of communication payload on the RC-IoT device. Compressive sensing and PCA has been shown to be useful [189] in reducing the payload.

Sequential learning (SL)

SL [190]–[192] uses intelligent distributed agents (RC-IoT devices) that sequentially learn about an underlying binary state of the system (such as a medical status, fire alarm, triggering event, or anomaly- and event-based transmission), and subsequently propagate it through the network, as shown in Figure 31. Depending on the number of previous agents from which information is gathered, SL is categorized into finite memory and infinite memory [5]. In infinite memory SL, agents collect information on the estimate from all other agents in the sequence and, hence, require more memory resources. Finite memory SL, on the other hand, collects information from a user-defined fixed number of previous agents

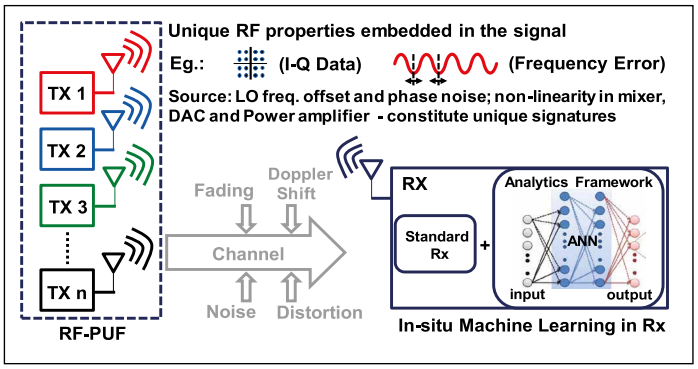

Figure 29. Visualization of RF-PUF in an asymmetric IoT network with multiple RC-IoT devices as transmitters and one resource-rich receiver [187].

and is more suited to RC-IoT devices. It was shown in [190] and [192] that it is possible to converge to an accurate underlying state using only two previous states (the tradeoff being higher convergence time and, hence, more latency, the allowable limit for which depends on the application). Also, unlike the traditional centralized ML architecture, SL can have a distributed implementation and does not require an extensive data set for learning. However, SL requires machine-to-machine (M2M) communication, which may increase energy consumption if it is not taken care of at the network implementation level. SL is particularly useful for event/anomaly detection applications, whereas ML is more suitable for data analytics applications with higher complexity and higher resource requirement.

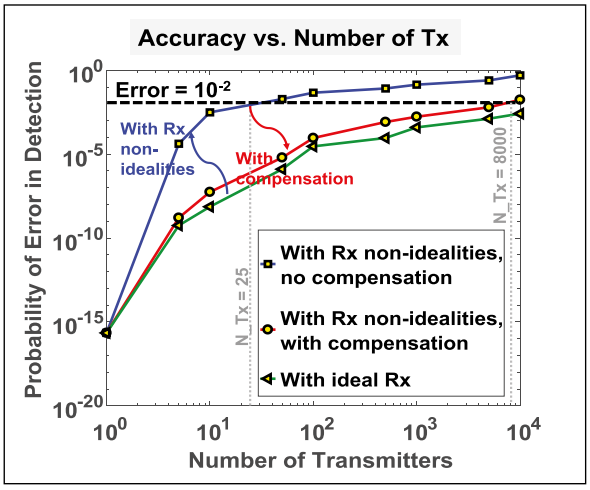

Figure 30. Probability of false detection as a function of the total number of transmitters in the system, with and without receiver signature compensation [187].

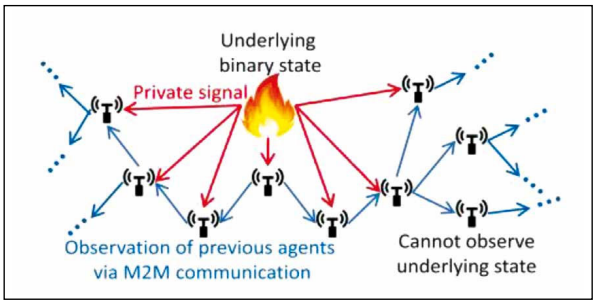

Figure 31. Sequential learning in IoT [5].

Reinforcement learning (RL)

RL implementations [188], [193]–[197] utilize the interaction between the agent and the environment in a method based on rewards and penalties. In RL, the agents can perform a predefined set of actions in an environment with a set of states and a state-transition function. The action of the agents changes the states, and based on the state transitions and the final goal, the environment rewards (or penalizes) the agent. The agent tries to maximize its immediate as well as future rewards and learns to converge to a steady state as explained in [193]. The action-reward combination works as a closed loop feedback system with a high chance of convergence, and can be implemented using computationally simple algebraic Q-learning algorithm [188]. As shown in Figure 32a, the agent receives the reward based on the state transition due to the action taken at a given state. The total cumulative reward (called Q value) of performing an action *at* at a given state s_t is given by the linear combination of the old Q value, the immediate reward and the total estimate of the future rewards as indicated by the Q-learning-equation in Figure 32a.

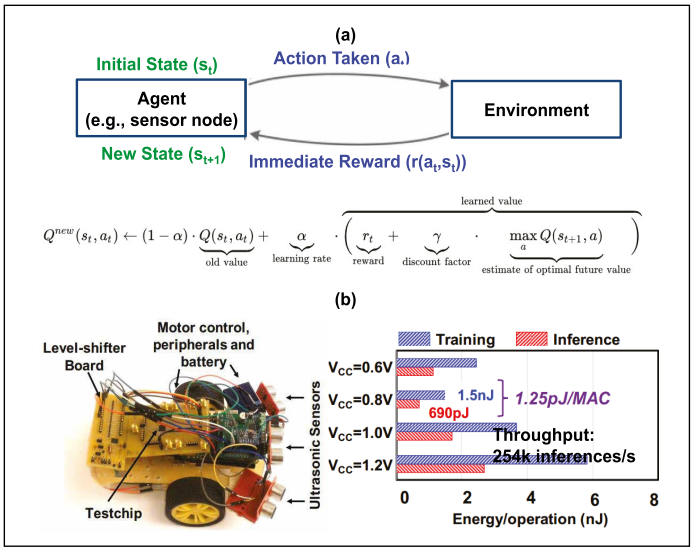

Figure 32. (a) Q-learning (RL framework) for RC-IoT devices. (b) Implementation showing an RL-enabled microrobot [197] with time-domain inputs and processing through a three-layer neural network and SVM loss function.

Amravati et al. [197] demonstrated a time-domain mixed-signal neuromorphic accelerator with embedded RL implemented using a three-layer neural network with 84 neurons. The test chip was built in 55-nm CMOS technology and was mounted on a mobile microrobot for autonomous exploration of the environment (Figure 32b). The peak power was only 690 μW at 1.2-V supply while operating at 3.12 TOPS/W. The peak energy efficiency was 690 pJ/Inference and 1.5 nJ/training (1.25 pJ/MAC), making it one of the highest performance and lowest power implementation till date. The low energy is attributed to: 1) time-domain mixed-signal MAC operations with time-domain inputs which do not need voltage to time or time to voltage conversions and consumes scaled energies based on the importance of the computation and 2) a relatively low 6-bit precision, which was shown to be enough for low-to-medium complexity applications [45] involving pattern/object recognition.

The convergence process for RL is slower than SL and the requirement to preemptively know the states and state-transition-matrix makes RL challenging for medium- and high-complexity applications. However, for low-complexity, high-latency-tolerant tasks such as resource management or power management [195], [196], RL can be an extremely relevant choice. Parallel Q-learning (PQL) algorithms (PCSP-8 [198], PQL-C [199], CS-RL, and CS-RL-EXT [200]) have also been developed for distributed, resource-constrained applications and for speeding up the RL convergence. Figure 33 shows a comparison of speed-up using these techniques.

In essence, SL- and RL-based learning techniques have shown enough promise through lightweight algorithms that can be implemented on the small IoT nodes. However, a network-wide full realization of these techniques for context discovery and assessment is still a wide open area of research. New devices and technologies such as mixed-signal neurons, memristors, spin-transfer-torque-based devices, opto-electronic and ferroelectric devices with in-memory and near-memory computation to reduce memory fetch, computation, and communication power in a neural network are areas of active research and hold

tremendous potential for future. Online/incremental learning is of paramount importance because of the variations in the manufacturing process and operating conditions. By fully utilizing the capabilities of devices, hardware, and algorithms together, the path toward more efficient context-aware systems needs to be paved.

The way ahead: Future of IoT systems

Where we stand: Systems with efficient components and promising applications

The preceding analysis puts us in perspective of the current state-of-the-art in adaptive, context-aware IoT hardware and machine intelligence. The number of applications is numerous, ranging from small-scale smart biosensing and smart cars to medium-scale smart homes and offices, and to large-scale smart cities. Significant research efforts have been put into optimizing the available resources for corresponding applications as pointed out in previous sections.

Where the future lies: Secure, context-aware, intelligent, and adaptive devices and systems

However, bigger IoT networks still remain suboptimal in terms of effective utilization of the distributed resources. The vision of secure, context-aware, intelligent, and adaptive devices and systems, as presented throughout this article, involves a holistic optimization of all the resource-constrained leaf devices within a network, in each of the following subareas.

- *Sensing:* As the sensing leaf nodes in an IoT network have the most stringent resource constraints, the sensing process itself should be made extremely low power through sub-Nyquist-rate CS [22], [30], [33] for sparse signals (such as audio and image), or made adaptive/energy-resolution scalable through time-/frequency-based sensing [40] for slowly varying signals with high DR (such as radiation and vibration). The adaptivity information/resolution requirement (context) should come from the cloud for latency-relaxed applications, and from in-senor/on-gateway learning hardware for latency-limited scenarios. Reconfigurability among Nyquist-rate sensing, CS and time-/frequency-based sensing can be an optional feature, depending on the applications and amount of resources available.
- *Computation:* The intelligent RC-IoT nodes should have the capability of locally extracting important information from the sensed data to reduce subsequent power consumption in communicating otherwise raw data bits to the cloud. Anomaly/Outlier/Event detection and data compression are the two most important forms of in-sensor/edge analytics that are required in today's systems and is an extremely promising research direction for bringing down the power consumption due to nonoptimal data handling and communication.
- *Communication:* As shown in the analysis presented in the "Intelligent Computing Platforms" and the Intelligent Communication" sections, this is the subsystem toward which a lot of research focus should be directed for practical feasibility of the context-aware vision. The numerous modalities available (proximity communication [83], HBC [84], NFC, ZigBee, ANT, BTLE, Wi-Fi and LoRA, among others) makes this a multidimensional and multilevel optimization problem with possibilities of intra-PHY and inter-PHY adaptability and tradeoffs. Techniques such as anomaly detection and channel quality estimation would determine when to communicate, and how much data are to be sent (e.g., burst-mode communication will bring in further energy efficiency and context awareness on top of data compression, through duty cycled intermittent communication, even with good channel quality [129]). Furthermore, short-range low-energy communication using HBC/ANT/BTLE should be explored to assess the possibility of spatial data compression for sensors within close proximity of each other. If the spatial data compression

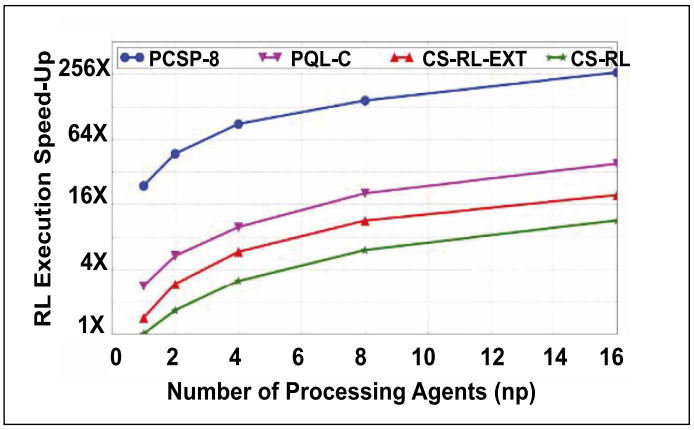

Figure 33. Convergence speed-up using parallel Q-learning algorithms (PQL with coallocation of storage and processing: PCSP-8 [198], PQL with local cache: PQL-C [199], constant share RL: CS-RL, and extended CS-RL: CS-RL-EXT [200]).

is possible, only one node in an RC-IoT cluster would take the responsibility to communicate the compressed data to the upper level gateway/cloud (possibly using a higher power modality like LoRa for long-range communication). Again, processing all the above information would require sophisticated learning algorithms to be implemented in different hierarchical levels of the IoT architecture which, by itself, is an involved optimization problem.

- *Energy management:* As shown in the "Intelligent Energy Management" section, high-dynamic-range and high-power conversion LDOs with low voltage droop/droop recovery time are one of the major requirements in a dynamic IoT scenario. High DR adaptation techniques such as sampling frequency based reconfigurable LDOs [124] and SMC LDOs [126] have been explored. However, challenges due to checkpointing and data consistency need to be looked into. Recent checkpointing schemes such as the one shown in [128] have demonstrated improved latencies in a medium-to-high-resource device—though similar and more lightweight techniques need to be developed for highly resource-constrained devices.
- *Adaptive security:* The RF-PUF [187] framework shown in the "Learning Frameworks for RC-IoT Devices" section, along with low-level metal routing for the encryption core [174] for EM-SCA resistance can be utilized as a baseline security feature at no additional power/area overhead in extremely resource-constrained C0 devices. Lightweight implementation of ASNI/IVR [168], [170] with minimal overhead should be placed as well in C0 devices, while nodes with more relaxed constraints can benefit from implementations with better signature attenuation (consuming higher power). These techniques should also be augmented with one or multiple traditional security features such as hash-based message authentication (HMAC) and mutual authentication/OAuth based on the context (application, importance of collected data) and resources available, and can be adaptive in nature.

Cognitive hierarchy theory (CHT)

To capture and exploit the multitude of reconfigurable modalities effectively in the IoT environment, it is necessary to model the heterogeneity, resource constraints, and distribution of the IoT devices within the architecture in a structured manner. CHT is an emerging tool to provide such a modeling framework using behavioral game theory, and is based on bounded rationalities [5], [201], [202]. The theory of bounded rationalities ensures that each node in the network tries to find its best strategy, bounded by information from the lower level nodes in the hierarchy, its own computational capacity, and time/resource available. CHT model (Figure 34) inherently takes care of the device heterogeneity in IoT as it considers the resources available for each device separately. References [201] and [202] present further details of the CHT techniques, while [5] demonstrates an example of the CHT theory in determining the type of learning algorithm (ML, SL, and RL) to be implemented on a particular IoT device based on its resource constraints. It must be noted that though CHT would define a structure in the heterogeneous IoT hierarchy, such an algorithm cannot be implemented in C0 and possibly C1 devices. However, the output of the algorithm can be passed on to the RC-IoT devices from higher level nodes which have higher computational power.

IoT NETWORKS are different from traditional networks in view of their specific challenges in device heterogeneity, resource constraints, context-variability, and security, thereby necessitating adaptive solutions for resource-aware operation. In this article, we have presented a broad review of the different areas that need to be looked into for holistic, system-level resource optimization for RC-IoT devices in a network. Various techniques in sensing (compressed-domain sensing/energy-resolution scalable frequency-domain

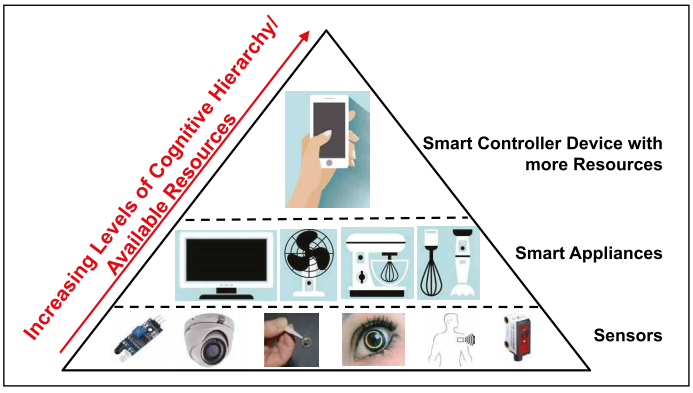

Figure 34. Distribution of IoT devices according to the levels of cognitive hierarchy theory [5].

sensing), computation (in-sensor/edge analytics in the form of outlier detection and data compression), communication (intra-PHY and inter-PHY adaptation with low and high-power modalities), power management and security were analyzed, and the vision for a secure, context-aware, adaptive, resource-constrained but intelligent IoT device was presented. However, numerous challenges (in the form of system-level controller design for adaptive architectures, reliability, security, latency limitations, intermittent powering/checkpointing and real-time/online learning) still exist in realizing a full implementation of the concepts demonstrated, indicating future research directions toward building smarter and more adaptive systems. In that context, the goal in this article has been to identify the current trends, foundations and components of the envisioned RC-IoT devices to enable the design of more efficient connected intelligent systems in the future. ■

Acknowledgments

This work was supported in part by the Semiconductor Research Corporation (SRC) under Grant 2720.001 and in part by the National Science Foundation (NSF) CRII Award under Grant CNS 1657455.

■ References

[1] R. Lloyd, *Mobile Traffic From Wearables Explodes as the Internet of Everything Accelerates*. Accessed: December 5, 2018. [Online]. Available: http://blogs.cisco.com/news/mobile-traffic-from-wearables-explodes-as-the-internet-of-everything-accelerates

[2] S. Sen, "Invited: Context-aware energy-efficient communication for IoT sensor nodes," in *Proc. 2016 53nd ACM/EDAC/IEEE Design Autom. Conf.*, Jun. 2016, pp. 1–6.

[3] J. Gubbi et al., "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Gener. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, 2013.

[4] F. Delicato, P. Pires, and T. Batista, "Resource management for Internet of Things," in *The Resource Management Challenge in IoT*, Cham, Switzerland: Springer, 2017.

[5] T. Park, N. Abuzainab, and W. Saad, "Learning how to communicate in the Internet of Things: Finite resources and heterogeneity," *IEEE Access*, vol. 4, pp. 7063–7073, 2016.

[6] C. Perera et al., "Context aware computing for the Internet of Things: A survey," *IEEE Commun. Surveys Tuts.*, vol. 16, pp. 414–454, First 2014.

[7] C. Perera et al., "A survey on Internet of Things from industrial market perspective," *IEEE Access*, vol. 2, pp. 1660–1679, 2014.

[8] *The 'Only' Coke Machine on the Internet*. Accessed: December 9, 2018 [Online]. Available: https://www.cs.cmu.edu/~coke/history_long.txt

[9] M. Weiser, "The computer for the 21st century," *Scientific American*, Sep. 1991.

[10] K. Ashton, *That 'Internet of Things' Thing*. Accessed: December 9, 2018. [Online]. Available: https://www.rfidjournal.com/articles/view?4986

[11] P. Langley and H. A. Simon, "Applications of machine learning and rule induction," *Commun. ACM*, vol. 38, pp. 54–64, Nov. 1995.

[12] A. S. Tanenbaum and A. S. Woodhull, *Operating Syst. Design and Implementation*, Upper Saddle River, NJ: Prentice-Hall, 2005, 3rd ed.

[13] C. Bormann, M. Ersue, and A. Keranen, *Terminology for Constrained-node Netw.*, *RFC 7228*. Accessed: January 21, 2019. [Online]. Available: https://www.rfc-editor.org/info/rfc7228

[14] B. N. Schilit and M. M. Theimer, "Disseminating active map information to mobile hosts," *IEEE Netw.*, vol. 8, pp. 22–32, Sep. 1994.

[15] G. D. Abowd et al., "Toward a better understanding of context and context-awareness," in *Proc. 1st Int. Symp. Handheld Ubiquitous Comput.'99*, London, UK: Springer-Verlag, 1999, pp. 304–307.

[16] A. K. Dey, G. D. Abowd, and D. Salber, "A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications," *Hum.-Comput. Interact.*, vol. 16, pp. 97–166, Dec. 2001.

[17] P. J. Brown, "The STICK-e document: A framework for creating context-aware applications," in *Electron. Pub.*, pp. 259–272, 1996.

[18] D. Franklin and J. Flachsbart, *All Gadget and No Representation Makes Jack a Dull Environment*, 1998.

[19] T. Rodden et al., "Exploiting context in HCI design for mobile systems," in *Proc. Workshop Hum. Comput. Interact. Mobile Dev.*, 1998.

[20] R. Hull, P. Neaves, and J. Bedford-Roberts, "Toward situated computing," in *Proc. Digest Papers First Int. Symp. Wearable Comp.*, Oct. 1997, pp. 146–153.

[21] K. Henricksen and J. Indulska, "Developing context-aware pervasive computing applications: Models and approach," *Pervasive Mob. Comput.*, vol. 2, pp. 37–64, Feb. 2006.

[22] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4203–4215, Dec. 2005.

[23] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, Apr. 2006.

[24] M. F. Duarte et al., "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, pp. 83–91, Mar. 2008.

[25] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.

[26] R. Baraniuk and P. Steeghs, "Compressive radar imaging," in *Proc. 2007 IEEE Radar Conf.*, Apr. 2007, pp. 128–133.

[27] A. Amaravati et al., "A 130nm 165nJ/frame compressed-domain smashed-filter based mixed-signal classifier for "In-sensor" analytics in smart cameras," *Trans. Circ. Syst. II*, no. 9, 2017.

[28] A. Anvesha et al., "A light-powered, always-on, smart camera with compressed domain gesture detection," in *Proc. 2016 Int. Symp. Low Power Electron. Design*, 2016, pp. 118–123.

[29] S. Xu et al., "Appearance-based gesture recognition in the compressed domain," in *2017 IEEE Int. Conf. Acoust., Speech Signal Process.*, pp. 1722–1726, Mar. 2017.

[30] A. Anvesha et al., "A 65nm compressive-sensing time-based ADC with embedded classification and INL-aware training for arrhythmia detection," in *2017 IEEE Biomedical Circ. and Syst. Conf.*, pp. 1–4, Oct. 2017.

[31] S. Qaisar et al., "Compressive sensing: From theory to applications, a survey," *J. Commun. Netw.*, vol. 15, pp. 443–456, Oct, 2013.

[32] H. Djelouat, A. Amira, and F. Bensaali, "Compressive sensing-based IoT applications: A review," *J. Sensor Actuator Netw.*, vol. 7, no. 4, 2018.

[33] A. Amaravati et al., "A light-powered smart camera with compressed domain gesture detection," *IEEE Trans. Circ. and Syst. for Video Technology*, vol. 28, pp. 3077–3085, Oct. 2018.

[34] J. Zhang, Z. Wang, and N. Verma, "18.4 A matrix-multiplying ADC implementing a machine-learning classifier directly with data conversion," in *Proc. ISSCC 2015*, Feb. 2015, pp. 1–3.

[35] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: A survey," *Artifi. Intell. Rev.*, vol. 43, no. 1, pp. 1–54, 2015.

[36] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 677–695, 1997.

[37] J. D. Roose et al., "Flexible and self-adaptive sense-and-compress for submicrowatt always-on sensory recording," in *ESSCIRC 2018—IEEE 44th Euro. Solid State Circ. Conf.*, pp. 282–285, Sep. 2018.

[38] M. Trakimas and S. R. Sonkusale, "An adaptive resolution asynchronous ADC architecture for data compression in energy constrained sensing applications," *IEEE Trans. Circ. Syst. I*, vol. 58, pp. 921–934, May 2011.

[39] C. Ieong et al., "A 0.45 V 147-375 nW ECG compression processor with wavelet shrinkage and adaptive temporal decimation architectures," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 25, pp. 1307–1319, Apr. 2017.

[40] B. Chatterjee et al., "A wearable real-time CMOS dosimeter with integrated zero-bias floating-gate sensor and an 861nW 18–bit energy-resolution scalable time-based radiation to digital converter," in *2019 IEEE Custom Integr. Circ. Conf.*, Apr. 2019.

[41] W. Li, J. Bao, and W. Shen, "Collaborative wireless sensor networks: A survey," in *2011 IEEE Int. Conf. Syst. Man Cybern.*, pp. 2614–2619, Oct. 2011.

[42] M. Singh and V. K. Prasanna, "A hierarchical model for distributed collaborative computation in wireless sensor networks," in *Proc. Int. Parallel Distrib. Process. Symp.*, Apr. 2003, p. 166.2.

[43] R. M. Buehrer, H. Wymeersch, and R. M. Vaghefi, "Collaborative sensor network localization: Algorithms and practical issues," in *Proc. IEEE*, vol. 106, Jun. 2018, pp. 1089–1114.

[44] N. Cao et al., "In-sensor analytics and energy-aware self-optimization in a wireless *sensor node," in Proc. 2017 IEEE MTT-S Int. Microw. Symp.*, pp. 200–203, Jun. 2017.

[45] B. Chatterjee et al., "Exploiting inherent error-resiliency of deep neural networks to achieve extreme energy-efficiency through mixed-signal neurons," *ArXiv:1806.05141 (CS)*, vol. abs/1806.05141, 2018.

[46] B. Chatterjee et al., "An energy-efficient mixed-signal neuron for inherently error-resilient neuromorphic systems," in *Proc. 2017 IEEE Int. Conf. Reboot. Comput.*, 2017, pp. 1–2.

[47] C. H. Bennett, "Notes on Landauer's principle, reversible computation, and Maxwell's Demon," *Studies Hist. Phil. Modern Phys.*, vol. 34, no. 3, pp. 501–510, 2003.

[48] H. T. Friis, "A note on a simple transmission formula," in *Proc. IRE*, vol. 34, May 1946, pp. 254–256.

[49] A. J. Johansson, "Performance of a radio link between a base station and a medical implant utilising the MICS standard," in *26th Ann. Int. Conf. IEEE Eng. Med. Biol. Soci.*, vol. 1, pp. 2113–2116, Sep. 2004.

[50] C. Salazar et al., "A 2.4 GHz interferer-resilient wake-up receiver using a dual-IF MultiStage N-path architecture," *IEEE J. Solid-State Circ.*, vol. 51, pp. 2091–2105, Sep. 2016.

[51] B. Razavi, *RF Microelectronics.* Upper Saddle River, NJ: Prentice Hall, 2011, 2nd ed..

[52] A. Ebrazeh and P. Mohseni, "30 pJ/b, 67 Mbps, centimeter-to-meter range data telemetry with an IR-UWB wireless link," *IEEE Trans. Biomed. Circ. Syst.*, vol. 9, pp. 362–369, Jun. 2015.

[53] W. Wu et al., "Localized outlying and boundary data detection in sensor networks," *IEEE Trans. Knowl. Data Eng.*, vol. 19, pp. 1145–1157, Aug. 2007.

[54] C.-C. J. K. Minwook et al., "Distributed spatiotemporal outlier detection in sensor networks," in *Proc. SPIE*, vol. 5819, 2005, pp. 5819–5819–12.

[55] T. Palpanas et al., "Distributed deviation detection in sensor networks," *SIGMOD Rec.*, vol. 32, pp. 77–82, Dec. 2003.

[56] S. Papadimitriou et al., "LOCI: Fast outlier detection using the local correlation integral," in *Proc. 19th Int. Conf. Data Eng.*, Mar. 2003, pp. 315–326.

[57] B. Sheng, Q. Li, W. Mao, and W. Jin, "Outlier detection in sensor networks," in *Proc. 8th ACM Int. Symp. Mobile Ad Hoc Netw. Comput. MobiHoc '07*, New York, NY, 2007, pp. 219–228.

[58] J. Branch et al., "In-network outlier detection in wireless sensor networks," in *26th IEEE Int. Conf. Distrib. Comput. Syst. '06*, pp. 51–51, Jul. 2006.

[59] Y. Zhuang and L. Chen, "In-network outlier cleaning for data collection in sensor networks," in *In CleanDB Workshop VLDB 2006*, pp. 41–48, 2006.

[60] S. Rajasegarar et al., "Distributed anomaly detection in wireless sensor networks," in *2006 10th IEEE Singapore Int. Conf. Commun. Syst.*, pp. 1–5, Oct. 2006.

[61] S. Rajasegarar et al., "Quarter sphere based distributed anomaly detection in wireless sensor networks," in *2007 IEEE Int. Conf. Commun.*, pp. 3864–3869, Jun. 2007.

[62] N. Shahid, I. H. Naqvi, and S. B. Qaisar, "One-class support vector machines: Analysis of outlier detection for wireless sensor networks in harsh environments," *Artif. Intell. Rev.*, vol. 43, pp. 515–563, Apr. 2015.

[63] D. Janakiram, V. A. Reddy, and A. V. U. P. Kumar, "Outlier detection in wireless sensor networks using bayesian belief networks," in *2006 1st Int. Conf. Commun. Syst. Softw. Middlew.*, pp. 1–6, Jan. 2006.

[64] E. Elnahrawy and B. Nath, "Context-aware sensors," in *Wireless Sensor Netw.*, EWSN 2004. Lecture Notes in Computer Science, Berlin, Heidelberg: Springer, Jan. 2004, vol. 2920.

[65] P. Malhotra et al., "Long short term memory networks for anomaly detection in time series," in *ESANN*, 2015.

[66] S. Ahmad et al., "Unsupervised realtime anomaly detection for streaming data," *Neurocomputing*, vol. 262, pp. 134–147, 2017.

[67] V. Chatzigiannakis et al., "Hierarchical anomaly detection in distributed large-scale sensor networks," in *11th IEEE Symp. Comput. Commun.'06*, pp. 761–767, Jun. 2006.

[68] Y. Lee, Y. Yeh, and Y. F. Wang, "Anomaly detection via online oversampling principal component analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 25, pp. 14601470, Jul. 2013.

[69] V. Barnett and T. Lewis, *Outliers in Statistical Data.* John Wiley & Sons Ltd., 1994, 2nd ed.

[70] Y. Zhang, N. Meratnia, and P. Havinga, "Outlier detection techniques for wireless sensor networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 12, pp. 159–170, Second 2010.

[71] J. Hochenbaum, O. S. Vallis, and A. Kejariwal, "Automatic anomaly detection in the cloud via statistical learning," ArXiv. 1704.07706, 2017.

[72] K. C. Barr and K. Asanovic, "Energy-aware lossless data compression," *ACM Trans. Comput. Syst.*, vol. 24, pp. 250–291, Aug. 2006.

[73] D. Garlan et al., "Project Aura: Toward distraction-free pervasive computing," *IEEE Pervasive Comput.*, vol. 1, pp. 22–31, Apr. 2002.

[74] C. Sadler and M. Martonosi, "Data compression algorithms for energy-constrained devices in delay tolerant networks," in *Proc. ACM Conf. Embedded Netw. Sensor Syst.*, 2006, pp. 265–278.

[75] A. Rooshenas et al., "Reducing the data transmission in wireless sensor networks using the principal component analysis," in *Proc. 2010 Sixth Int. Conf. Intell. Sensors Sensor Netw. Inf. Process.*, Dec. 2010, pp. 133–138.

[76] A. Morell et al., "Data aggregation and principal component analysis in WSNs," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 39083919, Jun. 2016.

[77] S. V. Macua, P. Belanovic, and S. Zazo, "Consensus-based distributed principal component analysis in wireless sensor networks," in *Proc. 2010 IEEE 11th Int. Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2010, pp. 1–5.

[78] D. Petrovic et al., "Data funneling: Routing with aggregation and compression for wireless sensor networks," in *Proc. First IEEE Int. Workshop Sensor Netw. Protocols Appl.*, May 2003, pp. 156–162.

[79] T. Arici et al., "PINCO: A pipelined in-network compression scheme for data collection in wireless sensor networks," in *Proc. 12th Int. Conf. Comput. Commun. Netw.*, Oct. 2003, pp. 539–544.

[80] E. Magli, M. Mancin, and L. Merello, "Low-complexity video compression for wireless sensor networks," in *Proc. 2003 Int. Conf. Multimedia Expo '03*, vol. 3, Jul. 2003, pp. III-585.

[81] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Process. Mag.*, vol. 19, pp. 51–60, Mar. 2002.

[82] S. Sen, J. Koo, and S. Bagchi, "TRIFECTA: Security, energy efficiency, and communication capacity comparison for wireless IoT devices," *IEEE Internet Comput.*, vol. 22, pp. 74–81, Jan. 2018.

[83] C. Thakkar et al., "A 32 Gb/s bidirectional 4-channel 4 pJ/b capacitively coupled link in 14 nm CMOS for proximity communication," *IEEE J. Solid-State Circ.*, vol. 51, pp. 3231–3245, Dec. 2016.

[84] S. Maity et al., "A 6.3pJ/b 30Mbps -30dB SIR-tolerant broadband interference-robust human body communication transceiver using time domain signal-interference separation," in *Proc. 2018 IEEE Custom Integr. Circ. Conf.*, Apr. 2018, pp. 1–4.

[85] D. Das et al., "Enabling covert body area network using electro-quasistatic human body communication," in *Sci. Rep.*, Jan. 2019.

[86] I. Akyildiz et al., "Wireless sensor networks: A survey," *Comput. Netw.*, vol. 38, no. 4, pp. 393–422, 2002.

[87] A. J. Goldsmith and S. Chua, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 595–602, May 1998.

[88] A. Srivastava et al., "Bio-WiTel: A low-power integrated wireless telemetry system for healthcare applications in 401–406 MHz band of MedRadio spectrum," *IEEE J. Biomed. Health Inf.*, vol. 22, pp. 483–494, Mar. 2018.

[89] S. Sen, R. Senguttuvan, and A. Chatterjee, "Environment-adaptive concurrent companding and bias control for efficient power-amplifier operation," *IEEE Trans. Circ. Syst. I*, vol. 58, pp. 607–618, Mar. 2011.

[90] S. Sen et al., "Process-variation tolerant channel-adaptive virtually zero-margin low-power wireless receiver systems," *IEEE Trans. Comput.-Aided Design Integr. Circ. and Syst.*, vol. 33, pp. 1764–1777, Dec. 2014.

[91] J. Zhu, H. Krishnaswamy, and P. R. Kinget, "Field-programmable LNAs with interferer-reflecting loop for input linearity enhancement," *IEEE J. Solid-State Circ.*, vol. 50, pp. 556–572, Feb. 2015.

[92] F. Behbahani et al., "Adaptive analog IF signal processor for a wide-band CMOS wireless receiver," *IEEE J. Solid-State Circ.*, vol. 36, pp. 1205–1217, Aug. 2001.

[93] D. T. Lin et al., "A low-power adaptive receiver utilizing discrete-time spectrum-sensing," *IEEE Trans. Microw. Theory Techn.*, vol. 61, pp. 1338–1346, Mar. 2013.

[94] P. Malla et al., "A 28mW spectrum-sensing reconfigurable 20MHz 72dB-SNR 70dB-SNDR DT AY ADC for 802.11n/WiMAX receivers," in *Proc. 2008 IEEE Int. Solid-State Circ. Conf.—Digest Tech. Pap.*, Feb. 2008, pp. 496–631.

[95] Aurangozeb et al., "Channel-adaptive ADC and TDC for 28 Gb/s PAM-4 digital receiver," *IEEE J. Solid-State Circ.*, vol. 53, pp. 772788, Mar. 2018.

[96] R. Senguttuvan, S. Sen, and A. Chatterjee, "VIZOR: Virtually zero margin adaptive RF for ultralow power wireless communication," in *Proc. 2007 25th Int. Conf. Comput. Design*, Oct. 2007, pp. 580–586.

[97] R. Senguttuvan, S. Sen, and A. Chatterjee, "Multidimensional adaptive power management for low-power operation of wireless devices," *IEEE Trans. Circ. and Syst. 11*, vol. 55, pp. 867–871, Sep. 2008.

[98] S. Sen, M. Verhelst, and A. Chatterjee, "Orthogonally tunable inductorless RF LNA for adaptive wireless systems," in *2011 IEEE Int. Symp. Circ. and Syst.*, pp. 285–288, May 2011.

[99] S. Sen et al., "A power-scalable channel-adaptive wireless receiver based on built-in orthogonally tunable LNA," *IEEE Trans. Circ. and Syst. I*, vol. 59, pp. 946–957, May 2012.

[100] D. Banerjee et al., "Realtime use-aware adaptive MIMO RF receiver systems for energy efficiency under BER constraints," in *Proc. 2013 50th ACM/EDAC/IEEE Design Autom. Conf.*, May 2013, pp. 17.

[101] D. Banerjee et al., "Real-time use-aware adaptive RF transceiver systems for energy efficiency under BER constraints," *IEEE Trans. Comput.-Aided Design Integr. Circ. Syst.*, vol. 34, pp. 1209–1222, Aug. 2015.

[102] S. Sen et al., "ProVIZOR: Process tunable virtually zero margin low power adaptive RF for wireless systems," in *Proc. 2008 45th ACM/IEEE Design Autom. Conf.*, Jun. 2008, pp. 492–497.

[103] S. Sen, "Channel-adaptive zero-margin and process-adaptive self-healing communication circuits/systems: Special session paper," in *Proc. 2014 IEEE/ACM Int. Conf. Comput.-Aided Design*, Nov. 2014, pp. 80–85.

[104] D. Banerjee et al., "Self-learning MIMO-RF receiver systems: Process resilient real-time adaptation to channel conditions for low power operation," in *Proc. 2014 IEEE/ACM Int. Conf. Comput.-Aided Design*, Nov. 2014, pp. 710–717.

[105] D. Banerjee, S. Sen, and A. Chatterjee, "Self learning analog/mixed-signal/RF systems: Dynamic adaptation to workload and environmental uncertainties," in *Proc. the IEEE/ACM Int. Conf. Comput.-Aided Design '15*, Piscataway, NJ, 2015, pp. 59–64.

[106] D. Banerjee et al., "Self-learning RF receiver systems: Process aware real-time adaptation to channel conditions for low power operation," *IEEE Trans. Circ. and Syst. 1*, vol. 64, pp. 195–207, Jan. 2017.

[107] S. Sen, "Channel-adaptive zero-margin & process-adaptive selfhealing communication circuits/systems: Special session paper," in *Proc. 2014 IEEE/ACM Int. Conf. Comput.-Aided Design*, 2014, pp. 80–85.

[108] C. Thakkar et al., "23.2 a 32gb/s bidirectional 4-channel 4pj/b capacitively coupled link in 14nm cmos for proximity communication," in *Solid-State Circ. Conference (ISSCC)*, *2016 IEEE Int.*, pp. 400–401, 2016.

[109] T. Tsukizawa et al., "A fully integrated 60GHz CMOS transceiver chipset based on WiGig/IEEE802.11ad with built-in self calibration for mobile applications," in *Proc. ISSCC 2013*, Feb. 2013, pp. 230–231.

[110] M. He et al., "A 40nm dual-band 3-stream 802.11a/b/g/n/ac mimo WLAN SoC with 1.1gb/s over-the-air throughput," in *Proc. ISSCC 2014*, Feb. 2014, pp. 350–351.

[111] S. Maity, D. Das, and S. Sen, "Wearable health monitoring using capacitive voltage-mode human body communication," in *Proc. 2017 39th Ann. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2017.

[112] S. Maity et al., "Secure human-internet using dynamic human body communication," in *Proc. 2017 IEEE/ACM Int. Symp. Low Power Electron. Design*, 2017, pp. 1–6.

[113] S. Maity et al., "Characterization and classification of human body channel as a function of excitation and termination modalities," arXiv preprint arXiv:1805.02492, 2018.

[114] S. Sen, "Socialhbc: Social networking and secure authentication using interference-robust human body communication," in *Proc. 2016 Int. Symp. Low Power Electron. Design*, 2016, pp. 34–39.

[115] S. Maity et al., "BioPhysical modeling, characterization and optimization of electro-quasistatic human body communication," *IEEE Trans. Biomed. Eng.*, pp. 1–1, 2018.

[116] S. Maity, K. Mojabe, and S. Sen, "Characterization of human body forward path loss and variability effects in voltage-mode HBC," *IEEE Microw. Wireless Compon. Lett.*, vol. 28, no. 3, pp. 266–268, 2018.

[117] S. Maity et al., "A sub-nw wake-up receiver for human body communication," in *Proc. 2018 IEEE Biomedical Circ. Syst. Conf.*, 2018, pp. 1–4.

[118] T. Musah et al., "A 4-32 Gb/s bidirectional link with 3-tap ffe/6-tap dfe and collaborative cdr in 22 nm cmos," *IEEE J. Solid-State Circ.*, vol. 49, no. 12, pp. 3079–3090, 2014.

[119] J. Jaussi et al., "26.2 a 205mw 32gb/s 3-tap ffe/6-tap dfe bidirectional serial link in 22nm CMOS," in *Proc. 2014 IEEE Int. Solid-State Circ. Conf. Digest Techn. Pap.*, 2014, pp. 440–441.

[120] T.-C. Hsueh et al., "26.4 a 25.6 gb/s differential and ddr4/gddr5 dual-mode transmitter with digital clock calibration in 22nm CMOS," in *Proc. 2014 IEEE Int. Solid-State Circ. Conf. Digest Techn. Pap.*, 2014, pp. 444–445.

[121] S. Maity, D. Das, and S. Sen, "Adaptive interference rejection in human body communication using variable duty cycle integrating DDR receiver," in *Proc. 2017 Design Autom. Test Euro. Conf. Exhibition*, 2017, pp. 1763–1768.

[122] S. Maity, P. Mehrotra, and S. Sen, "An improved update rate baud rate CDR for integrating human body communication receiver," in *Proc. 2018 IEEE Biomedical Circ. and Syst. Conf.*, 2018, pp. 1–4.

[123] S. Gangopadhyay, S. B. Nasir, and A. Raychowdhury, "Integrated power management in IoT devices under wide dynamic ranges of operation," in *Proc. 2015 52nd ACM/EDAC/IEEE Design Autom. Conf.*, Jun. 2015, pp. 1–6.

[124] S. B. Nasir, S. Gangopadhyay, and A. Raychowdhury, "A 0.13µm fully digital low-dropout regulator with adaptive control and reduced dynamic stability for ultra-wide dynamic range," in *Proc. ISSCC 2015*, Feb. 2015, pp. 1–3.

[125] S. B. Nasir, S. Gangopadhyay, and A. Raychowdhury, "All-digital low-dropout regulator with adaptive control and reduced dynamic stability for digital load circuits," *IEEE Trans. Power Electron.*, vol. 31, pp. 8293–8302, Dec. 2016.

[126] S. B. Nasir, S. Sen, and A. Raychowdhury, "Switched-mode-control based hybrid LDO for fine-grain power management of digital load circuits," *IEEE J. Solid-State Circ.*, vol. 53, pp. 569–581, Feb. 2018.

[127] A. Keshavarzi, "Embedded systems and innovative technologies for IoT applications," in *Proc. Tutorial 62nd Int. ELectron Dev. Meeting*, 2016.

[128] Z. Ghodsi, S. Garg, and R. Karri, "Optimal checkpointing for secure intermittently-powered IoT devices," in *Proc. 2017 IEEE/ACM Int. Conf. Comput.-Aided Design*, Nov. 2017, pp. 376–383.

[129] N. Cao et al., "Self-optimizing IoT wireless video sensor node with in situ data analytics and context-driven energy-aware real-time adaptation," *IEEE Trans. Circ. and Syst. I*, vol. 64, pp. 2470–2480, Sep. 2017.

[130] N. Cao et al., "Smart sensing for HVAC control: Collaborative intelligence in optical and IR cameras,"

IEEE Trans. Ind. Electron., vol. 65, pp. 9785–9794, Dec. 2018.

[131] J. King and A. I. Awad, "A distributed security mechanism for resource-constrained IoT devices," *Informatica 40*, vol. 40, pp. 133–143, 2016.

[132] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Comput. Netw.*, vol. 54, pp. 2787–2805, Oct. 2010.

[133] O. Saeed, "Towards Internet of Things: Survey and future vision," *Int. J. Comput. Netw.*, 2013.

[134] M. Yun and B. Yuxin, "Research on the architecture and key technology of Internet of Things (IoT) applied on smart grid," in *Proc. 2010 Int. Conf. Adv. Energy Eng.*, Jun. 2010, pp. 69–72.

[135] CISCO, *The Internet of Things Reference Model*. Accessed: December 5, 2018. [Online]. Available: https://www.cisco.com/c/dam/global/en_ph/assets/ciscoconnect/pdf/bigdata/jim_green_cisco_connect.pdf.

[136] A. Mosenia and N. K. Jha, "A comprehensive study of security of Internet-of-Things," *IEEE Trans. Emerg. Topics Comput.*, vol. 5, pp. 586–602, Oct. 2017.

[137] M. Burhan et al., "IoT elements, layered architectures and security issues: A comprehensive survey," *Sensors*, vol. 18, no. 9, 2018.

[138] H. Suo, J. Wan, C. Zou, and J. Liu, "Security in the Internet of Things: A review," in *Proc. 2012 Int. Conf. Comput. Sci. Electron. Eng.*, vol. 3, pp. 648–651, Mar. 2012.

[139] M. V. Bharathi et al., "Node capture attack in wireless sensor network: A survey," in *Proc. 2012 IEEE Int. Conf. Comput. Intell. Comput. Res.*, Dec. 2012, pp. 1–3.

[140] D. Puthal et al., "Threats to networking cloud and edge datacenters in the Internet of Things," *IEEE Cloud Comput.*, vol. 3, pp. 64–71, May 2016.

[141] D. Brumley and D. Boneh, "Remote timing attacks are practical," in *Proc. 12th Conf. USENIX Security Symp. '03*, Berkeley, CA, USENIX Association, 2003, vol. 12, pp. 1–1.

[142] M. Conti, N. Dragoni, and V. Lesyk, "A survey of man in the middle attacks," *IEEE Commun. Surveys Tuts.*, vol. 18, pp. 2027–2051, Thirdquarter 2016.

[143] R. H. Weber, "Internet of Things—new security and privacy challenges," *Comput. Law Security Rev.*, vol. 26, no. 1, pp. 23–30, 2010.

[144] B. V. Sundaram et al., "Encryption and hash based security in Internet of Things," in *Proc. 2015 3rd Int. Conf. Signal Process. Commun. Netw.*, Mar. 2015, pp. 1–6.

[145] F. Li and P. Xiong, "Practical secure communication for integrating wireless sensor networks into the Internet of Things," *IEEE Sensors J.*, vol. 13, pp. 3677–3684, Oct. 2013.

[146] Z. Li et al., "Research on PKI-like protocol for the Internet of Things," in *Proc. 2013 Fifth Int. Conf. Meas. Technol. Mechatron. Autom.*, Jan. 2013, pp. 915–918.

[147] E. Hammer-Lahav, *The OAuth 1.0 Protocol*. Accessed: December 12, 2018. [Online]. Available: https://tools.ietf.org/html/rfc5849

[148] *OAuth 2.0*. [Online]. Accessed: December 8, 2018. Available: https://oauth.net/2/

[149] S. Cirani, G. Ferrari, and L. Veltri, "Enforcing security mechanisms in the IP-based Internet of Things: An algorithmic overview," *Algorithms*, vol. 6, no. 2, pp. 197–226, 2013.

[150] L. Kulseng et al., "Lightweight mutual authentication and ownership transfer for RFID systems," in *Proc. IEEE INFOCOM*, Mar. 2010, pp. 1–5.

[151] G. Zhao et al., "A novel mutual authentication scheme for Internet of Things," in *Proc. Int. Conf. Modeling, Identif. Contr.*, Jun. 2011, pp. 563–566.

[152] T. Eisenbarth et al., "A survey of lightweight-cryptography implementations," *IEEE Des. Test*, vol. 24, pp. 522–533, Nov. 2007.

[153] S. Ravi et al., "Security in embedded systems: Design challenges," *ACM Trans. Embed. Comput. Syst.*, vol. 3, pp. 461–491, Aug. 2004.

[154] S. Babar et al., "Proposed embedded security framework for Internet of Things (IoT)," in *Proc. 2011 2nd Int. Conf. Wireless VITAE*, Feb. 2011, pp. 1–5.

[155] A. Sanzgiri and D. Dasgupta, "Classification of insider threat detection techniques," in *Proc. 11th Ann. Cyber Inf. Security Res. Conf. '16*, New York, NY, 2016, pp. 25:1–25:4.

[156] J. R. C. Nurse et al., "Smart insiders: Exploring the threat from insiders using the Internet-of-Things," in *Proc. 2015 Int. Workshop Secure Internet of Things*, Sep. 2015, pp. 5–14.

[157] S. Madakam, R. Ramaswamy, and S. K. Tripathi, "Internet of Things (IoT): A literature review," *J. Comput. Commun.*, 2015.

[158] R. Khan et al., "Future internet: The Internet of Things architecture, possible applications and key challenges," in *Proc. 2012 10th Int. Conf. Front. Inf. Technol.*, Dec. 2012, pp. 257–260.

[159] P. Sethi and S. R. Sarangi, "Internet of Things: Architectures, protocols, and applications," *J. Electr. Comput. Eng.*, 2017.

[160] Q. M. Ashraf and M. H. Habaebi, "Autonomic schemes for threat mitigation in Internet of Things," *J. Netw. Comput. Appl.*, vol. 49, pp. 112–127, 2015.

[161] R. Canzanese, M. Kam, and S. Mancoridis, "Toward an automatic, online behavioral malware classification system," in *Proc. 2013 IEEE 7th Int. Conf. Self-Adaptive Self-Organizing Syst.*, Sep. 2013, pp. 111–120.

[162] L. Bilge and T. Dumitra, "Before we knew it: An empirical study of zero-day attacks in the real world," in *Proc. 2012 ACM Conf. Comput. Commun. Security'12*, New York, NY, 2012, pp. 833–844.

[163] R. Kaur and M. Singh, "A survey on zero-day polymorphic worm detection techniques," *IEEE Commun. Surveys Tuts.*, vol. 16, pp. 1520–1549, Third 2014.

[164] A. Bar-On et al., "Improved key recovery attacks on reduced-round AES with practical data and memory complexities." Cryptology ePrint Archive, Report 2018/527, 2018.

[165] A. Biryukov et al., "Key recovery attacks of practical complexity on AES-256 variants with up to 10 rounds," in *Proc. 29th Ann. Int. Conf. Theory Appl. Cryptogr. Techn. '10*, Berlin, Heidelberg, Springer-Verlag, 2010, pp. 299–319.

[166] I. Thomson, *AES-256 Keys Sniffed in Seconds Using e200 Kit a Few Inches Away*. Accessed: November 11, 2018. [Online]. Available: https://www.theregister.co.uk/2017/06/23/aes_256_cracked_50_seconds_200_kit/

[167] C. C. E. Brier and F. Olivier, "Correlation power analysis with a leakage model," *Cryptogr. Hardw. Embedded Syst.*, pp. 16–29, 2004.

[168] D. Das et al., "ASNI: Attenuated signature noise injection for low-overhead power side-channel attack immunity," *IEEE Trans. Circ. Syst. I*, vol. 65, pp. 3300–3311, Oct. 2018.

[169] D. Das et al., "High efficiency power side-channel attack immunity using noise injection in attenuated signature domain," in *Proc. 2017 IEEE Int. Symp. Hardw. Oriented Security Trust*, 2017, pp. 62–67.

[170] M. Kar et al., "Reducing power side-channel information leakage of AES engines using fully integrated inductive voltage regulator," *IEEE J. Solid-State Circ.*, vol. 53, pp. 2399–2414, Aug. 2018.

[171] A. Singh et al., "Improved power/EM side-channel attack resistance of 128-bit AES engines with random fast voltage dithering," *IEEE J. Solid-State Circ.*, vol. 54, pp. 569–583, Feb. 2019.

[172] T. D. Cnudde, M. Ender, and A. Moradi, "Hardware masking, revisited," *Trans. Cryptogr. Hardw. Embedded Syst.*, vol. 2018, pp. 123–148, 2018.

[173] T.-H. Le et al., "A proposition for correlation power analysis enhancement," in *Proc. 8th Int. Conf. Cryptogr. Hardw. Embedded Syst. '06*, Berlin, Heidelberg, Springer-Verlag, 2006, pp. 174–186.

[174] D. Das et al., "STELLAR: A generic EM side-channel attack protection through ground-up root-cause analysis," in *Proc. 2019 IEEE Int. Symp. Hardw. Oriented Security Trust*, May 2019.

[175] D. Das et al., "X-DeepSCA: Cross-device deep learning side channel attack," in *Proc. 2019 ACM/IEEE Design Autom. Conf.*, 2019.

[176] C. Herder et al., "Physical unclonable functions and applications: A tutorial," *Proc. IEEE*, Aug. 2014, vol. 102, pp. 1126–1141.

[177] U. Ruhrmair and M. van Dijk, "PUFs in security protocols: Attack models and security evaluations," in *Proc. 2013 IEEE Symp. Security Privacy*, May 2013, pp. 286–300.

[178] B. Chatterjee, D. Das, and S. Sen, "RF-PUF: IoT security enhancement through authentication of wireless nodes using in situ machine learning," in *Proc. 2018 IEEE Int. Symp. Hardw. Oriented Security Trust*, Apr. 2018, pp. 205–208.

[179] G. E. Suh and S. Devadas, "Physical unclonable functions for device authentication and secret key generation," in *Proc. 2007 44th ACM/IEEE Design Autom. Conf.*, Jun. 2007, pp. 9–14.

[180] Z. Paral and S. Devadas, "Reliable and efficient PUF-based key generation using pattern matching," in *Proc. 2011 IEEE Int. Symp. Hardw.-Oriented Security Trust*, Jun. 2011, pp. 128–133.

[181] D. E. Holcomb, W. P. Burleson, and K. Fu, "Power-Up SRAM state as an identifying fingerprint and source of true random numbers," *IEEE Trans. Comput.*, vol. 58, pp. 11981210, Sep. 2009.

[182] D. E. Holcomb, W. P. Burleson, and K. Fu, "Initial SRAM state as a fingerprint and source of true random numbers for RFID tags," in *Proc. Conf. RFID Security*, 2007.

[183] Y. Su, J. Holleman, and B. P. Otis, "A Digital 1.6 pJ/bit chip identification circuit using process variations," *IEEE J. Solid-State Circ.*, vol. 43, pp. 69–77, Jan. 2008.

[184] S. Sutar et al., "D-PUF: An intrinsically reconfigurable DRAM PUF for device authentication and random number generation," *ACM Trans. Embed. Comput. Syst.*, vol. 17, pp. 17:1–17:31, Dec. 2017.

[185] U. Ruhrmair et al., "Modeling attacks on physical unclonable functions," in *Proc. 17th ACM Conf. Comput. Commun. Security '10*, New York, NY, 2010, pp. 237–249.

[186] C. Zhou, K. K. Parhi, and C. H. Kim, "Secure and reliable XOR Arbiter PUF design: An experimental study based on 1 trillion challenge response pair measurements," in *Proc. 54th Ann. Design Autom. Conf. '17*, New York, NY, 2017, pp. 10:1–10:6.

[187] B. Chatterjee et al., "RF-PUF: Enhancing IoT security through authentication of wireless nodes using in situ machine learning," *IEEE Internet of Things J.*, 2018.

[188] M. A. Alsheikh et al., "Machine learning in wireless sensor networks: Algorithms, strategies, and applications," *IEEE Commun. Surveys Tuts.*, vol. 16, pp. 1996–2018, Fourthquarter 2014.

[189] R. Masiero et al., "Data acquisition through joint compressive sensing and principal component analysis," in *Proc. GLOBECOM 2009—2009 IEEE Global Telecommun. Conf.*, Nov. 2009, pp. 1–6.

[190] T. M. Cover, "Hypothesis testing with finite statistics," *Ann. Math. Statist.*, vol. 40, pp. 828–835, Jun. 1969.

[191] T. Park and W. Saad, "Learning with finite memory for machine type communication," ArXiv. 1610.01723, vol. abs/1610.01723, 2016.

[192] K. Drakopoulos, A. E. Ozdaglar, and J. N. Tsitsiklis, "On learning with finite memory," ArXiv. 1209.1122, vol. abs/1209.1122, 2012.

[193] R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning*. Cambridge, MA: MIT Press, 1998, 1st ed.

[194] L. P. Kaelbling, M. L. Littma, and A. W. Moore, "Reinforcement learning: A survey," *J. Artif. Intell. Res.*, vol. 4, 1996.

[195] D. Athukoralage et al., "Regret based learning for UAV assisted LTE-U/WiFi public safety networks," in *Proc. 2016 IEEE Global Commun. Conf.*, Dec. 2016, pp. 1–7.

[196] Z. Liu and I. Elhanany, "RL-MAC: A QoS-aware reinforcement learning based MAC protocol for wireless sensor networks," in *Proc. 2006 IEEE Int. Conf. Netw. Sensing Contr.*, Apr. 2006, pp. 768–773.

[197] A. Amravati et al., "A 55nm time-domain mixed-signal neuromorphic accelerator with stochastic synapses and embedded reinforcement learning for autonomous micro-robots," in *Proc. 2018 IEEE Int. Solid-State Circ. Conf.*, Feb. 2018, pp. 124–126.

[198] M. Camelo and J. F. a. S. Latre, "A scalable parallel Q-learning algorithm for resource constrained decentralized computing environments," in *Proc. 2016 2nd Workshop Machine Learn. HPC Environ*, Nov. 2016, pp. 27–35.

[199] A. M. Printista, M. L. Errecalde, and C. I. Montoya, "A parallel implementation of Q-learning based on communication with cache," in *J. Comput. Sci. Techn.*, 2002.

[200] R. M. Kretchmar, "Reinforcement learning algorithms for homogenous multi-agent systems," in *Workshop Agent Swarm Program.*, 2003.

[201] C. F. Camerer, T.-H. Ho, and J.-K. Chong, "A cognitive hierarchy model of games*," *Quart. J. Econ.*, vol. 119, pp. 861–898, Aug. 2004.

[202] B. W. Rogers, T. R. Palfrey, and C. F. Camerer, "Heterogeneous quantal response equilibrium and cognitive hierarchies," *J. Econ. Theory*, vol. 144, no. 4, pp. 1440–1467, 2009.

Baibhab Chatterjee is currently pursuing a PhD degree with the School of Electrical Engineering, Purdue University, West Lafayette, IN. His research interests include low-power analog, RF, and mixed-signal circuit design for secure biomedical/IoT applications. Chatterjee has a BTech in electronics and communication engineering from National Institute of Technology Durgapur, Durgapur, India (2011), an MTech in electrical engineering from the IIT Bombay, Mumbai, India (2015). He is a Student Member of the IEEE.

Ningyuan Cao is currently pursuing a PhD degree in electrical engineering with the Georgia Institute of Technology, Atlanta, GA. His research interests include low-power machine-learning ASIC design, wireless sensor design and power management, and energy harvesting circuit design. Cao has a BS in electrical engineering from Shanghai Jiaotong University, Shanghai, China (2013), an MS in electrical engineering from Columbia University, New York, NY (2015). He is a Student Member of the IEEE.

Arijit Raychowdhury is an Associate Professor with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA. His research interests include low-power digital and mixed-signal circuit design, device–circuit interactions, and novel computing models and hardware realizations. Raychowdhury has a PhD in electrical and computer engineering from Purdue University, West Lafayette, IN (2007). He is a Senior Member of the IEEE.

Shreyas Sen is an Assistant Professor with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN. His research interests include mixed-signal circuits/systems for Human Body Communication, IoT, Biomedical, and Hardware Security. Sen has a PhD in electrical and computer engineering from Georgia Tech, Atlanta, GA (2011). He is a Senior Member of the IEEE.

■ Direct questions and comments about this article to Baibhab Chatterjee, Purdue University, West Lafayette, IN 47907, USA; bchatte@purdue.edu.