Energy-Autonomous Wireless Communication for Millimeter-Scale Internet-of-Things Sensor Nodes

Yajing Chen, Nikolaos Chiotellis, Li-Xuan Chuo, Carl Pfeiffer, Member, IEEE, Yao Shi, Student Member, IEEE, Ronald G. Dreslinski, Anthony Grbic, Fellow, IEEE, Trevor Mudge, Life Fellow, IEEE, David D. Wentzloff, Member, IEEE, David Blaauw, Fellow, IEEE, and Hun Seok Kim, Member, IEEE

Abstract—This paper presents an energy-autonomous wireless communication system for ultra-small Internet-of-Things (IoT) platforms. In the proposed system, all necessary components, including the battery, energy-harvesting solar cells, and the RF antenna, are fully integrated within a millimeter-scale form factor. Designing an energy-optimized wireless communication system for such a miniaturized platform is challenging because of unique system constraints imposed by the ultra-small system dimension. The proposed system targets orders of magnitude improvement in wireless communication energy efficiency through a comprehensive system-level analysis that jointly optimizes various system parameters, such as node dimension, modulation scheme, synchronization protocol, RF/analog/digital circuit specifications, carrier frequency, and a miniaturized 3-D antenna. We propose a new protocol and modulation schemes that are specifically designed for energy-scarce ultra-small IoT nodes. These new schemes exploit abundant signal processing resources on gateway devices to simplify design for energy-scarce ultra-small sensor nodes. The proposed dynamic link adaptation guarantees that the ultra-small IoT node always operates in the most energy efficient mode for a given operating scenario. The outcome is a truly energy-optimized wireless communication system to enable various classes of new applications, such as implanted smart-dust devices.

Index Terms—Ultra-small IoT node, ultra-low power wireless communication, energy optimized communication.

I. INTRODUCTION

ULTRA-SMALL Internet-of-Things (IoT) sensor nodes with perpetual energy harvesting have come into reality empowered by very-large-scale system integrated (VLSI) circuit innovations [1]–[5] and fabrication technology improvement. Leading into the realistic world of ‘smart dust’ [1]–[6], ultra-small (specifically in millimeter-scale) IoT platforms present a wide range of new applications such as biomedical implants [7], [8], security/safety surveillance, infrastructure monitoring [9], [10], and smart building [11], which are all extremely platform size sensitive.

Manuscript received January 31, 2016; revised May 14, 2016; accepted August 4, 2016. Date of publication September 20, 2016; date of current version December 29, 2016. This work was supported in part by the Samsung Global Research Outreach Program, in part by ARM Ltd., in part by NSF under Grant CNS-1111541, and in part by a Gift from Intel. (Corresponding author: Hun Seok Kim.)

The authors are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: hunseok@umich.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSAC.2016.2612041

Fig. 1. Millimeter-scale Michigan Micro Mote (M3), pressure sensing system (left) and imaging system (right).

We envision a millimeter-scale, general purpose computing platform (Fig. 1) [1]–[5] that is small enough to be seamlessly integrated into the real world without being noticed. Recent research in VLSI circuits demonstrates that realizing such an ultra-small system is indeed feasible by integrating a rechargeable battery, a solar energy harvester layer, various sensor (temperature, pressure, imager) layers, and a general purpose processor layer; all within a millimeter-scale form-factor (see Fig. 1). The extremely small form-factor of these systems imposes a critical challenge in system power and energy management. Since battery replacement is impractical, these systems target energy-autonomous operation, employing a millimeter-scale rechargeable thin-film battery that is continually trickle charged by harvested ambient energy [1]–[5].

In millimeter-scale IoT platforms, wireless communication dominates overall power/energy consumption [1]–[5], [14]–[16]. The power breakdown of the millimeter-scale, energy-autonomous Michigan Micro Mote (M3) sensor node (Fig. 1) [1]–[5] reveals that wireless communication consumes more than 65% of the overall power budget even with aggressive duty cycling. Therefore, enabling energy optimized wireless communication is the most critical issue in prolonging the lifespan of energy-constrained ultra-small IoT platforms, and realizing perpetual device operation powered only by energy harvesting.

In this paper, we present a truly energy-autonomous, fully self-contained wireless communication system that optimally utilizes the scarce energy/power resources available in ultra-small scale IoT platforms. To achieve this goal, a cross-layer, system-level optimization framework is proposed to jointly...
optimize various system parameters such as modulation scheme, synchronization protocol, RF/analog/digital circuit specifications, data rate, carrier frequency, miniaturized 3D antenna efficiency, etc. The proposed optimizations will be performed under stringent system constraints that are unique to millimeter-scale energy-harvesting IoT platforms. Given a millimeter-scale IoT node dimension, we seek to explore the rich tradeoff space between the miniaturized 3D antenna size, battery capacity, and the energy reservoir capacitor size to achieve the maximum energy efficiency for wireless communication. The ultra-small form-factor brings with it the significant challenge of realizing long distance communication especially for indoor, non-line-of-sight operation. Prior works in the mm-scale wireless communication systems had limited communication ranges. For example, the state-of-the-art mm-scale designs [12], [13] report \( \leq 7m \) and 0.5\text{m} using 8\text{GHz} and 24\text{GHz} carrier frequencies respectively in direct line-of-sight scenarios. In this work, we propose an adaptive communication solution that can support > 15\text{m} link distance in indoor, non-line-of-sight conditions.

We assume that each ultra-small IoT node is paired with a gateway such as a smartphone or a WiFi access point. Direct communication among ultra-small IoT sensor nodes is not the focus of this work. Since the energy/power efficiency of the gateway device is not a primary concern, it is reasonable to assume that the gateway is equipped with abundant resources for advanced signal processing. Exploiting this asymmetry, this paper will investigate an energy-optimized communication system for a power-constrained IoT sensor node with very dissimilar transmitter and receiver configurations. The proposed system will demonstrate that powerful signal processing at the gateway can mitigate circuit impairments on the ultra-low power (ULP, \( < 100\mu W \)) sensor node. Hence, the circuit specification on the sensor node can be significantly relaxed for better energy efficiency, benefiting from gateway signal processing. A new synchronization protocol is investigated to improve the sensor node receiver energy efficiency by utilizing accurate timing and frequency offset estimation attainable on the gateway. This synchronization scheme eliminates the need for a power-demanding phase-lock-loop (PLL) and an external frequency reference crystal for the millimeter-scale sensor node wireless transceiver.

Conventional modulation schemes, such as on-off-keying (OOK) and binary frequency shift keying (FSK), widely used in prior ULP transceiver works [4], [5] are far from optimal for the proposed system since these conventional simplistic modulation schemes lack the ability to dynamically adapt to various operating scenarios. We propose a new modulation-coding scheme that is specifically designed for millimeter-scale IoT sensor nodes to expand their connectivity in an energy-efficient way. In addition, we will show that gateway guided link adaptation provides an order(s)-of-magnitude improvement in data rate and/or link distance by automatically adjusting the modulation-coding scheme and other modulation parameters on the fly for the millimeter-scale IoT sensor node.

This paper is organized as follows. Section II discusses the energy-autonomous millimeter-scale communication system design considerations. A new synchronization protocol is discussed in Section II. Section III provides mathematical models of the modulation scheme, coding, synchronization, energy consumption, and data rate of the proposed system. In Section IV, we mathematically define the dynamic link adaptation problem for various objective functions, and quantify the optimum results and the impact of different system configurations. Section V concludes the paper.

II. SYSTEM DESIGN

In this section, we will introduce system constraints of the energy-autonomous millimeter-scale communication system. The proposed system design is driven by the objective of achieving maximum energy efficiency, longest link distance, and/or highest data rate. Critical system design decisions such as the antenna type and modulation scheme will be justified in this section.

A. Millimeter-Scale Wireless Communication

System Constraints

The proposed energy-autonomous wireless system configuration is depicted in Fig. 2. A cubic shape form-factor is considered to minimize the system volume and to constrain the length of the largest dimension. The transceiver circuits, thin-film battery, and other discrete components including the energy reservoir capacitor (\( C_{\text{res}} \)) are sandwiched between the solar energy harvester panel and the millimeter-scale 3D antenna. A stacked layer approach introduced in [1]–[5] shall be applied to integrate multiple functional layers such as a processor layer and sensor layers on top of each other with minimal area overhead.
The millimeter-scale dimension (largest dimension is less than $5\text{mm}$) imposes several unique constraints for energy-autonomous wireless systems. For example, the peak current of a millimeter-scale thin film battery cannot exceed $\approx 100\text{μA}$ [1]–[5] because of the internal battery resistance. We utilize a discrete $\leq 1\text{μF}$ capacitor (see Fig. 2) as the energy reservoir (or energy buffer) to address the mismatch between the instantaneous peak power requirement for transceiver operation and the peak battery current limitation. The impact of this energy reservoir capacitor will be analyzed in Section IV.C. The solar energy harvesting layer is exposed to make each of the two dipoles resonant. We will not concern ourselves with matching the antenna to a specific impedance because the antenna is part of the oscillator circuit, and does not have to be matched. Capacitive (inductive) nature of an electric (magnetic) dipole antenna requires a lumped inductor (capacitor) to make the antenna resonate. Since lumped capacitors exhibit much higher quality factor than inductors, the antenna efficiency, $\eta_{e}$ and $\eta_{m}$, of the electric and magnetic dipole antenna should be reevaluated considering realistic quality factors of the lumped series elements.

The required inductance for an electric dipole is $L^e = 1/(\omega^2 C^e)$, whereas a magnetic dipole resonates with a series capacitance $C^m = 1/(\omega^2 L^m)$, where $\omega$ is the angular frequency. It is reasonable to assume that the quality factor of the inductor and capacitor is $Q_L = (\omega L^e/R_L) = 50$ and $Q_C^m = 1/(\omega C^m R_C^m) = 250$, respectively. These values are typical among surface mount inductors and capacitors in the carrier frequency range of interest, $f_c < 6\text{GHz}$. This carrier frequency selection is justified in Section II.C.

The radiation efficiency for electric and magnetic dipoles is a function of antenna dimension and carrier frequency as shown in Table I and Fig. 3. Considering the realistic quality factor of lumped elements, magnetic dipoles can lead to higher overall radiation efficiencies in millimeter-scale designs, outperforming electric dipoles ($\hat{\eta}_m > \hat{\eta}_e$) as shown in Fig. 3 bottom, given that $f_c \leq 5\text{GHz}$. Notice the opposite is true ($\eta_m \ll \eta_e$) if lossless lumped components were used (Fig. 3 top). Based on this observation, we propose the use of magnetic dipole antennas for millimeter-scale communication systems.

Fig. 2 (on the right) shows the design of the millimeter-scale 3D antenna, which can be fabricated as a multi-layer...
efficiency

radius to represent the dimension of the antenna assuming the remainder of this paper, we use the effective antenna radiation efficiency as the 3D magnetic antenna design. For in the theoretical model in Table I that provides the same antenna radius in Table II corresponds to the radius in Table II with a 1GHz carrier frequency. Simulations were radiation efficiency of the magnetic antenna design is provided printed circuit board. For various dimensions, the simulated radiation efficiency of the magnetic antenna design is provided in Table II with a 1GHz carrier frequency. Simulations were performed using ANSYS Electronics Desktop, a commercially available full-wave electromagnetic solver. The ‘effective’ antenna radius in Table II corresponds to the radius $r_{\text{ant}}$ in the theoretical model in Table I that provides the same radiation efficiency as the 3D magnetic antenna design. For the remainder of this paper, we use the effective antenna radius to represent the dimension of the antenna assuming the efficiency $\hat{\eta}_m$ predicted by Table I.

It is worth noting that although the millimeter-scale magnetic dipole is overall more efficient than the electric dipole counterpart, its efficiency $\hat{\eta}_m$ is still very poor ($<1\%$ for $1\text{GHz}$ operation as shown in Table II) compared to conventional systems with centimeter-scale antennas. The modulation-coding scheme proposed in later sections is carefully designed to address this challenge, and to eventually achieve $>15m$ distance links for the millimeter-scale system.

C. Carrier Frequency Selection

The optimal carrier frequency of the proposed system is chosen to maximize the signal-to-noise ratio (SNR). The carrier frequency affects the pathloss of the signal as well as the antenna efficiency given its millimeter-scale dimensions. The RF circuit power efficiency is also affected by the carrier frequency selection.

We consider an indoor environment that includes one layer of wall/floor which blocks the line-of-sight (i.e., non-LOS) between the gateway and the sensor node. While a higher carrier frequency is preferred for higher antenna efficiency given the antenna’s millimeter-scale size (see Fig. 3), a lower frequency significantly reduces the pathloss of the signal in line-of-sight as well as the loss from wall/floor penetration. Therefore, the optimal carrier frequency selection needs to strike a balance in this non-trivial tradeoff. In our analysis, we use the modified ITU indoor average power propagation loss model [20] given by (1)

$$L(f_c, d) = 20 \log_{10} \left( \frac{4\pi f_c}{c} \right) + 30 \log_{10}(d) + FL(f_c) - g_{\text{sensor}}(f_c) - g_{\text{gateway}} [\text{in dB}]. \quad (1)$$

In (1), $c$ is the speed of light and $g_{\text{sensor}} = 10\log_{10}(1.5 \hat{\eta}(f_c))$ is the antenna gain that combines the antenna directivity ($= 1.5$) and the radiation efficiency $\hat{\eta}(f_c)$ of the 3D millimeter-scale antenna discussed in Section II.B. The gateway antenna dimension is not our primary concern, and we assume the gateway antenna gain $g_{\text{gateway}}$ is constant and equal to 3dB. The term $30\log_{10}(d)$ dictates the pathloss (using an exponent of 3.0 instead of 2.0 in the theoretical free-space pathloss) as a function of distance $d$. $FL(f_c)$ is the additional pathloss due to one layer of floor/wall that is optionally applied to non-LOS scenarios. The original ITU model [20] suggests $FL = 9dB, 14dB$ and $16dB$ at $915MHz, 2.4GHz,$ and $5.2GHz,$ respectively. We modified this term to be linear with carrier frequency as $FL(f_c) = 4 \times (f_c \text{ in GHz}) + 7$ in dB based on the measurement results from [21]–[23] (averaged in log domain). For line-of-sight (LOS) scenarios, we set $FL(f_c) = 0$. Fig. 4 shows the propagation model (1) evaluated for both LOS and non-LOS scenarios for various antenna dimensions and carrier frequencies. It is worth noting that, considering both pathloss and antenna efficiency, $\approx 1GHz$ operation is optimal in the non-LOS scenario, when the antenna dimension is limited to $\approx 2mm$. This is because the additional pathloss (including wall penetration) offsets the higher antenna efficiency at higher frequencies. Meanwhile, for LOS (outdoor) operation, a higher frequency ($\geq 5GHz$), possibly even at mm-wavelengths, [13] is desired to minimize the overall propagation loss (see Fig. 4, on the right). We have made a decision to design the system with a carrier frequency of $\approx 1GHz$ because 1) relative long range ($>15\text{ meter}$) operation in NLOS through a wall is desired, and 2) a significantly different model (for antenna efficiency, circuit power efficiency, etc.) is needed to analyze higher frequencies when the wavelength approaches mm-scale. The impact of carrier frequency on the millimeter-scale
communication system energy efficiency, data rate, and link distance will be quantified in Section IV.C.

D. Modulation Scheme

A conventional wireless transmitter uses a phase-locked loop (PLL) for carrier frequency synthesis to support coherent modulation and/or sub-channel tuning for non-coherent modulation. The power consumption of a conventional PLL architecture with the state-of-the-art technology is around $1 - 10 \text{mW}$ [24]–[26]. This much power overhead is unacceptable for power-constrained millimeter-scale sensor nodes where sustained battery power (and harvested power) is limited to 100s of $\mu W$. Furthermore, a PLL typically requires an external crystal oscillator as the phase reference. The commercial crystal has a volume larger than $1.6 \times 1.0 \times 0.5 \text{mm}^3$ [27], increasing overhead on the overall system dimension.

The proposed system integrates the RF transceiver with a high quality factor magnetic antenna as described in Section II.B. Replacing a conventional PLL, we propose a ‘power oscillator’ technique where the magnetic antenna acts as an inductor, and it forms a resonant tank together with a tunable capacitor bank to generate the desired carrier frequency. When it is oscillating, the power will be radiated from the antenna without a power amplifier [4], [28]. Such implementation has advantages of lower cost, inherent frequency generation, high transmit efficiency ($\approx 30\%$), and inherent antenna matching. The downside of the power oscillator architecture is that it only allows non-coherent modulation schemes.

Note that the transmit efficiency of the proposed power oscillator architecture is maximized at a certain (100s of $\mu A$) bias current given $\approx 4V$ power supply [4], [28]. This current consumption is much lower than that of a conventional architecture with a PLL and power amplifier, but it still exceeds the range of the peak battery current of the proposed millimeter-scale system. This implies that transmitting a continuous signal at a reasonable transmit power efficiency is infeasible when the current is directly drawn from the millimeter-scale thin-film battery [1]–[5].

Circumventing this issue, we employ a sparse pulse based non-coherent modulation scheme. In the proposed scheme, the transmit pulse energy on the sensor node is drawn from the energy reservoir capacitor, not directly from the battery. This capacitor is trickle charged by the peak-current limited millimeter-scale thin-film battery while the battery is constantly recharged by the harvested solar energy. This particular power management architecture implies that pulses cannot be repeated until the trickle charging time of the reservoir capacitor, $T_{\text{charge}}$, is elapsed between pulses. This $T_{\text{charge}}$ time is regarded as the ‘forced idle time’ in our modulation scheme.

Exploiting this inherent pulse sparsity, we propose M-ary pulse position modulation (PPM) [29], [30] as the sensor node transmission scheme. In a conventional system, M-ary PPM has been investigated with the purpose of enhancing energy efficiency (a single pulse can contain multiple information bits) [29]. The drawback of M-PPM in a conventional system is its lower bandwidth efficiency as the symbol duration is proportional to M whereas the number of bits per symbol is a function of $log_2(M)$. On the contrary, in our proposed millimeter-scale system, the recharging time of the capacitor is inevitable and usually much longer than the pulse duration. Thus, the symbol duration is dominated by $T_{\text{charge}}$ if M is in a reasonable range ($\leq 64$) as depicted in Fig. 5. The forced idle time, $T_{\text{charge}}$, motivates the usage of $M > 2$ to enable higher energy efficiency and to increase the symbol rate at the same time.

The M-PPM symbol error rate is governed by the energy in a symbol. Since the proposed system draws energy from the reservoir capacitor, its capacitance limits the maximum energy per pulse constraining the maximum distance. Addressing this issue, we allow N-repetition of pulses to represent a symbol (each pulse is separated by $T_{\text{charge}}$, see Fig. 5) to expand the link distance beyond the limit of the reservoir capacitance at the potential cost of degraded data throughput.

For the proposed modulation scheme, the bandwidth efficiency penalty of M-PPM over binary-PPM is relatively insignificant because of inherent sparsity. Thus, we utilize numerous possible pulse positions in the sparse transmit signal to absorb error correction coding redundancy. In Section III and IV, we exploit the sparsity of the modulated signal either (1) to convey multiple information bits per pulse to maximize energy efficiency (and/or data rate) in a short distance or, (2) to maximize coding gain for longer distance. The adaptive N-repetition M-PPM scheme combined with variable rate convolutional coding is a powerful technique to realize an optimized communication link addressing different use-case scenarios and link objectives. Section IV provides in-depth discussion on distance - energy efficiency tradeoffs with the proposed dynamic N-repetition M-PPM modulation scheme with a variable rate convolutional code.

E. Gateway Guided Synchronization and Link Adaptation

We assume that each ultra-small IoT node is paired with a gateway device such as a smartphone or an access point. Direct communication among ultra-small IoT nodes is not the focus of this work. Since the energy/power efficiency of the gateway is not a primary concern, it is reasonable to assume that the gateway is equipped with powerful signal processing while the sensor node is extremely power constrained. Exploiting this asymmetry leads toward very dissimilar transmitter and receiver configurations for a millimeter-scale sensor node.

The PLL-free crystal-less RF transceiver discussed in Section II.D is a key enabler to achieve ultra-small integration for an energy-autonomous ULP sensor node. However, it imposes a new challenge in frequency predictability.
To address this issue, we propose a gateway assisted synchronization protocol that is initiated by sensor node transmission. In the proposed protocol, the gateway estimates the RF carrier frequency offset (CFO) and the baseband sampling frequency offset (SFO) between the gateway and the sensor node via multi-hypotheses correlations. Once the gateway identifies CFO and SFO, it sends a customized packet that ‘pre-compensates’ carrier and baseband frequency offsets for a particular sensor node. This gateway-assisted synchronization allows PLL-free crystal-less sensor node implementation, enabling ultra-low power and ultra-small system integration. It also eliminates the need for timing and frequency synchronization at the millimeter-scale sensor node, which is often the most power-dominant processing in conventional wireless receiver baseband.

The proposed protocol works in the following procedure (illustrated in Fig. 6):

1) The sensor node initiates communication by sending a set of sparse pulses, Synch_HDR, with a pre-defined (node ID dependent) pseudo-random interval.

2) Gateway is always listening and it detects the Synch_HDR via the multi-hypotheses correlation (algorithm in Section III.B) covering all possible ranges of the CFO and SFO. The instantaneous SNR is obtained based on the peak correlation value. The CFO and SFO are estimated at the gateway as a result of Synch_HDR detection.

3) After transmitting the Synch_HDR, the sensor node enters the receive mode. Demodulation at the sensor node starts at a predefined delay (turnaround time, $T_{\text{turn}}$), calculated using the sensor node baseband clock (implemented without PLL and crystal reference; see [31] for an example). At the sensor node, the correlation process searching for the symbol boundary is unnecessary if the symbol from the gateway arrives at the precise timing.

4) Gateway calculates the turnaround time using the estimated SFO to synchronize with the sensor node. It also adjusts carrier frequency $f_c$, based on the CFO estimation. In parallel, gateway solves the link adaptation problem (Section IV) to identify modulation parameters for the sensor node to maximize an objective function (data rate or energy efficiency) given the instantaneous SNR. Gateway sends the message to the sensor node at the exact symbol boundary that the sensor node is expecting. The message contains a command to select the optimal operating mode.

5) The sensor node sends more messages using the optimal mode dictated by the gateway.

Dynamic modulation parameters such as the pulse width, coding rate, and modulation size (M) for M-PPM allow the proposed system exploiting the rich tradeoff space to maximize a specific objective function such as the data rate and/or link distance. The mathematical formulation and analysis of the link adaptation will be discussed in Section IV.

In this work, we assume the sensor node is the main source of the wireless communication data traffic. Typical messages from the gateway to the sensor node are assumed to be short commands that are less bandwidth demanding. This is a valid assumption for majority of IoT applications where distributed sensor nodes collect various sensing data including audio/image.

Therefore, our link adaptation strategy with dynamic sparse M-PPM modulation is only applied to sensor node transmission. The gateway transmit signal, on the other hand, uses a simple static modulation scheme such as on-off keying (OOK). Note that the gateway transmit signal does not have to be sparse. In the proposed scheme, the gateway adjusts its transmit power so that a certain minimum signal sensitivity level at the sensor node receiver operating with 10s of $\mu W$ power budget (sustainable with thin-film battery power) is always satisfied. The comprehensive survey on the state-of-the-art ultra-low power ($< 100 \mu W$) OOK receiver designs [32] suggests that this scheme is certainly feasible especially when the baseband processing is greatly simplified by the proposed gateway-guided synchronization. Thus, our proposed scheme focuses on the sensor node transmission, assuming the other direction (gateway transmission - sensor node reception) does not limit the system performance.

Table III summarizes system constraints and challenges that are specific for the millimeter-scale system and proposed solutions to tackle each of them.

![Fig. 6. Gateway guided synchronization and link adaptation.](image)

![Fig. 7. Baseband Processing at Sensor Node.](image)

III. SYSTEM MODELING

A. Sensor Node Transmission Signal Modeling

The proposed modulation and coding scheme for sensor node transmission is depicted in Fig.7. The information bit stream is fed into a multi-rate convolutional code encoder, then mapped into M-PPM pulse signals. The M-PPM pulse is optionally repeated N times, and pulse shaping is performed with tunable pulse width ($T_{\text{pulse}}$).

In the proposed modulation and coding scheme, the M-PPM modulation is tightly coupled with convolutional encoding. The modulation-coding rate ($C_r$) is dynamically chosen from the set; $C_r \in \{\ldots, 3, 2, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\}$. Convolutional encoding is bypassed when $C_r \geq 1$, and higher data rate
is achieved by using $M = 2^{Cr}$ to carry $C_r$ information bits per symbol. On the other hand, when convolutional encoding is enabled ($C_r < 1$), $M = 2^{1/C_r}$ is enforced to convey $1/C_r$ coded bits per M-PPM symbol and, equivalently, one information bit per symbol. $M$ and $C_r$ satisfy the relationship (2). The usage of $C_r < 1$ coding rate is motivated by the transmit signal sparsity, where numerous pulse positions are available without a significant data rate penalty

$$M = \begin{cases} 2^{C_r} & C_r \geq 1 \\ 2^{1/C_r} & C_r < 1 \end{cases}$$

(2)

$$c_j = \left[ b_1C_r, \ldots, b_{(j+1)C_r-1} \right] T_{sym} \\ G(\sigma_j, b_j, C_r) C_r \geq 1 \\ C_r < 1.$$

(3)

We denote the j-th output from the encoder as $c_j$, a vector of size $\log_2 M \times 1$, given input bit stream $(b_0, b_1, \ldots)$. $G(\sigma_j, b_j, C_r)$ is the convolutional code generator function that produces an $1/C_r \times 1$ output vector per single input bit $b_j$ given the convolutional code trellis state $\sigma_j$. The trellis state is updated by $\sigma_j = \sum_{l=1}^v b_{j-l}2^{l-1}$ where $v$ is the code constraint length. We use convolutional code generator functions specified in [33] for various coding rate $C_r$ and $v$. Note that each $c_j$ vector is mapped to a single M-PPM symbol regardless of $C_r$. The mapping between the M-PPM symbol index $m_j \in \{0, 1, \ldots, M-1\}$ and $c_j$ is given by $m_j = c_j^T \mathbf{p}$ where $\mathbf{p} = [2^{\log_2 M-1}, \ldots, 2^1, 2^0]^T$ is the M-PPM position mapping vector. Fig. 8 shows an example of $v = 3$ convolutional coding with $C_r = 1/3, M = 2^{1/C_r} = 8$ PPM.

When the energy per pulse is limited by the capacity of the energy reservoir, we increase the symbol energy by $\frac{\nu}{T_{sym}}$ units. The $n$-th repeated pulse position of the j-th symbol is denoted by $T_{p}(j,n)$

$$T_{p}(j,n) = T_{pulse}m_j + (j-1)T_{sym} + (n-1)T_{idle}, \quad n = 1, \ldots, N.$$  

(4)

In (4), $T_{sym}$ is the symbol duration that consists of $N$ pulses. $T_{pulse}$ is the pulse width, which can be dynamically configured as a result of link adaptation discussed in Section IV. $T_{idle} = \max\{T_{charge}, M T_{pulse}\}$ is the forced idle time between pulse repetition determined by the maximum of the energy reservoir capacitor charging time $T_{charge}$ and the non-overlapping spacing for M-PPM $M T_{pulse}$. In case the system’s thin-film battery current is very limited, the charging time dominates $T_{idle} = T_{charge} \gg M T_{pulse}$ for a reasonable $M$ ($\leq 64$). The symbol duration of an $N$ repetition M-PPM symbol is obtained by (5). Given a pulse shape function $p(t)$ with support $[0, T_{pulse}]$, the sensor node transmit signal is represented by (6), where $\ast$ stands for convolution and $\delta(x)$ is the Dirac-Delta function

$$T_{sym} = M T_{pulse} + (N-1)T_{idle} + T_{charge}$$

(5)

$$s(t) = p(t) \ast \sum_{j, n=1}^N \delta(t - T_{p}(j,n)).$$

(6)

### B. Synchronization at Gateway

The gateway is always searching for a Synch_HDR from a sensor node that initiates communication. The proposed Synch_HDR detection process at the gateway provides reliable timing and frequency offset synchronization, which is critical to enable PLL-free crystal-less implementation for the millimeter-scale sensor node. The gateway employs non-coherent demodulation and synchronization that is based on received sample power. Coherent demodulation is infeasible because the sensor node cannot maintain phase coherency without a PLL.
Fig. 9 depicts the Synch_HDR detection process at the gateway with carrier frequency offset (CFO) and sampling frequency offset (SFO) tracking. The $h_{CFO}$ and $h_{SFO}$ are the number of hypotheses for discretized CFO and SFO respectively. During Synch_HDR detection, the incoming baseband ADC samples are mixed with various CFO hypotheses $(f_{CFO}^{(1)}, f_{CFO}^{(2)}, \ldots, f_{CFO}^{(h_{CFO})})$. Each CFO mixer output is low-pass filtered, convoluted with a matched filter (MF), and then power converted. The instantaneous maximum output, $p_{\text{max}}(t)$ is selected among $h_{CFO}$ MF outputs until the Synch_HDR is detected. Each maximum output is associated with a specific CFO hypothesis, $f_{CFO}^{(1)}(t)$, at time instance $t$.

This instantaneous maximum output signal $p_{\text{max}}(t)$ is then correlated with $h_{SFO}$ impulse sequences. The $j$-th correlation sequence $\sum_{n=1}^{N_p} \delta(t - t_n^{(j)})$ has pulse positions $t_n^{(j)}$, $n = 1, 2, \ldots, N_p$ that are determined by the predefined pulse interval in the Synch_HDR that is adjusted according to the $j$-th SFO hypothesis. $N_p$ is the number of pulses in a Synch_HDR. The Synch_HDR is successfully detected when the maximum from multiple hypotheses correlations exceeds a certain threshold. Consequently, the SFO estimate $\hat{f}_{SFO}$ is obtained by a particular SFO hypothesis that maximizes the correlation as in (7). In addition, the CFO estimate $\hat{f}_{CFO}$ is computed by (7), taking the average of $f_{CFO}^{\text{max}}(t)$ sampled at the pulse positions given by CFO estimation

$$\hat{f}_{SFO} = \underset{j=1,\ldots,h_{SFO}}{\arg \max} \int_{\mathbb{R}} \sum_{n=1}^{N_p} p_{\text{max}}(t) \delta(t - t_n^{(j)}) dt,$$

$$\hat{f}_{CFO} = \frac{1}{N_p} \sum_{n=1}^{N_p} f_{CFO}^{\text{max}}(t_n^{(w)}). \tag{7}$$

The left plot in Fig. 10 shows the performance of this CFO estimation scheme. Simulated with $T_{\text{pulse}} = 4 \mu s$, $C_r = 1$ and $N = 1$, 0.5dB performance degradation is observed from the ideal (no CFO) case when CFO is set to 800ppm (= 800kHz at $f_c = 1GHz$). This 800ppm CFO is 3.2× larger than the $T_{\text{pulse}} = 4 \mu s$ signal bandwidth. Once the Synch_HDR is detected, the mixing path that corresponds to $\hat{f}_{CFO}$ remains active while all other mixers are disabled. The matched filter output $p_{\text{max}}(t)$ is resampled using the SFO estimate $\hat{f}_{SFO}$.

After successful Synch_HDR detection, multiple hypotheses correlations to evaluate (7) remain active during the data modulation process tracking the residual SFO that might affect the system performance. Since $p_{\text{max}}(t)$ is resampled with $\hat{f}_{SFO}$, the SFO hypotheses $(f_{SFO}^{(1)}, f_{SFO}^{(2)}, \ldots, f_{SFO}^{(h_{SFO})})$ can now be realigned with finer granularity. As data pulse demodulation continues, expected pulse positions $t_n^{(j)}$ that originally all belong to Synch_HDR are sequentially replaced by detected data pulse positions for the residual offset tracking in a decision feedback fashion. The oldest pulse position in the hypothesis is replaced by the latest detected pulse position, and the $\hat{f}_{SFO}$ tracking continues until the end of the packet.

The right plot in Fig. 10 shows the simulation performance of the proposed CFO estimation and tracking algorithm. The simulation results confirm that performance degradation due to CFO is limited to an acceptable range ($\leq 1dB$ SNR loss) when $\leq 10000$ppm (1%) CFO is tested. This CFO requirement is very reasonable for ultra-low power ($< 1\mu W$) clock design [31] that does not require a reference crystal. Based on the synchronization algorithm performance shown in Fig.10, we argue that the proposed scheme mitigates CFO and SFO well, and the impact of residual CFO and SFO is insignificant. Hence, we will assume perfect gateway guided synchronization for the remaining sections of the paper.

### C. Demodulation Performance Modeling

In this section, we derive analytical packet error rate performance expressions for the proposed modulation-coding scheme.

1) Uncoded ($C_r \geq 1$) Cases: We employ a non-coherent energy detector at the receiver (i.e., gateway) to demodulate the N-repetition M-PPM signal transmitted from the sensor node. A channel with complex additive white Gaussian noise $C\mathcal{N}(0, N_0)$ is assumed throughout the performance analysis. An $N \times 1$ vector $r$ denotes the set of matched filter outputs sampled at the correct $N$ pulse positions for a N-repetition M-PPM symbol. Similarly, let $e$ be an $N \times 1$ vector, the set of matched filter outputs sampled at incorrect symbol pulse positions. The pulse energy is normalized to one without loss of generality throughout the analysis. Therefore, the symbol is correctly detected when $||r||^2 > ||e||^2$, for all $e$’s that correspond to $M - 1$ possible error positions. Assuming all symbols are equally probable, the probability of correct symbol detection $P_c$ is given by (8) where $P[]$ denotes probability. Note that $X_e = \frac{||e||^2}{N_0}$ is a non-centralized chi-square distributed random variable with a degree of freedom $2N$ and non-centralized parameter $s = \frac{N_0}{N_0}$. $X_c = \frac{||r||^2}{N_0}$ is centralized chi-square distributed with a degree of freedom $2N$.

$$P_c(N, M, N_0) = (P(X_r > X_e))^{M-1}. \tag{8}$$

Since $X_e$ has an even degree of freedom, it has a closed-form expression cdf [34]. The analytical expression of $X_r$’s pdf is available in [30]. Therefore, $P_c$ can be rewritten as (9) where $SNR = \frac{1}{N_0}$, $s = 2N \cdot SNR$ and $I_{N-1}$ is the modified Bessel function of the first kind

$$P_c(N, M, SNR) = \int_0^\infty \left( 1 - e^{-\frac{1}{2}} \sum_{j=0}^{N-1} \frac{1}{j!} \left( \frac{s}{2} \right)^j \right)^{M-1} \times \frac{1}{2} \left( \frac{N_0}{N_0} \right)^{\frac{N_0}{2}} e^{-\frac{x^2}{2}} I_{N-1}(s\sqrt{x}) dx. \tag{9}$$
When convolutional encoding is unused ($C_r \geq 1$), the number of symbols in a packet containing $L$ information bits is $L/C_r$. Using (9), the PER is obtained by (10)

$$\text{PER} = 1 - P_c(N, M, SNR)^{L/C_r}. \quad (10)$$

2) Convolutional Coded ($C_r < 1$) Cases: When convolutional coding is enabled ($C_r < 1$), the non-coherent energy detection is performed along with the optimal maximum likelihood sequence estimation (MLSE) [34] at the gateway. Unlike a conventional soft-input Viterbi decoding where the log likelihood ratio is used as the branch metric [35], the matched filter output power sampled at the expected pulse position is directly used as the branch metric in our scheme. The likelihood of each symbol sequence is represented by the accumulated metric for each state.

To arrive at an analytical expression of the convolution coded packet error rate, we use the union bound (11), which provides a strict but tight upper bound of the actual PER [34]. The trellis length $L$ of MLSE is equal to number of information bits (convolutional code input) in our modulation-coding scheme when $C_r < 1$. Without loss of generality, we assume the correct MLSE sequence corresponds to the all zero input sequence for our PER analysis

$$\text{PER} \leq 1 - \left( \prod_{j=1}^{L} \prod_{k=1}^{l} (1 - P_{e,\text{pair}}(l, k, N))^{A(l,k)} \right)^L \leq 1 - \left( \prod_{j=1}^{L} \prod_{k=1}^{l} \left( 1 - P_c(Nk, l/C_r, N_0) \right)^{A(l,k)} \right)^L. \quad (11)$$

In (11), $A(l, k)$ is the number of ‘length-$l$ distance-$k$ simple error’ events that diverge from the all-zero sequence from the beginning of the trellis and merge (for the first time) to the all-zero sequence after $l$ branch transitions. A length-$l$, distance-$k$ simple error event has $k$ different pulse positions from the all-zero sequence over $l$ trellis transitions. $P_{e,\text{pair}}(l, k, N)$ is the probability of the pairwise error event, which occurs when the all-zero sequence has a less accumulated branch metric than a length-$l$ distance-$k$ simple error given $N$ repetition modulation and noise power of $N_0$. In fact, it is straightforward to show that $P_{e,\text{pair}}(l, k, N) = P_c(Nk, l/C_r, N_0)$ using (9) when $M = 2^{1/C_r}$ and $C_r \in \{ \frac{1}{2}, \frac{1}{3}, \ldots \}$. Note that $A(l, k)$ is dictated by the convolutional code generator function $G(\sigma_j, b_j, C_r)$ as well as the M-PPM encoding. We empirically evaluate $A(l, k)$ for all convolutional codes considered in this work.

Fig. 11 shows side-by-side comparisons between simulated PER and analysis results given by (9) and (11) for various uncoded ($C_r \geq 1$) and coded ($C_r < 1$) cases. The packet length $L$ is 128 bits for all cases. The average pulse SNR (on x-axis of Fig. 11) is $1/N_0$ assuming normalized pulse energy. The uncoded PER analysis exactly matches with the simulation (with triangle dot), while the union bound of the coded PER (11) is proven to be tight for all coded cases (simulation results with circle dot). Hence, for the remaining sections, we use the union bound (11) with equality to represent the PER when convolutional coding is enabled. As Fig. 11 shows, the proposed modulation-coding scheme enables the system operating at low SNRs ($\approx 0\text{dB}$ per pulse) when $C_r \leq 1/2$ convolutional coding is combined with $N \geq 1$ repetition.

D. Data Rate and Energy Efficiency Modeling

The data rate of the proposed system is defined by the number information bits transmitted per unit time. The proposed system supports a wide range of data rates by changing modulation-coding parameters; $N$, $C_r$, and $T_{\text{pulse}}$ dynamically. The number of information bits contained in a symbol is $[C_r]$. That is, a single symbol conveys a single information bit
when the convolutional coding is enabled ($C_r < 1$). Without error correction coding, on the other hand, $C_r (\geq 1)$ bits are transmitted per symbol using $M = 2^{C_r}$ PPM. Using (5) for the symbol duration, the system data rate $R$ is obtained by (12). Recall that $T_{\text{idle}} = \max (MT_{\text{pulse}}, T_{\text{charge}})$ and $T_{\text{charge}}$ is a function of $T_{\text{pulse}}$ given battery current limitation. $M$ is a function of $C_r$ as given in (2). Therefore, the data rate $R$ is fully determined by three modulation-coding parameters: $N$, $C_r$, and $T_{\text{pulse}}$

$$R = \left[ \frac{C_r}{T_{\text{sym}}} \right] = \left[ \frac{N C_r}{MT_{\text{pulse}} + (N - 1)T_{\text{idle}} + T_{\text{charge}}} \right].$$

(12)

Achieving the maximum energy efficiency is one of the primary objectives of the proposed link adaptation system. The energy per information bit for millimeter-scale sensor node transmission has the expression (13)

$$E_b = \frac{\text{Energy in reservoir cap}}{\text{Number of info bits per symbol}} = \frac{P_{\text{ckt}} N T_{\text{pulse}}}{[C_r]} = \frac{P_{\text{TX}} N T_{\text{pulse}}}{\eta_{\text{ckt}}[C_r]}.$$  

(13)

In (13), $P_{\text{ckt}}$ is the constant power consumption of the ‘power oscillator’ circuit proposed in section II.D. Recall that the efficiency of the circuit is maximized at a certain constant bias condition. We assume the efficiency $\eta_{\text{ckt}} = 0.15$ is achieved when $P_{\text{ckt}} = 3.5 mW$ using an architecture similar to [28]. While a constant transmit power level $P_{\text{TX}} = \eta_{\text{ckt}} P_{\text{ckt}}$ is delivered to the antenna maintaining the maximum circuit efficiency, we adjust $T_{\text{pulse}}$(i.e., signal bandwidth) and/or $C_r$ for link adaptation in various SNR conditions. Note that the transmitter consumes near zero power during the idle time ($T_{\text{charge}}$ and $T_{\text{idle}}$) between pulses when only the ULP oscillator [31] is active to control the N-repetition M-PPM pulse timing.

### IV. System Optimization and Link Adaptation

In this section, we will formulate link adaptation optimization problems for the proposed energy-aware ultra-small IoT communication system. We first introduce system constraints of the millimeter-scale sensor node, and then formulate formal link adaptation optimization problems for 1) the maximum link distance given a data rate target, and 2) the maximum energy efficiency.

The design parameters and system constants for a realistic millimeter-scale sensor node communication system are specified in Table IV. These constants are used throughout the system link adaptation study, unless specified otherwise. The proposed system employs three modulation-coding parameters that can be dynamically adjusted for link adaptation; $C_r$, $N$, and $T_{\text{pulse}}$. These are discrete variables as shown in Table IV. All other parameters such as $T_{\text{charge}}$, $T_{\text{idle}}$, $E_b$ are all implicitly specified by the selection of $C_r$, $N$, and $T_{\text{pulse}}$.

Since the pulse energy is drawn from the capacitor, the energy per pulse is limited by the energy stored in the reservoir capacitor $C_{\text{res}}$. The upper bound on the pulse width, $T_{\text{pulse}}$, is obtained by (14) where $V_{\text{min}}$ is the minimum voltage required for transmitter circuit functionality. The recharging time $T_{\text{charge}}$ (15) is required between pulses to restore charge in the reservoir capacitor

$$T_{\text{pulse}} \leq \frac{\text{Energy in reservoir cap}}{2P_{\text{ckt}}} = \frac{C_{\text{res}}(V_{\text{DD}}^2 - V_{\text{min}}^2)}{2P_{\text{ckt}}}$$

(14)

$$T_{\text{charge}}(C_{\text{res}}, T_{\text{pulse}}) = \frac{C_{\text{res}}(V_{\text{DD}} - \sqrt{V_{\text{DD}}^2 - \frac{2P_{\text{ckt}} T_{\text{pulse}}}{C_{\text{res}}}})}{I_{\text{bat}}}.$$  

(15)

In Section III.C, the packet error rate of the proposed system was analyzed as a function of the pulse SNR. Given a transmit pulse width $T_{\text{pulse}}$, the pulse SNR at the gateway can be obtained by (16), where $NF_{\text{gateway}}$ is the noise figure of the gateway receiver

$$\text{SNR}(T_{\text{pulse}}, d) = 10 \log_{10}(P_{\text{TX}} N T_{\text{pulse}}) - N_0 - L(d, f_c) - NF_{\text{gateway}} \text{ [in dB]}.$$  

(16)

The inverse function of the PER expression (10) and (11) is difficult to obtain. Therefore, to satisfy a certain PER performance requirement $PER_{\text{target}}$, we numerically evaluate PER expressions (10) (11) and identify the target PER $SNR_{\text{target}}(N, T_{\text{pulse}}, C_r)$ as a function of $N$, $T_{\text{pulse}}$ and $C_r$, given $PER_{\text{target}}$. This mapping can be obtained off-line and

### Table IV: System Design Parameters, Constants and Adaptive Modulation-Coding Parameters

<table>
<thead>
<tr>
<th>System Design Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antenna Dimension and Gain</td>
<td>$r_{\text{ant}} = 1.3 mm$, $g_{\text{sensor}} = -31.2 dB$, $g_{\text{gateway}} = 3 dB$</td>
</tr>
<tr>
<td>Carrier Frequency, $f_c$</td>
<td>1 GHz</td>
</tr>
<tr>
<td>Reservoir Capacitor, $C_{\text{res}}$</td>
<td>150 nF</td>
</tr>
<tr>
<td>Battery Current, $I_{\text{bat}}$</td>
<td>30 $\mu A$</td>
</tr>
<tr>
<td>Circuit Operating Condition</td>
<td>$V_{\text{min}} = 2.6 V$, $V_{\text{DD}} = 3.6 V$, $P_{\text{ckt}} = 3.5 mW$, $\eta_{\text{ckt}} = 0.15$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Adaptive Modulation-Coding Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coding Rate $C_r$</td>
<td>${\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, 1, 2, ..., 8}$</td>
</tr>
<tr>
<td>Pulse Repetition $N$</td>
<td>${1, 2, ..., 16}$</td>
</tr>
<tr>
<td>Pulse Width $T_{\text{pulse}}$ in $\mu s$</td>
<td>$(0.05, 0.1, 0.2, 0.4, ..., 51.2, 102.4, ...)$</td>
</tr>
</tbody>
</table>
stored in the gateway memory for real-time link adaptation. For all feasible link adaptation solutions, $SNR(T_{\text{pulse}}, d) \geq SNR_{\text{target}}(N, T_{\text{pulse}}, C_r)$ has to be satisfied.

The constraint for the target data rate, $R_{\text{target}}$ shall be given as (17). Note that $M$ and $T_{\text{idle}}$ are determined by $C_r$ and $T_{\text{pulse}}$, respectively

$$R(N, T_{\text{pulse}}, C_r) = \frac{[C_r]}{MT_{\text{pulse}} + (N - 1)T_{\text{idle}} + T_{\text{charge}}} \geq R_{\text{target}}. \quad (17)$$

In the proposed system, the Synch_HDR (in Fig. 6) that initiates the communication between the sensor node and the gateway is designed for the worst case distance. Once the gateway receives this Synch_HDR, it analyzes the correlation output and estimates the channel state information such as the link distance $d$ to evaluate the instantaneous SNR (16). The gateway then solves the link adaptation problem for a given objective function, and notifies the sensor node the optimum mode selection result.

A. Maximum Distance - Data Rate Tradeoff

The maximum link distance between the millimeter-scale sensor node and the gateway is obtained by solving the optimization problem (18) for dynamic modulation-coding parameters $N$, $T_{\text{pulse}}$ and $C_r$. Quantifying the solution of this link adaptation problem is essential to verify system feasibility for a target application scenario

Maximize: $d$

Subject to:

$$T_{\text{pulse}} \leq \frac{C_{\text{res}}(V_{DD}^2 - V_{min}^2)}{2P_{ckt}},$$

$$T_c(C_{\text{res}}, T_{\text{pulse}}) = \frac{C_{\text{res}}(V_{DD} - \sqrt{V_{DD}^2 - \frac{2P_{ckt}T_{\text{pulse}}}{C_{\text{res}}}})}{I_{\text{bat}}},$$

$$M T_{\text{pulse}} + (N - 1)T_{\text{idle}} + T_{\text{charge}} \geq R_{\text{target}},$$

$$SNR(T_{\text{pulse}}, d) \geq SNR_{\text{target}}(N, T_{\text{pulse}}, C_r). \quad (18)$$

Fig. 12 shows the optimization result for the maximum link distance when the target minimum data rate $R_{\text{target}}$ ranges from $10^2$ to $10^6$ bit/s. The target PER is set to $10^{-3}$ with packet length of 128 bits. In Fig. 12 on the left, $x$-axis is the variable target data rate and $y$-axis is the maximum link distance attainable by operating the system at the optimal $N$, $T_{\text{pulse}}$, and $C_r$. The result is shown for the NLOS where one layer of wall penetration loss is considered using (1) in computing SNR. The maximum achievable link distance of the system will improve when it operates in less-challenging outdoor/LOS conditions because the floor/wall penetration loss term $FL$ will be removed from (1) and consequently from the link adaptation problem formulation.

The dash black line on Fig. 12 on the left is the distance that can be supported by a static scheme using $N = 1$, $C_r = 1$ (binary PPM without coding), and $T_{\text{pulse}} = 1 \mu s$. The data rate of this static scheme is fixed to $29 k\text{bit/s}$ while it can operate up to $3.4m$ distance in the NLOS scenario. For a slightly higher (31kbit/s) data rate, the optimization result indicates 60% ($2m$) distance gain over this particular static scheme by operating with $N = 1$, $C_r = 1/3$, and $T_{\text{pulse}} = 0.8 \mu s$. At this data point ($R_{\text{target}} = 30 k\text{bit/s}$), the optimal $C_r = 1/3$ is lower than the static scheme coding rate ($C_r = 1$) but data rate degradation is avoided using a shorter pulse width (thus shorter $T_{\text{charge}}$), while the longer distance is achieved by the error correction coding. The optimal link adaptation result demonstrates graceful tradeoffs in link distance vs. data rate as Fig. 12 (left) shows. For the proposed millimeter-scale system, the optimal distance ranges from $1m$ to $> 30m$ when the target data rate is set to $10^6$ and $10^2$ respectively in the NLOS scenario.

Fig.12 (on the right) depicts the optimal mode ($C_r^*, N^*, T_{\text{pulse}}^*$) selection results for the maximum distance objective link adaptation. Note that we consider discretized $T_{\text{pulse}} \in \{0.05, 0.1, 0.2, \ldots \} \mu s$ to render more realistic hardware implementation. When the target data rate is low, longer pulse widths and smaller coding rates are preferred to maximize the distance. But it is worth noting non-monotonic behavior in selection of $C_r$ when the system is allowed to adjust $N$ and $T_{\text{pulse}}$ optimally. When the target data rate is high (> 500kbit/s), un-coded ($C_r \geq 1$) M-PPM modulation is selected as the optimal. To maximize the link distance, the optimal system often selects $N > 1$ in addition to convolutional coding, while the optimal $T_{\text{pulse}}$ monotonically decreases in general to meet a higher data rate target.
The link adaptation for the maximum energy efficiency has the form of (19), where the energy per information bit is minimized for a given link distance target \( d_{\text{target}} \) satisfying the minimum data rate constraint \( R_{\text{target}} \).

The optimization results for the maximum energy efficiency is shown in Fig. 13. The results correspond to the NLOS scenario. Again, the dash black line on Fig. 13 (on the left) corresponds to a static scheme that uses \( N = 1 \), \( C_r = 1 \), and \( T_{\text{pulse}} = 1 \mu s \). Unlike this static scheme, the proposed system can provide graceful tradeoffs in energy efficiency as a function of link distance when the optimal mode is selected by the proposed gateway guided link adaptation strategy. For a target data rate of 10 kbit/s, the proposed millimeter-scale system achieves the optimal energy efficiency in the range of 0.01 – 5.5 nJ/bit depending on the operating link distance.

From the maximum energy efficiency link adaptation result, Fig. 13, one can compute the ‘sustainable data rate’. To make the data rate sustainable, the energy consumption for communication has to be below the harvested energy level in average. The state-of-the-art solar energy harvester [17] reports \( P_{\text{harvest}} \approx 1 \mu W \) harvested power per \( mm^2 \) solar panel area. The ‘sustainable data rate’ is computed by \( P_{\text{harvest}}/E_b \) bit/s where \( E_b^* \) is the optimal energy efficiency per information bit obtained by solving (19). Any instantaneous data rate higher than the ‘sustainable data rate’ would deplete the energy from the battery and thus require duty-cycled operation. Results in Fig. 13 and \( P_{\text{harvest}} = 5 \mu W \) (5 mm\(^2\) solar panel area) indicate the sustainable data rate is about 7.1 kbit/s for a 4 m distance link in a NLOS scenario.

\[
\text{Minimize: } E_b = \frac{P_{TXN} T_{\text{pulse}}}{\eta_{ckt} \eta(C_r)}
\]

Subject to:
\[
T_{\text{pulse}} \leq \frac{C_{\text{res}}(V_{DD}^2 - V_{min}^2)}{2P_{ckt}}
\]
\[
T_c(C_{\text{res}}, T_{\text{pulse}})
\]
\[
C_{\text{res}} \left( V_{DD} - \sqrt{V_{DD}^2 - \frac{2P_{ckt} T_{\text{pulse}}}{C_{\text{res}}}} \right)
\]
\[
I_{\text{bat}} \geq \frac{I_{\text{bat}}}{C_r}
\]
\[
M T_{\text{pulse}} + \frac{(N-1)T_{\text{idle}} + T_{\text{charge}}}{SNR(T_{\text{pulse}}, d) \geq SNR_{\text{target}}(N, T_{\text{pulse}}, C_r)}.
\]

(19)

The optimization results for the maximum energy efficiency in the NLOS scenario. Notice non-monotonic behavior in the optimal \( C_r \) selection and the result that \( N > 1 \) is rarely selected for the maximum energy efficiency objective. It indicates, for the same total energy per symbol, continuous energy draw (i.e., a single long pulse) is more energy efficiency than short pulse repetition.

### C. Impact of System Parameters

In this section, we look into system parameters that need to be optimized at the system design time. Four system parameters are analyzed; the antenna size \( (r_{\text{ant}}) \), carrier frequency \( (f_c) \), battery current \( (I_{\text{bat}}) \), and reservoir capacitor size \( (C_{\text{res}}) \). Since these parameters are not adjustable for dynamic link adaptation, a special attention has to be paid to specify these parameters considering their impact on overall system performance. For this study, modulation-coding parameters \( (C_r, N, T_{\text{pulse}}) \) are dynamically adapted for the optimal performance. All other parameters are set to default values specified in Table IV, unless specified otherwise.

The antenna size is the most critical system parameter that dominates the overall system volume (see Fig. 2). Smaller antenna sizes are certainly attractive to keep the system volume minimized. However, the radiation efficiency rapidly drops (see Fig. 3) as the millimeter-scale antenna size decreases. The left plot in Fig. 14 shows that \( r_{\text{ant}} = 0.86, 1.3, \) and 1.7 mm antennas can provide the maximum distance of 21.4, 31.3, and 41 m respectively for the target data rate of 100 bit/s in a NLOS scenario with link adaptation when all other parameters are fixed as in Table IV. The right plot in Fig. 14 also confirms that increasing the antenna size is an obvious way to significantly improve the energy efficiency.

The carrier frequency \( (f_c) \) affects the system performance via the antenna efficiency as well as wall penetration/pathloss characteristic. The impact of the carrier frequency in LOS and NLOS settings highly depends on how the wall penetration/pathloss is modeled. Fig. 15 shows the maximum distance link adaptation results obtained from different carrier frequency settings in both LOS and NLOS settings using our model (1). It implies that \( \geq 5 \text{GHz} \) operation is preferred if the application mostly targets LOS scenarios. For NLOS indoor operation that follows our propagation loss model,
$1\text{GHz}$ operation outperforms higher carrier frequency options, although the millimeter-scale antenna efficiency at $1\text{GHz}$ is very poor (less than 1%, see Fig. 3).

In the proposed system, the thin-film battery continually charges the reservoir capacitor. The battery current determines the recharging time $T_{\text{charge}}$ via (15). A larger battery typically has a lower internal resistance, thus allows a higher battery current. Since $T_{\text{charge}}$ is reduced with a higher battery current, higher energy per pulse can be drawn from the capacitor to increase the link distance while maintaining the same data rate. For a given link distance target, the link adaptation strategy can utilize the additional battery current to reduce the energy per data bit. The impact of various battery current levels is shown in Fig. 16.

In summary, for the maximum utilization of the millimeter-scale system dimension, the system designer must quantify the impact of each design parameter and purposefully determine the size of the antenna, battery, and energy reservoir capacitor along with the transceiver/sensor/processor integrated circuits to be integrated in an ultra-small form-factor (Fig. 2). When the overall system dimension is fixed, increasing the size of one component would inevitably limit the size of the other. The link adaptation optimization framework and its results
shown in this section provide a guideline for the designer to foresee the impact of critical design parameters in realizing a highly energy-optimized wireless communication system with a millimeter-scale form-factor constraint.

V. CONCLUSION

In this work, an energy-autonomous, self-contained wireless communications system that optimally utilizes the scarce energy/power resource in an ultra-small millimeter-scale sensor node is presented. A cross-layer system-level optimization framework is proposed to jointly optimize various system parameters including modulation, coding scheme, synchronization protocol, RF/analog/digital circuit specifications, data rate, carrier frequency, antenna efficiency, etc. Based on the comprehensive system model, the dynamic link adaptation problems are formulated to maximize transmission distance, throughput, and energy efficiency for various operating scenarios. The simulation results of the link adaptation protocol show the significant benefit of dynamically adapting to optimum system parameters. The impact of pre-silicon system parameters including antenna dimension, carrier frequency, battery current, and reservoir capacitor size is presented to guide system designers toward the energy-optimized communication system for ultra-small IoT sensor nodes.

REFERENCES

Yajing Chen is currently pursuing the Ph.D. degree in computer science engineering with the University of Michigan, Ann Arbor, MI, USA. Her research interests include wireless communications systems, software-defined radio, digital signal processing, and low power architecture design.

Nikolaos Chiotellis received the B.Sc. degree in electrical engineering from the National Technical University of Athens, Greece, in 2012, and the M.Sc. degree in applied electromagnetics and RF circuits from the University of Michigan, Ann Arbor, MI, USA, in 2016. Since 2013, he has been pursuing the Ph.D. degree. He has been the author/co-author of three journal papers and three conference papers. He was a finalist in the 2015 IEEE AP-S Student Paper Competition. His research interests include electromagnetics, metamaterials, metasurfaces, non-diffracting waves, and electrically small antennas for RF circuits.

Li-Xuan Chuo was born in Taipei, Taiwan, in 1991. He received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 2013. He is currently pursuing the Ph.D. degree with the University of Michigan, Ann Arbor, MI, USA. His research interests include low power RF front-end design and indoor localization.

Carl Pfeiffer (S’08–M’15) received the B.S.E., M.S.E., and Ph.D. degrees in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 2009, 2011, and 2015, respectively. In 2015, he became a Post-Doctoral Research Fellow with the University of Michigan. In 2016, he joined Defense Engineering Corporation as an onsite contractor for the Air Force Research Laboratory, Wright-Patterson Air Force Base, OH, USA. His research interests include engineered electromagnetic structures, such as metamaterials, metasurfaces, and frequency selective surfaces, antennas, microwave circuits, plasmonics, optics, and analytical electromagnetics/optics.

Ronald G. Dreslinski received the Ph.D. degree in computer science and engineering from the University of Michigan. He is currently an Assistant Professor with the Computer Science and Engineering Department, University of Michigan, Ann Arbor, MI, USA. His research interests include near-threshold computing, architectural simulator development, and high-radix on-chip interconnects. He was the winner of the ISSCC 2011 Student Design Contest and a recipient of the Young Computer Architect Award from the IEEE Computer Society’s Technical Committee on Computer Architecture.

Anthony Grbic (S’00–M’06–SM’14–F’16) received the B.A.Sc., M.A.Sc., and Ph.D. degrees in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 1998, 2000, and 2005, respectively. In January 2006, he joined the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA, where he is currently a Professor. His research interests include engineered electromagnetic structures (meta-materials, metasurfaces, electromagnetic band-gap materials, frequency-selective surfaces), plasmonics, antennas, analytical electromagnetics/optics, microwave circuits, and wireless power transmission systems. Dr. Grbic served as Technical Program Co-Chair in 2012 and Topic Co-Chair in 2016 for the IEEE International Symposium on Antennas and Propagation and USNC-URSI National Radio Science Meeting (AP-S/USNC-URSI). He was an Associate Editor for IEEE Antennas and Wireless Propagation Letters from 2010 to 2015. He is currently the Vice Chair of AP-S Technical Activities, Trident Chapter, IEEE Southeastern Michigan Section. Dr. Grbic was the recipient of AFOSR Young Investigator Award as well as NSF Faculty Early Career Development Award in 2008, the Presidential Early Career Award for Scientists and Engineers in January 2010, an Outstanding Young Engineer Award from the IEEE Microwave Theory and Techniques Society, a Henry Russel Award from the University of Michigan, and a Booker Fellowship from the United States National Committee of the International Union of Radio Science in 2011, the inaugural recipient of the Ernest and Bettine Kuh Distinguished Faculty Scholar Award in the Department of Electrical and Computer Science, University of Michigan in 2012. Prof. Grbic is an IEEE Fellow.

Trevor Mudge (S’74–M’77–SM’84–F’95–LF’13) received the Ph.D. degree in computer science from the University of Illinois at Urbana-Champaign, Urbana, IL, USA. He is currently a faculty member with the University of Michigan, Ann Arbor, MI, USA, where he is the Bredt Family Professor of Computer Science and Engineering. He has authored numerous papers on computer architecture, programming languages, VLSI design, and computer vision. He chaired 51 theses in these areas. In 2014, he received the ACM/IEEE CS Eckert-Mauchly Award and the University of Illinois Distinguished Alumni Award. He is a life fellow of the IEEE and a member of the ACM, the IET, and the British Computer Society.
David D. Wentzloff (S’02–M’07) received the B.S.E. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 1999, and the S.M. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2002 and 2007, respectively. Since 2007, he has been with the University of Michigan, where he is currently an Associate Professor of electrical engineering and computer science. His research interests include RF integrated circuits, with an emphasis on ultralow power design. In 2012, he co-founded PsiKick, a fabless semiconductor company developing ultralow power wireless SoCs. He has served on the Technical Program Committee for ICUWB from 2008 to 2010, ISLPED from 2011 to 2015, S3S from 2013 to 2015, and RFIC from 2013 to 2015, and as a Guest Editor for the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES, the IEEE Communications Magazine, and the Signal Processing: Image Communication. He is a member of the IEEE Circuits and Systems Society, the IEEE Microwave Theory and Techniques Society, the IEEE Solid-State Circuits Society, and Tau Beta Pi. He was a recipient of the 2009 DARPA Young Faculty Award, theEta Kappa Nu Professor of the Year Award in 2009 and 2010, the 2011 DAC/ISSCC Student Design Contest Award, the 2012 IEEE Subthreshold Microelectronics Conference Best Paper Award, the 2012 NSF CAREER Award, the 2014 ISSCC Outstanding Forum Presenter Award, theEta Kappa Nu ECE Professor of the Year Award in 2014 and 2015, the 2012 NSF CAREER Award, the 2014 ISSCC Outstanding Forum Presenter Award, theEta Kappa Nu ECE Professor of the Year Award in 2014 and 2015, and the 2015 Joel and Ruth Spira Excellence in Teaching Award.

David Blaauw (M’94–SM’07–F’12) received the B.S. degree in physics and computer science from Duke University in 1986, and the Ph.D. degree in computer science from the University of Illinois at Urbana–Champaign, Urbana, in 1991. He joined Motorola, Inc., Austin, TX, USA, where he was the Manager of the High Performance Design Technology Group. Since 2001, he has been on the faculty with the University of Michigan, where he is currently a Professor. He has authored over 500 papers and holds 50 patents. His research interests include VLSI design with particular emphasis on ultralow power and high performance design for ultralow power sensor nodes. He was the Technical Program Chair and the General Chair for the International Symposium on Low Power Electronic and Design. He was also the Technical Program Co-Chair of the ACM/IEEE Design Automation Conference and a member of the ISSCC Technical Program Committee.

Hun Seok Kim (S’10–M’11) received the B.S. degree from the Seoul National University, South Korea, and the M.S. and Ph.D. degrees from the University of California, Los Angeles (UCLA), all in electrical engineering. He currently holds nine granted patents and has over ten pending applications in the areas of digital communication, signal processing, and low-power integrated circuits. He is currently an Assistant Professor with the University of Michigan, Ann Arbor, MI, USA. His research interests include system novel algorithms and efficient VLSI architectures for low-power/high-performance signal processing, wireless communication, computer vision, and machine learning systems. Before joining the University of Michigan, he was a technical staff member with Texas Instruments Inc., from 2010 to 2014, while serving as an industry liaison for multiple university projects funded by the Semiconductor Research Corporation and Texas Instruments Inc. He was a recipient of multiple fellowships from the Ministry of Information and Telecommunication, South Korea, Seoul National University, and UCLA.

Hun Seok Kim (S’10–M’11) received the B.S. degree from the Seoul National University, South Korea, and the M.S. and Ph.D. degrees from the University of California, Los Angeles (UCLA), all in electrical engineering. He currently holds nine granted patents and has over ten pending applications in the areas of digital communication, signal processing, and low-power integrated circuits. He is currently an Assistant Professor with the University of Michigan, Ann Arbor, MI, USA. His research interests include system novel algorithms and efficient VLSI architectures for low-power/high-performance signal processing, wireless communication, computer vision, and machine learning systems. Before joining the University of Michigan, he was a technical staff member with Texas Instruments Inc., from 2010 to 2014, while serving as an industry liaison for multiple university projects funded by the Semiconductor Research Corporation and Texas Instruments Inc. He was a recipient of multiple fellowships from the Ministry of Information and Telecommunication, South Korea, Seoul National University, and UCLA.

Hun Seok Kim (S’10–M’11) received the B.S. degree from the Seoul National University, South Korea, and the M.S. and Ph.D. degrees from the University of California, Los Angeles (UCLA), all in electrical engineering. He currently holds nine granted patents and has over ten pending applications in the areas of digital communication, signal processing, and low-power integrated circuits. He is currently an Assistant Professor with the University of Michigan, Ann Arbor, MI, USA. His research interests include system novel algorithms and efficient VLSI architectures for low-power/high-performance signal processing, wireless communication, computer vision, and machine learning systems. Before joining the University of Michigan, he was a technical staff member with Texas Instruments Inc., from 2010 to 2014, while serving as an industry liaison for multiple university projects funded by the Semiconductor Research Corporation and Texas Instruments Inc. He was a recipient of multiple fellowships from the Ministry of Information and Telecommunication, South Korea, Seoul National University, and UCLA.