# Wireless Power Transfer for Future Networks: Signal Processing, Machine Learning, Computing, and Sensing

Bruno Clerckx, *Senior Member, IEEE*, Kaibin Huang, *Fellow, IEEE*, Lav R. Varshney, *Senior Member, IEEE*, Sennur Ulukus, *Fellow, IEEE*, and Mohamed-Slim Alouini, *Fellow, IEEE*

*(Overview Paper)*

***Abstract*—Wireless power transfer (WPT) is an emerging paradigm that will enable using wireless to its full potential in future networks, not only to convey information but also to deliver energy. Such networks will enable trillions of future low-power devices to sense, compute, connect, and energize anywhere, anytime, and on the move. The design of such future networks brings new challenges and opportunities for signal processing, machine learning, sensing, and computing. The objective is to make the best use of the RF radiations, spectrum, and network infrastructure to provide cost-effective and real-time power supplies to wireless devices and enable wireless-powered applications. In this paper, we first review recent signal processing techniques to make WPT and wireless information and power transfer (WIPT) as efficient as possible. Topics include high-power amplifier and energy harvester nonlinearities, active and passive beamforming, intelligent reflecting surfaces, receive combining with multi-antenna harvester, modulation, coding, waveform, large-scale (massive) multiple-input multiple-output (MIMO), channel acquisition, transmit diversity, multi-user power region characterization, coordinated multipoint, and distributed antenna systems. Then, we overview two different design methodologies: the *model and optimize* approach relying on analytical system models, modern convex optimization, and communication/information theory, and the *learning* approach based on data-driven end-to-end learning and physics-based learning. We discuss the pros and cons of each approach, especially when accounting for various nonlinearities in wireless-powered networks, and identify interesting emerging opportunities for the approaches to complement each other. Finally, we identify new emerging wireless technologies where WPT may play a key role—wireless-powered mobile edge computing, wireless-powered sensing, and wireless-powered federated learning—arguing WPT, communication, computation, sensing, and learning must be jointly designed.**

***Index Terms*—Wireless power transfer, wireless powered networks, wireless information and power transfer, wireless powered communications, wireless energy harvesting communications, signal processing, beamforming, intelligent reflecting surface, waveform, modulation, multi-antenna, optimization, information theory, machine learning, data-driven, end-to-end learning, physics-based learning, sensing, edge computing, federated learning.**

B. Clerckx is with the Electrical and Electronic Engineering Department at Imperial College London, London SW7 2AZ, UK (email: b.clerckx@imperial.ac.uk).

K. Huang is with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong (e-mail: huangkb@eee.hku.hk).

L. R. Varshney is with the Coordinated Science Laboratory and the Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign, Urbana, IL 61801 USA (e-mail: varshney@illinois.edu).

S. Ulukus is with the Department of Electrical and Computer Engineering, University of Maryland at College Park, College Park, MD 20742 USA (e-mail: ulukus@umd.edu).

M.-S. Alouini is with the Computer, Electrical, and Mathematical Science and Engineering (CEMSE) Division, King Abdullah University of Science, and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia (e-mail: slim.alouini@kaust.edu.sa).

This work has been partially supported by the EPSRC of UK under grant EP/P003885/1 and EP/R511547/1.

## I. Introduction

TWENTY years from now, according to Koomey's law [1], devices will require 10000 times less energy to compute a given task, due to the reduction in power requirements of their electronics. Moreover, trillions of Internet-of-Things (IoT) devices will emerge. This explosion of low-power devices demands a re-thinking of future network design, where wireless will be used to its full potential, not only to convey information but also to deliver energy. Wireless power will bring new opportunities, namely proactive and controllable energy supply with genuine mobility—no wires, no contact, no or reduced batteries—and therefore small, light, and compact devices. This will not only yield environmental benefits by eliminating the need to produce, maintain, or dispose of trillions of batteries, but also enable a myriad of new wireless applications such as autonomous low-power sensing and computing. This is all thanks to the prolonged lifetime and the long-term, predictable, and reliable energy supply offered by wireless power unlike ambient energy-harvesting technologies such as solar, thermal, or vibration.

Wireless power and wireless communications have, however, evolved as two separate fields in academia and industry [2]. This separation has consequences: first, current wireless networks broadcast radio-frequency (RF) energy into air (for communication purposes) but do not use it for charge devices; second, providing ubiquitous wireless power would require deploying a separate network of dedicated energy transmitters. Imagine instead a future network where information and energy flow together through the wireless medium. Wireless Information Transfer (WIT) and Wireless Power Transfer (WPT) would refer to two extreme strategies respectively targeting communication-only and power-only. A unified Wireless Information and Power Transfer (WIPT) design would be able to softly evolve between the two extremes to best use the RF spectrum/radiation and network infrastructure to communicate and energize, thereby outperforming traditional systems that separate communications and power.

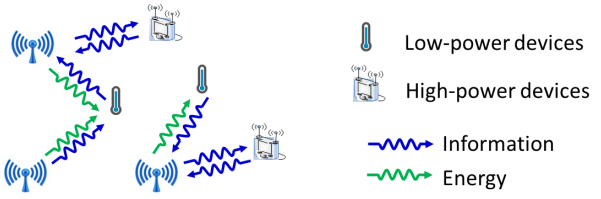



Fig. 1. Future of wireless powered networks

Such a network will enable the creation of highly efficient wireless power resources, such that low-power devices (e.g., sensors) with or without a communication capability can be wirelessly powered anywhere, anytime and on the move. Consequently low-power devices with communication capabilities can experience a true ubiquitous wireless connectivity. It will also enable low-power and high-power devices to co-exist in such a way that transmit signals simultaneously charge remote low-power devices and carry information to high-power devices (e.g. smartphones, tablets), as illustrated in Fig. 1. Wireless power will also enable emerging wireless applications, such as wireless-powered edge intelligence, wireless-powered computing, wireless-powered sensing, and wireless-powered autonomous systems.

The design of efficient wireless power resources, the integration of wireless power and communications, sensing and computing, and the positioning of wireless power as a key enabler of new wireless applications brings new challenges, ideas and opportunities, and calls for a paradigm shift in wireless system and network design. Numerous research problems must be addressed that cover a wide range of disciplines, including circuit and systems, sensors, antenna and propagation, microwave theory and techniques, communication, signal processing, machine learning, sensing, computing, and information theory.

### *A. Wireless Power for Future Networks: Overview of Challenges and Technologies*

Wireless power, especially in its most promising form of WPT, will be a fundamental building block of future wireless networks. WPT research over the past decades has largely focused on RF theories and techniques regarding the energy receptor with the design of efficient RF solutions, circuits, antennas, rectifiers and power management units [3]–[6]. Nevertheless, more recently, a new complementary line of research on communications and signal design for WPT has attracted significant attention in the communication and signal processing literature [7]. Additionally, there has been growing interest in bridging RF, signal, and system designs to bring these two communities closer together and better understand the fundamentals of an effective wireless powered network architecture [8]. This has resulted in a new understanding of signal and system design for WPT and WIPT [9].

There are numerous design challenges of the envisioned future network : 1) *Range*: Deliver wireless power at distances of 5-100s meters (m) for energizing low-power devices in indoor/outdoor settings; 2) *Efficiency*: Boost the end-to-end power transfer efficiency (up to a fraction of a percent/a few percent) or equivalently the DC power level at the energy harvester for a given transmit power; 3) *Non-line of sight (NLoS)*: Support Line of sight (LoS) and NLoS to widen real-world applications of future WIPT networks; 4) *Mobility support*: Support mobile devices at least for those at pedestrian speed; 5) *Ubiquitous accessibility*: Provide power ubiquitously within the network coverage area; 6) *Safety and health*: Make RF system safe and comply with the regulations; 7) *Energy consumption*: Limit the energy consumption of wireless powered devices; 8) *Seamless integration of wireless communication and wireless power*: Unify wireless communication and wireless power into WIPT; 9) *Integrated WPT, sensing, computing, and communication*: Integrate WPT with sensing/computing and communication in 5G-and-beyond systems with virtualization and network slicing.

Challenges (1)–(7) are being studied in various communities [6]–[8], [10], [11]. Solutions cover a wide range of areas spanning sensors, devices, RF, communication, signal and system designs for WPT. Typical WPT scenarios under study are illustrated in Fig. 2 and include:

- *Single-user (point-to-point) WPT*: The focus here is on a single energy transmitter (ET) and a single energy receiver (ER). Both ET and ER may be equipped with multiple co-located antennas. This scenario is the fundamental building block of future wireless networks since most of the challenges (1)–(7) must be tackled for this setup before considering multi-user scenarios.
- *Multi-user WPT*: The focus here is on transmit antennas being either co-located or distributed and delivering energy to multiple ERs equipped with one or multiple antennas.

Challenge (8) has recently been reviewed in [9] in an attempt to lay the fundamentals of WIPT from energy harvester modeling to signal and system designs. In contrast to WPT and WIT, where the emphasis of the system design is to exclusively deliver energy and information, respectively, in WIPT, both energy and information are to be delivered. The challenge is therefore to understand how to make the best use of the RF radiation and the RF spectrum to provide both information and energy, and requires the characterization of the fundamental trade-off between the amount of information and the amount of energy that can be delivered in a wireless network and how signals should be designed to achieve this trade-off.

As illustrated in Fig. 3, WIPT can be categorized into three different types.

- *Simultaneous Wireless Information and Power Transfer (SWIPT)*: Energy and information are simultaneously transmitted from one or multiple transmitter(s) to one or multiple receiver(s) [12]–[32]. The information receiver(s) (IR) and ER can be co-located or separated. With co-located receivers, each receiver is a single (typically low-power) device that is simultaneously being charged and receiving data. With separate receivers, ER and IR are different devices, the former being a low-power device being energized, the latter being a device receiving data.
- *Wirelessly Powered Communication Networks (WPCNs)*:

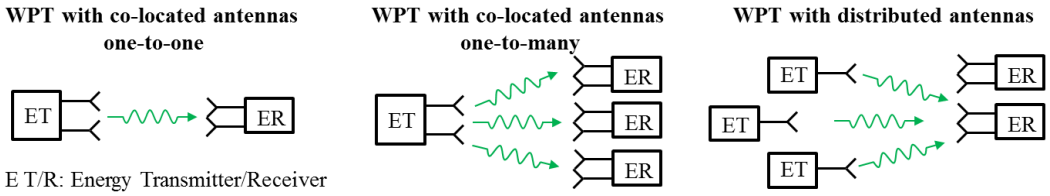



Fig. 2. WPT deployment scenarios.

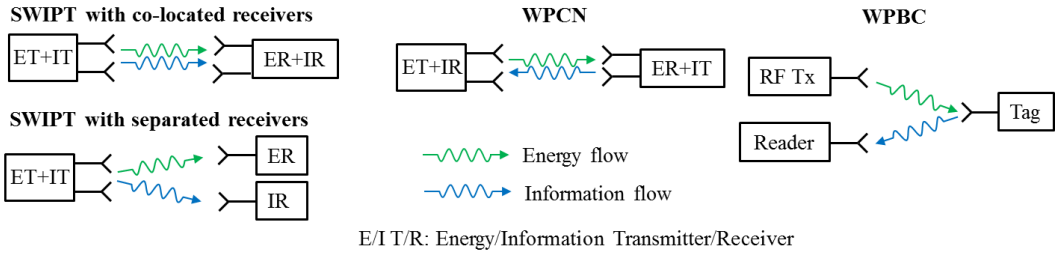



Fig. 3. Different WIPT scenarios and architectures.

Energy is transmitted in the downlink from an access point to a receiver and information is transmitted in the uplink [33]–[35]. The receiver is a device that harvests energy in the downlink and uses the harvested energy to transmit data in the uplink.

- *Wirelessly Powered Backscatter Communication (WPBC)*: Energy is transmitted in the downlink and information is transmitted in the uplink using backscatter modulation at a tag to reflect and modulate the incoming RF signal for communication with a reader [36]–[38]. Backscatter communications benefit from several orders-of-magnitude lower power consumption than conventional wireless communications because tags do not require oscillators to generate carrier signals [39].

Moreover, a network could have a mixture of all of these types of transmissions with multiple co-located and/or distributed ETs and information transmitter(s) (IT).

Challenge (9) is new and arises since next-generation Internet-of-Things (IoT) that build on the 5G/6G platform are seeing an increasing level of integration between storage, computing, and communication so as to efficiently enable a wide range of new applications in distributed sensing, edge computing and artificial intelligence (AI). Thus, wirelessly powering next-generation IoT calls for the joint control of WPT, sensing, computing, and communication so as to optimize efficiency of a system supporting specific applications. In particular, there exist trade-offs between transferred energy and energy consumption of sensing/computing (e.g., on-device AI model training) and communication (e.g., mobile computation offloading). Quantifying and exploiting such trade-offs can substantially improve system performance.

### *B. Objectives and Organization*

Various review papers have appeared in past years on WPT, emphasizing separately RF, circuit and antenna solutions [4]–[6], [10], [11], and communications, signal and system design solutions [7]. More recently attempts have been made to bridge RF, signal and system designs to get a better understanding of the fundamental building blocks of an efficient WPT network architecture [8]. This synthesis of work in different areas of WPT has yielded critical observations and given a fresh new look to promising avenues for WPT signal and system design. As an example, [8] shows that the *nonlinear* nature of the WPT design problem, both for the ET and the ER, must be accounted for at the signal and the circuit-level design.

Similarly, review papers on WIPT have also appeared [40]–[50]. Emphasis was put at that time on characterizing the fundamental tradeoff between conveying information and energy, so-called rate-energy (R-E) tradeoff, under the assumption of a very simple *linear* model of the ET and ER. In recent years, the validity of this linear model has been questioned and there has been an increasing departure from simple linear assumptions in the WIPT literature. It turns out that the linear model is inaccurate and leads to inefficient WIPT designs, and that WIPT design radically changes once we adopt more realistic *nonlinear* models of the energy harvester (EH) [9]. Recently, [9] showed how crucial the EH model is to WIPT signal and system designs and how WIPT signal and system designs revolve around the underlying EH model. It highlighted different linear and nonlinear EH models, and showed in a systematic way how WIPT designs and R-E tradeoff differ for each of them. In particular, the paper showed how the modeling of the EH can have tremendous influence on the design of the physical and higher layers of WIPT networks.

This paper overviews recent advances and emerging opportunities for signal processing, machine learning, computing, and sensing in the broad area of future wireless powered networks (including WPT, WIPT, and other emerging wireless powered applications). The objectives are threefold.

*First*, this paper aims to provide a review of recent signal processing techniques to tackle the challenges of WPT and WIPT and make them a reality. Topics discussed include

high power amplifier (HPA) and EH nonlinearities, transmit active and passive beamforming and intelligent reflecting surfaces (also called reconfigurable intelligent surfaces), receive combining with multi-antenna harvester, modulation, coding, waveform, joint beamforming, combining and waveform, large-scale (massive) multiple-input multiple-output (MIMO), channel acquisition, transmit diversity, power region characterization in multi-user WPT, coordinated multipoint and distributed antenna systems for wireless powered networks. A particular emphasis is on how the design of those techniques is deeply rooted in the EH nonlinearity and contrasts with a previous tutorial [7] where nonlinearity was highlighted only as part of the waveform design.

*Second*, this paper aims to provide an overview of various design methodologies. All past tutorial and review papers rely exclusively on the traditional model-and-optimize approach where they derive an analytical system model (under some assumptions) and use modern convex optimization tools to optimize it. Instead here we also discuss the role machine learning, in the form of model-based and data-based end-to-end learning and physics-based learning, can play to design future wireless-powered networks. This is particularly relevant due to the importance of accounting for various sources of nonlinearity in wireless power. We identify the pros and cons of the *model and optimize* approach and the *learning* approach and identify interesting emerging opportunities for machine learning to complement human expertise.

*Third*, this paper aims to identify emerging wireless technologies where WPT will play a key role. In particular, we discuss and study how WPT can enable wireless-powered computing, wireless-powered sensing, and wireless-powered edge/federated learning.

*Organization:* In Section II, we introduce the system model of WPT, discuss the HPA and EH nonlinearity and EH architecture, and review various signal processing techniques used to increase the end-to-end power transfer efficiency of single-user and multi-user WPT. Section III builds upon previous section and introduces the system model of WIPT before reviewing various signal processing techniques to achieve the best R-E tradeoff of WIPT. Section IV discusses and contrasts the pros and cons of two major design methodologies to design WPT and WIPT, namely the model-and-optimize approach and the learning approach. Section V discusses how wireless power will enable new and emerging scenarios and applications in future wireless powered networks such as wireless-powered computing, sensing, and federated learning. Section VI concludes the paper and discusses future works.

*Notation:* In this paper, scalars are denoted by italic letters. Boldface lower- and upper-case letters denote vectors and matrices, respectively. $\mathbb{C}^{M\times N}$ denotes the space of $M \times N$ complex matrices. $j$ denotes the imaginary unit, i.e., $j^2 = -1$. $\mathbb{E}[\cdot]$ denotes statistical expectation and $\Re\{\cdot\}$ represents the real part of a complex number. $\mathbf{I}_M$ denotes an $M \times M$ identity matrix and $\mathbf{0}$ denotes an all-zero vector/matrix. $|.|$ and $\|.\|$ refer to the absolute value of a scalar and the 2-norm of a vector. For an arbitrary-size matrix $\mathbf{A}$, its complex conjugate, transpose, Hermitian transpose, and Frobenius norm are respectively denoted as $\mathbf{A}^*$, $\mathbf{A}^T$, $\mathbf{A}^H$, and $\|\mathbf{A}\|_F$. $[\mathbf{A}]_{im}$ denotes the $(i,m)$th element of matrix $\mathbf{A}$. For a square Hermitian matrix $\mathbf{S}$, $\mathrm{Tr}(\mathbf{S})$ denotes its trace, while $\lambda_{\max}(\mathbf{S})$ and $\mathbf{v}_{\max}(\mathbf{S})$ denote its largest eigenvalue and the corresponding eigenvector, respectively. In the context of random variables, i.i.d. stands for independent and identically distributed. The distribution of a Circularly Symmetric Complex Gaussian (CSCG) random variable with zero-mean and variance $\sigma^2$ is denoted by $\mathcal{CN}(0,\sigma^2)$; hence with the real/imaginary part distributed as $\mathcal{N}(0,\sigma^2/2)$. $\sim$ stands for "distributed as". We use the notation $\mathrm{sinc}\,(t) = \frac{\sin(\pi t)}{\pi t}$. $\mathrm{diag}(\mathbf{A}_1,\ldots,\mathbf{A}_N)$ refers to a block diagonal matrix with blocks being $\mathbf{A}_1$, …, $\mathbf{A}_N$.




Fig. 4. The block diagram of a generic WPT system [7].

## II. Wireless Power Transfer: Key Technologies to Increase Efficiency

In the past decade, there has been a significant interest in WPT and ambient wireless energy harvesting (WEH) for low-power (e.g., from $\mu$W to a few W) delivery over distances of a few m to hundreds of m [51], [52]. This originates from the increasing need to build reliable and convenient wireless power systems for remotely energizing low-power devices, such as sensors, RFID tags, and consumer electronics [8], [53], [54].

Fig. 4 shows a generic WPT system that consists of an RF ET and an ER. A DC power source is used to generate a signal that is upconverted to the RF domain at the ET, then transmitted over the air, and collected at an ER in the RF domain before being converted to DC. The ER is made of an antenna combined with a rectifier (rectenna) and a power management unit (PMU). Since the majority of the electronics requires a DC power source, a rectifier is required to convert RF to DC. The recovered DC power then either supplies a low power device directly, or is stored in a battery or a super capacitor for high power low duty-cycle operations. The recovered DC power can also be managed by a DC-to-DC converter before being stored. In WPT, the entire link, including ET and ER, of Fig. 4 can be fully optimized. Therefore, in contrast to ambient WEH, WPT offers full control of the design and room to enhance the end-to-end power transfer efficiency $e$

$$e = \frac{P^r_{\rm dc}}{P^t_{\rm dc}} = \underbrace{\frac{P^t_{\rm rf}}{P^t_{\rm dc}}}_{e_1}\underbrace{\frac{P^r_{\rm rf}}{P^t_{\rm rf}}}_{e_2}\underbrace{\frac{P^r_{\rm dc}}{P^r_{\rm rf}}}_{e_3}, \tag{1}$$

where $e_1$, $e_2$, and $e_3$ denote the DC-to-RF, RF-to-RF, and RF-to-DC power conversion/transmission efficiency, respectively.

### A. Signal and System Model

We consider a single-user point-to-point MIMO WPT system in a general multipath environment. This setup is referred to as "WPT with co-located antennas one-to-one" in Fig. 2. The ET is equipped with $M$ antennas that transmit power to a

ER equipped with $Q$ receive antennas. We consider the general setup of a multi-subband transmission (with a single subband being a special case) employing $N$ orthogonal subbands where the $n^{\text{th}}$ subband has carrier frequency $f_n$ and all subbands employ equal bandwidth $f_{\text{w}}$, $n = 0, ..., N-1$. The carrier frequencies (also called tones) are evenly spaced such that $f_n = f_0 + n\Delta_f$ with the inter-carrier frequency spacing $\Delta_f$ (with $f_{\text{w}} \leq \Delta_f$).

The WPT signal transmitted on antenna $m$, $x_{\text{rf},m}(t)$, is a multi-carrier modulated waveform with frequencies $f_n$, $n = 0, ..., N-1$, carrying independent symbols on subband $n = 0, ..., N-1$. The input WPT signal at time $t$ to the HPA of antenna $m = 1, ..., M$ is given by

$$x_{\text{in},m}(t) = \sqrt{2}\Re\left\{\sum_{n=0}^{N-1} x_{m,n}(t)e^{j2\pi f_n t}\right\} \tag{2}$$

with the baseband equivalent signal $x_{m,n}(t)$ given by

$$x_{m,n}(t) = \sum_{k=-\infty}^{\infty} x_{m,n,k}\,\text{sinc}(f_{\text{w}}t - k) \tag{3}$$

where $x_{m,n,k}$ denotes the complex-valued power carrying symbol at time index $k$, modeled as a random variable generated in an i.i.d. fashion. $x_{m,n}(t)$ has bandwidth $[-f_{\text{w}}/2, f_{\text{w}}/2]$. For the special case of unmodulated WPT, $x_{m,n}(t)$ is constant across $t$, i.e., $x_{m,n}(t) = x_{m,n} = s_{m,n}e^{j\varphi_{m,n}}$, $\forall t$. In this case, $x_{\text{in},m}(t)$ is a summation of $N$ sinewaves inter-separated by $\Delta_f$ Hz, and hence essentially occupies zero bandwidth.

The power at the transmitter before HPA is written as

$$P_{\text{dc}}^t = \sum_{n=0}^{N-1} \text{Tr}(\mathbf{Q}_n) = \text{Tr}(\mathbf{Q}), \tag{4}$$

with $\mathbf{Q} = \text{diag}\{\mathbf{Q}_0, ..., \mathbf{Q}_{N-1}\}$ where the positive semidefinite input covariance matrix $\mathbf{Q}_n$ at subband $n$ is defined as $\mathbf{Q}_n \triangleq \mathbb{E}\left[\mathbf{x}_n(t)\mathbf{x}_n^H(t)\right] \in \mathbb{C}^{M\times M}$ and $\mathbf{x}_n(t) \triangleq \left[x_{1,n}(t), ..., x_{M,n}(t)\right]^T$ denotes the signal vector across the $M$ antennas in subband $n$. For convenience, we also define $P_n = \text{Tr}(\mathbf{Q}_n)$ as the transmit power in subband $n$ such that $P_{\text{dc}}^t = \sum_{n=0}^{N-1} P_n$.

The input signal $x_{\text{in},m}(t)$ on each antenna $m$ is then amplified by a HPA and filtered using a band-pass filter (BPF) into the transmit WPT signal $x_{\text{rf},m}(t)$

$$x_{\text{rf},m}(t) = \sqrt{2}\Re\left\{\sum_{n=0}^{N-1} x_{\text{rf},m,n}(t)e^{j2\pi f_n t}\right\}, \tag{5}$$

with

$$x_{\text{rf},m,n}(t) = \sum_{k=-\infty}^{\infty} x_{\text{rf},m,n,k}\,\text{sinc}(f_{\text{w}}t - k). \tag{6}$$

Realistically, the relationship between $x_{\text{in},m}(t)$ and $x_{\text{rf},m}(t)$ is nonlinear and accounts for coupling across frequencies as well as magnitude and phase distortions induced by the HPA and BPF. The transmit WPT signal $x_{\text{rf},m}(t)$ is then transmitted over the air by antenna $m$. The total average transmit power is expressed as $P_{\text{rf}}^t = \sum_{m=1}^{M} \mathbb{E}[x_{\text{rf},m}(t)^2]$ and is subject to the constraint $P_{\text{rf}}^t \leq P$.

The transmit WPT signal propagates through a multipath channel characterized by $L$ paths. Let $\tau_l$ and $\alpha_l$ be the delay and amplitude gain of the $l^{\text{th}}$ path, respectively. Further, denote by $\zeta_{q,m,n,l}$ the phase shift of the $l^{\text{th}}$ path between transmit antenna $m$ and receive antenna $q$ for subband $n$. The signal received at antenna $q$ ($q = 1, ..., Q$) from transmit antenna $m$ can be expressed as

$$\begin{aligned} y_{\text{rf},q,m}(t) &= \sqrt{2}\Re\Bigg\{\sum_{l=0}^{L-1}\sum_{n=0}^{N-1} \alpha_l x_{\text{rf},m,n}(t-\tau_l) \\ &\qquad e^{j2\pi f_n(t-\tau_l)+\zeta_{q,m,n,l}}\Bigg\}, \\ &\approx \sqrt{2}\Re\left\{\sum_{n=0}^{N-1} h_{q,m,n}x_{\text{rf},m,n}(t)e^{j2\pi f_n t}\right\}. \end{aligned} \tag{7}$$

We have assumed $\max_{l\neq l'}|\tau_l - \tau_{l'}| < 1/f_{\text{w}}$ so that, for each subband, $x_{\text{rf},n,m}(t)$ are narrowband signals, thus $x_{\text{rf},m,n}(t-\tau_l) = x_{\text{rf},m,n}(t)$, $\forall l$. Variable $h_{q,m,n} = \sum_{l=0}^{L-1} \alpha_l e^{j(-2\pi f_n \tau_l + \zeta_{q,m,n,l})}$ is the baseband channel frequency response between transmit antenna $m$ and receive antenna $q$ at frequency $f_n$.

The total signal and noise received at antenna $q$ is the superposition of the signals received from all $M$ transmit antennas, i.e.,

$$y_{\text{rf},q}(t) = \sqrt{2}\Re\left\{\sum_{n=0}^{N-1} \mathbf{h}_{q,n}\mathbf{x}_{\text{rf},n}(t)e^{j2\pi f_n t}\right\} + w_{\text{A},q}(t), \tag{8}$$

where $w_{\text{A},q}(t)$ is the antenna noise, $\mathbf{h}_{q,n} \triangleq \left[h_{q,1,n}, ..., h_{q,M,n}\right]$ denotes the channel vector from the $M$ transmit antennas to receive antenna $q$, and $\mathbf{x}_{\text{rf},n}(t) \triangleq \left[x_{\text{rf},1,n}(t), ..., x_{\text{rf},M,n}(t)\right]^T$.

Ignoring the noise power, the total RF power received by all $Q$ antennas of the receiver can be expressed as

$$P_{\text{rf}}^r = \sum_{q=1}^{Q} \mathbb{E}\left[y_{\text{rf},q}(t)^2\right] = \sum_{q=1}^{Q}\sum_{n=0}^{N-1} \mathbb{E}\left[|\mathbf{h}_{q,n}\mathbf{x}_{\text{rf},n}(t)|^2\right]. \tag{9}$$

Finally, unless stated explicitly, we assume perfect Channel State Information at the Transmitter (CSIT).

Next, the output DC power $P_{\text{dc}}^r$ depends on the exact ER architecture to be discussed in subsequent sections.

*Remark 1:* The system model is written using a general form assuming that complex-valued symbols are random variables and occupy a non-zero bandwidth. This is used to ease and harmonize the system model with WIPT discussed in Section III. It is nevertheless to be noted that if the aim is to design WPT without any consideration for communications, one would strictly speaking not need complex-valued symbols to be random, and one could assume them deterministic (with zero bandwidth), therefore transforming the above system model into unmodulated WPT with multisine waveforms (with $N$ sinewaves) transmitted from each antenna.

### *B. Transmitter (HPA) Nonlinearity*

We here discuss the modeling of the HPA. The HPA input-output relationship is realistically nonlinear, though this source of nonlinearity is commonly ignored.

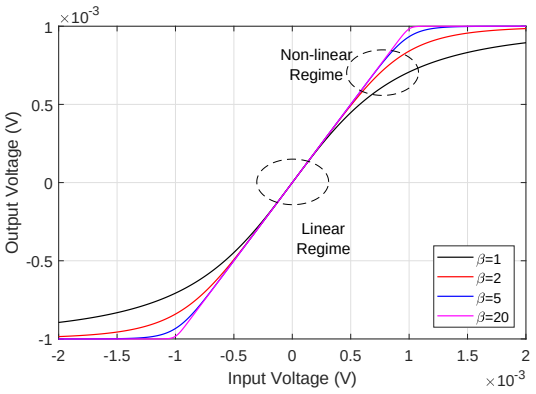



Fig. 5. Input-output voltage characteristics for solid state HPA with $A_{\rm s} = 10^{-3}$V and $G = 1$.

*1) Linear HPA:* : If we ignore the HPA nonlinearity and assume the relationship between $x_{{\rm in},m}(t)$ and $x_{{\rm rf},m}(t)$ is linear such that $x_{{\rm rf},m}(t) = G x_{{\rm in},m}(t)$ with $G$ the amplification gain, and taking $G = 1$ for simplicity of exposure, we have $P^t_{\rm rf} = \sum_{m=1}^{M} \mathbb{E}[x_{{\rm rf},m}(t)^2] = P^t_{\rm dc}$. In other words, referring to Fig. 4, the DC-to-RF conversion efficiency $e_1$ is equal to 1. More realistically, under the linear regime of the HPA, $P^t_{\rm rf} = e_1 P^t_{\rm dc}$ with $e_1$ a constant strictly smaller than 1 and independent of the input signal (but whose exact value depends on the HPA technology).

The total RF power received by all $Q$ antennas can then be expressed more easily as

$$P^r_{\rm rf} = e_1 \sum_{n=0}^{N-1} {\rm Tr}\left(\mathbf{H}_n^H \mathbf{H}_n \mathbf{Q}_n\right), \tag{10}$$

where $\mathbf{H}_n \triangleq \left[\mathbf{h}_{1,n}^H, \cdots, \mathbf{h}_{Q,n}^H\right]^H \in \mathbb{C}^{Q\times M}$ denotes the MIMO channel matrix from the $M$ transmit antennas to the $Q$ receive antennas at subband $n$.

*2) Nonlinear HPA:* : The HPA has a nonlinear characteristics that distorts its input signal and makes it challenging to analyze. Indeed, real HPAs do not exhibit a pure linear behavior and $x_{{\rm rf},m}(t) = f_{\rm HPA}\left(x_{{\rm in},m}(t)\right)$ where $f_{\rm HPA}$ is a nonlinear function, which leads to $P^t_{\rm rf} = e_1(x_{{\rm in},m}(t)) P^t_{\rm dc}$, i.e., $e_1(x_{{\rm in},m}(t))$ is itself a nonlinear function of $x_{{\rm in},m}(t)$. A common model for solid state HPA [55], [56] is written as

$$f_{\rm HPA}(x_{{\rm in},m}(t)) = \frac{G}{\left(1 + \left(\frac{G|x_{{\rm in},m}(t)|}{A_{\rm s}}\right)^{2\beta}\right)^{\frac{1}{2\beta}}} x_{{\rm in},m}(t) \tag{11}$$

where $A_{\rm s}$ is the output saturation voltage, $G$ is the amplification gain, and $\beta$ represents the smoothness of the transition from the linear regime to the saturation. In Fig. 5, (11) is illustrated for $A_{\rm s} = 10^{-3}$V and $G = 1$. The HPA would operate in the linear regime if the input voltage is significantly smaller than $A_s$ and would operate in the nonlinear regime (leading to saturation) otherwise.

### C. Energy Receiver Nonlinearity and Architecture

We here discuss the architecture and related nonlinearity of single-antenna and multi-antenna ER.

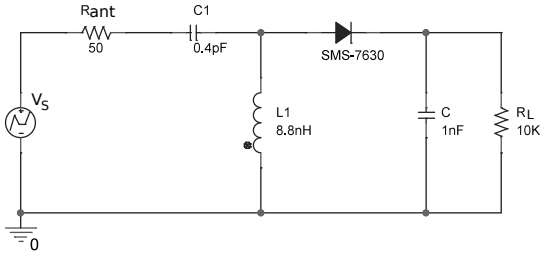



Fig. 6. Single series rectifier designed for an average RF input power of -20dBm (10$\mu$W) at 2.45GHz [58]. $v_{\rm s}$ is the voltage source of the antenna. R1 models the antenna impedance. C1 and L1 form the matching network. SMS-7630 refers to the type of Schottky diode. C and $R_{\rm L}$ form the low-pass filter with $R_{\rm L}$ being the output load.

*1) Single-Antenna Energy Receiver:* The key building block of the ER is the rectenna [57]. A rectenna harvests electromagnetic energy, then rectifies and filters it using a low pass filter. The rectenna can be optimized for the specific operating frequencies, input power level and input waveforms. Various rectifier technologies (including the popular Schottky diodes) and topologies (with single and multiple diode rectifier) have been studied [4]–[6]. The simplest form of rectifier, so-called single series rectifier, is illustrated by the circuit in Fig. 6 [58]. It is made of a matching network (to match the antenna impedance to the rectifier input impedance) followed by a single diode and a low-pass filter. This circuit was designed for 10$\mu$W input power at 2.45GHz.

Using circuit simulations and the single-series rectifier from Fig. 6, Fig. 7 illustrates the dependency of the RF-to-DC conversion efficiency $e_3$ to the signal average RF power and shape at the input of the rectifier when a continuous wave (CW), i.e., a single sinewave, and a multisine waveform (with $N = 8$ equispaced frequencies) are used for excitation [58]. Those two excitations have the same average RF input power $P^r_{\rm rf}$, but their shape is different.

We note that $e_3$ is particularly low at low input power for both types of excitations. This is due to the rectifier sensitivity with the diode not being easily turned on at low input power. Nevertheless, the multisine waveform manages to boost $e_3$ in the low power regime much better than CW. Importantly, for a given waveform, be it CW or multisine, $e_3$ increases with $P^r_{\rm rf}$ in the normal region of operation of the rectifier, namely whenever the diode is not in the breakdown region. Beyond a few hundreds of $\mu$W input power, irrespectively of the input signal shape, the output DC power saturates and $e_3$ suddenly significantly drops when the rectifier enters the diode breakdown region[1], which is not the intended region of operation of the rectifier.

The key observation of Fig. 7 is that due to the EH nonlinearity, $e_3$ is clearly not a constant, but depends on 1) the input power level and 2) the shape of the input signal $y_{\rm rf}$ [60]–[64]. Mathematically, this is reflected by the fact that the output DC voltage $v_{\rm out} = f_{\rm EH}\left(y_{\rm rf}(t)\right)$ where $f_{\rm EH}\left(y_{\rm rf}(t)\right)$ is a nonlinear function of $y_{\rm rf}(t)$, which has as consequence that $P^r_{\rm dc} = e_3\left(y_{\rm rf}(t)\right) P^r_{\rm rf}$, i.e. $e_3$ is not a constant but rather a

[1]The diode SMS-7630 becomes reverse biased at $P^r_{\rm rf} \approx 500\mu$W to 1mW for CW. To operate beyond such input power, multiple diode rectifier is preferred to avoid the saturation problem [4], [6], [59].

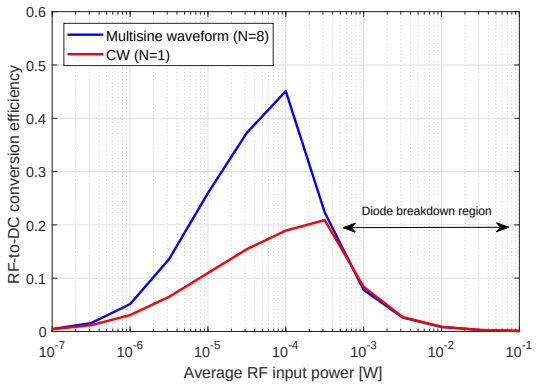



Fig. 7. RF-to-DC conversion efficiency $e_3$ vs average RF input power $P_{\rm rf}^r$ with rectifier from Fig. 6 obtained from circuit simulations [58]. The input signal is a CW at 5.18 GHz and rectifier is designed for -20dBm input power.

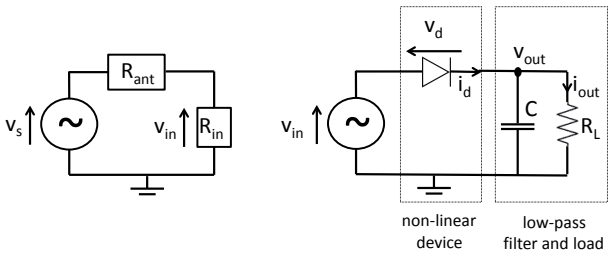



Fig. 8. Antenna equivalent circuit (left) and a single diode rectifier (right) [64]. The rectifier comprises a non-linear device (diode) and a low-pass filter (consisting of a capacitor C and a load $R_L$).

nonlinear function of the input signal to the rectenna. Note the importance of writing $f_{\rm EH}\left(y_{\rm rf}(t)\right)$ and $e_3\left(y_{\rm rf}(t)\right)$ instead of simply $f_{\rm EH}\left(P_{\rm rf}^r\right)$ and $e_3\left(P_{\rm rf}^r\right)$. $f_{\rm EH}$ and $e_3$ are not simply a nonlinear function of the average RF input power $P_{\rm rf}^r$ of the input waveform $y_{\rm rf}(t)$, but also of the shape of this input waveform!

The first and only model available in the WPT signal design literature that captures power and shape dependency on output DC power was derived in [64]–[66], and is briefly summarized in the sequel. Let us abstract the rectifier in Fig. 6 into the simplified representation in Fig. 8. We consider for simplicity a rectifier with input impedance $R_{\rm in}$ composed of a single series diode followed by a low-pass filter with a load. We consider this setup as it is the simplest rectifier configuration[2]. As per the system model, the RF signal $y_{\rm rf}(t)$ impinging on the receive antenna has an average power $P_{\rm rf}^r$. The receive antenna is assumed lossless and modeled as an equivalent voltage source $v_{\rm s}(t)$ in series with an impedance $R_{\rm ant}$ as shown in Fig. 8. With perfect matching ($R_{\rm in} = R_{\rm ant}$), the input voltage of the rectifier $v_{\rm in}(t)$ can be related to the received signal $y_{\rm rf}(t)$ by $v_{\rm in}(t) = y_{\rm rf}(t)\sqrt{R_{\rm ant}}$. A rectifier is always made of a nonlinear rectifying component such as diode followed by a low pass filter with load as shown in Fig. 8.

The current $i_d(t)$ flowing through an ideal diode (neglecting its series resistance) relates to the voltage drop across the diode $v_{\rm d}(t) = v_{\rm in}(t) - v_{\rm out}(t)$ as

$$i_d(t) = i_{\rm s}\left(e^{\frac{v_{\rm d}(t)}{n v_{\rm t}}} - 1\right), \tag{12}$$

where $i_{\rm s}$ is the reverse bias saturation current, $v_{\rm t}$ is the thermal voltage, $n$ is the ideality factor (assumed equal to 1.05).

Taking the polynomial (Taylor) expansion of the diode I-V characteristics $i_d(t)$, truncating it at the $n_o^{th}$ order, making use of some physical assumptions on an ideal low-pass filter that removes the non-DC components in $i_d(t)$ and the rectenna output voltage $v_{\rm out}(t)$, the output DC voltage of the rectifier $v_{\rm out}$ can be approximated as the following nonlinear function of $y_{\rm rf}(t)$

$$v_{\rm out} = f_{\rm EH}\left(y_{\rm rf}(t)\right) = \sum_{i \text{ even}, i\geq 2}^{n_o} \beta_i \mathbb{E}\left[y_{\rm rf}(t)^i\right] \tag{13}$$

where $\beta_i = \frac{R_{\rm ant}^{i/2}}{i!(n v_{\rm t})^{(i-1)}}$ [64], [67]. The operator $\mathbb{E}[\cdot]$ in (13) has the effect of taking the DC component of the diode current $i_{\rm d}(t)$ but also averaging over the potential randomness carried by the input signal $y_{\rm rf}(t)$. Consequently, the harvested DC power $P_{\rm dc}^r$ of the single-antenna receiver is then given by

$$P_{\rm dc}^r = \frac{v_{\rm out}^2}{R_{\rm L}}. \tag{14}$$

We clearly see that $f_{\rm EH}\left(y_{\rm rf}(t)\right)$ is a nonlinear function of $y_{\rm rf}(t)$. Specifically, it is a function of the input signal average power $P_{\rm rf}^r = \mathbb{E}\left[y_{\rm rf}(t)^2\right]$ (i.e. the second moment of $y_{\rm rf}(t)$) but also of its higher order moments $\mathbb{E}\left[y_{\rm rf}(t)^i\right]$ for $i$ even and $i > 2$. This dependency on the second and higher order moments of $y_{\rm rf}(t)$ explains why multisine outperforms CW in Fig. 7 [64], but also explains why $e_3$ is an increasing function of $P_{\rm rf}^r$. Indeed, due to the convexity of the I-V characteristics and the polynomial expansion, using Jensen's inequality, we have

$$\mathbb{E}\left[y_{\rm rf}(t)^i\right] \geq \left(\mathbb{E}\left[y_{\rm rf}(t)^2\right]\right)^{\frac{i}{2}} = \left(P_{\rm rf}^r\right)^{\frac{i}{2}} \tag{15}$$

for $i$ even and $i \geq 2$, so that

$$v_{\rm out} \geq \sum_{i \text{ even}, i\geq 2}^{n_o} \beta_i \left(P_{\rm rf}^r\right)^{\frac{i}{2}}. \tag{16}$$

Taking for instance as $y_{\rm rf}(t)$ a multisine waveform with average power $P_{\rm rf}^r$ uniformly distributed across the $N$ sinewaves, we can easily show that $\mathbb{E}\left[y_{\rm rf}(t)^4\right]$ scales proportionally to $N\left(P_{\rm rf}^r\right)^2$, therefore demonstrating that $\mathbb{E}\left[y_{\rm rf}(t)^4\right] > \left(P_{\rm rf}^r\right)^2$ for sufficiently large $N$ and explaining mathematically why multisine (and other types of signals) can outperform CW ($N = 1$) [64]. We can draw two crucial observations from relationships (15) and (16), respectively.

*Observation 1:* Relationship (15) highlights the key role of choosing input signals with large $\mathbb{E}\left[y_{\rm rf}(t)^i\right]$. Two input signals may indeed have the same $\mathbb{E}\left[y_{\rm rf}(t)^2\right] = P_{\rm rf}^r$ but very different $\mathbb{E}\left[y_{\rm rf}(t)^4\right]$. This explains mathematically the dependence of $e_3$ (and $P_{\rm dc}^r$) on the shape of the input signal in Fig. 7.

*Observation 2:* The lower bound (16) highlights that $e_3$ increases with $P_{\rm rf}^r$. This explains mathematically the dependence of $e_3$ on the input power level in Fig. 7 for the practical operation regime of the rectifier (not in breakdown),

[2]The model is not limited to a single series diode but also holds for more general rectifiers with many diodes as per [68].

and highlights that the strategy that maximizes $P_{\rm rf}^r$ *does not* maximize $P_{\rm dc}^r$, but only maximizes a lower bound on $P_{\rm dc}^r$.

Those two observations highlight that a signal theory, design and processing of basic building blocks of wireless powered networks such as modulation, waveform, and input distribution, are influenced by the EH nonlinearity and motivate efficient signal and system designs that leverage the EH nonlinearity. The crucial role played by this EH nonlinearity in the signal designs and evaluations of WPT, SWIPT, and WPBC was first highlighted in [64], [66], and [37], respectively.

*Remark 2:* Slightly different formulations of the above EH model are available in [64], [67], [69]–[71], where the output is expressed in terms of DC current instead of voltage, or where operator $\mathbb{E}[.]$ is applied without performing the polynomial expansion.

*Remark 3:* Other models for $P_{\rm dc}^r$ are available in the literature as discussed in greater details in [9]. Those models either assume $e_3$ constant (so-called linear model [7]) or only capture the dependency of $e_3$ on $\mathbb{E}\left[y_{\rm rf}(t)^2\right] = P_{\rm rf}^r$ (e.g., so-called saturation nonlinear model [72]). The linear model is very inaccurate [64], [73]. The saturation model is more accurate since it is based on curve fitting, but does not capture the dependency of the rectification process on the shape of the input signal and arguably over-emphasizes the importance of saturation in the EH. Saturation is unlikely a major problem in wireless powered networks since the typical input RF power levels (below $100\mu$W) are smaller than the saturation level, as demonstrated by over the air measurements with various types of signals in [73], [74] (and also in Fig. 11 and 12 below). Moreover, if saturation happens to lead to a significant performance loss, it implies that the rectifier was not designed carefully enough for the expected range of input power levels. Saturation can indeed be avoided by a proper design of the rectifier [4], [6], [9], [59], [66]. The interested reader is referred to [9], [66] and references therein for more discussions on EH models.

*2) Multi-Antenna Energy Receiver:* Two main combining strategies exist, namely DC combining and RF combining, as illustrated in Fig. 9 and Fig. 10, respectively [75]. In DC combining, each receive antenna is connected to a rectifier and the number of rectifiers increases with the number of receive antennas. However, in RF combining, the RF signals from all receive antennas are first combined in the RF domain before being fed to a single rectifier used to rectify the combined RF signal. A combination of those two architectures is also possible, as well as other variants based on the use of power splitters and power combiners [76].

In the DC combiner architecture,

$$P_{\rm dc}^r = \sum_{q=1}^{Q} \frac{v_{{\rm out},q}^2}{R_{\rm L}}, \tag{17}$$

where $v_{{\rm out},q} = f_{\rm EH}\left(y_{{\rm rf},q}(t)\right)$ is the output DC voltage of the rectifier connected to receive antenna $q$.

In the RF combiner architecture, a frequency-dependent analogue combiner $\mathbf{w}_{{\rm R},n}$ is applied to the received signals (8) such that the received signal after combining fed to the single rectifier is given by

$$\tilde{y}(t) = \sqrt{2}\Re\left\{\sum_{n=0}^{N-1} \mathbf{w}_{{\rm R},n}^H \mathbf{H}_n \mathbf{x}_{{\rm rf},n}(t) e^{j2\pi f_n t}\right\} + \tilde{w}_{\rm A}(t), \tag{18}$$

where $\tilde{w}_{\rm A}$ is the effective combined noise. Note that in practice it may be difficult to design frequency-dependent combiner, in which case $\mathbf{w}_{\rm R} = \mathbf{w}_{{\rm R},n}$ is constant across frequency. The combiner is subject to the constraint $\|\mathbf{w}_{\rm R}\|^2 \leq 1$ originating from the fact that since the RF combining circuit is passive, the output power of the RF combining circuit should be no larger than its input power. Additionally, the combiner may be subject to constant modulus constraint so as to be implemented by $Q$ phase shifters of the form

$$\mathbf{w}_{\rm R} = \frac{1}{\sqrt{Q}}\left[e^{-j\theta_1}, e^{-j\theta_2}, \ldots, e^{-j\theta_Q}\right]^T, \tag{19}$$

where $\theta_q$ denotes the $q$th phase shift for $q = 1, \ldots, Q$. Finally, the output DC power is given by $P_{\rm dc}^r = \frac{v_{\rm out}^2}{R_{\rm L}}$ where $v_{\rm out} = f_{\rm EH}\left(\tilde{y}(t)\right)$.

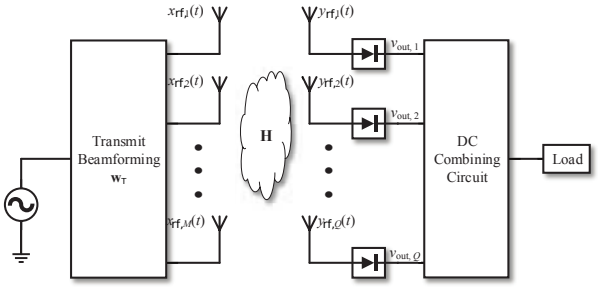



Fig. 9. Schematic of the MIMO WPT system with DC combining at the receiver [75].

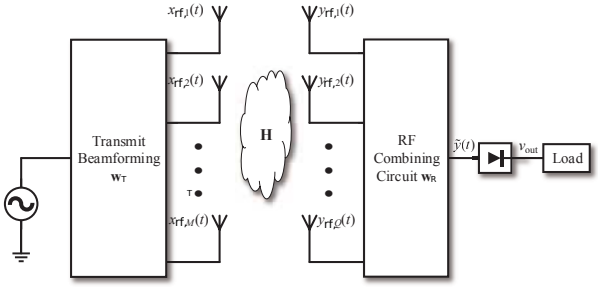



Fig. 10. Schematic of the MIMO WPT system with RF combining at the receiver [75].

## D. End-to-End Efficiency, Energy Maximization and Problem Formulation

A major and interesting technical challenge in WPT system design is that the maximization of $e$ is not achieved by maximizing $e_1$, $e_2$, $e_3$ independently from each other. This is because $e_1$, $e_2$, $e_3$ are coupled due to the aforementioned nonlinearities, especially at practical input RF power range 1 $\mu$W -1 mW. Indeed, since $e_3$ is a function of the input signal shape and power to the rectifier and therefore a function of the transmit signal and the wireless channel state. Similarly, $e_2$ depends on the transmit signal and the channel state and so is $e_1$, since it is a function of the HPA nonlinearity.

One possible problem formulation is therefore to find the signaling strategies that maximizes $e$. This can be written as

$$\max_{p(\mathbf{x}_0,\ldots,\mathbf{x}_{N-1})} \quad e(\mathbf{x}_0,\ldots,\mathbf{x}_{N-1}) \tag{20}$$

$$\text{subject to} \quad P_{\text{rf}}^t \leq P. \tag{21}$$

where maximization is performed over the input distributions $p(\mathbf{x}_0,\ldots,\mathbf{x}_{N-1})$ that satisfies the average transmit power constraint $P_{\text{rf}}^t \leq P$. An alternative formulation that is more common consists in maximizing the harvested DC output power

$$\max_{p(\mathbf{x}_0,\ldots,\mathbf{x}_{N-1})} \quad P_{\text{dc}}^r(\mathbf{x}_0,\ldots,\mathbf{x}_{N-1}) \tag{22}$$

$$\text{subject to} \quad P_{\text{rf}}^t \leq P, \tag{23}$$

In those two formulations, if the ER is equipped with an RF combiner, the optimization would have to be additionally performed over $\mathbf{w}_{\text{R}}$ subject to the constraint $\|\mathbf{w}_{\text{R}}\|^2 \leq 1$ or structure as in (19). Note that in the event power bearing symbols are deterministic, $\max_{p(\mathbf{x}_0,\ldots,\mathbf{x}_{N-1})}$ can be replaced by $\max_{\mathbf{x}_0,\ldots,\mathbf{x}_{N-1}}$.

Note that those two formulations are not equivalent. The main difference is that (20) specifically accounts for $e_1$, while (22) does not. To account for $e_1$ and HPA efficiency, additional constraints can be added to problem (22) for instance in the form of peak-to-average power (PAPR) constraints [64].

### *E. Signal Processing Techniques for Single-User WPT*

In this section, we review recent signal processing techniques developed to tackle the challenges of WPT, increase its efficiency and its range in a single-user setting. Techniques discussed include transmit active beamforming, transmit passive beamforming and intelligent reflecting surfaces, receive combining with multi-antenna harvester, waveform, joint beamforming, combining and waveform, large-scale (massive) multiple-input multiple-output (MIMO), channel acquisition, transmit diversity, time-reversal, and retrodirective arrays. Importantly, while some of those techniques focus on enhancing $P_{\text{rf}}^r$ and $e_2$ (and therefore a lower bound on $P_{\text{dc}}^r$), others such as waveform, transmit diversity, receive combiner, joint waveform and beamforming are deeply rooted in the EH nonlinearity (and therefore the maximization of $P_{\text{dc}}^r$ itself) and only appeared to light once the nonlinearity is accounted for in the signal design.

*1) Transmit Active Beamforming:* Leveraging the presence of multiple antennas at the transmitter, each equipped with an RF chain, the simplest strategy is transmit active beamforming to increase $P_{\text{rf}}^r$. Considering a multiple-input single-output (MISO) setup ($Q = 1$) with $N = 1$ and a linear HPA, (8) boils down to $y_{\text{rf}}(t) = \sqrt{2}\Re\left\{\mathbf{h}\mathbf{w}_{\text{T}}x(t)e^{j2\pi ft}\right\} + w_{\text{A}}(t)$, with $\mathbf{w}_{\text{T}}$ the transmit beamformer. The transmitter simply performs conventional Maximum Ratio Transmission (MRT) $\mathbf{w}_{\text{T}} = \sqrt{P}\bar{\mathbf{h}}^H$, with $\bar{\mathbf{h}} = \mathbf{h}/\|\mathbf{h}\|$ and $x(t)$ being any chosen random input with unit power (with $x(t) = 1$ corresponding to a CW). Fig. 11 and 12 illustrate the benefits in terms of output DC power $P_{\text{dc}}^r$ and range of WPT by adopting MRT beamforming with 1, 2, 4, 8 transmit antennas and continuous wave ($N = 1$, 1 tone), based on experimental data gathered

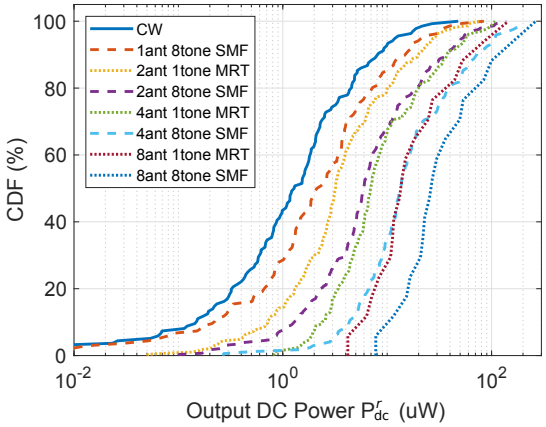



Fig. 11. CDF of output DC power ($P_{\text{dc}}^r$) measurement results at different distances from 0.6 to 5.4 m with the number of transmit antennas $M$=1, 2, 4, 8, and the number of carriers (tones) $N$=1 and 8 [74].

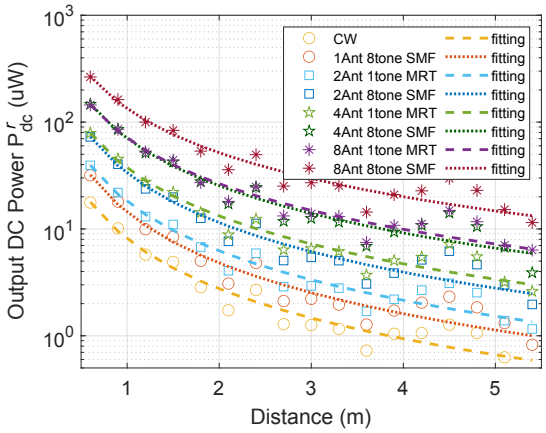



Fig. 12. Curve fitting and measurement results of output DC power ($P_{\text{dc}}^r$) versus distance with $N$=1, 8 and $M$=1,2,4,8 [74].

in a typical indoor environment at 2.4GHz with a rectenna similar to Fig. 6 under an Effective Isotropic Radiated Power (EIRP) of 36dBm [74]. Other experimental results of such beamforming technique can be found in [77], [78].

*2) Transmit Passive Beamforming:* Transmit Passive Beamforming through *intelligent reflecting surface* (IRS), also known as *reconfigurable intelligent surface*, has gained popularity as an emerging technology for wireless networks [79]–[81]. IRS consists of a large number of $L$ reconfigurable passive elements (without any need for an RF chain) integrated into the propagation environment. By collaboratively adjusting the impedance of all passive elements at the IRS, the reflected signals add coherently with the signals from other paths at the desired receiver to increase the received RF signal power, therefore enabling a passive beamforming gain. Owing to the passive structure, IRS has several advantages including low cost, low profile, light weight, conformal geometry, low power consumption and no additive thermal noise during the reflection.

Considering a SISO setup ($M = 1$, $Q = 1$) with $N = 1$ and a linear HPA, $y_{\text{rf}}(t) = \sqrt{2}\Re\left\{hx(t)e^{j2\pi ft}\right\} + w_{\text{A}}(t)$ with $h = g_{\text{d}} + \mathbf{g}_{\text{r}}\boldsymbol{\Theta}\mathbf{g}_{\text{i}}$ where $g_{\text{d}}$ refers to the direct channel between the ET and ER, $\mathbf{g}_{\text{r}}$ is a $1 \times L$ vector channel

between the IRS (equipped with $L$ elements) and the ER, and $\mathbf{g}_\mathrm{i}$ is the $L \times 1$ vector channel between the ET and the IRS. $\mathbf{\Theta}$ is the scattering matrix of the $L$-port reconfigurable impedance network and is subject to the constraints $\mathbf{\Theta} = \mathbf{\Theta}^T$ and $\mathbf{\Theta}^H\mathbf{\Theta} = \mathbf{I}_L$ [82]. The $L$-port reconfigurable impedance network is constructed with reconfigurable and passive elements so that it can reflect the incident signal with a reconfiguration that can be adapted to the channel. Three different architectures are possible, namely single connected reconfigurable impedance network characterized by a diagonal $\mathbf{\Theta}$ (the simplest and most common one [79]–[81]), group connected reconfigurable impedance network characterized by a block diagonal $\mathbf{\Theta} = \mathrm{diag}(\mathbf{\Theta}_1, \mathbf{\Theta}_2, \ldots, \mathbf{\Theta}_G)$ where the $L$ elements have been divided into $G$ groups with each group having $L_G = L/G$ elements, and fully connected reconfigurable impedance network characterized by a full $\mathbf{\Theta}$ [82].

Considering a group connected reconfigurable impedance network, the design of $\mathbf{\Theta}$ that maximizes $e_2$ is the solution of the optimization problem

$$\max_{\mathbf{\Theta}} \quad |h_\mathrm{d} + \mathbf{h}_\mathrm{r}\mathbf{\Theta}\mathbf{h}_\mathrm{i}|^2 \tag{24}$$

$$\text{subject to} \quad \mathbf{\Theta} = \mathrm{diag}(\mathbf{\Theta}_1, \mathbf{\Theta}_2, \ldots, \mathbf{\Theta}_G), \tag{25}$$

$$\mathbf{\Theta}_g^H\mathbf{\Theta}_g = \mathbf{I}_{L_G}, \forall g, \tag{26}$$

$$\mathbf{\Theta}_g = \mathbf{\Theta}_g^T, \forall g. \tag{27}$$

The single and fully connected reconfigurable impedance networks can be designed similarly by noting that they are two special cases of the group connected reconfigurable impedance network, i.e. with $G = L$ ($L_G = 1$) and $G = 1$ ($L_G = L$), respectively. One way to solve (24)-(27) is by reformulating it as an unconstrained optimization problem [82]. It has been shown in [82] that for a given $L$, the larger $L_G$ (and the smaller $G$) the higher the received RF power (and therefore the higher $e_2$). In other words, the received RF power $P^r_{\mathrm{rf,full}}$ of fully connected networks is larger than that of the group connected network ($P^r_{\mathrm{rf,group}}$) and single connected network ($P^r_{\mathrm{rf,single}}$), at the cost of a higher implementation complexity. Group connected network exhibit a nice trade-off between complexity and performance. This is illustrated in Fig. 13 where the power gains $P^r_{\mathrm{rf,group}}/P^r_{\mathrm{rf,single}}$ and $P^r_{\mathrm{rf,full}}/P^r_{\mathrm{rf,single}}$ of the group connected and fully connected reconfigurable impedance networks over the single connected reconfigurable impedance network are displayed as a function of $L$ for several values of group size $L_G$ [82]. Compared with the single connected reconfigurable impedance network, fully connected reconfigurable impedance network can increase the received signal power by up to 62%. For group connected with $L_G = 2, 3, 4, 6, 8$, gains of 26%, 37%, 43%, 49%, 52% are achieved over the single connected network, respectively.

In contrast to active antenna arrays where the amplitudes and phases can be adjusted freely at each antenna and at each frequency, the elements in IRS are subject to less flexibility due to the passive nature of the IRS and the hardware constraints. Specifically, taking a single-connected network, due to constraints $\mathbf{\Theta}^H\mathbf{\Theta} = \mathbf{I}$, the amplitude of the diagonal entries is fixed to unity and only the phases of those entries are optimized. Moreover the IRS is commonly assumed frequency flat in the sense that the phases of the passive elements are kept constant across frequency. Despite those constraints, the passive beamforming gain can be significant and IRS brings some natural benefits to WPT since IRS can help increasing the RF power level $P^r_\mathrm{rf}$ at the input of the rectenna [83]–[88]. The presence of active antennas and passive IRS leads to a joint design and optimization of active and passive beamforming.

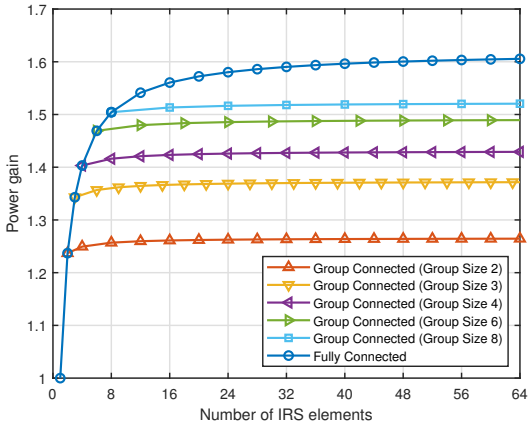



Fig. 13. Power gain of the group connected and fully connected reconfigurable impedance networks over the single connected reconfigurable impedance network. [82].

*3) Receive Combining:* Beamforming is not limited to the ET and can also be used at the ER subject to a proper design of the DC and RF combiner schemes of Fig. 9 and 10 [75]. Assuming $N = 1$ and a linear HPA, $y_{\mathrm{rf},q}(t) = \sqrt{2}\Re\left\{\mathbf{h}_q\mathbf{w}_\mathrm{T}x(t)e^{j2\pi ft}\right\} + w_{\mathrm{A},q}(t)$. In MIMO WPT with DC combiner, only the transmit beamformer $\mathbf{w}_\mathrm{T}$ is optimized and problem (22) is equivalent to

$$\max_{\mathbf{w}_\mathrm{T}} \frac{1}{R_\mathrm{L}} \sum_{q=1}^{Q} \left( \sum_{i \text{ even}, i\geq 2}^{n_o} \beta_i \zeta_i \left|\mathbf{h}_q\mathbf{w}_\mathrm{T}\right|^i \right)^2 \tag{28}$$

$$\text{subject to} \quad \left\|\mathbf{w}_\mathrm{T}\right\|^2 \leq P, \tag{29}$$

where $\zeta_2 = 1/2$, $\zeta_4 = 3/8$, $\zeta_6 = 5/16$ [75]. With RF combiner, $\tilde{y}_\mathrm{rf}(t) = \sqrt{2}\Re\left\{\mathbf{w}_\mathrm{R}^H\mathbf{H}\mathbf{w}_\mathrm{T}x(t)e^{j2\pi ft}\right\} + \tilde{w}_\mathrm{A}(t)$, and $\mathbf{w}_\mathrm{T}$ and $\mathbf{w}_\mathrm{R}$ need to be jointly optimized. Subject to combiner structure (19), Problem (22) is equivalent to

$$\max_{\mathbf{w}_\mathrm{T}, \{\theta_q\}} \left|\mathbf{w}_\mathrm{R}^H\mathbf{H}\mathbf{w}_\mathrm{T}\right|^2 \tag{30}$$

$$\text{subject to} \quad \left\|\mathbf{w}_\mathrm{T}\right\|^2 \leq P, \tag{31}$$

$$\mathbf{w}_\mathrm{R} = \frac{1}{\sqrt{Q}}\left[e^{-j\theta_1}, e^{-j\theta_2}, \ldots, e^{-j\theta_Q}\right]^T, \tag{32}$$

$$-\pi \leq \theta_q \leq \pi, 1 \leq q \leq Q. \tag{33}$$

Those non-convex optimization problems can be solved by involving geometric program (GP) and semi-definite relaxation (SDR) [75].

Interestingly, due to the rectenna nonlinearity that induces a higher $e_3$ for higher input power level (recall Fig. 7 for power levels lower than saturation and Observation 2), it turns out that RF combining outperforms DC combining since the rectifier in RF combining operates on a higher

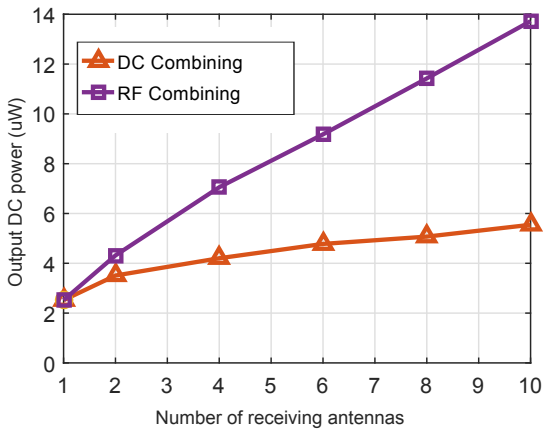



Fig. 14. Average output DC power $P_{dc}^{r}$ versus the number of receive antennas $Q$ for $M = 10$ transmit antennas based on circuit simulations [75].

RF power input signal. In other words, RF combining can leverage the nonlinearity more efficiently than DC combining. The performance gains of RF combining methods over DC combining can be quite significant as shown in the circuit simulations of Fig. 14. We can see that $P_{\rm dc}^{r}$ increases with the number of receive antennas $Q$ (this is in a way reminiscent of increasing $M$ in transmit beamforming), but the increase is faster with RF combining than DC combining. Hence, in MIMO WPT, while increasing the number of transmit antennas $M$ and receive antennas $Q$ helps to increase $P_{\rm rf}^{r}$ and therefore $e_2$, a suitable choice of the combiner at the receiver further helps by increasing $e_3$.

The challenge with RF combining is that it not only needs CSIT but also Channel State Information at the Receiver (CSIR) for the joint transmit beamforming and receive combiner optimization. In contrast, DC combining only needs CSIT for transmit beamforming optimization.

*4) Waveform:* Another promising strategy is the design of transmit multi-carrier ($N > 1$) waveform to utilize the nonlinear characteristic of the rectenna so as to boost $e_2 \times e_3$ [64]. Such design originates from the fact that the output of the EH and $e_3$ are a nonlinear function of the rectenna input signal shape, as shown in Fig. 7 and discussed in Observation 1. The transmit waveform design has a significant influence on $P_{\rm dc}^{r}$, namely it not only affects $e_2$ and $P_{\rm rf}^{r}$, but also $e_3$.

In [64], a systematic methodology was derived to design and optimize waveforms for WPT. The optimal waveform design in [64] is adaptive to the frequency selective channel (with frequency flat channel being a special case) and is rooted in the tradeoff between allocating the power to the strongest carrier so as to leverage the frequency diversity/selectivity and maximize $e_2$ and allocate power across $N$ carriers so as to leverage the rectifier nonlinearity and maximize $e_3$. As a result, the optimal waveform allocates power non-uniformly across the $N$ carriers, with the carriers corresponding to stronger channel gain allocated more power. Due to the EH nonlinearity, the waveform design results from a non-convex and computationally involved optimization problem. Assuming $M = 1$, $Q = 1$, linear HPA and deterministic multisine waveform, $y_{\rm rf}(t) = \sqrt{2}\Re\left\{\sum_{n=0}^{N-1} h_n x_n e^{j2\pi f_n t}\right\} + w_{\rm A}(t)$. Denoting $x_n = s_n e^{j\varphi_n}$ and $h_n = A_n e^{j\varphi_n}$, the optimal set of phases $\{\varphi_n\}$ and magnitudes $\{s_n\}$ that are solutions of problem (22) are given by $\varphi_n^{\star} = -\varphi_n$ and by the solutions of the optimization problem (for $n_o = 4$)

$$\max_{\{s_n\}} \ \alpha\left[\sum_{n=0}^{N-1} s_n^2 A_n^2\right] + \sum_{\substack{n_0,n_1,n_2,n_3 \\ n_0+n_1=n_2+n_3}} \prod_{j=0}^{3} s_{n_j} A_{n_j} \quad (34)$$

$$\text{subject to} \ \sum_{n=0}^{N-1} s_n^2 \le P \quad (35)$$

where $\alpha = \beta_2\zeta_2/(\beta_4\zeta_4)$. The first term in (34) relates to $P_{\rm rf}^{r} = \mathbb{E}\left[y_{\rm rf}(t)^2\right]$ and the second term to $\mathbb{E}\left[y_{\rm rf}(t)^4\right]$. The challenge is due to the nonlinear coupling across frequency captured by the second term. The first term will favor a single-sinewave power allocation strategy, i.e., allocating all the power to sinewave corresponding to $\max_n A_n$. However due to the presence of the second term, such a single-sinewave strategy is in general sub-optimal. Indeed, the optimal solution results from a tradeoff between maximizing the first term (and therefore maximize $P_{\rm rf}^{r}$) by allocating power to a single sinewave and leveraging the nonlinearity of the second term (and therefore maximize $e_3$) by allocating power across multiple sinewaves. Consequently, the optimal solution, obtained using reverse GP, reveals that the power is allocated across all sinewaves but more power is allocated to frequencies corresponding to larger channel gains [64]. This is illustrated in Fig. 15 where the upper graph is the magnitude of the channel frequency response and the lower figure illustrates the solution of problem (34) ("opt") at $N = 16$ uniformly spaced frequencies. Doing so, the waveform exploits a channel frequency diversity gain and the EH nonlinearity.

GP does not lend itself easily to implementation due to the high complexity. Other optimization frameworks to design waveforms have therefore been proposed in [67], [69]. Suboptimal low complexity methods, called SMF, have also been proposed in [68]. A simple way to allocate power across frequencies is as follows $s_n^2 = cA_n^{2\beta}$ where $c$ is a constant satisfying the average transmit power constraint. By scaling the channel gain using an exponent proportional to $\beta > 1$, the waveform allocates more (resp. less) power to the frequency components corresponding to large (resp. weak) channel gains and replicates the main behavior of the "opt" solution. This is illustrated in Fig. 15 with $\beta = 1, 3$ [68]. By adjusting $\beta > 1$, we amplify the strong frequency components and attenuate the weak ones, so as to come close to the optimal power allocation. Though suboptimal, the SMF design was shown to perform close to the "opt" GP design.

Such optimized and low complexity waveforms were shown using circuit simulations to provide significant benefits of 100%-200% over conventional continuous-wave signal and non-optimized waveforms in a wide range of rectifier topologies by leveraging the channel frequency diversity gain and a gain originating from the rectifier nonlinearity [64], [68]. They have been successfully experimentally validated, demonstrating gains of 105%-170% in real-time over-the-air experimentation, in [73]. In Fig. 11 and 12, the benefit in terms of output

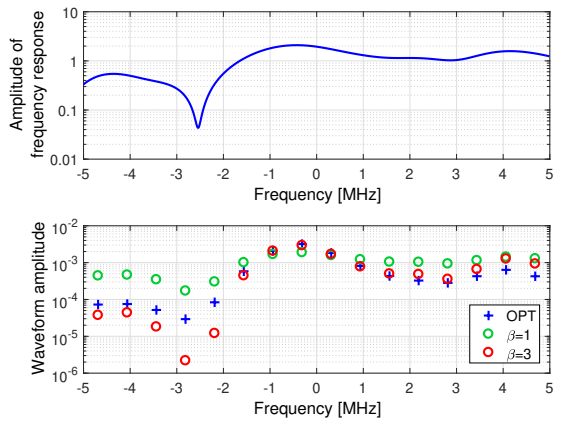



Fig. 15. Frequency response of the wireless channel and WPT waveform magnitudes (N = 16) for 10 MHz bandwidth [68].

DC power and range with using $N = 8$ over conventional CW ($N = 1$) is illustrated.

In the presence of HPA nonlinearity, PAPR constraints could be added to problem (34) as in [64], though a more involved and interesting problem would be to revise problem (34) accounting for HPA nonlinearity (11).

*5) Joint Beamforming, Combining and Waveform:* Remarkably, such waveforms can also be designed for a multi-antenna transmitter so as to additionally exploit a beamforming gain [64], [67]. Joint waveform and beamforming enables to simultaneously harvest three different gains, namely a beamforming gain, a frequency diversity gain and a gain related to the rectifier nonlinearity, and therefore offers additional opportunities over spatial domain processing/beamforming-only or over frequency domain waveform-only to boost $e_2 \times e_3$.

Though the optimal design of joint waveform and beamforming results from the solution of an optimization problem [64], [67], a simple combination of the MRT beamforming and the low-complexity SMF waveform was demonstrated experimentally in [74] to significantly boost the output DC power and the range of WPT, as illustrated in Fig. 11 and Fig. 12. We observe that WPT performance gains can be obtained by exploiting either the frequency domain, the spatial domain, or both domains jointly. Besides, the 8-antenna 1-tone waveform shows a similar performance to that of the 4-antenna 8-tone waveform. In the same manner, 4-antenna single-tone and 2-antenna 8-tone, and 2-antenna single-tone and 1-antenna 8-tone show similar performance. Such behavior demonstrates that one can trade the spatial domain (number of antennas) processing with the frequency domain (number of tones) processing and inversely, and the gains in terms of output DC power and range can be accumulated using a joint beamforming and waveform strategy.

The accumulated gains of beamforming and waveform also applies to MIMO WPT where a joint waveform, transmit beamforming and receive combining was shown to provide significant gains over individual techniques [89]. It was shown that the joint waveform and beamforming design provides a higher output DC power than the beamforming only design with a relative gain exceeding 180% when $M = 2$, $N = 16$, and $Q = 2$. Moreover, RF combining was shown to provide a higher output DC power than DC combining with a relative gain which can be up to 550% when $M = 2$, $N = 8$, and $Q = 10$.

Similarly, the transmit waveform and active beamforming can be jointly designed together with the passive beamforming at the IRS so as to efficiently exploit frequency and spatial domain gains [87], [88]. Note that those gains were demonstrated despite the frequency flat constraints of the passive elements of the IRS (the scattering matrix $\Theta$ of the IRS is constant across frequency).

*6) Large-Scale (Massive) MIMO:* The results in [64] also highlight the potential of a *large-scale multisine multiantenna* ($M >> 1$, $N >> 1$) closed-loop WPT architecture, reminiscent of Massive MIMO with orthogonal frequency division multiplexing (OFDM) in communications. In [67], such a promising architecture was studied and shown to enhance $e$ and increase the range of WPT. It enables highly efficient WPT by jointly optimizing transmit signals over a large number of frequency components and transmit antennas, thereby combining the benefits of pencil beams (as in Massive MIMO) and waveform design to exploit the large beamforming gain of the transmit antenna array and the nonlinearity of the rectifier at long distances. The challenge is the large number of dimensions $N$ and $M$ that requires a reformulation of the optimization problem. The new design offers significantly lower complexity in signal design compared to the GP approach [67].

Interestingly in the limit of large $M$, the design of the joint multiantenna multisine waveform is simplified thanks to the channel hardening. Indeed, with $y_{\text{rf}}(t) = \sqrt{2}\Re\left\{\sum_{n=0}^{N-1} \mathbf{h}_n \mathbf{x}_n e^{j2\pi f_n t}\right\} + w_{\text{A}}(t)$, and writing $\mathbf{x}_n = s_n \bar{\mathbf{h}}_n^H$ (with $\bar{\mathbf{h}}_n = \mathbf{h}_n / \|\mathbf{h}_n\|$ and $\sum_{n=0}^{N-1} s_n^2 \leq P$), $\lim_{M\to\infty} \|\mathbf{h}_n\| / \sqrt{M} = 1$ and the channel after beamforming becomes effectively frequency flat due to channel hardening on all frequencies. In the limit of large $M$, $\{s_n\}$ is therefore simply obtained as the solution (34) over an effective frequency-flat channel ($A_n = A$). A good (close to optimum) strategy is to allocate power uniformly across frequencies, i.e. $s_n = \sqrt{P/N}$.

*7) Channel Acquisition:* The aforementioned techniques have been designed assuming perfect CSIT. In practice, the CSI should be acquired by the ET and several strategies have been proposed, including forward-link training with CSI feedback, reverse-link training via channel reciprocity, power probing with limited feedback, and channel estimation based on backscatter communications [7], [90]–[97]. The first two are similar to strategies used in modern communication systems, but incur too high energy consumption and/or too complex processing for low power nodes. The third is more promising and tailored to WPT because it is implementable with very low communication and signal processing requirements at the ER. The fourth one is also promising and is based on the idea that the ET exploits its observed backscatter signals to estimate the backscatter-channel (i.e., ET-to-ER-to-ET) state information (BS-CSI) directly instead of estimating the forward channel ET-ER as in previous three techniques. The BS-CSI is then

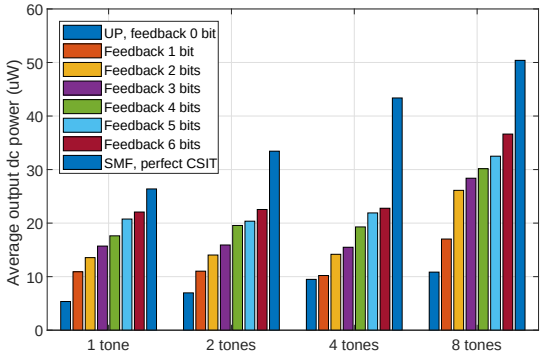



Fig. 16. Measurement results of output DC power $P_{\rm dc}^r$ of joint waveform and beamforming with limited feedback for $M = 4$, $N = 1, 2, 4, 8$ and 1 to 6 bits of feedback [98].

used in the transmit signal design.

The framework for power probing with limited feedback of [96] focuses on general setup of multi-antenna multi-carrier WPT over frequency-selective channels. It demonstrates that one can jointly exploit a beamforming gain, the channel frequency selectivity, and the EH nonlinearity through a joint waveform and beamforming based on limited feedback. To that end, it relies on the output DC power measurement and a limited number of feedback bits for the selection or the refinement of the joint waveform and beamforming. In the selection strategy, the ET transmits over multiple time slots with a different (joint waveform and beamforming) precoder within a codebook at each time slot, and the ER reports the index of the precoder in the codebook that offers the largest $P_{\rm dc}^r$. In the refinement strategy, the ET sequentially transmits using two precoders in each stage, and the ER reports one feedback bit, indicating an increase or a decrease in $P_{\rm dc}^r$ during this stage. Based on multiple one-bit feedback, the ET successively refines precoders (across space and frequency) in a tree-structured codebook over multiple stages. The optimization of the codebook of joint waveform and beamformers is pretty challenging and employs the framework of the generalized Lloyd's algorithm. Fig. 16 illustrates the experimental results obtained with such a strategy for 1 to 6 bits of feedback with $M = 4$ and $N = 1, 2, 4, 8$ [98]. We note the significant increase in $P_{\rm dc}^r$ as $N$ increases with perfect CSIT, and the need for larger codebook sizes to come closer to the perfect CSIT performance.

*8) Transmit Diversity:* Another WPT signal strategy, denoted as transmit diversity [58], relies on $M$ dumb transmit antennas to induce fast fluctuations of the wireless channel through a simple phase sweeping method consisting of the transmission of a signal $x(t)$ on each antenna with an antenna dependent time-varying phase $\psi_m(t)$, namely $x_{{\rm in},m}(t) = \sqrt{2}\Re\left\{x(t)e^{j(2\pi ft+\psi_m(t))}\right\}$. Those fluctuations are shown to boost $P_{\rm dc}^r$. This is another consequence of EH nonlinearity in Observation 1, namely that fading and fast fluctuations of the wireless channel do not increase $P_{\rm rf}^r$ but increase $\mathbb{E}\left[y_{\rm rf}(t)^4\right]$ and therefore $e_3$. In contrast to the beamforming strategies, transmit diversity does not rely on any form of CSIT. Interestingly, this highlights that multiple transmit antennas are useful in WPT even in the absence of CSIT. In [58], [73] real-time over-the-air measured gains of 50% to 100% were demonstrated with a two-antenna transmit diversity strategy over single-antenna setup without any need for CSIT.

Transmit diversity has a number of practical benefits leading to low cost deployments, namely the use of dumb antennas fed with a low PAPR CW (hence making a better use of $e_1$), no need for synchronization among transmit antennas, applicable to co-located and distributed antenna deployments, transparent to the ERs (which eases the system implementation), applicable to deployments with a massive number of devices (massive IoT deployments) for which CSIT acquisition is unpractical. Another related CSI-free multiantenna techniques for WPT has been proposed in [99].

*9) Other Techniques:* Another technique that can be seen as an alternative to multi-antenna beamforming to enable directional/energy focusing transmission for WPT, is *time-reversal* [100], [101]. With time-reversal, the multipaths in the wireless channel are used as virtual antennas to enable spatial-temporal focusing effect and enhance $e_2$. Upon acquiring the channel impulse response, the transmitter sends a time-reversed conjugate waveform, using the principle of match filtering, in order to leverage the multipath channel and focus the signal power at the receiver input. Time-reversal can be applied to a single-antenna or multi-antenna transmitters and requires large bandwidth in order to distinguish as many paths in the channel as possible. Note that time reversal waveform design is different from aforementioned waveform design and is not rooted in the EH nonlinearity. It may be promising to investigate how those two types of waveforms can be designed in a unified manner.

Another low complexity (without the need for sophisticated digital signal processing) alternative to enable beamforming gain in multi-antenna settings is by using *retrodirective arrays*. Upon receiving a signal from any direction, retrodirective arrays exploit channel reciprocity to transmit a signal response, in the form of a phase-conjugated version of the received signal, back to the same direction without the need of knowing the source direction or performing explicit channel estimation/feedback [102], [103]. Two well known retrodirective array structures are Van Atta arrays and the heterodyne retrodirective arrays with phase-conjugating circuits. WPT using retrodirective techniques have been studied in [104] and experimentally demonstrated in different setups [105]–[107]. It would be interesting to explore how waveform and retrodirective array could be jointly designed so as to exploit the beamforming and waveform gains.

### *F. Signal Processing Techniques for Multi-User WPT*

WPT is not limited to a single ET and ER. In a multiuser WPT setting with one ET and $K$ ERs, with each ER having one rectenna (Fig. 2), the output DC power $P_{{\rm dc},k}^r$ at a given rectenna $k$ depends on $P_{{\rm dc},j}^r$ at another rectenna $j \neq k$; i.e., a given signal, e.g., waveform or beamformer, may be suitable for one rectenna but inefficient for another. Therefore, a tradeoff exists between the output DC power of the different

rectennas. The energy region formulates this tradeoff by expressing the set of output DC power at all rectennas that can be achieved simultaneously, which is written mathematically as a weighted sum of output DC power as

$$\max_{p(\mathbf{x}_0,\ldots,\mathbf{x}_{N-1})} \quad \sum_{k=1}^{K} v_k P^r_{\mathrm{dc},k}(\mathbf{x}_0,\ldots,\mathbf{x}_{N-1}) \tag{36}$$

$$\text{subject to} \quad P^t_{\mathrm{rf}} \leq P, \tag{37}$$

where, by changing the weights $v_k$, we can operate on a different point of the energy region boundary and therefore favor one user over another one. An alternative problem formulation can be written as a maximization of minimum energy among all $K$ users

$$\max_{p(\mathbf{x}_0,\ldots,\mathbf{x}_{N-1})} \min_{k=1,\ldots,K} \quad P^r_{\mathrm{dc},k}(\mathbf{x}_0,\ldots,\mathbf{x}_{N-1}) \tag{38}$$

$$\text{subject to} \quad P^t_{\mathrm{rf}} \leq P. \tag{39}$$

Those two problems have been studied in [67].

All aforementioned single-user techniques can be designed for multiuser WPT. Among the most advanced ones leveraging combined frequency-domain and spatial-domain gains, we note the joint beamforming and waveform for multiuser WPT of [64], [67], and the joint passive beamforming and waveform design for multiuser IRS-aided WPT of [88]. Fig. 17 illustrates the energy region for a two user MISO WPT scenario with a joint beamforming and waveform spanning twenty transmit antennas ($M = 20$) and ten frequencies ($N = 10$), obtained by solving problem (36) [67]. The challenge is that solving (36) in this two user setup results in a coupled optimization of the frequency and the spatial domains. Indeed, while decoupling the spatial and frequency domains by first designing the spatial beamformer in each frequency and then designing the power allocation across subbands is optimal in the single-user/rectenna case, it is clearly suboptimal in the multi-user/rectenna case [67]. The key takeaway here is that, by optimizing the waveform to jointly transfer power to multiple users simultaneously, we obtain an energy region ("weighted sum") that is larger than that achieved by a time-sharing approach, such as time-division multiple access (TDMA), where the transmit waveform and beamforming is optimized for a single user at a time and each user is scheduled to receive energy during a fraction of the time. Other multi-user waveform designs have appeared in [108], [109].

Considering an entire network consisting of many ETs and ERs (Fig. 2), [7] defines various network architectures. All ETs can cooperate jointly to design the transmit signals for multiple ERs (in the form of a coordinated multipoint (CoMP)-based WPT) or locally coordinate their efforts such that a given ER is served by a subset of ETs (or, in the simplest scenario, where each ER is served by a single ET). As a consequence, different resource allocation strategies (centralized versus distributed) in terms of CSI sharing and acquisition at the different ETs can be considered.

In [7], it was shown that distributing antennas (DAS) across a coverage area (as in Fig. 2) and enabling cooperation among them distributes energy more evenly in space and, therefore, enhances the ubiquitous accessibility of wireless power as compared to a co-located deployment. Strong energy beams in the direction of users are also avoided, which is desirable from a safety perspective.

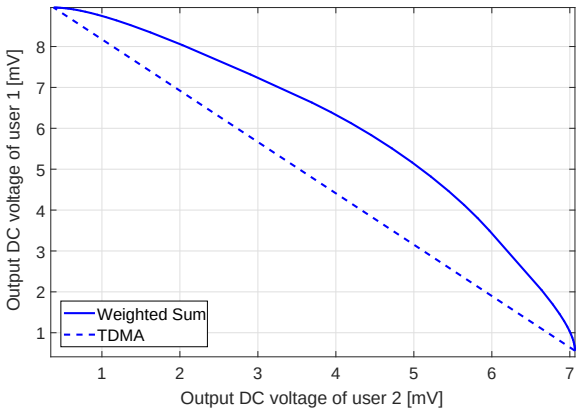



Fig. 17. The two-user energy region with joint beamforming and waveform ($M$ = 20 and $N$ = 10) [67].

Recent experimental results of WPT architecture based on DAS in [110] show that that WPT DAS can boost $P^r_{\mathrm{dc}}$ by up to 30 dB in a single-user setting and the sum of output DC power by up to 21.8 dB in a two-user setting and broaden the service coverage area in a low cost, low complexity, and flexible manner by suitably and dynamically selecting transmit antenna and frequency. Other DAS WPT studies have been reported in [111]–[116].

## III. Wireless Information and Power Transfer: Achieving the Best Rate-Energy Tradeoff

Building upon WPT signal design and processing techniques, we can study how to integrate communications and power into WIPT. The objective is here to achieve the best R-E tradeoff.

### A. Signal and System Model

We focus on a single-user (point-to-point) multi-subband MIMO SWIPT system (referred to as "SWIPT with co-located receivers" in Fig. 3). The system model is the same as in Section II-A though $x_{m,n,k}$ now denotes the complex-valued information and power carrying symbol, instead of just a power carrying symbol, since both information and power are transmitted simultaneously. The information is captured in $S$ possible messages $s \in \mathcal{M} = \{1, 2, ..., S\}$, where $\mathcal{M}$ denotes the set of messages. The mapping from $\mathcal{M}$ to the sequence of complex-valued transmitted information and power carrying symbols $x_{m,n,k}$ is denoted by $g_{\theta_{\mathrm{T}}}$, where $\theta_{\mathrm{T}}$ refers to the set of transmitter design parameters.

The processing then depends on the exact SWIPT receiver architecture. One commonality nevertheless exists among all considered types of receivers. Namely, from an energy perspective, $y_{\mathrm{rf},q}(t)$ (or a fraction of it) is conveyed to an ER, where energy is harvested directly from the RF-domain signal. From an information perspective, an IR downconverts $y_{\mathrm{rf},q}(t)$

(or a fraction of it) and filters it to produce the baseband signal for subband $n$

$$y_{q,n}(t) = \mathbf{h}_{q,n}\mathbf{x}_n(t) + w_{q,n}(t), \tag{40}$$

where $w_{q,n}(t)$ is the downconverted received filtered noise and accounts for both the antenna and the RF-to-baseband processing noise. After sampling at a frequency $f_{\rm w}$, the sampled outputs at time instants $k$ (multiples of the sampling period) can be expressed as

$$y_{q,n,k} = \mathbf{h}_{q,n}\mathbf{x}_{n,k} + w_{q,n,k} \tag{41}$$

with $\mathbf{x}_{n,k} \triangleq \left[x_{1,n,k}, ..., x_{M,n,k}\right]^T$. Following the i.i.d. channel inputs and the discrete memoryless channel assumptions, we drop the time index $k$ such that

$$y_{q,n} = \mathbf{h}_{q,n}\mathbf{x}_n + w_{q,n}. \tag{42}$$

We model $w_{q,n}$ as an i.i.d. and CSCG random variable with variance $\sigma^2$, i.e., $w_{q,n} \sim \mathcal{CN}(0, \sigma^2)$, where $\sigma^2 = \sigma_A^2 + \sigma_P^2$ is the total Additive White Gaussian Noise (AWGN) power originating from the antenna ($\sigma_A^2$) and the RF-to-baseband processing ($\sigma_P^2$).

The observations from all receive antennas can then be stacked to obtain

$$\mathbf{y}_n = \mathbf{H}_n\mathbf{x}_n + \mathbf{w}_n, \tag{43}$$

where $\mathbf{y}_n \triangleq \left[y_{1,n}, ..., y_{Q,n}\right]^T$, $\mathbf{w}_n \triangleq \left[w_{1,n}, ..., w_{Q,n}\right]^T$.

The estimated message $\hat{s}$ is then produced by the information decoder (ID) which maps the received noisy sequence $\mathbf{y}_n$ to $\hat{s} \in \mathcal{M}$ through a parametric function denoted by $h_{\theta_{\rm R}}$, where $\theta_{\rm R}$ refers to the set of receiver design parameters at the ID.

Finally, we assume perfect Channel State Information at the Transmitter (CSIT) and perfect Channel State Information at the Receiver (CSIR).

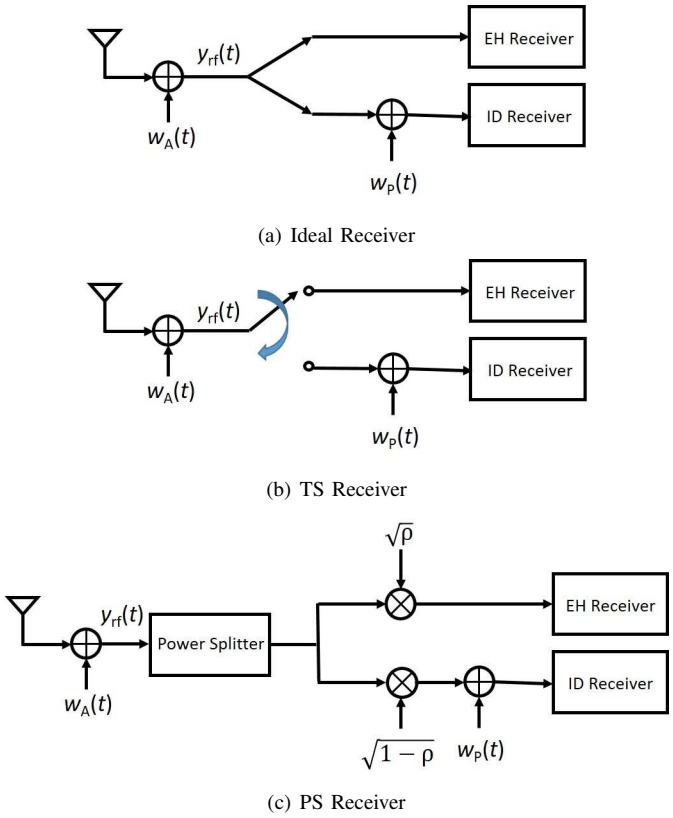
### *B. Receiver Architectures*

Several receiver architectures for SWIPT have been proposed in Fig. 3.

An *Ideal Receiver* (Fig. 18(a)) assumes the same signal $y_{{\rm rf},q}(t)$ is conveyed to the EH and also simultaneously RF-to-baseband downconverted and conveyed to the ID [12], [13]; however, no practical circuits can currently realize this operation. Different R-E tradeoffs could be realized by varying the design of the transmit signals to favor rate or energy.

A *Time Switching (TS) Receiver* (Fig. 18(b)) consists of a conventional ID and an EH (following the structure in Section II-C) that are co-located [14], [17], [19]. The transmission block is divided into two orthogonal time slots, one for power transfer and the other for data transfer. In each time slot, the transmit waveforms are optimized for either WPT or WIT. Accordingly, the receiver switches its operation periodically between harvesting energy and decoding information in the two time slots. Then, different R-E tradeoffs are realized by varying the length of the WPT time slot, jointly with the transmit signals [117].

In a *Power Splitting (PS) Receiver* (Fig. 18(c)), the EH and ID receiver components are the same as those of a TS receiver.




(a) Ideal Receiver

(b) TS Receiver

(c) PS Receiver

Fig. 18. Three single-antenna receiver architectures for SWIPT: (a) Ideal receiver (using the same signal for both the ID and EH receivers); (b) TS receiver (switching the signal to either ID or EH receiver); and (c) PS receiver (splitting a portion of the signal to ID receiver and the rest to EH receiver) [9].

The transmitted signals are jointly optimized for information and energy transfer and the received signal is split into two streams, where one stream with PS ratio $0 \leq \rho \leq 1$ is exploited for EH, and the other with power ratio $1-\rho$ is utilized for ID [14], [17], [18]. Hence, assuming perfect matching (as in Section II-C), the input voltage signals $\sqrt{\rho R_{\rm ant}}y_{\rm rf}(t)$ and $\sqrt{(1-\rho)R_{\rm ant}}y_{\rm rf}(t)$ are respectively conveyed to the EH and the ID. Different R-E tradeoffs are realized by adjusting the value of $\rho$ jointly with the transmit signals.

It is common to assume that the power of the processing noise is much larger than that of the antenna noise, i.e., $\sigma_P^2 \gg \sigma_A^2$, such that $\sigma^2 = \sigma_A^2 + \sigma_P^2 \approx \sigma_P^2$. As explained in [14], the above setting results in the worst-case R-E region for the practical PS receiver.

### *C. Rate-Energy Region and Problem Formulation*

The R-E region $\mathcal{C}_{\rm R-E}$ is defined as the set of all pairs of rate $R$ and energy $E$ such that simultaneously the receiver can communicate at rate $R$ and harvested energy $E$. The R-E region in general is obtained through a collection of input distributions $p(\mathbf{x}_0, ..., \mathbf{x}_{N-1})$ that satisfies the average transmit

power constraint $P_{\rm rf}^t \leq P$. Mathematically, we can write

$$\mathcal{C}_{\rm R-E}(P) \triangleq \bigcup_{\substack{p(\mathbf{x}_0,\ldots,\mathbf{x}_{N-1}):\\ P_{\rm rf}^t \leq P}} \left\{ (R,E) : R \leq \sum_{n=0}^{N-1} I\left(\mathbf{x}_n, \mathbf{y}_n\right), \right. \\ \left. E \leq P_{\rm dc}^r\left(\mathbf{x}_0, ..., \mathbf{x}_{N-1}\right) \right\} \quad (44)$$

where $I\left(\mathbf{x}_n, \mathbf{y}_n\right)$ refers to the mutual information between the channel input $\mathbf{x}_n$ and the channel output $\mathbf{y}_n$ on subband $n$ and $P_{\rm dc}^r$ is a nonlinear function of $\mathbf{x}_0, ..., \mathbf{x}_{N-1}$. For multi-antenna harvesters based on RF combining, the receive combiner would also have to be jointly optimized with the input distribution.

One approach to identify the R-E region is to calculate the capacity (supremization of the mutual information over all possible input distributions $p(\mathbf{x}_0, ..., \mathbf{x}_{N-1})$) of a complex AWGN channel subject to an average RF power constraint $P_{\rm rf}^t \leq P$ and a receiver delivered power constraint $P_{\rm dc}^r(\mathbf{x}_0, ..., \mathbf{x}_{N-1}) \geq \bar{E}$, for different values of $\bar{E} \geq 0$. Namely,

$$\sup_{p(\mathbf{x}_0,\ldots,\mathbf{x}_{N-1})} \quad \sum_{n=0}^{N-1} I(\mathbf{x}_n; \mathbf{y}_n) \quad (45)$$

$$\text{subject to} \quad P_{\rm rf}^t \leq P, \quad (46)$$

$$P_{\rm dc}^r(\mathbf{x}_0, ..., \mathbf{x}_{N-1}) \geq \bar{E}, \quad (47)$$

where $\bar{E}$ is interpreted as the minimum required or target delivered power.

### *D. Signal Processing Techniques for WIPT*

Designing SWIPT[3] requires the transmit signals to carry information and therefore to be subject to some randomness and this randomness has an impact on the amount of harvested DC power. This raises interesting questions on how modulated signals perform in comparison to deterministic signals for WPT, and consequently on how to design modulation, waveform, coding and multi-antenna processing for SWIPT. Due to space limitations, emphasis is put on key single-user (point-to-point) techniques, though they can be extended to multi-user settings. Readers are invited to consult [9] for more discussions on multi-user WIPT.

*1) Modulation and Input Distribution:* Let us first assume a SISO ($M = Q = 1$) single-subband ($N = 1$) transmission with a linear HPA (with $e_1 = 1$) and the ideal receiver. The system model in (42) simplifies to $y = hx + w$. Problem (45)-(47) becomes

$$\sup_{p(x)} \quad I(x; y) \quad (48)$$

$$\text{subject to} \quad \mathbb{E}\left[|x|^2\right] \leq P, \quad (49)$$

$$P_{\rm dc}^r(x) \geq \bar{E}. \quad (50)$$

From (13) and Observation 1, $P_{\rm dc}^r(x)$ is not only a function of $\mathbb{E}\left[|x|^2\right]$ but also a function of $\mathbb{E}\left[|x|^4\right]$ and higher order moments [66], [118].

[3]Even though emphasis is put on SWIPT in this section, the analysis and ideas reviewed in the paper also find applications in WPCN and WPBC.

Interestingly, the higher moments of the input distribution have a significant impact on the selection of the input distribution $p(x)$. It was shown in [66] that modulation using CSCG inputs leads to a higher $P_{\rm dc}^r(x)$ compared to an unmodulated input, despite presenting the same average power $P_{\rm rf}^r$ at the rectenna input. This gain originates from the large higher order ($> 2$) moment of CSCG inputs, which is leveraged by the rectifier nonlinearity and modeled by the higher order terms in (16). Indeed CSCG inputs $x \sim \mathcal{CN}(0, P)$ have $\mathbb{E}\left[|x|^4\right] = 2P^2$ while unmodulated CW inputs with the same average RF power only achieve $\mathbb{E}\left[|x|^4\right] = P^2$ [66].

Even larger gains can be obtained using asymmetric Gaussian inputs [118] and on-off keying (or also called flash signaling) [119]. Indeed, real Gaussian modulation outperforms CSCG modulation despite the same average input power $P_{\rm rf}^r$ to the rectifier. In [118], assuming general non-zero mean Gaussian inputs $\Re\{x\} \sim \mathcal{N}(\mu_r, P_r)$ and $\Im\{x\} \sim \mathcal{N}(\mu_i, P_i)$ with $P_r + P_i \leq P$, it is found that zero mean asymmetric Gaussian inputs with $P_r + P_i = P$ achieve the supremum in Problem (48)-(50). CSCG input obtained by equally distributing power between the real and the imaginary dimensions, i.e., $\Re\{x\} \sim \mathcal{N}(0, P/2)$ and $\Im\{x\} \sim \mathcal{N}(0, P/2)$ is optimal for rate maximization. However, as $\bar{E}$ increases, the input distribution becomes asymmetric with $P_r$ increasing and $P_i = P - P_r$ decreasing (or inversely) till the rate is minimized and the energy is maximized by allocating the transmit power to only one dimension, e.g. $\Re\{x\} \sim \mathcal{N}(0, P)$. This is because a higher fourth moment is obtained by allocating all power to one dimension. Indeed, $\mathbb{E}\left[x^4\right] = 3P^2$ for $x \sim \mathcal{N}(0, P)$ in contrast to $\mathbb{E}\left[|x|^4\right] = 2P^2$ with $x \sim \mathcal{CN}(0, P)$.

In Fig. 19, the information rate $I(x; y)$ and (normalized) output DC power $P_{\rm dc}^r(x)$ for complex Gaussian inputs is shown versus $P_r$, assuming an ideal receiver. It is observed that the information rate and output DC power are indeed maximized and minimized, respectively, for $P_i = P_r = P/2$. Alternatively the information rate and output DC power are minimized and maximized, respectively when $P_i = 0$, $P_r = P$ or $P_i = P$, $P_r = 0$. This shows that there is a fundamental R-E tradeoff even in the simplest SISO AWGN scenario. It is important to recall that the tradeoff is induced by the presence of the fourth and higher moments of the received signal $y_{\rm rf}(t)$ in $v_{\rm out}$. Had we accounted only for the second term in $v_{\rm out}$, $P_{\rm dc}^r(x)$ in Fig. 19 would have been replaced by a flat curve. Ignoring the nonlinearity brought by the higher order terms, there is no tradeoff between R-E and $P_i = P_r = P/2$ simultaneously maximizes the rate and energy [9], [12], [13].

Relaxing the constraints on Gaussian inputs, it is remarkably shown in [119] that the capacity of an AWGN channel under transmit average power and target delivered power constraints as characterized by Problem (48)-(50) is obtained by adopting a combination of CSCG and on-off-keying inputs. Let $E_{\rm G}$ denote the output DC power with the input $x \sim \mathcal{CN}(0, P)$. For $\bar{E} \leq E_{\rm G}$, the capacity is achieved via the unique input $x \sim \mathcal{CN}(0, P)$. For $\bar{E} > E_{\rm G}$, the capacity is approached by using time sharing between CSCG distribution and on-off keying, reminiscent of flash signaling, exhibiting a low probability of high amplitude signals. Such a combination of CSCG and on-off-keying inputs achieves a larger R-E region

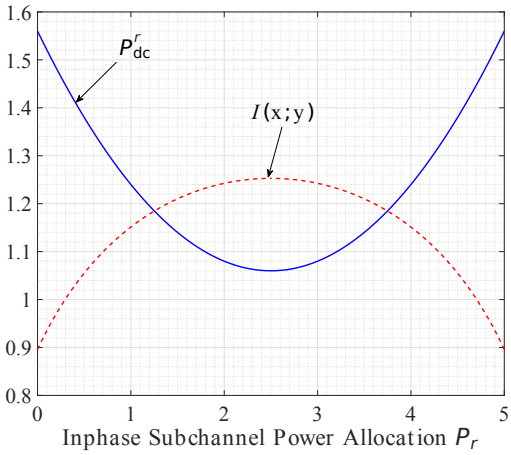



Fig. 19. Mutual information $I(x;y)$ (Red dashed line) and (normalized) output DC power $P^r_{\rm dc}(x)$ (blue solid line) corresponding to the complex Gaussian inputs with asymmetric power allocation [119]. The transmitted information rate is maximized for $P_i = P_r = P/2$ and delivered power is maximized when $P_i = 0$, $P_r = P$ or $P_i = P$, $P_r = 0$.

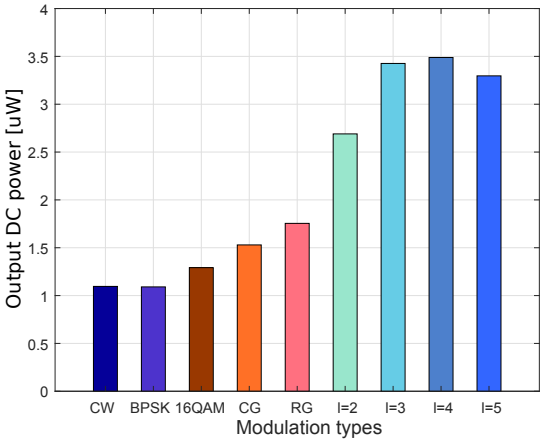



Fig. 20. Output DC power ($P^r_{\rm dc}$) of conventional and WPT-optimized modulations [74]. CG refers to CSCG, and RG to real Gaussian. $l$=1,2,3,4,5 refer to on-off keying modulations.

than asymmetric Gaussian inputs. Writing the complex input as $x = re^{j\theta}$ with its phase $\theta$ uniformly distributed over $[0, 2\pi)$, such a on-off keying distribution is given by the following probability mass function

$$p_r(r) = \begin{cases} 1 - \frac{1}{l^2}, & r = 0, \\ \frac{1}{l^2}, & r = l\sqrt{P}, \end{cases} \tag{51}$$

with $l \geq 1$. Such a distribution has indeed a low probability of high amplitude signals since $r = l\sqrt{P}$ increases and $p_r(l\sqrt{P})$ decreases as $l$ increases. We note that $\mathbb{E}\left[|x|^4\right] = l^2P^2$ and $\mathbb{E}\left[|x|^2\right] = \mathbb{E}\left[r^2\right] = P$, hence achieving large higher moments as $l$ increases while satisfying the average RF power constraint. Choosing $l > \sqrt{3}$ makes the fourth order moment higher than the $2P^2$ and $3P^2$ obtained respectively with real Gaussian and CSCG inputs. Here again, the benefits of departing from Gaussian inputs originate from Observation 1 that favors the use of distributions with large higher moments of the channel input $x$.

On-off keying has been shown to significantly boost $e_3$ conversion efficiency over various baselines [74], [119]. In Fig. 20, we display circuit simulation results of various modulations using the rectifier of Fig. 6. We note that on-off keying ($l$=1,2,3,4,5) provides three times (i.e. gain of over 200%) more output DC power ($P^r_{\rm dc}$) than conventional communication modulations such BPSK and 16-QAM despite having the same average RF input power ($P^r_{\rm rf}$).

The optimal distribution resulting from the use of on-off keying discussed so far assumed a linear HPA. In the presence of nonlinear HPA response as in Fig. 5, high amplitude signals would be distorted and the optimal distribution would need to be re-assessed. One solution is to introduce an additional amplitude constraint in Problem (48)-(50). It was shown in [119] that under average transmit power, amplitude, and (nonlinear) delivered output DC power constraints, the optimal capacity achieving distributions are discrete with a finite number of mass points for the amplitude and continuous uniform for the phase (see also [2]). Other SWIPT design to account for HPA nonlinearities have been discussed in [120], [121] and is further discussed in Section IV-B1.

Though the above discussion assumed an ideal receiver, conclusions on the modulation and input distribution are applicable to TS and PS receiver. With a TS receiver, the transmitter would transmit asymmetric Gaussian or preferably on-off keying and CSCG alternatively on the two orthogonal time slots and the receiver would switch accordingly. With a PS receiver, the asymmetry ($P_r$ and $P_i$ values) in the Gaussian input would have to be optimized jointly with the PS ratio $\rho$. Similarly a combination of CSCG and on-off keying jointly optimized with the PS ratio could be used. Given two fixed distributions, one based on CSCG and the other based on on-off keying, an ideal receiver would achieve a larger R-E region than a PS receiver, which itself has a larger R-E region than a TS receiver.

Other information theoretical studies of SWIPT with nonlinear EH models have appeared in [71], [122], [123].

*2) Waveform:* Let us now consider the SISO multi-subband transmission such that (42) becomes $y_n = h_n x_n + w_n$ in subband $n$. The capacity achieving waveform and input distribution remains an open problem. Nevertheless, interesting results are known assuming Gaussian inputs. Assuming a linear HPA (with $e_1 = 1$) and a SISO multi-carrier waveform with a general non-zero mean Gaussian inputs $\Re\{x_n\} \sim \mathcal{N}(\mu_{nr}, P_{nr} - \mu^2_{nr})$ and $\Im\{x_n\} \sim \mathcal{N}(\mu_{ni}, P_{ni} - \mu^2_{ni})$ on each carrier/subband $n = 0, \ldots, N-1$, with $P_n = P_{nr} + P_{ni}$ and $\sum_{n=0}^{N-1} P_n \leq P$, problem (45)-(47) becomes

$$\sup_{\{\mu_{nr},\mu_{ni},P_{nr},P_{ni}\}_{n=0}^{N-1}} \quad \sum_{n=0}^{N-1} I(x_n; y_n) \tag{52}$$

$$\text{subject to} \quad \text{Tr}\left(\mathbf{Q}\right) \leq P, \tag{53}$$

$$P^r_{\rm dc}(\mathbf{x}_0, ..., \mathbf{x}_{N-1}) \geq \bar{E}. \tag{54}$$

In [66], [124], problem (52)-(54) was investigated for such Gaussian inputs. It was shown that, while single-subband favors asymmetric inputs with a zero mean as described in previous section, multi-subband favors non-zero mean and

asymmetric inputs.

In Fig. 21, the R-E regions for *Asymmetric Non-zero mean Gaussian* (ANG), *Symmetric Non-zero mean Gaussian* (SNG) and *Zero mean Gaussian* (ZG) are illustrated for $N = 9$ over a frequency selective channel for $n_o = 4$ in (13) [124]. We also compare with the R-E region corresponding to the optimal power allocations under the linear model assumption with $n_o = 2$, denoted by *Zero mean Gaussian for Linear model* (ZGL). As it is observed in Fig. 21, due to the asymmetric power allocation in ANG, there is an improvement in the R-E region compared to SNG. This gain is reminiscent of the gain observed for single-carrier modulation. Additionally, it is observed that ANG and SNG achieve larger R-E regions compared to optimized ZG and that ANG and SNG perform better than ZGL. This highlights the fact that under EH nonlinearity ($n_o > 2$), ZGL is far from optimal and the R-E region enhancement offered by ANG over ZGL in Fig. 21 illustrates the gain obtained by accounting for the EH nonlinearity in SWIPT signal and system design.

The reason why ANG and SNG lead to larger R-E regions is due to the fact that the fourth order term in (13) (and therefore the output DC power) is boosted by allowing the mean of the channel inputs to be non-zero [66], [124]. Hence, in contrast to ZGL, we note that the EH nonlinearity impacts not only the power allocation strategy across subbands but also the input distribution in each subband.

The superiority of non-zero mean inputs over zero mean inputs can be qualitatively explained by the fact that a multi-carrier unmodulated waveform, e.g. multisine, is more efficient in exploiting the EH nonlinearity and therefore boosting $P_{\text{dc}}^r$ compared to a multi-carrier modulated waveform with CSCG inputs. From analysis and circuit simulations in [64], [66], $P_{\text{dc}}^r$ was shown to scale linearly with $N$ for an unmodulated multisine waveform. This originates from all the carriers being in phase, which turns on the rectifier diode (and therefore boosts its sensitivity) in the low power level in a periodic manner by sending high energy pulses every $1/\Delta_f$. On the other hand, $P_{\text{dc}}^r$ scales at most logarithmically with $N$ for a modulated waveform due to the independent CSCG randomness (and therefore random fluctuations of the amplitudes and phases) of the information-carrying symbols across subbands.

In Fig. 22, from (a) to (d), the optimized inputs in terms of their complex mean $\mu_n,\ l = 0, \ldots, 8$ (represented as dots) and their corresponding variances $\sigma_{nr}^2, \sigma_{ni}^2,\ n = 0, \ldots, 8$ (represented as ellipses) are shown for points $A,\ B,\ C$ and $E$ in Fig. 21, respectively. Point $A$ represents the maximum output DC power. Hence the waveform obtained in Point $A$ corresponds to the optimal deterministic multisine WPT waveform of (34) [64]. Point $B$ represents the performance of a typical input used for power and information transfer. Point $C$ represents the performance of an input obtained via the conventional water-filling strategy in multi-subband communications (the delivered power constraint is inactive). These three plots show that as we move from point $A$ to point $C$, the means of the different inputs decrease. Also, as we move to point $C$, the means get to zero with their variances increasing asymmetrically until the power allocation reaches the water-filling solution (where the power allocation between the real and imaginary components are symmetric).

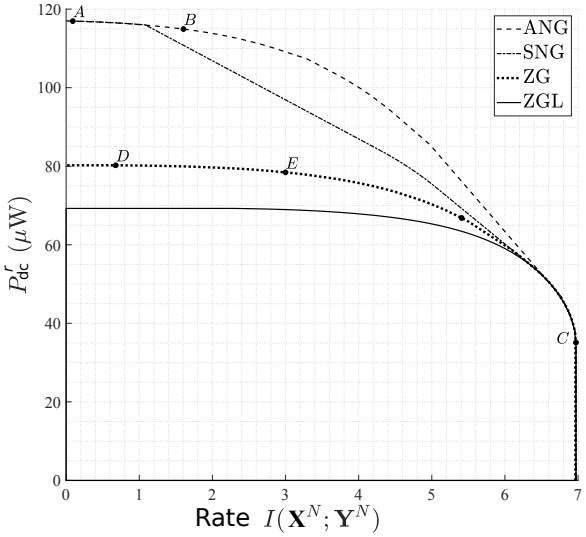



Fig. 21. The optimized R-E regions corresponding to ANG, SNG, ZG and ZGL with an average power constraint $P = 100$ $\mu$W and SNR 20 dB for an ideal receiver [124].

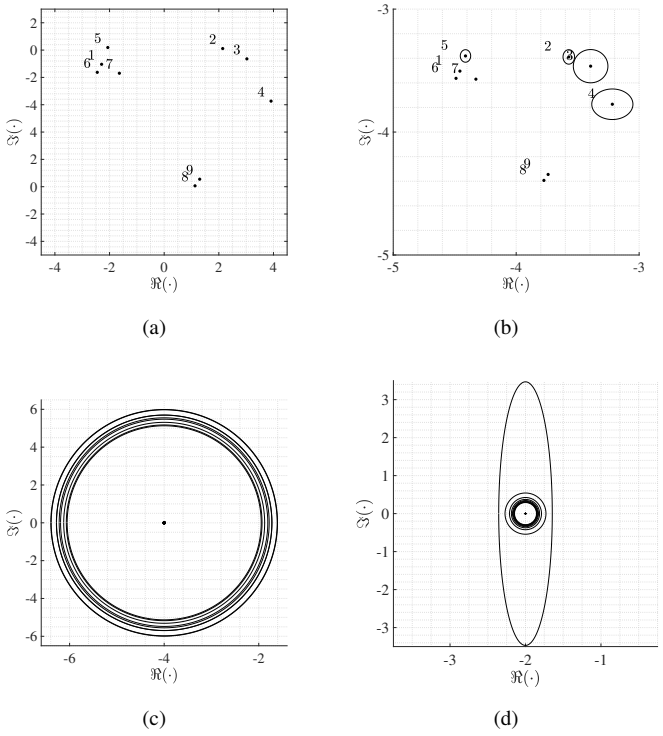



Fig. 22. From (a) to (d), the mean and variance of different inputs corresponding to the points $A$, $B$, $C$ and $E$, respectively [124].

Point $D$ in Fig. 21 corresponds to the input where all of the subbands other than the strongest one (in terms of the $\max_{n=0,\ldots,N-1} |h_n|^2$) are allocated zero power. All the transmit power is allocated to the strongest subband in an asymmetric manner to either real or imaginary component of the input [118], [124]. Note that this is different from the power allocation under the EH linear model assumption (i.e. ZGL),

for which all the transmit power is allocated to the strongest subband but equally divided between the real and imaginary components of the input. Fig. 22 illustrates the variances of the inputs on the different subbands corresponding to the point $E$ in Fig. 21. As we move from point $D$ to point $C$ (increasing the information demand at the receiver) in Fig. 21, the variance of the strongest subband varies asymmetrically (in its real and imaginary components).

SWIPT with non-zero mean Gaussian inputs translates into an architecture based on the superposition of two waveforms at the transmitter: a power waveform comprising a deterministic multisine (as in section II-E4) and an information waveform comprising multi-carrier modulated (with CSCG or asymmetric Gaussian inputs) waveforms. The complex-valued information-power symbol on subband $n$ can then be explicitly written as $x_n = x_{\mathrm{P},n} + x_{\mathrm{I},n}$ with $x_{\mathrm{P},n} = \mu_{nr} + j\mu_{ni}$ representing the deterministic power symbol of the multisine waveform on subband $n$ and $x_{\mathrm{I},n} \sim \mathcal{N}(\mu_{nr}, P_{nr} - \mu_{nr}^2) + j\mathcal{N}(\mu_{ni}, P_{ni} - \mu_{ni}^2)$ representing the zero-mean Gaussian distributed information symbol of the modulated waveform on subband $n$.

Since $x_{\mathrm{P},n}$ is deterministic, the receiver could operate with and without waveform cancellation. In the former case, the contribution of the power waveform is subtracted from the received signal after down-conversion from RF-to-baseband (BB) and Analog-to-Digital Conversion (ADC). In the latter case, a power waveform cancellation operation is not needed and the baseband receiver decodes the translated version of the symbols.

The above waveform and input distribution discussion, though introduced for the ideal receiver in Fig. 18, also holds for TS and PS receivers. Nevertheless, one interesting implication of the non-zero mean Gaussian inputs is that TS can outperform PS in multi-subband transmission. It is indeed shown in [66] that for a sufficiently large $N$ (e.g. $N = 16$), TS is preferred at low SNR and PS at high SNR, but in general the largest convex hull is obtained by a combination of PS and TS. This contrasts with the single-subband case where PS outperforms TS.

The above discussion relies on Gaussian inputs. Leveraging the above observations on input distribution, the design of of multicarrier waveform with finite constellation under EH nonlinearity was studied in [125]. The authors adapted PSK modulation to SWIPT requirements and showed the benefits of asymmetric PSK modulation to enable a larger R-E region compared to that obtained with conventional symmetric PSK constellations. Another SWIPT architecture that leverages the waveform design can be found in [126].

*3) Coding:* As we have indicated, there is much insight into information-theoretic, communication-theoretic, and system design aspects of WIPT. Yet, there has been limited study of practical codes for the WIPT problem: practical codes that approach or achieve R-E limits have only recently been developed [127]. A coding theory viewpoint on simultaneously transmitting information and energy is important in practice.

Constrained codes have been developed for the WIPT problem [128]–[130], especially in the noiseless setting. When requiring smooth energy delivery, e.g., due to finite-sized battery at the receiver, run-length limited (RLL) codes may not be best for WIPT [131], but subblock energy-constrained codes (SECCs) [132], [133] and skip-sliding window codes [134], [135] may be better. Initial analyses of constrained code performance over noisy channels indicate they do not achieve information-theoretic limits.

In establishing fundamental limits of WIPT, notice that a certain optimal distribution over the transmitted symbols is required for the receiver to extract a given amount of expected energy. Further, in a number of important cases, this distribution on the signaling constellation is not uniform [2], [12], [119]. Non-uniform input distribution requirements for WIPT rules out linear codes since linear codes can achieve Shannon capacity only when the optimal input distribution is uniform over the input alphabet. One might consider nonlinear algebraic codes like Kerdock codes and Preparata codes [136], but there is limited understanding of their performance for asymmetric input distributions and would likely not achieve capacity.

WIPT was also explored using the concatenation of an inner nonlinear trellis code with an outer LDPC code, showing performance $\sim 0.8$ dB away from the capacity of an AWGN channel [137], but did not achieve the R-E limit.

To achieve information-theoretic R-E limits using practical codes, [127] built on polar codes for asymmetric channels [138], [139], and proposed two polar coding techniques that achieve R-E limits: one technique involved concatenating nonlinear mappings with linear polar codes, whereas the other involves using randomized rounding [127]. The work focused on discrete memoryless channels and the AWGN channel with peak output power constraint, but investigating optimal polar codes for realistic nonlinear WPT system models is of interest.

*4) Multi-Antenna and Intelligent Reflecting Surface:* In a MISO setup $y_n = \mathbf{h}_n \mathbf{x}_n + w_n$, it can be shown for a general multi-band transmission that MRT in each subband is optimal [66]. Hence, the optimal input symbol vector can be written as $\mathbf{x}_n = \bar{\mathbf{h}}_n^H x_n$ with $\bar{\mathbf{h}}_n = \mathbf{h}_n / \left\|\bar{\mathbf{h}}_n\right\|$ and $x_n$ designed according to the optimal input distribution/waveform of a SISO transmission.

In a multi-band IRS-aided SWIPT with linear HPA (with $e_1 = 1$), $M = 1$ and $L$ elements, $y_n = (h_{\mathrm{d}_n} + \mathbf{h}_{\mathrm{r},n} \boldsymbol{\Theta} \mathbf{h}_{\mathrm{i},n})\, x_n + w_n$, and thanks to (24)-(27), the R-E region of (52)-(54) is expanded as

$$\sup_{\boldsymbol{\Theta}, \{\mu_{nr}, \mu_{ni}, P_{nr}, P_{ni}\}_{n=0}^{N-1}} \quad \sum_{n=0}^{N-1} I(x_n; y_n) \tag{55}$$

$$\text{subject to} \quad \mathrm{Tr}\left(\mathbf{Q}\right) \leq P, \tag{56}$$

$$P_{\mathrm{dc}}^{r}(x_0, ..., x_{N-1}, \boldsymbol{\Theta}) \geq \bar{E}, \tag{57}$$

$$\boldsymbol{\Theta} = \mathrm{diag}(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \ldots, \boldsymbol{\Theta}_G), \tag{58}$$

$$\boldsymbol{\Theta}_g^H \boldsymbol{\Theta}_g = \mathbf{I}_{L_G}, \forall g, \tag{59}$$

$$\boldsymbol{\Theta}_g = \boldsymbol{\Theta}_g^T, \forall g. \tag{60}$$

This problem is addressed in [87] under the simpler case of single connected IRS (i.e. diagonal $\boldsymbol{\Theta}$) and symmetric non-zero mean Gaussian inputs.

## IV. WPT and WIPT Signal and System Design Methodologies

We now reflect on the methodology to design WPT and WIPT. Two approaches are discussed, namely the "model and optimize" approach and the "learning" approach.

### A. The "Model and Optimize" Approach

All the above communications and signal design/processing techniques for WPT and WIPT were developed following a systematic "model and optimize" approach, namely an analytical (physics-based) nonlinear model of the EH (Section II-C) was first derived and tools from signal processing, optimization, communication and information theories were then developed to design those techniques. As demonstrated in the previous sections, the "model and optimize" approach that accounts for the EH nonlinearity provides significant potential and gains towards efficient WPT/WIPT design. Importantly, all experimental results in [73], [74], [98] so far have validated the theory and the designs developed in [7], [9], [58], [64], [66]–[68], [75], [89], [96], [118], [119], [122], [124].

The EH nonlinearity has now appeared through this approach to play a crucial role and have a profound impact on WPT and WIPT signal and system designs [9]. Nonlinearity leads to a WIPT design quite different from that of conventional wireless communication, and changes the basic characteristics of the PHY and MAC layers such as input distribution and modulation, waveform, RF spectrum use, beamforming and multi-antenna, resource allocation as well as the transmitter and receiver architecture. Moreover, ignoring the nonlinearity leads to inefficient designs of WPT and WIPT [9], [64], [73], [75], [89].

The benefits of the "model and optimize" approach are its ability 1) to identify optimal solution and reliable analytic performance guarantees on the accuracy of the solution and 2) to interpret the results and get insights into the signal and system design. Indeed, the approach relies on optimization and communication/information theories to derive optimal input signal for WPT/WIPT accounting for nonlinearity, which led to further insights into implementable approaches. Nevertheless, the "model and optimize" approach also comes with two severe *limitations*.

*First*, although nonlinearity is incorporated in the harvester model of Section II-C and can capture the effect of input signal power and shape, several simplifying assumptions are made on perfect impedance matching, ideal low pass filter, diode modeling, parasitics, etc [64]. Departing from those assumptions makes the model analytically not tractable, preventing to formulate the optimization of WPT/WIPT. This is illustrated by two examples on WPT waveform design, namely 1) in the presence of imperfect matching: the formulation leads to a chicken-and-egg problem where the waveform design is a function of the impedance mismatch and the impedance mismatch is itself a function of the waveform design due to nonlinearity; 2) in the presence of a non-ideal low pass filter: the choice of the load impacts WPT signal design and one may or may not benefit from using multisine/modulated signals over CW depending on the load and input power, though pinpointing the best set of waveform, input power and load is currently challenging [140]–[142].

*Second*, even if mathematical modelling of the EH nonlinearity is feasible using approximations, finding the fundamental limits of WIPT is mathematically extremely challenging and computationally intensive, as evidenced by works [119], [124]. Generally speaking, identifying optimal inputs of nonlinear channels is known in information theory to be a complicated problem [143]. This often leads to situations where no efficient algorithms are available to solve the problem at hand, and the typical and straightforward approaches to tackle such problems are either considering linearized models or obtaining approximations and lower bounds on capacity [144].

### B. The "Learning" Approach

Despite advances in communication and optimization theories, many systems subject to nonlinear responses are unknown in terms of their optimal behaviour (e.g. capacity) and signal/system design. WPT/WIPT are instances of such systems. Their designs face problems that cannot be mathematically formulated and/or for which no efficient solutions exist. The lack of tractable mathematical models of the rectenna, and more generally of the entire WPT chain (transmitter-receiver, including HPA), as well as algorithms to solve WPT/WIPT signal/system optimizations in general settings is a bottleneck towards efficient WPT/WIPT designs. In this regard, machine learning (ML) is instrumental since they can be used to circumvent these modelling and algorithmic challenges. This calls for a "learning" approach.

*First*, let us consider WPT. By collecting data from circuit simulations and prototypes, we can investigate two different WPT design strategies. The first strategy consists in fixing the transmit waveform hypothesis set and learn the rectenna parameters (e.g. input power level, load). By considering load and input power levels as inputs to a deep neural network (NN), the output would be the best waveform taken in a predefined hypothesis class. For instance, consider two possible waveforms $\mathrm{WF}_1$ and $\mathrm{WF}_2$, and a training set consisting of inputs (power level, load) and output (the best waveform among $\mathrm{WF}_1$ and $\mathrm{WF}_2$). Supervised learning (logistic regression) enables to classify under which input power level and load, $\mathrm{WF}_1$ (resp. $\mathrm{WF}_2$) is preferred. This can be generalized to larger hypothesis classes and rectenna parameters, and tackles the limitations of the current approaches used in the RF and microwave literature [140]–[142]. The second strategy consists in fixing the rectenna parameters and learn the best transmit waveform. For a given load and input power, reinforcement learning can then be used to design waveform based on sequential feedback (measurements) from the environment (circuit and prototype), despite the presence of nonlinearity, impedance mismatch, etc. This would tackle the limitations of the "model and optimize" approach used in [7], [64], [67], [68], [96].

*Second*, let us consider WIPT. By drawing data samples from the model or from measurements, learning can be used to obtain well performing achievable strategies. For instance, a combination of supervised learning and reinforcement learning

could be used to learn efficient input distributions for general nonlinear WIPT. This would tackle the limitations of the "model and optimize" approach used in [9], [66], [72], [118], [119], [122], [124].

A "learning" approach towards WPT and WIPT will provide different perspectives on signal and system design, fundamental limits, channel estimation and feedback, low-complexity and efficient strategies. Learning communications systems [145], [146] seeks to optimize transmitter and receiver jointly without any artificially introduced block structure, and over any type of channel without the need for prior mathematical modeling and analysis. Using ML to address communications-related problems has been studied in the past [147] and has received renewed interests due to the development of deep learning (DL) software libraries [148]–[150], and specialized hardware. Several groups have in the past few years investigated DL applications in communications, e.g. interpreting a communication system as an autoencoder (when the system can be properly modeled) [145], [146], channel decoding [151]–[153], compression [154], CSI feedback [155], and optical communications [156], [157]. NNs can also be trained for RF design to learn the behaviour of passive/active components and circuits [158].

In view of the above and that large training datasets can be created from circuit simulations and experiments, learning appears well suited for WPT/WIPT. Nevertheless, learning is not a one-size-fits-all solution. Indeed, despite the potential to tackle the aforementioned limitations of WPT/WIPT, the drawback of the learning approach, in contrast with the model-and-optimize approach, is the lack of performance guarantees on the accuracy of the solutions and the lack of human-interpretability. This also calls for developing an interplay between model-and-optimize and learning approaches to get the best of both worlds.

A learning approach towards WPT and WIPT signal and system design, and an *integrated* approach that leverages the complementarity and synergy of the learning and the model-and-optimize approaches, remain largely uncharted research territories. Some early and promising results on this learning approach for WPT and WIPT design have appeared in [70], [159]–[162]. In the sequel, we further discuss some learning techniques for WPT and WIPT relying on model-based and data-based end-to-end learning and physics-based learning.

*1) Model-based and Data-based End-to-End Learning:* Considering a SISO single-subband transmission, the goal is to design the channel inputs $x$ such that the receiver demand of information and power is satisfied. Following [70], [159], [160], we consider the SWIPT system as an NN-based denoising autoencoder, where both the transmitter and receiver are implemented as NNs in order to perform the modulation and demodulation processes, respectively. A general NN-based implementation of the point-to-point system with ideal receiver is illustrated in Fig. 23. At the transmitter, the message $s \in \mathcal{M}$ is converted into a binary vector of length $\lceil \log_2(S) \rceil$ denoted by $\mathbf{s}_b$ ($\lceil x \rceil$ returns the smallest integer larger than $x$). The vector $\mathbf{s}_b$ is then processed by the NN and is converted into a codeword $x = g_{\theta_\mathrm{T}}(s)$. Accordingly, the set of transmitter design parameters $\theta_\mathrm{T}$ are related to the weights and biases

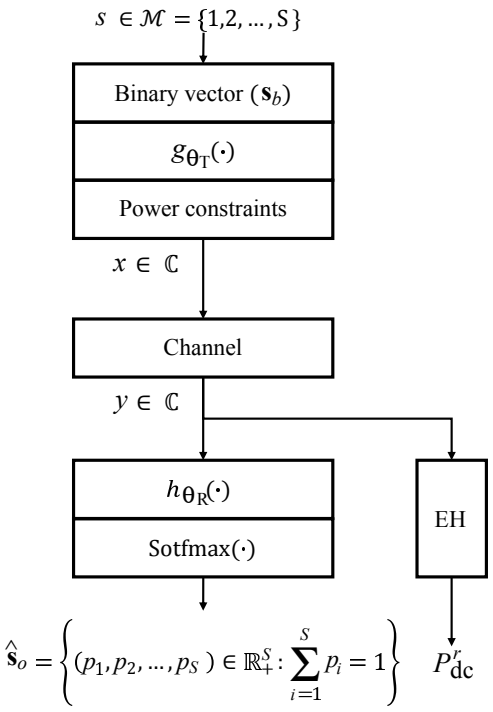



Fig. 23. NN-based implementation of the SISO single-subband SWIPT [160].

across the encoder module. To satisfy the power constraint, a power normalization is performed as the last layer of the transmitter module.

The symbol $x$ is corrupted by the channel noise. At the receiver, the estimation is performed by mapping the received noisy observation $y$ to an $S$-dimensional output probability vector denoted by $\hat{\mathbf{s}}_o$ (and estimating the message by returning the index corresponding to the maximum probability). Accordingly, $\theta_\mathrm{R}$ refers to the set of receiver parameters in terms the weights and biases across the decoder module.

The delivered DC power at the receiver can be modelled in two different ways, either based on an analytical model e.g., (13) and (14) [70], [159], or by learning its input-signal/output-power relationship from measurement data [160]. The latter is particularly powerful, and the larger the amount and diversity (in terms of a large range of input signal power and shape, load, etc) of the collected data, the higher the accuracy of the learned model. An EH model can obtained by applying non-linear regression over collected real data [160]. In particular, we study the data collected from the EH circuit from Fig. 6. The function we consider for modelling the EH is given as

$$f_{\mathrm{LNM}}(P_{\mathrm{in}}^r) = \sigma(W_3\sigma(W_2\sigma(W_1 P_{\mathrm{in}}^r + w_1) + w_2) + w_3), \quad (61)$$

where $P_{\mathrm{in}}^r$ denotes EH instantaneous input power, i.e., $P_{\mathrm{in}}^r = |y|^2$. $\mathcal{W} = \{W_1^{3\times1}, W_2^{2\times3}, W_3^{1\times2}, w_1^{3\times1}, w_2^{2\times1}, w_3^{1\times1}\}$ is the set of parameters to be optimized and $\sigma(\cdot) = \tanh(\cdot)$. In order to train the parameters we use Gradient Descent optimization applied over the following objective function

$$L_{\mathrm{EH}}(\mathcal{W}) = \frac{1}{m}\sum_{i=1}^{m}(f_{\mathrm{LNM}}(P_{\mathrm{in}}^r) - P_{\mathrm{dc}}^r)^2, \quad (62)$$

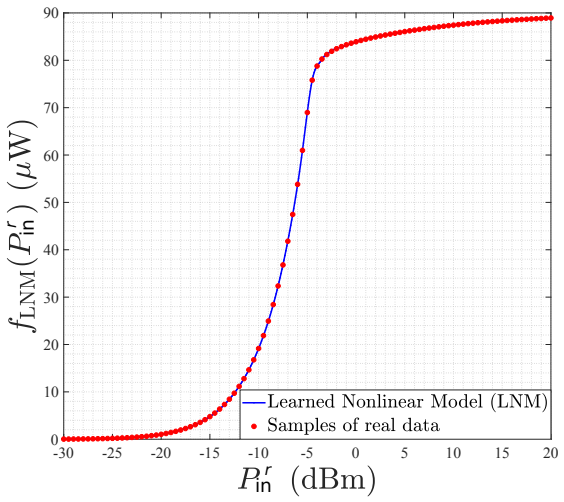



Fig. 24. Model of an EH based on applying nonlinear regression over collected real data from rectifier in Fig. 6 with a CW input waveform [160].

where $P_{\text{dc}}^r$ is the collected output DC power corresponding to an instantaneous input power $P_{\text{in}}^r$ and $m$ is the number of collected data used for training. In Fig. 24, the learned model (solid blue line) and the collected data (red dots) are illustrated[4].

Note that, since for power delivery purposes, the received RF signal is directly fed into the EH, the signal is not processed through the NN. We aim at following a learning approach and training the structure in Fig. 23 to minimize the objective function $L(\theta_{\text{T}}, \theta_{\text{R}})$ given as

$$L(\theta_{\text{T}}, \theta_{\text{R}}) = \frac{1}{|\mathcal{B}|} \sum_{l \in \mathcal{B}} \mathcal{L}(s_o^{(l)}, \hat{s}_o^{(l)}) + \frac{\lambda}{P_{\text{dc}}^r(s^{(l)})}, \tag{63}$$

where $\mathcal{B}$ is a randomly drawn minibatch of training data, which is assumed to be generated i.i.d. with a uniform distribution over the message set $\mathcal{M}$ and $|\mathcal{B}|$ is the cardinality of that batch. $\mathcal{L}(\mathbf{s}_o^{(l)}, \hat{\mathbf{s}}_o^{(l)}) = -\sum_{i=1}^{S} s_{o,i}^{(l)} \log \hat{s}_{o,i}^{(l)}$ is the cross entropy function between the one-hot representation of the $l^{th}$ training sample (denoted by $\mathbf{s}_o^{(l)}$) and its corresponding output probability vector $\hat{\mathbf{s}}_o^{(l)}$ at the receiver. $s_{o,i}^{(l)}$ and $\hat{s}_{o,i}^{(l)}$ indicate the $i^{\text{th}}$ entry of the vectors $\mathbf{s}_o^{(l)}$ and $\hat{\mathbf{s}}_o^{(l)}$, respectively.

The approach can be extended to coded modulation, in which case the output of the mapping function is of higher dimension, i.e., a sequence of symbols $x$. Similarly, the approach is extendable to multi-user settings, for instance the broadcast channel, multiple access channel and the interference channel. The end-to-end learning of SWIPT in all those types of channels has been studied in [160]. Further extension to more general multi-carrier settings remains an open area.

In the sequel, we illustrate the benefit of the above approach through two examples of modulation design under EH nonlinearity under linear and nonlinear HPA regime.

[4]The measurements are obtained using Continuous Wave (CW) signals and assuming that the circuit is operating in steady state (since the effect of the transient state is negligible). Modulation learning using CW measurements is justified because amplitude modulation effectively corresponds to a CW with different power levels.

*Example 1 - Modulation under Nonlinear EH and Linear HPA*: In Fig. 25, we illustrate the set of constellation points (with $M = 16$) obtained via end-to-end learning considering the EH as shown in Fig. 6 and using data-based model (61) for $P_{\text{rf}}^r = 10\mu$W. We assume that a linear HPA with $G = 1$ and $A_s = \infty$, and therefore do not need to model it.

By increasing $\lambda$, the demand for power at the receiver increases. Accordingly, the modulation loses its symmetry around the origin in a way that some of the transmitted symbols are getting away from the origin. As the receiver power demand increases, the transmit signal modulation is reformed. In the extreme scenario, where the receiver demand for power is at its maximum, the symbols possess only two amplitudes (one away from the origin and the other equal to zero) and becomes an on-off keying signalling. In this example, since only one symbol is shooting away, the probability of the on (high amplitude) signal is 1/16 and the probability of the off (zero amplitude) signal is 15/16.

An interesting observation about the learned modulations in Fig. 25 (in particular focusing on the last sub-figure) is that the channel input empirical distribution approaches to a distribution with two mass points for the amplitude, one with "low-probability/high-amplitudes" and the other with "high-probability/zero-amplitudes". This result is inline with (51) where we recall that the information-theoretic optimal channel input distributions of problem (48)-(50) for large power delivery (accounting for EH nonlinearity) follow the same behaviour, i.e., "low-probability/high-amplitudes" and "high-probability/zero-amplitudes".

Though modulation has been computed here using Fig. 24 and illustrated for $P_{\text{rf}}^r = 10\mu$W, the same learning approach can be applied to model (13) as in [70], [159] and to the curve-fitting model of [72] as in [159], as well as to other input power levels $P_{\text{rf}}^r$ as in [160]. The shape of the constellation would change depending on the model and the input power level. Interestingly, a major conclusion of [160] is to utilize learning-based results to design non learning-based algorithms that perform as well. In particular, inspired by the results obtained via learning, an algorithmic approach for coded modulation design has been proposed, which performs very close to its learning counterparts, and is significantly superior due to its high real-time adaptability to new system design parameters.

*Example 2 - Modulation under Nonlinear EH and Nonlinear HPA*: In Fig. 26, we revisit the results of Fig. 25 but account for the HPA nonlinearity of (11) assuming $\beta = 10$ and the output saturation voltage $A_s = 4.47mV$. In Fig. 25, the high amplitude symbol has an amplitude of 12.64mV and was left undistorted by the HPA because its nonlinearity was not modeled. Under HPA nonlinearity, there is no benefit to increase the voltage of one symbol above $A_s$. Instead, we note from Fig. 26 that another symbol is moved away from the origin to increase the EH output power. Comparing Fig. 25 and 26, we also note that HPA nonlinearity leads to a decrease of $P_{\text{dc}}^r$. This illustrates how deep learning can be used for end-to-end SWIPT design accounting for both transmit and receive nonlinearities.

*2) Physics-based Learning:* Physics-based learning has recently emerged as a way to bring together the benefits of

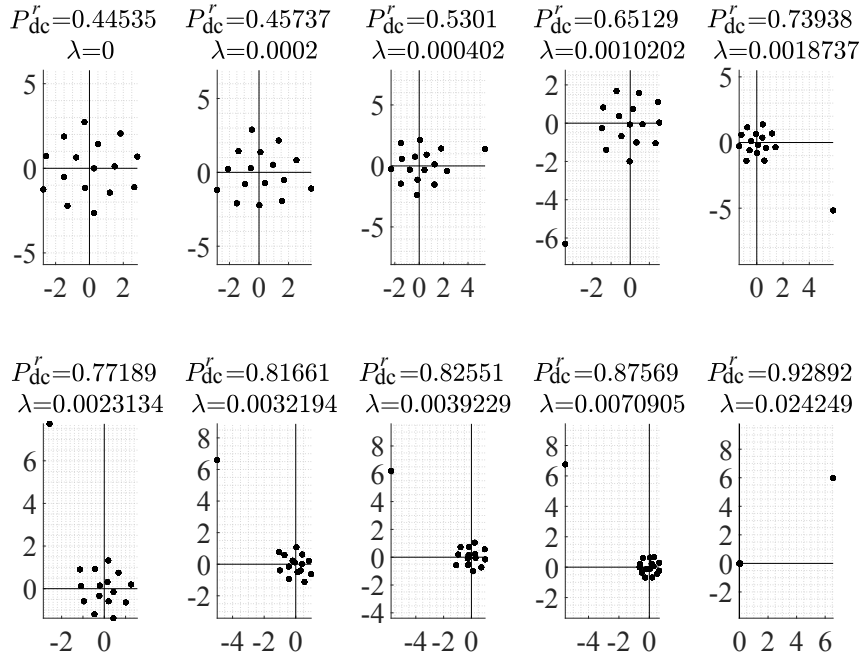



Fig. 25. Illustration of learned modulations under EH nonlinearity and linear HPA, respectively, with $M = 16$, $P_{\rm rf}^r = 10\mu$W and SNR$= 50$ (16.98 dB) for different values of $\lambda$ with corresponding $P_{\rm dc}^r$.

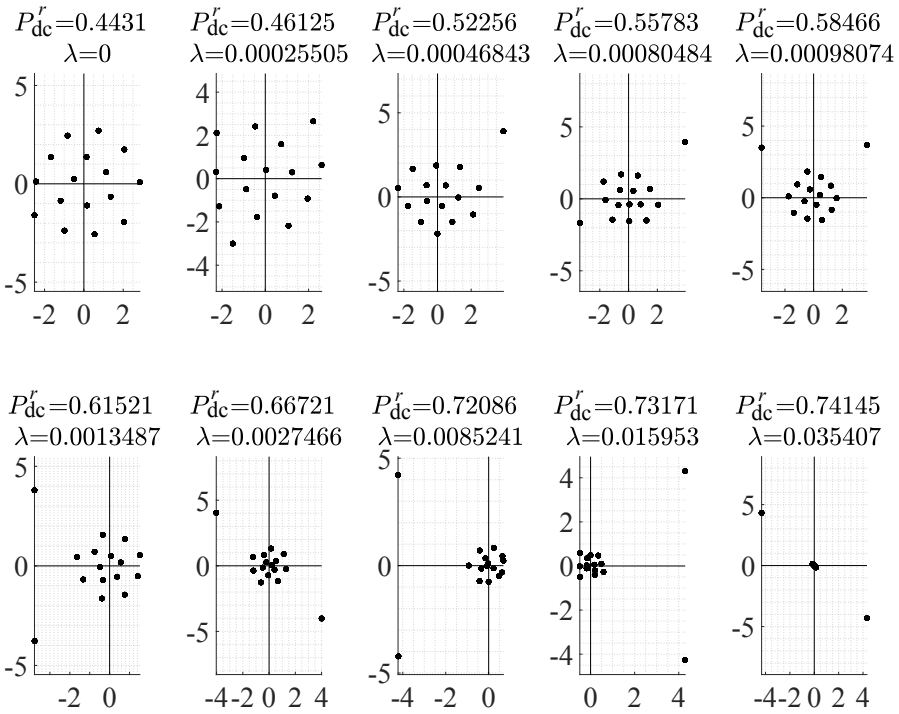



Fig. 26. Illustration of learned modulations (output constellation of HPA) under EH nonlinearity and HPA nonlinearity, with $M = 16$, $P_{\rm rf}^r = 10\mu$W and SNR$= 50$ (16.98 dB) for different values of $\lambda$ with corresponding $P_{\rm dc}^r$.

learning and model-and-optimize approaches, especially in scientific discovery and engineering system design. Though it has not been used much yet in WPT settings, various manifestations as surveyed in [165] hold promise as compelling research directions. These manifestations include physics-guided loss functions for machine learning algorithms, physics-guided initialization of model parameters as a starting state for data-driven training, physics-guided neural network architecture design, residual modeling where the residuals of a physical model are characterized by a data-driven model, and hybrid modeling where physics-based and data-driven models are used together.

One example where physical knowledge has been used in WPT is in computing the R-E function for settings where the EH function is only known from simulation/experimental samples [25]. To bound the Shannon capacity in this setting, the known physical smoothness of energy harvesting circuits is used to reduce the sample complexity needed in learning. A related idea to reduce sample complexity is to constrain deep learning algorithms to yield models that are scientifically consistent with known physics [163], [164].

Another physics-based learning approach is to learn symbolic expressions (a first stage in model-and-optimize) using data-driven techniques. The standard technique for learning such physical laws from data is symbolic regression, which typically yields expressions using basic trigonometric and polynomial functions [166], [167], and may incorporate prior physical knowledge [169]. Information lattice learning is related to symbolic regression and also performs human-interpretable knowledge discovery from data, but based on physical knowledge of basic symmetries [171], [172]. Alternatively, one can first train black-box deep learning models and then distill simple scientific laws from them [170]. One can imagine learning the nonlinear phenomenology of WPT circuits and systems using these techniques and then optimizing system design.

In machine learning, there is often limited training data available in certain parts of the space. Moreover, typical black-box machine learning models lack appropriate metacognition and are wildly confident in those regions since they are unconstrained by data. To address this issue, one can combine data-driven models with models from physical knowledge via *algorithm fusion*, whether using a hard switch or soft Bayesian weights determined by confidence levels estimated from density of training data. In a way, the physical model acts as a backup to the data-driven model [168]. This hybrid approach would allow safe optimization of WPT systems.

## V. Wireless Powered Internet-of-Things

The next-generation Internet-of-Things (IoT) is expected to connect tens of billions of edge devices (e.g., sensors, smartphones, and wearable computing devices) to automate a wide range of services, such as environmental monitoring, transportation, healthcare, traffic monitoring, and public-safety surveillance [173]. Among others, a key challenge is the high maintenance cost of recharging the batteries of an enormous number of sensors and devices. WPT is a promising solution. In this section, we discuss the design of three specific wireless-powered IoT systems: wireless powered mobile edge computing, wireless powered crowd sensing, and wireless powered distributed machine learning.

As a basic operation of mobile edge computing, *mobile computation offloading* (MCO) refers to offloading computation-intensive tasks from mobiles to the cloud, thereby reducing the former's energy consumption and enriching their capabilities and features [174]. On the other hand, realizing MCO involves the transmission of data and messages across the air interface [175]. To rein in the incurred transmission-power cost has been driving researchers to jointly design algorithms for MCO and adaptive transmission under the criterion of maximum mobile energy savings (see e.g., the survey in [176] and references therein). On the other hand, active research is also being carrierd out on energy-efficient mobile (local) computing. In the area, a wide range of techniques have been proposed to decrease mobile energy consumption by e.g., task scheduling [177], dynamic power management [178], and control of CPU-cycle frequencies [179]–[182]. In Section V-A, we demonstrate the design of wireless powered mobile edge computing by jointly designing three relevant technologies: 1) WPT, 2) MCO, and 3) local computing using the mobile CPU.

Traditional solutions for wireless sensor networking are limited in their coverage and scalability, as well as suffer from high maintenance costs [183]. Recently, leveraging the massive number of sensors on handheld and wearable equipment has led to the emergence of *mobile crowd sensing* (MCS) [184]. The key MCS challenges include how to prolong devices' battery lives and incentivizing users' involvement. In Section V-B, we discuss a MCS design that uses WPT as an incentivization mechanism to recharge the batteries of *mobile sensors* (MSs) in return for their participation in MCS.

Recently, the existence of enormous mobile data has been driving the deployment of algorithms for distributed machine learning at the network edge to distill the data into intelligence by training AI models [196]. *Federated learning* (FL) is arguably the most popular distributed-learning framework in practice. Its popularity arises from its preservation of data privacy by distributing a learning task over multiple edge devices (e.g., sensors and smartphones), which keeps data local [197]. To endow on edge devices the power of running AI algorithms, a recent focus in the semiconductor industry is to develop low-power mobile GPUs. The advancements in this direction and the area of low-complexity mobile AI make it possible to wire power devices to participate in FL in the near future. This leads to the emergence a new research topic, namely wireless-powered distributed learning, which is discussed Section V-C.

The discussion in the preceding sections targets generic WPT and WIPT systems. Their designs use generic metrics (e.g., communication rates or end-to-end WPT efficiencies) and have not discussed the energy consumption of the receivers. In contrast, the wireless-powered IoT systems discussed in this section target specific applications. As a result, their designs feature more detailed energy consumption models for more elaborate computation and operations (e.g., CPU-cycle frequency control, sensing, and data compression) and

application-specific performance metrics (e.g., data utility for crowd sensing). Due to limited space, we consider only three types of wireless-powered IoT systems, namely MEC, MCS and FL systems. However, the design approach of jointly optimizing WPT, communication, and computing is general and can be applied to other types of IoT systems.

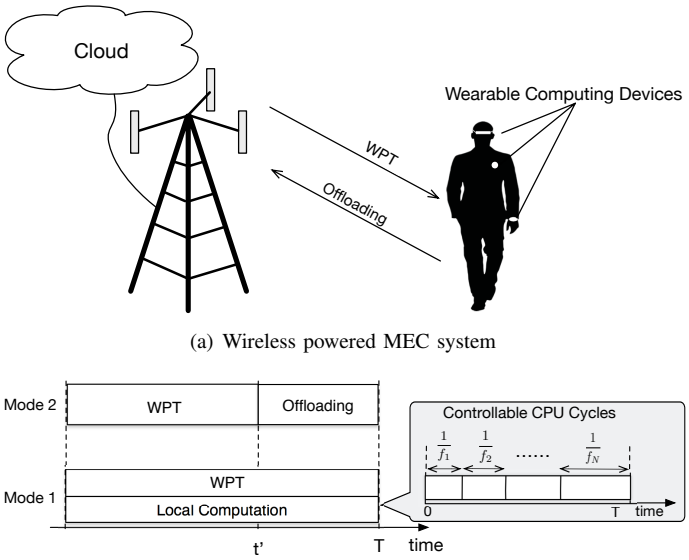
### A. Wireless-Powered Mobile Edge Computing

A simple wireless-powered MEC system is illustrated in Fig. 27(a). The mobile with milliwatt power consumption (e.g., mobile sensor and wearable computing device) is equipped with a single antenna and served by a multi-antenna BS that is connected to a cloud. The function of the BS is twofold: 1) performing WPT to the mobile or 2) offloads a computation task from it. Alternatively, the task can be also executed locally at the mobile. On one hand, local computing and WPT are allowed to coexist in time since only the later need use the antenna. On the other hand, MCO and WPT requires time sharing of the antenna given its half-duplex transmission. It is worth mentioning that a more complex design for the case of executing a multi-task program will involve partitioning of the program into multiple tasks with some offloaded and the rest executed locally. Considering the BS, energy beamforming is applied for WPT and receive beamforming for receiving the signal during MCO. For simplicity, we assume the existence of channel reciprocity. That allows the channel power gain to be represented as a scalar $g = |h|^2$. Last, the mobile harvests $P_{\rm dc}^r$ energy per time unit.

*1) Edge Device Operations:*

- **Local computing**: The task of the edge device is to process a fixed number of input data within a duration of $T$ seconds. Adopting an existing model [181], [182], the required number of CPU cycles, denoted as $X$, can be modeled as a random variable with a given distribution. To specify the distribution, let $p_k$ denote the probability that the data processing is not completed after $k$ CPU cycles and $N$ the allowed maximum number of CPU cycles. Mathematically, $p_k = \Pr(X \geq k)$ with $k = 1, 2, \cdots, N$. It is assumed that mobile has the knowledge of the distribution so as to make a decision on if the computing task should be offloaded or not.
- **CPU control**: The CPU-clock frequency can be controlled and its value directly determines the computation energy consumption. A single CPU cycle consumes a certain amount of energy, denoted as $E(f^{\sf clk})$, which is a function of the frequency $f^{\sf clk}$. Based on the model in [185], $E(f^{\sf clk}) = \gamma (f^{\sf clk})^2$ where the given parameter $\gamma$ depends on the switched capacitance. The CPU-cycle frequencies for the CPU cycle $1, 2, \ldots, N$ are represented by $f_1^{\sf clk}, f_2^{\sf clk}, \ldots, f_N^{\sf clk}$, respectively.
- **Computation offloading**: The device offloads computation by transmitting data to the BS to be processed in the cloud. After that, the computation result is downloaded to the device. The capacity of the uplink channel (in bit/s) is denoted as $C$ and can be written as

$$C = W \log \left( 1 + \frac{P_{\rm rf}^t g}{\sigma^2} \right) \tag{64}$$




(a) Wireless powered MEC system

(b) Mobile operation modes

Fig. 27. (a) Wirelessly powered MEC system and (b) the mobile operation modes.

where $\sigma^2$ is the variance of complex white Gaussian channel noise and $W$ the channel bandwidth. As the cloud has practically infinite computational resources, the time for computing in the cloud is much smaller than offloading time. The same is downloading time of computation result because the high BS transmission power keeps the downlink transmission delay small. In addition, the computation result usually has a small size and decoding it at the mobile consumes negligible energy consumption with respect to that for offloading and local computing.

*2) Wireless Powered Local Computing:* Consider the case of executing the computation task using the local CPU. The objective of energy efficient local computing is to control the CPU-cycle frequencies, $\{f_k^{\sf clk}\}$, so as to minimize the expected mobile energy consumption, denoted as $\bar{E}_{\rm mob}$. As the computation task is likely to be completed with no more than $N$ CPU cycles,

$$\bar{E}_{\rm mob} \left( \{f_k^{\sf clk}\}_{k=1}^X \right) \approx \sum_{k=1}^N \gamma p_k (f_k^{\sf clk})^2. \tag{65}$$

The frequency control must satisfy two constraints. The first constraint is the computation deadline: $\sum_{k=1}^N \frac{1}{f_k^{\sf clk}} \leq T$. The other is the well known energy harvesting constraint accounting for the fact that the consumed energy cannot exceed the harvested energy at any time instant. The constraint can be decomposed into $N$ sub-constraints given as

$$\sum_{k=1}^m \gamma (f_k^{\sf clk})^2 \leq P_{\rm dc}^r \sum_{k=1}^m \frac{1}{f_k^{\sf clk}}, \quad m = 1, 2, \cdots, N. \tag{66}$$

At the left side of the above inequality is the total energy consumed by the first $m$ CPU cycles while the total energy harvested by the end of the $m$th cycle is at the right side. Based on above discussion, the design of energy efficient local

computing can be formulated as the following optimization problem:

$$\min_{\{f_k^{\text{clk}}\}} \quad \sum_{k=1}^{N} \gamma p_k (f_k^{\text{clk}})^2 \tag{67}$$

$$\text{s.t.} \quad \sum_{k=1}^{m} \gamma (f_k^{\text{clk}})^2 \le P_{\text{dc}}^r \sum_{k=1}^{m} \frac{1}{f_k^{\text{clk}}}, \qquad \forall m, \tag{68}$$

$$\sum_{k=1}^{N} \frac{1}{f_k^{\text{clk}}} \le T, \tag{69}$$

$$f_k^{\text{clk}} > 0, \qquad \forall k. \tag{70}$$

Though the problem is non-convex, it can be transformed into an equivalent convex problem by replacing $\{f_k^{\text{clk}}\}$ with a new set of variables $\{y_k\}$ with $y_k = \frac{1}{f_k^{\text{clk}}}$. The equivalent problem allows the application of Lagrange multiplier theory to shed light on the structure of the optimal policy. To this end, define two positive constants $a$ and $a'$ as

$$a = \frac{\gamma N^3}{T^3} \quad \text{and} \quad a' = \frac{\gamma}{T^3} \left( \sum_{k=1}^{N} p_k^{\frac{1}{3}} \right)^2 \left( \sum_{k=1}^{N} p_k^{-\frac{2}{3}} \right). \tag{71}$$

Then the optimal CPU-cycle frequencies $\{f_1^{\text{clk}\star}, f_2^{\text{clk}\star}, \cdots, f_N^{\text{clk}\star}\}$ that solve the optimization problem in (67)-(70) have the following properties [187].

1) **Low-power regime**: If the recovered DC power $P_{\text{dc}}^r < a$, the harvested energy is insufficient for accomplishing the computation task.
2) **Medium-power regime**: If the recovered DC power $P_{\text{dc}}^r \in [a, a')$, the optimal CPU-cycle frequencies should be set as

$$f_k^{\text{clk}\star} = \left[ \frac{1}{T} \sum_{m=1}^{N} (p_m + \lambda)^{\frac{1}{3}} \right] (p_k + \lambda)^{-\frac{1}{3}}, \qquad \forall k \tag{72}$$

where the positive constant $\lambda$ is the Lagrange multiplier associated with the energy harvesting constraint.
3) **High-power regime**: If the recovered DC power $P_{\text{dc}}^r \ge a'$, $\{f_k^{\text{clk}\star}\}$ are independent of $P_{\text{dc}}^r$:

$$f_k^{\text{clk}\star} = \left( \frac{1}{T} \sum_{m=1}^{N} p_m^{\frac{1}{3}} \right) p_k^{-\frac{1}{3}}, \qquad \forall k. \tag{73}$$

*3) Wireless Powered Computation Offloading:* Consider the case of offloading the computation task to the cloud. The objective of energy efficient computation offloading is to maximize the mobile energy savings, referring to the difference between harvested energy and transmission energy consumption. As shown in Fig. 27(b), the time interval $[0, T]$ is divided into two parts: $[0, t']$ and $(t', T]$, corresponding to WPT and offloading, respectively. The amount of energy harvested over the interval $[0, t']$ can be written as $E_{\text{WPT}}(t') = P_{\text{dc}}^r t'$. Consider offloading. The most energy-efficient data transmission policy under a deadline constraint as proved in [186] to be the fixed-rate transmission over the interval $(t', T]$. Given the result, let $E_{\text{off}}(t')$ represent the offloading energy consumption and it can be expressed as $E_{\text{off}}(t') = [2^{\frac{L}{B(T-t')}} - 1] \frac{\sigma^2}{h} (T - t')$. It follows that $[E_{\text{WPT}}(t') - E_{\text{off}}(t')]$ gives the energy savings. As the function need not be monotone over $t'$, its maximum can be found by optimizing the WPT/offloading time partitioning. To simplify notation, define the offloading duration $t = T - t'$. We can write $E_{\text{WPT}}(t) = P_{\text{dc}}^r (T - t)$ and $E_{\text{off}}(t) = (2^{\frac{L}{Bt}} - 1) \frac{\sigma^2}{g} t$. This allows the objective function to be simplified as

$$E_{\text{WPT}}(t) - E_{\text{off}}(t) = P_{\text{dc}}^r T + \left( \frac{\sigma^2}{g} - P_{\text{dc}}^r \right) t - \frac{\sigma^2}{g} t 2^{\frac{L}{Bt}}. \tag{74}$$

It follows that the design of energy efficient offloading can be formulated as the following optimization problem:

$$\begin{aligned} \max_{t} \quad & E_{\text{WPT}}(t) - E_{\text{off}}(t) \\ \text{s.t.} \quad & 0 < t < T, \\ & E_{\text{WPT}}(t) - E_{\text{off}}(t) \ge 0. \end{aligned}$$

The problem is convex since the objective function can be easily shown to be a concave function for $t \in (0, \infty)$. This allows us to obtain the optimal offloading duration, denoted as $t^\star$, in closed form. To this end, define a positive constant $a''$ as

$$a'' = \frac{\sigma^2}{e_1} \left\{ 1 + \left[ \frac{L \ln 2}{WT} + \tilde{W}(-e^{-1-\frac{L \ln 2}{WT}}) \right] \times \exp\left( \frac{L \ln 2}{WT} + \tilde{W}(-e^{-1-\frac{L \ln 2}{BT}}) + 1 \right) \right\} \tag{75}$$

where $\tilde{W}(x)$ is the Lambert function defined as the solution for $\tilde{W}(x) e^{\tilde{W}(x)} = x$. Then the optimal offloading duration $t^\star$ has the following properties.

1) **Low-power regime**: If the recovered DC power $P_{\text{dc}}^r < a''$, the harvested energy is insufficient for offloading.
2) **Sufficient-power regime**:If $P_{\text{dc}}^r \ge a''$, the optimal offloading duration

$$t^\star = \frac{\ln 2 \times L}{W\left[ 1 + \tilde{W}(\frac{P_{\text{dc}}^r}{\sigma^2 e} - \frac{1}{e}) \right]}. \tag{76}$$

*4) Optimal Offloading Decision:* The optimal offloading policy aims at maximizing the mobile energy savings. Using the results obtained in preceding subsections, we are ready to derive the optimal policy as follows.

- If either offloading or local computing is feasible but not both, i.e., $P_{\text{dc}}^r \ge a$ for local computing or $P_{\text{dc}}^r \ge a''$ for offloading, then this mode is chosen.
- If both are feasible, we can select the preferred mode by comparing their energy savings corresponding to the optimal polices derived previously.

### *B. Wireless Powered Crowd Sensing*

Fig. 28 illustrates a multi-user WPCS system. There exist in the system multiple single-antenna mobile sensors (MSs) connected to a single multi-antenna BS. Consider crowd-sensing within some time window of interest. It is divided into three phases: *message passing*, *WPT*, and *crowd sensing* (see Fig. 28), which are described as follows.

- **Message-passing phase**: Via feedback, the BS acquires knowledge of the parameters of each sensor including the channel state, sensing and compression power. Then the BS applies the knowledge to jointly control the allocation

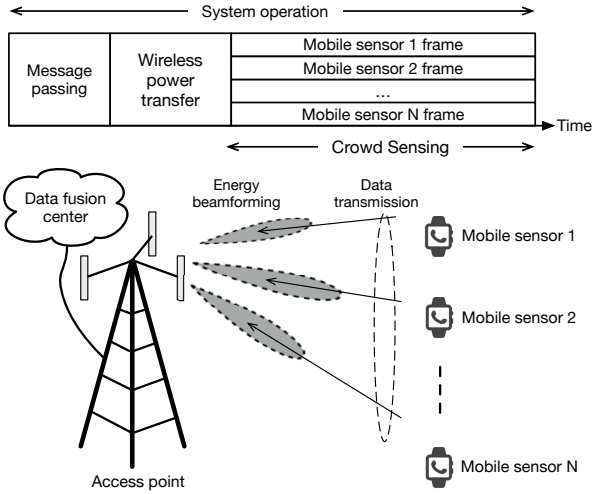



Fig. 28. Wireless powered crowd sensing system.

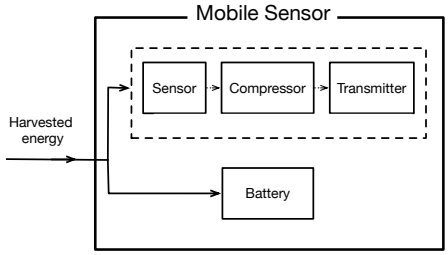



Fig. 29. Architecture and operations of a mobile sensor.

of transferred power to MS's and their operations (i.e., data sensing, compression, and transmission). Thereby, the data utility is maximized while the energy consumption is minimized. After solving the multi-objective optimization problem (elaborated in the sequel), the BS applies the optimal policy to inform each individual MS to control their compression ratios, sensing-data sizes, and the time sharing of its operations. We assume that receiving these control parameters with small sizes consumes a sensor negligible energy.

- **WPT phase**: WPT is adopted by the BS to incentivize MS's to participate in crowd sensing (see Fig. 28). For the energy transferred by the BS to each MS, a part is saved as a reward and the rest is used to execute the sensing task and transmit sensing data to the BS.
- **crowd-sensing phase** Adopting the settings communicated by the BS, the MS's perform sensing, data compression, and transmission simultaneously. The data transmission by MS's are over parallel OFDM sub-channels assigned by the BS. The crowd-sensing duration, represented by $T$, is separated into three slots for sensing, compression, and transmission. Their durations are denoted as $t_n^{(s)}$, $t_n^{(c)}$, and $t_n$, respectively. This introduces the following time constraint:

$$\text{(Time constraint)} \quad t_n^{(s)} + t_n^{(c)} + t_n \leq T. \tag{77}$$

Consider the WPT phase lasting $T_0$ (seconds). The BS points $K$ sharp beams to perform simultaneous WPT to $K$ MSs. The transmit power of the $n$-th beam is denoted as $P_{\mathrm{rf},n}^t$. The total transmit power of the BS should not exceed $P$. Mathematically,

$$\text{(Power constraint)} \quad \sum_{n=1}^{K} P_{\mathrm{rf},n}^t \leq P. \tag{78}$$

The channel between the BS and MS $n$ has a power gain denoted as $g_n$. Then the amount of energy harvested by MS $n$ can be written as $E_n^{(h)}(P_{\mathrm{rf},n}^t) = e_3 g_n P_{\mathrm{rf},n}^t T_0 = P_{\mathrm{dc},n}^r T_0$.

*1) Mobile Sensor Operations:* The operations of an arbitrary sensor, say MS $n$, are illustrated in Fig. 29 and described as follows.

- **Data sensing**: Based on experiments, the total energy consumption for sensing, denoted as $E_n^{(s)}(\ell_n^{(s)})$, is approximately proportional to the sensing-data size. The size, denoted as $\ell_n^{(s)}$, can written as $\ell_n^{(s)} = s_n t_n^{(s)}$, where $s_n$ denotes the output data rate and $t_n^{(s)}$ denotes the sensing duration. Then we can model the sensing energy as $E_n^{(s)} = q_n^{(s)} \ell_n^{(s)} = q_n^{(s)} s_n t_n^{(s)}$, where $q_n^{(s)}$ denotes the sensing energy consumption per bit.
- **Data compression**: Sensing data is compressed using lossless method. Some typical ones include Huffman, run-length, or Lempel-Ziv encoding. To simplify exposition, all MSs are assumed to adopt an identical compression method. It has a maximum compression ratio $R_{\max}$. The MS's choose their own compression ratios and the choice of the $n$-th MS is represented by $R_n \in [1, R_{\max}]$. The gives the size of compressed data as $\ell_n = \ell_n^{(s)}/R_n$. Some measurement results based on popular compression techniques, i.e., XZ compression, Zlib, and Zstandard [190] suggest that the number of CPU cycles needed to compress 1-bit of data can be fitted to an exponential function of the compression ratio $R_n$:

$$C(R_n, \epsilon) = e^{\epsilon R_n} - e^{\epsilon}, \tag{79}$$

where the positive constant $\epsilon$ depends on the specific compression algorithm. For a sanity check, $C(R_n, \epsilon) = 0$ for $R_n = 1$ in the special case of no compression. The CPU-cycle frequency at MS $n$ is fixed to be $f_n^{\mathrm{clk}}$. Then we can obtain the compression time duration as $t_n^{(c)} = (\ell_n^{(s)} C(R_n, \epsilon))/f_n^{\mathrm{clk}}$. As in the preceding subsections, each CPU cycle consumes the energy of $q_n^{(c)} = \gamma (f_n^{\mathrm{clk}})^2$. The energy consumption for data compression, denoted by $E_n^{(c)}(\ell_n^{(s)}, R_n)$, is given as $E_n^{(c)}(\ell_n^{(s)}, R_n) = q_n^{(c)} \ell_n^{(s)} C(R_n, \epsilon)$ with $C(R_n, \epsilon)$ in (79). As a result,

$$E_n^{(c)}(\ell_n^{(s)}, R_n) = q_n^{(c)} \ell_n^{(s)} (e^{\epsilon R_n} - e^{\epsilon}). \tag{80}$$

A similar model can be developed for lossy compression e.g., data truncation after discrete cosine transform.

- **Data transmission**: After compression, each scheduled MS uploads sensing data to the BS. For the $n$-th MS, the transmission power and time duration are represented by $\tilde{P}_{\mathrm{rf},n}^t$ and $t_n$, respectively. Assuming channel reciprocity, the achievable transmission rate (in bits/s), denoted by $v_n$, can be given as $v_n = \ell_n/t_n = W \log_2 \left(1 + \frac{g_n \tilde{P}_{\mathrm{rf},n}^t}{\sigma^2}\right)$. It follows that the transmission energy consumption is

$E_n^{(t)}(\ell_n) = \tilde{P}_{\mathrm{rf},n}^t t_n = \dfrac{t_n}{g_n} z(\ell_n/t_n)$, where we define the function $z(x)$ as $z(x) = \sigma^2(2^{\frac{x}{W}} - 1)$.

*2) System Optimization:* Both MSs and (system) operator receive rewards. The reward requested by a MS is in terms of energy transferred from the BS minus that for supporting the MS operations. Let $E_n^{(r)}$ denote the energy reward requested by MS $n$. It is proportional to the size of sensing data denoted as $\ell_n^{(s)}$, namely, $E_n^{(r)} = q_n^{(r)} \ell_n^{(s)}$ with $q_n^{(r)}$ being a fixed scaling factor. On the other hand, the net operator's reward is the utility of the data collected from MSs minus the energy cost. A commonly model of data utility is as follows (see e.g., [192]). We can measure the utility of $\ell_n$-bit data provided by MS $n$ using the logarithmic function $a_n \log(1 + b_n \ell_n^{(s)})$, where $b_n$ is the loss of information caused by compression and the weight factor $a_n$ depends on the data type. The monotonicity of the function captures the fact that more information-bearing data gives a higher level of data utility, e.g., higher machine-learning accuracy [194] or better image/video resolution [193]. On the other hand, the logarithmic function has the property of reflecting the diminishing return as the data size increases, resulting in more repeated and redundant data. With the model, we write the the sum data utility for the operator as

$$U(\boldsymbol{\ell}^{(s)}) = \sum_{n=1}^{K} a_n \log(1 + b_n \ell_n^{(s)}). \tag{81}$$

The operator's reward, denoted by $R(\boldsymbol{\ell}^{(s)}, \boldsymbol{P})$, can be modelled as

$$R(\boldsymbol{\ell}^{(s)}, \boldsymbol{P}) = \sum_{n=1}^{K} a_n \log(1 + b_n \ell_n^{(s)}) - c \sum_{n=1}^{K} P_{\mathrm{rf},n}^t T_0, \tag{82}$$

where $c$ denotes the price of unit energy as measured against unit data utility.

The objective of system optimization is to maximize the operator's reward under two constraints. One is that the totally energy harvested by a MS should exceed its spent energy:

$$E_n^{(r)} + E_n^{(s)} + E_n^{(c)} + E_n^{(t)} \le P_{\mathrm{dc},n}^r T_0, \qquad \forall n. \tag{83}$$

The other is the time constraint per round:

$$\frac{\ell_n^{(s)}}{s_n} + \frac{\ell_n^{(s)} C(R_n, \epsilon)}{f_n^{\mathrm{clk}}} + t_n \le T. \tag{84}$$

We can make the observation that if the sensing rate $s_n$ and the CPU-cycle frequency $f_n^{\mathrm{clk}}$ are fixed, the partitioning of crowd-sensing time of sensor $n$ for sensing, compression, and transmission can be determined by the *sensing-data size* $\ell_n$, *compression ratio* $R_n$, and *transmission time* $t_n$. Then the system operations can be optimized over all or a subset of these variables.

For example, the problem of jointly optimizing joint power allocation, sensing, and transmission can be formulated as

$$\begin{aligned} \max_{\substack{P_{\mathrm{rf},n}^t \ge 0, \ell_n^{(s)} \ge 0, \\ t_n \ge 0}} \quad & \sum_{n=1}^{K} a_n \log(1 + \ell_n^{(s)}) - c \sum_{n=1}^{K} P_{\mathrm{rf},n}^t T_0 \\ \text{s.t.} \quad & \sum_{n=1}^{K} P_{\mathrm{rf},n}^t \le P, \\ & \beta_n \ell_n^{(s)} + t_n \le T, \qquad \forall\, n, \\ & \xi_n \ell_n^{(s)} + \frac{t_n}{g_n} f\left(\frac{\ell_n^{(s)}}{R_n t_n}\right) \le P_{\mathrm{dc},n}^r T_0, \qquad \forall\, n, \end{aligned}$$

where $\xi_n = q_n^{(r)} + q_n^{(s)} + q_n^{(c)} C(R_n, \epsilon)$ and $\beta_n = \frac{1}{s_n} + \frac{C(R_n,\epsilon)}{f_n^{\mathrm{clk}}}$. This can be shown to be a convex program and thus can be solved efficiently using an existing algorithm. Based on the well-known KKT conditions, some light can be shed on the optimal policy [188]. In particular, the optimal sensor-transmission duration is

$$t_n^\star \propto \frac{T}{W R_n \beta_n}. \tag{85}$$

On the other hand, the optimal wireless-power allocation policy has a *threshold*-based structure:

$$P_{\mathrm{rf},n}^{t\star} = \begin{cases} \dfrac{1}{e_3 g_n T_0} \left[ \dfrac{t_n^\star}{g_n} f\left(\dfrac{T - t_n^\star}{R_n \beta_n t_n^\star}\right) + \dfrac{\xi_n (T - t_n^\star)}{\beta_n} \right], & \varphi_n \ge \lambda^\star, \\ 0, & \varphi_n < \lambda^\star, \end{cases}$$

where $\lambda^\star$ is a Lagrange multiplier and $\varphi_n$ represents the *MS-scheduling priority function* given as

$$\varphi_n = \frac{a_n e_3 g_n}{q_n^{(r)} + q_n^{(s)} + q_n^{(c)} C(R_n, \epsilon) + \frac{N_0 \ln 2}{g_n B R_n}} - c. \tag{86}$$

The result suggests that only the MSs with their priority functional values above the threshold $\lambda^\star$ should be scheduled for participating in the crowd sensing operation and being rewarded with energy in return. Finally, the optimal sensing-data sizes $\{(\ell_n^{(s)})^\star\}$ are proportional to the corresponding sensing durations:

$$(\ell_n^{(s)})^\star = \begin{cases} \dfrac{T - t_n^\star}{\beta_n}, & \varphi_n \ge \lambda^\star, \\ 0, & \varphi_n < \lambda^\star. \end{cases} \tag{87}$$

The problem of optimizing all system variables is non-convex. One low-complexity algorithm for finding a local optimal point is to iterate between solving convex sub-problems, each of which is over a subset of variables. For instance, one sub-problem can be that discussed above and the other being the joint optimization of compression and transmission, both of which are convex [188].

### C. Wireless-Powered Federated Learning

A wireless-powered FL system in a circular cell is illustrated in Fig. 30(a). In the system, a server equipped with an access point coordinates multiple devices to perform the task of training an AI model. Dense power-beacon, dedicated WPT stations, are deployed to reliably delivery energy to the devices. The randomly distributed beacons can be suitably

modeled as a Poisson point process with a density $\lambda_{\rm pb}$. If the density is high, beacons deliver to each device a stable amount of energy within an arbitrary communication round (or window), say one with the duration $T$, which is approximately given as

$$P_{\rm dc}^r = \frac{\pi\alpha'\rho P_{\rm rf}^t \lambda_{\rm pb}}{(\alpha'-2)\nu^{\alpha'-2}} \tag{88}$$

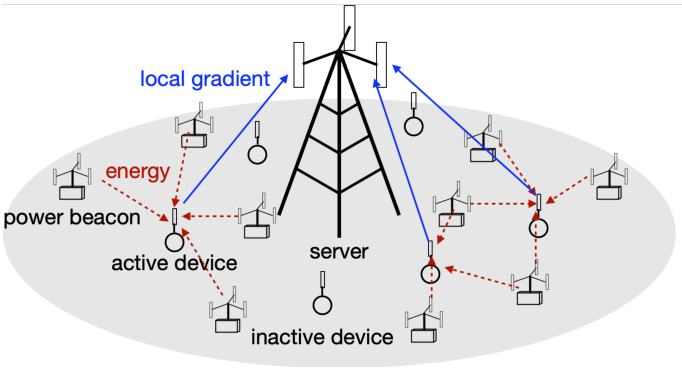
where $\nu \geq 1$ is a given minimum propagation distance avoiding singularity in path loss, $\alpha' > 2$ is the path-loss exponent of a WPT link, $P_{\rm rf}^t$ is the transmission power of power-beacons, and $\rho$ represents the product of energy-conversion efficiency and energy-beamforming gain. The efficiency of the beacon network can be characterized by a single parameter, termed the *spatial-energy density*, which is defined as $\lambda_{\rm energy} \triangleq P_{\rm rf}^t \lambda_{\rm pb} T$. It can be interpreted as the amount of energy delivered by the power-beacon network to an arbitrarily located device in a single round. The transfer energy is then applied by each device to computation and communication in the round. The detailed system operations that implement the FL algorithm are described in Section V-C1. Essentially, the SGD based algorithm supports distributed stochastic-gradient estimation using local data at devices. The processor (clock) frequencies as well as their sizes of mini-batch data used in estimation can be optimized to maximize the estimation accuracy under the WPT constraint. The details are presented in Section V-C2. Given the optimal control of local computation at devices, there exists a fundamental trade-off in the system between the WPT network parameter (i.e., delivered energy per unit area) and the learning performance. It is termed the WPT-learning trade-off and discussed in Section V-C3.

*1) System Operations and Models:* A typical FL algorithm is the distributed implementation of the classic iterative *stochastic gradient descent* (SGD) algorithm (see e.g., [195]). Each iteration involves broadcasting by the server and uploading by devices and thus is called a communication round. The FL algorithm repeats the rounds until the model under training converges.

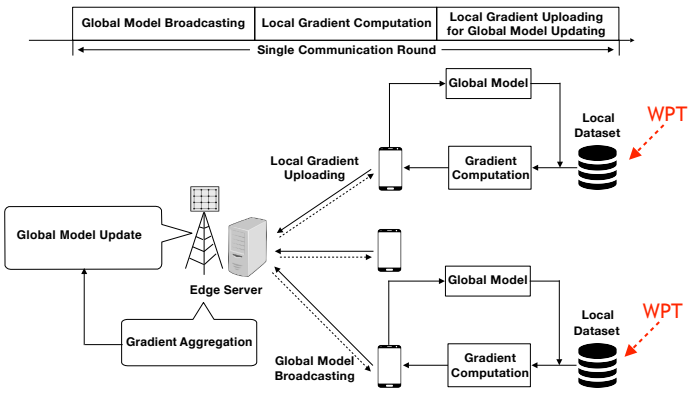
To facilitate exposition, some basic definitions are given as follows. The learning task aims at training a global model, represented by the high-dimensional parametric vector $\mathbf{w}$. The accuracy of the model is measured using a so called *global loss function* evaluated on the global dataset (aggregation of all distributed data), denoted as $F(\mathbf{w})$. Its gradient $\nabla F(\mathbf{w}^{(i)})$ is referred to as the *ground-truth gradient*. The distributed training process involves training of models at devices using local data, which are called *local models*. Each device attempts to estimate the ground-truth gradient using a mini-batch of local data under constraints on computation capacity and energy. The results are called *local gradients* denoted as $\mathbf{g}_k^{(i)}$ for device $k$ and round $(i)$. Without the constraints, all devices can use all available local data in estimation; then combining their estimations give the ground-truth gradient.

As illustrated in Fig. 30(b), the operations of each round, say round $i$, and relevant mathematical models are described as follows.

1) The server broadcasts the current global model, $\mathbf{w}^{(i)}$, to all devices for computing local gradients.




(a) Wireless-powered FL System




(b) System Operations

Fig. 30. (a) Wireless-powered FL system and (b) the system operations.

2) Assume that the model is a deep neural network. Using the received model, each device computes a local gradient using a data mini-batch and the well known back-propagation algorithm in the deep-learning literature. The computation is powered by part of the harvested energy given in (88). As in the preceding sub-sections, the processor clock frequency of each device, denoted as $f_k^{{\rm clk},(i)}$ for device $k$ and round $i$, is assumed controllable. The previous computation-energy model can be modified for a GPU to be [189]

$$E_{{\rm cmp},k}^{(i)} = b_k^{(i)} C_k W \left(f_k^{{\rm clk},(i)}\right)^2, \ \forall k \tag{89}$$

where $b_k^{(i)}$ denotes the mini-batch size, the coefficient $C_k$ is a constant characterizing the computation-energy efficiency of the processor, $W$ is the number of *floating point operations* (FLOPs) needed for processing each data sample in gradient computation. The computation time is

$$t_{{\rm cmp},k}^{(i)} = \frac{b_k^{(i)} W}{f_k^{{\rm clk},(i)}}, \ \forall k. \tag{90}$$

3) Allocated a bandwidth of $B$ Hz, each device transmits its local gradient at a data rate of

$$S_k^{(i)} = W \log_2 \left(1 + \frac{g_k^{(i)} P_{{\rm rf},k}^{t,(i)}}{\left(r_k^{(i)}\right)^{\alpha} \sigma^2}\right), \ \forall k, \tag{91}$$

where $g_k^{(i)}$ is the channel gain, $P_{\text{rf},k}^{t,(i)}$ represents the transmission power, $N_0$ the power spectrum density of the additive white Gaussian noise, $r_k^{(i)}$ the propagation distance, and $\alpha$ the path-loss exponent. Assume that each coefficient is quantized into $Q$ bits and part of the duration $T$, denoted as $t_k^{\text{cmm}}$, is allocated for transmission. Then the resultant transmission-energy consumption can be shown to be given as [189]

$$E_{\text{cmm},k}^{(i)} = \frac{\left(r_k^{(i)}\right)^{\alpha}}{g_k^{(i)}} \varphi(t_{\text{cmm},k}^{(i)}), \ \forall k, \tag{92}$$

where the function $\varphi(t) \triangleq BN_0 t \left(2^{\frac{qQ}{Bt}} - 1\right)$.

4) It is desirable for all devices to participate in model updating in each round. However, some devices may encounter deep channel fading and as a result do not have sufficient energy for the purpose. Thus, only a subset of devices, denoted as a subset $\mathcal{M}^{(i)} \subseteq \mathcal{K}$ with size $M^{(i)} = |\mathcal{M}^{(i)}|$, can participate in learning in this specific round. The server computes the global gradient by aggregating local gradients received from active devices:

$$\mathbf{g}^{(i)} = \begin{cases} \frac{1}{M^{(i)}} \sum\limits_{k \in \mathcal{M}^{(i)}} \mathbf{g}_k^{(i)}, & M^{(i)} > 0 \\ \mathbf{0}, & M^{(i)} = 0 \end{cases}. \tag{93}$$

Subsequently, the global model is then updated using SGD as

$$\mathbf{w}^{(i+1)} = \mathbf{w}^{(i)} - \eta \mathbf{g}^{(i)}, \tag{94}$$

where $\eta$ is the given learning rate (step size). This completes one round.

*2) Local Computation Optimization:* Given fixed computation time and energy, what should be the optimal mini-batch size and processor (clock) frequency? Note that from the perspective of an individual variable, a large value is desired for enhancing the gradient-estimation accuracy, but their effects on time and energy need to be balanced. The earlier question can be answered by formulating an optimization problem for maximizing the accuracy of local gradient estimation. It is shown in [189] that the expected deviation of the local gradient from its ground truth is inversely proportional to the sampled batch size, $b_k$. Then the desired optimization problem can be formulated as

$$\begin{aligned} \max_{\{b_k, f_k^{\text{clk}}\}} \quad & b_k \\ \text{s.t.} \quad & 0 < b_k C_k W (f_k^{\text{clk}})^2 \leq E_{\text{cmp},k}, \\ & 0 < \frac{b_k W}{f_k^{\text{clk}}} \leq T_{\text{cmp}}, \end{aligned}$$

where $T_{\text{cmp}}$ refers to the duration of the local gradient computation phase. By integer relaxation of $b_k$ and solving the above problem, the optimal policy is found to be

$$b_k^{\star} = \frac{1}{W} \left( \frac{E_{\text{cmp},k} T_{\text{cmp}}^2}{C_k} \right)^{\frac{1}{3}} \quad \text{and} \quad f_k^{\star} = \left( \frac{E_{\text{cmp},k}}{C_k T_{\text{cmp}}} \right)^{\frac{1}{3}}. \tag{95}$$

*3) Optimal Learning-WPT Tradeoff:* Given the optimal control of local computation in the preceding sub-section, we can study the trade-off between the spatial energy density provided by the beacon network and the learning performance, thereby establishing the optimal learning-WPT trade-off.

Essentially, the spatial energy density affects the probability a device fails to participate in learning. Specifically, the event, called an outage, occurs when the harvested energy by the device is insufficient for computation and communication. Given the transferred power in (92) and assuming server-beamforming over Rayleigh fading channels, the outage probability can be shown to be [189]

$$P_{\text{out}} = \frac{\gamma(L, \xi) - \xi^{-\frac{2}{\alpha}} \gamma(L + \frac{2}{\alpha}, \xi)}{\Gamma(L)}, \tag{96}$$

where $\gamma(\cdot, \cdot)$ is the lower incomplete Gamma function, and the parameter $\xi$ is defined as

$$\xi = \frac{(\alpha' - 2) \nu^{\alpha'-2} R^{\alpha} \varphi(T_{\text{cmm}})}{\pi \alpha' \rho \lambda_{\text{energy}}} \tag{97}$$

where $R$ is the cell radius and $T^{\text{cmm}}$ is the duration of the gradient uploading phase. Note that the the parameter is inversely proportional to the spatial energy density $\lambda_{\text{energy}}$. It follows that the number of active devices, $M$, has the following binomial distribution

$$\Pr(M = m) = \frac{1}{1 - P_{\text{out}}^K} \binom{K}{m} (1 - P_{\text{out}})^m P_{\text{out}}^{K-m}, \tag{98}$$

where $1 \leq m \leq K$.

One common metric of measuring the FL (or SGD) performance is the model-convergence speed that is specified as follows. A sample of the stochastic learning process can be described by the model realizations $\{\mathbf{w}^{(i)}\}$. Given $N$ rounds, the convergence speed is commonly measured in the literature using the expected average gradient norms, $\mathbb{E}\left[\frac{1}{N} \sum_{i=0}^{N-1} \|\nabla F(\mathbf{w}^{(i)})\|^2\right]$. Its value, a function of $N$, is expected to be sufficiently small if the model is likely to converge within $N$ rounds. The convergence analysis of FL in the literature (see e.g., [195]) is modified in [189] to account for the random number of active devices in each round. As a result, the optimal learning-WPT trade-off is esablished as follows

$$\begin{aligned} \mathbb{E}\left[\frac{1}{N} \sum_{i=0}^{N-1} \|\nabla F(\mathbf{w}^{(i)})\|^2\right] \leq & \frac{2\left[F(\mathbf{w}^{(0)}) - F_*\right]}{\eta N} \\ & + \frac{\delta W \overline{C}_{\sigma} (1 - P_{\text{out}})}{\rho^{\frac{1}{3}} K (T_{\text{cmp}})^{\frac{2}{3}} \lambda_{\text{energy}}^{\frac{1}{3}}} + R_{\text{es}}, \end{aligned} \tag{99}$$

where the parameter $\overline{C}_{\sigma}$ is explained shortly, $K$ the total number of devices, the constant $\delta \triangleq \frac{4}{\alpha} B(\frac{2}{3}, \frac{2}{\alpha}) \left(\frac{(\alpha'-2)\nu^{\alpha'-2}}{\pi \alpha'}\right)^{\frac{1}{3}}$, and the residue term $R_{\text{es}}$ given as

$$R_{\text{es}} = 2 \left( \sum_{m=2}^{K} \frac{P_{\text{out}}^{m-1} - P_{\text{out}}^K}{K - m + 1} + P_{\text{out}}^2 \right) \Phi. \tag{100}$$

The constant $\overline{C}_\sigma$ is a weighted sum of the computation-energy efficiencies of all devices: a larger value indicates a high power-growth rate as the processor frequency is scaled up and vice versa. From the above result, we can infer the effect of WPT as follows. Increasing the spatial-energy density leads to a linear growth of energy harvested by each device. As a result, active devices can estimate the local gradient with higher accuracy by using larger mini-batch sizes. Next, due to the gradient aggregation at the server, increasing $K$ reduces the expected gradient norm following the scaling law of $O\left(\frac{1}{K}\right)$. Last, the term $R_{\text{es}}$ captures the loss of convergence rate caused by those devices in computation-outage. If $P_{\text{out}}$ is small, $R_{\text{es}}$ scales as $O(P_{\text{out}})$.

## VI. Conclusions

This article has first provided a tutorial overview of various signal processing techniques for WPT and WIPT. It then discussed the benefits of two different design methodologies based on *model and optimize* and *learning* approaches. Finally it showed how WPT, computing, sensing, and communication need to be jointly designed in future wireless powered applications.

One first conclusion of the paper is to highlight that signal processing and machine learning techniques have an important role to play in WPT for future networks, but the techniques need to be developed in light of the physics and hardware constraints of WPT. This calls for abandoning naive and oversimplified linear models and accounting for nonlinearities and non-idealities at the transmitter and receiver ends.

A second conclusion is that WPT will act an important enabler for future networks and opens the door to new challenges and opportunities where communications, computing, and sensing have to be jointly designed together with WPT. It is hoped that the signal processing, machine learning, computing, and sensing techniques presented here will help inspire future research in this promising area and pave the way for designing and implementing efficient WPT, WIPT, and wireless-powered systems and networks in the future.

It is our hope that the ideas and tools discussed and reviewed in this overview paper will stimulate and encourage further research in the broad area of wireless power (and information) transfer for future networks at microwave, millimeter, terahertz, and visible light frequencies.

## References


[1] J. G. Koomey, S. Berard, M. Sanchez, and H. Wong, "Implications of Historical Trends in the Electrical Efficiency of Computing," *IEEE Ann. History Comput.*, vol. 33, no. 3, p. 46-54, Mar. 2011.

[2] L. R. Varshney, "On Energy/Information Cross-Layer Architectures," in *Proc. IEEE Int. Symp. Information Theory*, pp. 1356-1360, Jul. 2012.

[3] S. Hemour and K. Wu, "Radio-Frequency Rectifier for Electromagnetic Energy Harvesting: Development Path and Future Outlook," *Proc. of IEEE*, vol. 102, no. 11, Nov. 2014.

[4] A. Boaventura, A. Collado, N. B. Carvalho, and A. Georgiadis, "Optimum Behavior: Wireless Power Transmission System Design through Behavioral Models and Efficient Synthesis Techniques," *IEEE Microw. Mag.*, vol. 14, no. 2, pp. 26-35, Mar./Apr. 2013.

[5] C. R. Valenta and G. D. Durgin, "Harvesting Wireless Power: Survey of Energy-Harvester Conversion Efficiency in Far-Field, Wireless Power Transfer Systems," *IEEE Microw. Mag.*, vol. 15, pp. 108-120, Jun. 2014.

[6] A. Costanzo and D. Masotti, "Smart solutions in smart spaces: Getting the most from far-field wireless power transfer," *IEEE Microw. Mag.*, vol. 17, no. 5, pp. 30-45, May 2016.

[7] Y. Zeng, B. Clerckx, and R. Zhang, "Communications and Signals Design for Wireless Power Transmission," *IEEE Trans. Commun.*, Vol 65, No 5, pp 2264 – 2290, May 2017.

[8] B. Clerckx, A. Costanzo, A. Georgiadis, and N. B. Carvalho, "Toward 1G Mobile Power Networks: RF, Signal and System Designs to Make Smart Objects Autonomous" *IEEE Microw. Mag.*, Sept./Oct. 2018.06088.

[9] B. Clerckx, R. Zhang, R. Schober, D. W. K. Ng, D. I. Kim, and H. V. Poor, "Fundamentals of Wireless Information and Power Transfer: From RF Energy Harvester Models to Signal and System Designs," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 1, pp. 4-33, Jan 2019.

[10] N. Shinohara, *Wireless Power Transfer via Radiowaves*, Hoboken, NJ, USA: Wiley, 2014.

[11] N. B. Carvalho et al., "Wireless Power Transmission: R&D Activities within Europe," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 4, pp. 1031–1045, Apr. 2014.

[12] L. R. Varshney, "Transporting Information and Energy Simultaneously," in *Proc. Int. Symp. Inf. Theory*, Jul. 2008, pp. 1612-1616.

[13] P. Grover and A. Sahai, "Shannon Meets Tesla: Wireless Information and Power Transfer," in *Proc. Int. Symp. Inf. Theory*, Jun. 2010.

[14] R. Zhang and C. K. Ho, "MIMO Broadcasting for Simultaneous Wireless Information and Power Transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 1989-2001, May 2013.

[15] J. Xu, L. Liu, and R. Zhang, "Multiuser MISO Beamforming for Simultaneous Wireless Information and Power Transfer," *IEEE Trans. Signal Process.*, vol.62, no.18, pp.4798-4810, Sept. 2014.

[16] H. Son and B. Clerckx, "Joint Beamforming Design for Multi-User Wireless Information and Power Transfer," *IEEE Trans. Wireless Commun.*, vol. 13, no. 11, pp. 6397-6409, Nov. 2014.

[17] X. Zhou, R. Zhang, and C. K. Ho, "Wireless Information and Power Transfer: Architecture Design and Rate-Energy Tradeoff," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4754-4767, Nov 2013.

[18] L. Liu, R. Zhang, and K. C. Chua, "Wireless Information and Power Transfer: A Dynamic Power Splitting Approach," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3990-4001, September, 2013.

[19] J. Park and B. Clerckx, "Joint Wireless Information and Energy Transfer in a Two-User MIMO Interference Channel," *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 4210-4221, Aug. 2013.

[20] J. Park and B. Clerckx, "Joint Wireless Information and Energy Transfer in a K-User MIMO Interference Channel," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5781-5796, Oct. 2014.

[21] J. Park and B. Clerckx, "Joint Wireless Information and Energy Transfer with Reduced Feedback in MIMO Interference Channels," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 8, pp. 1563-1577, Aug. 2015.

[22] K. Huang and E. G. Larsson, "Simultaneous Information and Power Transfer for Broadband Wireless Systems," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5972-5986, Dec. 2013.

[23] X. Zhou, R. Zhang, and C. K. Ho, "Wireless Information and Power Transfer in Multiuser OFDM Systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 2282-2294, Apr. 2014.

[24] D. W. K. Ng, E. S. Lo, and R. Schober, "Wireless Information and Power Transfer: Energy Efficiency Optimization in OFDMA Systems,"*IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6352-6370, Dec. 2013.

[25] D. Seo and L. R. Varshney, "Information and Energy Transmission With Experimentally Sampled Harvesting Functions," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4479-4490, Jun. 2019.

[26] A. A. Nasir, X. Zhou, S. Durrani, and R. A. Kennedy, "Relaying Protocols for Wireless Energy Harvesting and Information Processing," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3622-3636, Jul. 2013.

[27] Y. Huang and B. Clerckx, "Joint Wireless Information and Power Transfer for an Autonomous Multiple-Antenna Relay System," *IEEE Comm. Lett.*, vol. 19, no. 7, pp. 1113-1116, July 2015.

[28] Z. Xiang and M. Tao, "Robust Beamforming for Wireless Information and Power Transmission," *IEEE Wireless Commun. Lett.*, vol. 1, no. 4, pp. 372–375, 2012.

[29] S. Timotheou, I. Krikidis, G. Zheng, and B. Ottersten, "Beamforming for MISO Interference Channels with QoS and RF Energy Transfer," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2646-2658, May 2014.

[30] M. R. A. Khandaker and K. K. Wong, "SWIPT in MISO Multicasting Systems," *IEEE Wireless Commun. Lett.*, vol. 3, no. 3, pp. 277-280, Jun. 2014.

[31] A. M. Fouladgar and O. Simeone, "On the Transfer of Information and Energy in Multi-User Systems," *IEEE Commun. Lett.*, vol. 16, no. 11, pp. 1733-1736, Nov. 2012.

[32] Z. Ding, I. Krikidis, B. Sharif, and H. V. Poor, "Wireless Information and Power Transfer in Cooperative Networks with Spatially Random Relays," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4440-4453, Aug. 2014.

[33] H. Ju and R. Zhang, "Throughput Maximization in Wireless Powered Communication Networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 418-428, Jan. 2014.

[34] K. Huang and V. K. N. Lau, "Enabling Wireless Power Transfer in Cellular Networks: Architecture, Modeling and Deployment," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 902-912, Feb. 2014.

[35] H. Lee et al., "Resource Allocation Techniques for Wireless Powered Communication Networks with Energy Storage Constraint," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2619-2628, Apr. 2016.

[36] K. Han, and K. Huang, "Wirelessly Powered Backscatter Communication Networks: Modeling, Coverage, and Capacity," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2548-2561, Apr. 2017.

[37] B. Clerckx, Z. Bayani Zawawi and K. Huang "Wirelessly Powered Backscatter Communications: Waveform Design and SNR-Energy Tradeoff," *IEEE Commun. Lett.*, vol. 21, no. 10, pp 2234-2237, Oct 2017.

[38] Z.B. Zawawi, Y. Huang and B. Clerckx, "Multi-User Wirelessly Powered Backscatter Communications: Non-linearity, Waveform Design and SINR-Energy Tradeoff," *IEEE Trans. on Wireless Commun.*, vol. 18, no. 1, pp. 241 – 253, Jan 2019.

[39] C. Boyer and S. Roy, "Backscatter Communication and RFID: Coding, Energy, and MIMO Analysis," *IEEE Trans. Commun.*, vol. 62, pp. 770-785, Mar. 2014.

[40] S. Bi, C. K. Ho, and R. Zhang, "Wireless Powered Communication: Opportunities and Challenges," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 117-125, Apr. 2015.

[41] H. Tabassum, E. Hossain, A. Ogundipe, and D. I. Kim, "Wireless Powered Cellular Networks: Key Challenges and Solution Techniques," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 63-71, Jun. 2015.

[42] K. Huang and X. Zhou, "Cutting the Last Wires for Mobile Communications by Microwave Power Transfer," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 86-93, Jun. 2015.

[43] S. Bi, Y. Zeng, and R. Zhang, "Wireless Powered Communication Networks: an Overview," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 10-18, Apr. 2016.

[44] I. Krikidis, S. Timotheou, S. Nikolaou, G. Zheng, D. W. K. Ng, and R. Schober, "Simultaneous Wireless Information and Power Transfer in Modern Communication Systems," *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 104-110, Nov. 2014.

[45] Z. Ding, C. Zhong, D. W. Ng, M. Peng, H. A. Suraweera, R. Schober, and H. V. Poor, "Application of Smart Antenna Technologies in Simultaneous Wireless Information and Power Transfer," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 86-93, Apr. 2015.

[46] X. Chen, Z. Zhang, H. H. Chen, and H. Zhang, "Enhancing Wireless Information and Power Transfer by Exploiting Multi-Antenna Techniques," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 133-141, Apr. 2015.

[47] X. Lu, P. Wang, D. Niyato, D. I. Kim, and Z. Han, "Wireless Networks with RF Energy Harvesting: a Contemporary Survey," *IEEE Commun. Surveys Tuts.*, pp. 757-789, 2015.

[48] K. Huang, C. Zhong, and G. Zhu, "Some New Research Trends in Wirelessly Powered Communications," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 19-27, Apr. 2016.

[49] S. Ulukus, A. Yener, E. Erkip, O. Simeone, M. Zorzi, P. Grover, and K. Huang, "Energy Harvesting Wireless Communications: A Review of Recent Advances," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 3, Mar. 2015.

[50] D. Niyato, D. I. Kim, M. Maso, and Z. Han, "Wireless Powered Communication Networks: Research Directions and Technological Approaches," *IEEE Wireless Commun.*, vol. 24, no. 6, pp. 88-97, Dec. 2017.

[51] E. Falkenstein, M. Roberg, and Z. Popovic, "Low-power wireless power delivery," *IEEE Trans. Microwave Theory Tech.*, vol. 60, no. 7, pp. 2277–2286, Jun. 2012.

[52] Z. Popovic, "Cut the cord: Low-power far-field wireless powering," *IEEE Microwave Mag.*, vol. 14, no. 2, pp. 55–62, Mar. 2013.

[53] H.J. Visser, R.J.M. Vullers, "RF Energy Harvesting and Transport for Wireless Sensor Network Applications: Principles and Requirements," *Proc. IEEE*, Vol. 101, No. 6, June 2013.

[54] Z. Popovic, E. A. Falkenstein, D. Costinett, and R. Zane, "Low-power far-field wireless powering for wireless sensors," *Proc. IEEE*, vol. 101, no. 6, pp. 1397–1409, Jun. 2013.

[55] C. Rapp, "Effects of HPA nonlinearity on a 4-DPSK/OFDM signal for a digital sound broadcasting system," in Proc. Eur. Conf. Satellite Commun., Belgium, vol. 1, pp. 179–184, Oct. 1991.

[56] C. Dehos and T. C. Schenk, "Digital compensation of amplifier nonlinearities in the receiver of a wireless system," in *14th IEEE Symposium on Communications and Vehicular Technology in the Benelux*, 2007.

[57] W. Brown, "The history of power transmission by radio waves," *IEEE Trans. Microw. Theory Tech.*, vol. 32, no. 9, pp. 1230–1242, Sep. 1984.

[58] B. Clerckx and J. Kim, "On the Beneficial Roles of Fading and Transmit Diversity in Wireless Power Transfer with Nonlinear Energy Harvesting," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7731 – 7743, Nov. 2018.

[59] H. Sun, Z. Zhong, and Y.-X. Guo, "An Adaptive Reconfigurable Rectifier for Wireless Power Transmission," *IEEE Microw. Wireless Comp. Lett.*,vol. 23, no. 9, Sept 2013.

[60] M. S. Trotter, J. D. Griffin, and G. D. Durgin, "Power-Optimized Waveforms for Improving the Range and Reliability of RFID Systems," in *Proc. 2009 IEEE International Conference on RFID*.

[61] A. S. Boaventura and N. B. Carvalho, "Maximizing DC Power in Energy Harvesting Circuits Using Multisine Excitation," in *Proc. 2011 IEEE MTT-S*.

[62] A. Collado and A. Georgiadis, "Optimal Waveforms for Efficient Wireless Power Transmission," *IEEE Microw. Wirel. Compon. Lett.*, May 2014.

[63] C. R. Valenta, M. M. Morys, and G. D. Durgin, "Theoretical Energy-Conversion Efficiency for Energy-Harvesting Circuits Under Power-Optimized Waveform Excitation," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 5, May 2015.

[64] B. Clerckx and E. Bayguzina, "Waveform Design for Wireless Power Transfer," *IEEE Trans. Signal Process.*, vol. 64, no. 23, Dec 2016.

[65] B. Clerckx, E. Bayguzina, D. Yates, and P. D. Mitcheson, "Waveform Optimization for Wireless Power Transfer with Nonlinear Energy Harvester Modeling," in *Proc. IEEE ISWCS 2015*, Aug. 2015, Brussels.

[66] B. Clerckx, "Wireless Information and Power Transfer: Nonlinearity, Waveform Design and Rate-Energy Tradeoff," *IEEE Trans. Signal Process.*, vol 66, no 4, pp 847-862, Feb. 2018.

[67] Y. Huang and B. Clerckx, "Large-Scale Multi-Antenna Multi-Sine Wireless Power Transfer," *IEEE Trans. Signal Process.*, vol. 65, no. 21, pp 5812-5827, Nov. 2017.

[68] B. Clerckx and E. Bayguzina, "A Low-Complexity Adaptive Multisine Waveform Design for Wireless Power Transfer," *IEEE Ant. Wireless Prop. Lett.*, Vol. 16, pp 2207 – 2210, 2017.

[69] M. R. V. Moghadam, Y. Zeng, and R. Zhang, "Waveform Optimization for Radio-Frequency Wireless Power Transfer," in *Proc. IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC)*, 2017.

[70] M. Varasteh, E. Piovano and B. Clerckx, "A Learning Approach to Wireless Information and Power Transfer Signal and System Design," IEEE ICASSP 2019.

[71] R. Morsi, V. Jamali, D. W. K. Ng, and R. Schober, "On the Capacity of SWIPT Systems with a Nonlinear Energy Harvesting Circuit," in *Proc. IEEE Int. Conf. Commun.* 2018.

[72] E. Boshkovska, D. W. K. Ng, N. Zlatanov, and R. Schober, "Practical Non-Linear Energy Harvesting Model and Resource Allocation for SWIPT Systems," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2082-2085, Dec 2015.

[73] J. Kim, B. Clerckx, and P.D. Mitcheson "Signal and System Design for Wireless Power Transfer: Prototyping, Experiment and Validation," *IEEE Trans. on Wireless Comm.*, vol. 19, no. 11, pp. 7453-7469, Nov. 2020.

[74] J. Kim and B. Clerckx, "Range Expansion for Wireless Power Transfer: A Joint Beamforming and Waveform Architecture," https://arxiv.org/abs/2010.01680.

[75] S. Shen and B. Clerckx, "Beamforming Optimization for MIMO Wireless Power Transfer with Nonlinear Energy Harvesting: RF Combining versus DC Combining," *IEEE Trans. on Wireless Comm*, in press.

[76] G. Ma, J. Xu, Y. Zeng, and M. R. V. Moghadam, "A generic receiver architecture for MIMO wireless power transfer with nonlinear energy harvesting," IEEE Signal Processing Letters, vol. 26, no. 2, pp. 312–316, Feb 2019.

[77] K. W. Choi, L. Ginting, P. A. Rosyady, A. A. Aziz, and D. I. Kim, "Wireless-Powered Sensor Networks: How to Realize," *IEEE Trans. Wireless Commun.*, vol. 16, pp. 221-234, Jan. 2017.

[78] K. W. Choi, P. A. Rosyady, L. Ginting, A. A. Aziz, D. Setiawan, and D. I. Kim, "Theory and Experiment for Wireless-Powered Sensor Networks: How to Keep Sensors Alive," *IEEE Trans. Wireless Commun.*, vol. 17, pp. 430-444, Jan. 2018.

[79] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116 753–116 773, 2019.

[80] S. Gong, X. Lu, D. T. Hoang, D. Niyato, L. Shu, D. I. Kim, and Y.-C. Liang, "Towards smart radio environment for wireless communications via intelligent reflecting surfaces: A comprehensive survey," arXiv preprint arXiv:1912.07794, 2019.

[81] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface aided wireless communications: A tutorial," arXiv preprint arXiv:2007.02759, 2020.

[82] S. Shen, B. Clerckx, and R. Murch, "Modeling and Architecture Design of Intelligent Reflecting Surfaces using Scattering Parameter Network Analysis," arXiv:2011.11362.

[83] C. Pan, H. Ren, K. Wang, M. Elkashlan, A. Nallanathan, J. Wang, and L. Hanzo, "Intelligent reflecting surface aided MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE Journal on Selected Areas in Communications*, 2020.

[84] Q. Wu and R. Zhang, "Weighted sum power maximization for intelligent reflecting surface aided SWIPT," *IEEE Wireless Communications Letters*, 2019.

[85] Y. Tang, G. Ma, H. Xie, J. Xu, and X. Han, "Joint transmit and reflective beamforming design for IRS-assisted multiuser MISO SWIPT systems," arXiv:1910.07156, 2019.

[86] Q. Wu and R. Zhang, "Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints," arXiv:1910.06220, 2019.

[87] Zhao et al., "Intelligent Reflecting Surface-Aided SWIPT: Joint Waveform, Active and Passive Beamforming Design," arXiv:2012.05646.

[88] Feng et al., "Waveform and Beamforming Design for Intelligent Reflecting Surface Aided Wireless Power Transfer," in submission.

[89] S. Shen and B. Clerckx, "Joint Waveform and Beamforming Optimization for MIMO Wireless Power Transfer," arXiv:2009.07500

[90] J. Xu and R. Zhang, "Energy beamforming with one-bit feedback," *IEEE Trans. Signal Process.*, vol. 62, no. 20, pp. 5370-5381, Oct., 2014.

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

[92] Y. Zeng and R. Zhang, "Optimized training for net energy maximization in multi-antenna wireless energy transfer over frequency-selective channel," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2360-2373, Jun. 2015.

[93] J. Xu and R. Zhang, "A general design framework for MIMO wireless energy transfer with limited feedback," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2475-2488, May 2016.

[94] S. Abeywickrama, T. Samarasinghe, C. K. Ho, and C. Yuen, "Wireless Energy Beamforming Using Received Signal Strength Indicator Feedback," *IEEE Trans. Signal Process.*, vol. 66, no. 1, pp. 224-235, Jan. 2018.

[95] K. W. Choi, D. I. Kim, and M. Y. Chung, "Received Power-Based Channel Estimation for Energy Beamforming in Multiple-Antenna RF Energy Transfer System," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1461-1476, Mar. 2017.

[96] Y. Huang and B. Clerckx, "Waveform Design for Wireless Power Transfer with Limited Feedback," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp 415-429, Jan. 2018.

[97] G. Yang, C. K. Ho, and Y. L. Guan, "Multi-Antenna Wireless Energy Transfer for Backscatter Communication Systems," *IEEE J. Sel. Areas Commun.*, Vol. 33, No. 12, Dec 2015.

[98] S. Shen, J. Kim, and B. Clerckx, "Closed-Loop Wireless Power Transfer with Adaptive Waveform and Beamforming: Design, Prototype and Experiment," in submission.

[99] O. L. A. López, H. Alves, R. D. Souza, and S. Montejo-Sánchez, "Statistical Analysis of Multiple Antenna Strategies for Wireless Energy Transfer," *IEEE Trans. Wireless Commun.*, vol. 67, no. 10, pp. 7245–7262, Oct. 2019.

[100] M.-L. Ku, Y. Han, H.-Q. Lai, Y. Chen, and K. J. Ray Liu, "Power waveforming: Wireless power transfer beyond time-reversal," *IEEE Trans. Signal Process.*, vol. 64, no. 22, pp. 5819–5834, Nov. 2016.

[101] M.-L. Ku, Y. Han, B. Wang, and K. J. R. Liu, "Joint Power Waveforming and Beamforming for Wireless Power Transfer," *IEEE Trans. Signal Process.*, vol. 65, no. 24, pp. 6409-6422, Dec. 2017.

[102] R. Y. Miyamoto and T. Itoh, "Retrodirective arrays for wireless communications," *IEEE Microw. Mag.*, vol. 3, no. 1, pp. 71–79, Mar. 2002.

[103] C. Y. Pon, "Retrodirective array using the heterodyne technique," *IEEE Trans. Antennas Propag.*, vol. 12, no. 2, pp. 176–180, Mar. 1964.

[104] S. Lee, Y. Zeng, and R. Zhang, "Retrodirective multi-user wireless power transfer with massive MIMO," IEEE Wireless Communications Letters, vol. 7, no. 1, pp. 54–57, 2017.

[105] Y.-J. Ren and K. Chang, "New 5.8-GHz circularly polarized retrodirective rectenna arrays for wireless power transmission," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 7, pp. 2970–2976, Jul. 2006.

[106] Y. Li and V. Jandhyala, "Design of retrodirective antenna arrays for short-range wireless power transmission," *IEEE Trans. Antennas Propag.*, vol. 60, no. 1, pp. 206–211, Jan. 2012.

[107] X. Wang, S. Sha, J. He, L. Guo, and M. Lu, "Wireless power delivery to low-power mobile devices based on retro-reflective beamforming," *IEEE Antennas Wireless Propag. Lett.*, vol. 13, pp. 919–922, May 2014

[108] K. W. Kim, H. S. Lee, and J. W. Lee, "Waveform design for fair wireless power transfer with multiple energy harvesting devices," *IEEE J. Sel. Areas Commun.*, vol. 37, pp. 34–47, Jan. 2019.

[109] K. W. Kim, H. S. Lee, and J. W. Lee, "Opportunistic waveform scheduling for wireless power transfer with multiple devices," *IEEE Trans. Wireless Commun.*, vol. 19, pp. 5651–5665, Sep. 2020.

[110] S. Shen, J. Kim, C. Song, and B. Clerckx, "Wireless Power Transfer with Distributed Antennas: System Design, Prototype, and Experiments," *IEEE Trans. on Industrial Electronics*, in press.

[111] K. W. Choi, A. A. Aziz, D. Setiawan, N. M. Tran, L. Ginting, and D. I. Kim, "Distributed Wireless Power Transfer System for Internet-of-Things Devices," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2657 - 2671, Aug. 2018.

[112] J. Zhou, Q. Zhang, Q. Li, and J. Qin, "Effect of oscillator jitters on distributed energy beamforming for wireless energy transfer," *IEEE Commun. Letters*, vol. 19, no. 11, pp. 2045–2048, 2015.

[113] S. Lee and R. Zhang, "Distributed wireless power transfer with energy feedback," *IEEE Trans. Signal Process.*, vol. 65, no. 7, pp. 1685–1699, 2017.

[114] C. Zhang and G. Zhao, "On the deployment of distributed antennas of power beacon in wireless power transfer," *IEEE Access*, vol. 6, pp. 7489–7502, 2018.

[115] Y. Huang, Y. Liu, and G. Y. Li, "Energy efficiency of distributed antenna systems with wireless power transfer," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 1, pp. 89–99, 2019.

[116] A. Salem, L. Musavian, and K. A. Hamdi, "Wireless power transfer in distributed antenna systems," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 737–747, 2019.

[117] L. R. Varshney, *Unreliable and Resource-Constrained Decoding*, Ph.D. thesis, Massachusetts Institute of Technology, 2010.

[118] M. Varasteh, B. Rassouli, and B. Clerckx, "Wireless Information and Power Transfer over an AWGN channel: Nonlinearity and Asymmetric Gaussian Signaling," in *Proc. IEEE Inf. Theory Wkshp.* 2017.

[119] M. Varasteh, B. Rassouli, and B. Clerckx, "On Capacity-Achieving Distributions for Complex AWGN Channels under Nonlinear Power Constraints and their Applications to SWIPT," *IEEE Trans. on Inf. Theory*, vol. 66, no. 10, pp. 6488-6508, Oct. 2020.

[120] H. H. Jang, K. W. Choi, and D.I. Kim, "Novel Frequency-Splitting SWIPT for Overcoming Amplifier Nonlinearity," *IEEE Wireless Communications Letters*, vol. 9, no. 6, pp. 826-829, June 2020.

[121] I. Krikidis, "Information-energy capacity region for swipt systems with power amplifier nonlinearity," arXiv preprint arXiv:2004.05428, 2020.

[122] M. Varasteh, B. Rassouli, H. Joudeh, and B. Clerckx, "SWIPT Signalling over Complex AWGN Channels with Two Nonlinear Energy Harvester Models," in *Proc. Int Symp. Inf. Theory*, 2018. 2018.

[123] R. Morsi, V. Jamali, A. Hagelauer, D. W. K. Ng, and R. Schober, "Conditional capacity and transmit signal design for SWIPT systems with multiple nonlinear energy harvesting receivers," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 582–601, Jan 2020.

[124] M. Varasteh, B. Rassouli, and B. Clerckx, "SWIPT Signaling over Frequency-Selective Channels with a Nonlinear Energy Harvester: Non-Zero Mean and Asymmetric Inputs," *IEEE Trans. on Commun.*, vol. 67, no. 10, pp. 7195-7210, Oct 2019.

[125] E. Bayguzina and B. Clerckx, "Asymmetric Modulation Design for Wireless Information and Power Transfer with Nonlinear Energy Harvesting," *IEEE Trans. on Wireless Comm.*, vol 18, no 12, Dec 2019.

[126] D. I. Kim, J. H. Moon, and J. J. Park, "New SWIPT using PAPR: How it works," *IEEE Wireless Commun. Lett.*, vol. 5, no. 6, Dec. 2016.

[127] S. Basu and L. R. Varshney, "Polar Codes for Simultaneous Information and Energy Transmission," in *Proc. IEEE 20th Int. Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*," Jul. 2019.

[128] Á. I. Barbero, E. Rosnes, G. Yang, and O. Ytrehus, "Constrained codes for passive RFID communication," in *Proc. 2011 Inf. Theory Appl. Workshop*, Feb. 2011.

[129] A. M. Fouladgar, O. Simeone, and E. Erkip, "Constrained codes for joint energy and information transfer," *IEEE Trans. Commun.*, vol. 62, no. 6, pp. 2121–2131, Jun. 2014.

[130] A. Tandon, M. Motani, and L. R. Varshney, "On code design for simultaneous energy and information transfer," in *Proc. 2014 Inf. Theory Appl. Workshop*, Feb. 2014.

[131] ——, "Are RLL codes suitable for simultaneous energy and information transfer?" arXiv:1807.09040 [cs.IT]., Jul. 2018.

[132] ——, "Subblock-constrained codes for real-time simultaneous energy and information transfer," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 4212–4227, Jul. 2016.

[133] A. Tandon, H. M. Kiah, and M. Motani, "Bounds on the size and asymptotic rate of subblock-constrained codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 10, pp. 6604–6619, Oct. 2018.

[134] T.-Y. Wu, A. Tandon, L. R. Varshney, and M. Motani, "Skip-sliding window codes," *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2018, pp. 1799–1803.

[135] T.-Y. Wu, A. Tandon, M. Motani, and L. R. Varshney, "Outage-Constrained Rate of Skip-Sliding Window Codes," *IEEE Trans. Green Commun. Netw.*, vol. 4, no. 2, pp. 506–514, Jun. 2020.

[136] J. H. van Lint, *Introduction to Coding Theory*. Springer, 1998.

[137] M. Dabirnia and T. M. Duman, "On code design for joint energy and information transfer," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2677–2688, Jun. 2016.

[138] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7829–7838, Dec. 2013.

[139] D. Sutter, J. M. Renes, and F. Dupuis, "Achieving the capacity of any dmc using only polar codes," in *Proc. IEEE Inf. Theory Workshop (ITW'12)*, Sep. 2012, pp. 114–118.

[140] F. Bolos, J. Blanco, A. Collado, and A. Georgiadis, "RF energy harvesting from multi-tone and digitally modulated signals," *IEEE Trans. Microwave Theory Tech.*, vol. 64, no. 6, pp. 1918–1927, June 2016.

[141] M. H. Ouda, P. Mitcheson and B. Clerckx, "Optimal Operation of Multi-Tone Waveforms in Low RF-Power Receivers," in *Proc. IEEE Wireless Power Transfer Conference (WPTC)*, 2018.

[142] N. Pan, D. Belo, M. Rajabi, D. Schreurs, N. B. Carvalho and S. Pollin, "Bandwidth Analysis of RF-DC Converters Under Multisine Excitation," *IEEE Trans. on Microwave Theory and Techn.*, vol. 66, no. 2, pp. 791-802, Feb. 2018.

[143] J. J. Fahs and I. C. Abou-Faycal, "Using hermite bases in studying capacity-achieving distributions over AWGN channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5302–5322, Aug. 2012.

[144] Y. Lomnitz and M. Feder, "Communication over individual channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7333–7358, Nov. 2011.

[145] T. J. O'Shea, K. Karra, and T. C. Clancy, "Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention," in *2016 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, Dec 2016, pp. 223–228.

[146] T. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563-575, Dec. 2017.

[147] M. Ibnkahla, "Applications of neural networks to digital communications - a survey," Signal Processing, vol. 80, no. 7, pp. 1185 – 1215, 2000.

[148] R. Al-Rfou and et al, "Theano: A python framework for fast computation of mathematical expressions," CoRR, vol. abs/1605.02688, 2016. [Online]. Available: http://arxiv.org/abs/1605.02688

[149] M. Abadi and et al, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," CoRR, vol. abs/1603.04467, 2016. [Online]. Available: http://arxiv.org/abs/1603.04467

[150] T. Chen and et al, "Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems," CoRR, vol. abs/1512.01274, 2015. Available: http://arxiv.org/abs/1512.01274.

[151] E. Nachmani, Y. Beery, and D. Burshtein, "Learning to decode linear codes using deep learning," CoRR, vol. abs/1607.04793, 2016. Available: http://arxiv.org/abs/1607.04793

[152] E. Nachmani, E. Marciano, D. Burshtein, and Y. Beery, "RNN decoding of linear block codes," CoRR, vol. abs/1702.07560, 2017. Available: http://arxiv.org/abs/1702.07560

[153] T. Gruber, S. Cammerer, J. Hoydis, and S. ten Brink, "On deep learning-based channel decoding," CoRR, vol. abs/1701.07738, 2017. Available: http://arxiv.org/abs/1701.07738

[154] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Unsupervised representation learning of structured radio communication signals," CoRR, vol. abs/1604.07078, 2016. Available: http://arxiv.org/abs/1604.07078

[155] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive mimo csi feedback," *IEEE Wireless Communications Letters*, 2018.

[156] B. Karanov, M. Chagnon, F. Thouin, T. A. Eriksson, H. Bulow, D. Lavery, P. Bayvel, and L. Schmalen, "End-to-end deep learning of optical fiber communications," arXiv:1804.04097, 2018.

[157] R. T. Jones, T. A. Eriksson, M. P. Yankov, D. Zibar, "Deep Learning of Geometric Constellation Shaping including Fiber Nonlinearities," arXiv:1805.03785, 2018.

[158] Q.-J. Zhang, K. C. Gupta, and V. K. Devabhaktuni, "Artificial Neural Networks for RF and Microwave Design—From Theory to Practice," *IEEE Trans. on Microwave Theory and Techniques*, Vol. 51, No. 4, April 2003.

[159] M. Varasteh, J. Hoydis, and B. Clerckx, "Learning Modulation Design for SWIPT with Nonlinear Energy Harvester: Large and Small Signal Power Regimes," IEEE SPAWC 2019.

[160] M. Varasteh, J. Hoydis, and B. Clerckx, "Learning to Communicate and Energize: Modulation, Coding and Multiple Access Designs for Wireless Information-Power Transmission," *IEEE Trans. on Commun.*, vol. 68, no. 11, pp. 6822-6839, Nov. 2020.

[161] J.-M. Kang , C.-J. Chun, and I.-M. Kim, "Deep-Learning-Based Channel Estimation for Wireless Energy Transfer," *IEEE Commun. Letters*, Vol. 22, No. 11, pp. 2310-2313, Nov 2018.

[162] J.-M. Kang, "Reinforcement Learning Based Adaptive Resource Allocation for Wireless Powered Communication Systems," *IEEE Commun. Letters*, vol. 24, no. 8, pp. 1752-1756, Aug. 2020.

[163] A. Karpatne, W. Watkins, J. Read, and V. Kumar, "Physics-guided Neural Networks (PGNN): An Application in Lake Temperature Modeling," arXiv:1710.11431 [cs.LG], Oct. 2017.

[164] X. Jia, J. Willard, A. Karpatne, J. Read, J. Zwart, M. Steinbach, and V. Kumar, "Physics guided RNNs for modeling dynamical systems: A case study in simulating lake temperature profiles," in *Proc. 2019 SIAM Int. Conf. Data Mining*, pp. 558-566, May 2019.

[165] J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar, "Integrating Physics-Based Modeling with Machine Learning: A Survey," arXiv:2003.04919 [physics.comp-ph], Mar. 2020.

[166] S.-M. Udrescu and M. Tegmark, "AI Feynman: A Physics-Inspired Method for Symbolic Regression," *Science Advances*, vol. 6, no. 16, eaay2631, Apr. 2020.

[167] S.-M. Udrescu, A. Tan, J. Feng, O. Neto, T. Wu, and M. Tegmark, "AI Feynman 2.0: Pareto-Optimal Symbolic Regression Exploiting Graph Modularity," in *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, 2020.

[168] N. Kshetry and L. R. Varshney, "Safety in the Face of Unknown Unknowns: Algorithm Fusion in Data-driven Engineering Systems," IEEE ICASSP 2019, pp. 8162-8166.

[169] C. Hillar and F. Sommer, "Comment on the article 'Distilling free-form natural laws from experimental data'," arXiv:1210.7273 [nlin.AO], Oct. 2012.

[170] M. Cranmer, A. Sanchez-Gonzalez, P. Battaglia, R. Xu, K. Cranmer, D. Spergel, and S. Ho, "Discovering Symbolic Models from Deep Learning with Inductive Biases," in *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, 2020.

[171] H. Yu, T. Li, and L. R. Varshney, "Probabilistic Rule Realization and Selection," in *Advances in Neural Information Processing Systems 30*, 2017.

[172] H. Yu, I. Mineyev, and L. R. Varshney, "A Group-Theoretic Approach to Computational Abstraction: Symmetry-Driven Hierarchical Clustering," arXiv:1807.11167 [cs.LG], Jul. 2018.

[173] M. Swan, "Sensor mania! The internet of things, wearable computing, objective metrics, and the quantified self 2.0," *J. Sens. Actuator Networks*, vol. 1, pp. 217–253, 2012.

[174] K. Kumar, J. Liu, Y.-H. Lu, and B. Bhargava, "A survey of computation offloading for mobile systems," *Mobile Networks and Applications*, vol. 18, no. 1, pp. 129–140, 2013.

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

[176] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective," *IEEE Commun. Surveys & Tutorials*, vol. 19, no. 4, pp. 2322-2358, 2017.

[177] F. Yao, A. Demers, and S. Shenker, "A scheduling model for reduced CPU energy," in *Proc. IEEE Ann. Symp. Foundations of Computer Science*, pp. 374–382, 1995.

[178] L. Benini, A. Bogliolo, G. Paleologo, and G. De Micheli, "Policy optimization for dynamic power management," *IEEE Trans. Computer-Aided Design of Integrated Cir. and Sys.*, vol. 18, no. 6, pp. 813–833, 1999.

[179] P. Pillai and K. G. Shin, "Real-time dynamic voltage scaling for low-power embedded operating systems," in *Proc. ACM SIGOPS Operating Sys. Review*, vol. 35, pp. 89–102, 2001.

[180] J. R. Lorch and A. J. Smith, "Improving dynamic voltage scaling algorithms with pace," in *Proc. ACM SIGMETRICS*, vol. 29, pp. 50–61, Jun. 2001.
[181] W. Yuan and K. Nahrstedt, "Energy-efficient soft real-time CPU scheduling for mobile multimedia systems," *ACM Trans. Computer Sys.*, vol. 37, no. 5, pp. 149–163, 2003.
[182] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, "Energy-optimal mobile cloud computing under stochastic wireless channel," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4569–4581, Sep. 2013.
[183] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Comput. netw.*, vol. 38, no. 4, pp. 393–422, 2002.
[184] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: Current state and future challenges," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 32–39, 2011.
[185] T. D. Burd and R. W. Brodersen, "Processor design for portable systems," *J. VLSI Signal Process*, pp. 203–221, Aug. 1996.
[186] B. Prabhakar, E. Uysal Biyikoglu, and A. El Gamal, "Energy-efficient transmission over a wireless link via lazy packet scheduling," in *Proc. IEEE INFOCOM*, vol. 1, pp. 386–394, 2001.
[187] C. You, K. Huang and H. Chae, "Energy Efficient Mobile Cloud Computing Powered by Wireless Energy Transfer," *IEEE JSAC*, vol. 34, No. 5, pp 1757-1771, May 2016.
[188] X. Li, C. You, S. Andreev, Y. Gong, and K. Huang, "Wirelessly Powered Crowd Sensing: Joint Power Transfer, Sensing, Compression, and Transmission," *IEEE JSAC*, vol. 37, no. 2, pp. 391-406, 2019.
[189] Q. Zeng, Y. Du, and K. Huang, "Wirelessly Powered Federated Edge Learning: Optimal Tradeoffs Between Convergence and Power Transfer," *submitted to IEEE Trans. on Wireless Commun.*, 2021.
[190] A. V. D. Ven, "Linux OS data compression options: Comparing behavior," *https://clearlinux.org/blogs/linux-os-data-compression-options-comparing-behavior*, 2017.
[191] A. P. Chandrakasan, S. Sheng, and R. W. Brodersen, "Low-power CMOS digital design," *IEEE J. Solid-State Circuits*, vol. 27, no. 4, pp. 473–484, 1992.
[192] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing," in *Proc. MobiCom*, 2012, pp. 173–184.
[193] J. C. Russ, *The Image Processing Handbook*. CRC press, 2016.
[194] N. M. Nasrabadi, "Pattern recognition and machine learning," *Journal of Electronic Imaging*, vol. 16, no. 4, p. 049901, 2007.
[195] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Int. Conf. Artif. Intell. Statist. (AISTATS)*, Fort Lauderdale, USA, Apr 20-22, 2017.
[196] G. Zhu, D. Liu, Y. Du, C. You, J. Zhang, and K. Huang, "Toward an intelligent edge: Wireless communication meets machine learning," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 19–25, 2020.
[197] W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y. C. Liang, Q. Yang, D. Niyato, and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 2031–2063, 2020.



**Bruno Clerckx** is a (Full) Professor, the Head of the Wireless Communications and Signal Processing Lab, and the Deputy Head of the Communications and Signal Processing Group, within the Electrical and Electronic Engineering Department, Imperial College London, London, U.K. He received the M.S. and Ph.D. degrees in applied science from the Université Catholique de Louvain, Louvain-la-Neuve, Belgium, in 2000 and 2005, respectively. From 2006 to 2011, he was with Samsung Electronics, Suwon, South Korea, where he actively contributed to 4G (3GPP LTE/LTE-A and IEEE 802.16m) and acted as the Rapporteur for the 3GPP Coordinated Multi-Point (CoMP) Study Item. He has authored two books, over 200 peer-reviewed international research papers, and 150 standards contributions, and is the inventor of 80 issued or pending patents among which 15 have been adopted in the specifications of 4G standards and are used by billions of devices worldwide. His research area is communication theory and signal processing for wireless networks. He has been a TPC member, a symposium chair, or a TPC chair of many symposia on communication theory, signal processing for communication and wireless communication for several leading international IEEE conferences. He was an Elected Member of the IEEE Signal Processing Society SPCOM Technical Committee. He served as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He has also been a (lead) guest editor for special issues of the EURASIP Journal on Wireless Communications and Networking, IEEE ACCESS, the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING. He was an Editor for the 3GPP LTE-Advanced Standard Technical Report on CoMP. He is an IEEE Distinguished Lecturer of the IEEE Communications Society.



**Kaibin Huang** (Fellow, IEEE) received the B.Eng. and M.Eng. degrees from the National University of Singapore, and the Ph.D. degree from The University of Texas at Austin, all in electrical engineering. Presently, he is an Associate Professor and an Associate Head in the Dept. of Electrical and Electronic Engineering at The University of Hong Kong, Hong Kong. He received the IEEE Communication Society's 2021 Best Survey Paper, 2019 Best Tutorial Paper, 2019 Asia Pacific Outstanding Paper, 2015 Asia Pacific Best Paper Awards as well as Best Paper Awards at IEEE GLOBECOM 2006 and IEEE/CIC ICCC 2018. Moreover, he received an Outstanding Teaching Award from Yonsei University in S. Korea in 2011. He has served as the lead chairs for the Wireless Comm. Symp. of IEEE Globecom 2017 and the Comm. Theory Symp. of IEEE GLOBECOM 2014 and the TPC Co-chairs for IEEE PIMRC 2017 and IEEE CTW 2013. He is an Executive Editor for IEEE Transactions on Wireless Communications, an Associate Editor for Journal on Selected Areas in Communications (JSAC), and an Area Editor for IEEE Transactions on Green Communications and Networking. Previously, he has also served on the editorial board of IEEE Wireless Communications Letters. Moreover, he has guest edited special issues for IEEE JSAC, IEEE Journal on Selected Topics on Signal Processing, and IEEE Communications Magazine. He is an IEEE Fellow and an IEEE Distinguished Lecturer of both the IEEE Communications and Vehicular Technology Societies. He has been named a Highly Cited Researcher by Clarivate Analytics in 2019 and 2020.



**Lav R. Varshney** (Senior Member, IEEE) received the B.S. degree (magna cum laude) with honors in electrical and computer engineering from Cornell University, Ithaca, NY, USA, in 2004, and the S.M., E.E., and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, in 2006, 2008, and 2010, respectively. He is currently an Associate Professor with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana–Champaign. From 2019 to 2020, he was on leave as a Principal Research Scientist with Salesforce Research, Palo Alto, CA, USA. From 2010 to 2013, he was a Research Staff Member with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA. His research interests include information and coding theory, statistical signal processing, neuroscience, and artificial intelligence. Dr. Varshney is a member of Eta Kappa Nu, Tau Beta Pi, and Sigma Xi. He was a recipient of the IBM Faculty Award in 2014 and a Finalist for the Bell Labs Prize in 2014 and 2016. He and his students have received several best paper awards. He currently serves on the advisory board of the AI XPRIZE.



**Sennur Ulukus** (Fellow, IEEE) received the B.S. and M.S. degrees in electrical and electronics engineering from Bilkent University and the Ph.D. degree in electrical and computer engineering from the Wireless Information Network Laboratory (WINLAB), Rutgers University. She is the Anthony Ephremides Professor of Information Sciences and Systems with the Department of Electrical and Computer Engineering, University of Maryland at College Park, where she also holds a joint appointment with the Institute for Systems Research (ISR). Prior to joining UMD, she was a Senior Technical Staff Member with AT&T Labs-Research. Her research interests are in information theory, wireless communications, machine learning, signal processing and networks, with recent focus on private information retrieval, age of information, distributed coded computation, machine learning for wireless, energy harvesting communications, physical layer security, and wireless energy and information transfer. She is a Distinguished Scholar-Teacher of the University of Maryland. She received the 2003 IEEE Marconi Prize Paper Award in Wireless Communications, the 2019 IEEE Communications Society Best Tutorial Paper Award, the 2005 NSF CAREER Award, the 2011 ISR Outstanding Systems Engineering Faculty Award, and the 2012 ECE George Corcoran Outstanding Teaching Award. She was a Distinguished Lecturer of the IEEE Information Theory Society from 2018 to 2019. She is an Area Editor for the IEEE Transactions on Wireless Communications (2019-present) and a Senior Editor for the IEEE Transactions on Green Communications and Networking (2020-present). She was an Area Editor for the IEEE Transactions on Green Communications and Networking (2016-2020), an Editor for the IEEE Journal on Selected Areas in Communications-Series on Green Communications and Networking (2015-2016), an Associate Editor for the IEEE Transactions on Information Theory (2007-2010), and an Editor for the IEEE Transactions on Communications (2003-2007). She was a Guest Editor for the IEEE Journal on Selected Areas in Communications (2008, 2015, 2021), Journal of Communications and Networks (2012), and the IEEE Transactions on Information Theory (2011). She is the TPC chair of 2021 IEEE Globecom, and was a TPC co-chair of 2019 IEEE ITW, 2017 IEEE ISIT, 2016 IEEE Globecom, 2014 IEEE PIMRC, and 2011 IEEE CTW.



**Mohamed-Slim Alouini (S'94-M'98-SM'03-F'09)** was born in Tunis, Tunisia. He received the Ph.D. degree in electrical engineering from the California Institute of Technology (Caltech), Pasadena, CA, USA, in 1998. He served as a Faculty Member for the University of Minnesota, Minneapolis, MN, USA, and also for Texas A&M University at Qatar, Education City, Doha, Qatar. In 2009, he joined as a Professor of electrical engineering with the King Abdullah University of Science and Technology (KAUST), Thuwal, Makkah Province, Saudi Arabia. His current research interest includes the modeling, design, and performance analysis of wireless communication systems.