On Capacity-Achieving Distributions for Complex AWGN Channels Under Nonlinear Power Constraints and their Applications to SWIPT

Morteza Varasteh, Borzoo Rassouli and Bruno Clerckx

Department of Electrical and Electronic Engineering, Imperial College London, London, U.K.

{m.varasteh12; b.rassouli12; b.clerckx}@imperial.ac.uk.

Abstract

The capacity of a complex and discrete time memoryless Additive White Gaussian Noise (AWGN) channel under three constraints, namely, input average power, input amplitude and delivered power at the output is studied. The delivered power constraint is modelled as a linear combination of even-moment statistics of the channel input being larger than a threshold. It is shown that the capacity of an AWGN channel under transmit average power and receiver delivered power constraints is the same as the capacity of an AWGN channel under an average power constraint, however, depending on the two constraints, it can be either achieved (via Circular Symmetric Complex Gaussian (CSCG) input) or arbitrarily approached (via time sharing between inputs with high amount of information, e.g. CSCG, and inputs with high amount of power, exhibiting a low probability of high amplitude signals). As an application, a simultaneous wireless information and power transfer (SWIPT) problem is studied, where an experimentally-validated nonlinear model of the harvester is used. It is shown that the delivered power depends on higher order statistics of the channel input. Two inner bounds, one based on complex Gaussian inputs and the other based on convexifying the optimization probability space, are obtained for the Rate-Power (RP) region. For Gaussian inputs, the optimal inputs are zero mean and a tradeoff between information and power is recognized by considering asymmetric power allocations between Inphase and Quadrature subchannels. Through numerical algorithms, it is observed that the numerically obtained input (NOI) distributions attain larger RP region compared to Gaussian input counterparts. The benefits of the newly developed and optimized input distributions are also confirmed and validated through realistic circuit simulations. The results reveal the crucial role played by the energy harvester nonlinearity on SWIPT and provide new engineering guidelines on how to exploit this nonlinearity in the design of SWIPT modulation, signal and architecture.

I. INTRODUCTION

Radio-Frequency (RF) waves can be utilized for transmission of both information and power simultaneously. As one of the primary works in the information theory literature, Varshney studied this problem in [2], in which he characterized the capacity-power function for a point-to-point discrete memoryless channel (DMC). He showed the existence of a tradeoff between the information rate and the delivered power for some channels, such as, point-to-point binary channels and amplitude constraint Gaussian channels. Recent results in the literature have also revealed that in many scenarios, there is a tradeoff between information rate and delivered power. Just to name a few, frequency-selective channel [3], MIMO broadcasting [4], interference channel [5].

One of the major efforts in a Simultaneous Wireless Information and Power Transfer (SWIPT) architecture is to increase the Direct-Current (DC) power at the output of the harvester without increasing transmit power. The harvester, known as rectenna, is composed of an antenna followed by a rectifier.[1] In [6], [7], it is shown that the RF-to-DC conversion efficiency is a function of rectenna's structure, as well as its input waveform (power and shape). Accordingly, in order to maximize rectenna's DC power output, a systematic waveform design is crucial to make the best use of an available RF spectrum [7]. In [7], an analytical model for the rectenna's output is introduced via the Taylor expansion of the diode characteristic function and a systematic design for multisine waveform is derived. The nonlinear model and the design of the waveform was validated using circuit simulations in [7], [8] and recently confirmed through prototyping and experimentation in [9]. Those works also confirm the inaccuracy and inefficiency of a linear model of the rectifier obtained by truncating the approximation of the diode characteristic function to the second order[2]. As one of the main conclusions, it is shown that the rectifier's nonlinearity is beneficial to the system performance and has a significant impact on the design of signals and systems involving wireless power.

The design of an efficient SWIPT architecture fundamentally relies on designing an efficient Wireless Power Transfer (WPT) structure as an important building block of SWIPT. The SWIPT literature has so far focused on the linear model of the rectifier, e.g., [3]–[5], whereas, it is expected that considering nonlinearity effect changes the SWIPT design, signalling and architecture significantly. Indeed, in [11], [12], the design of SWIPT waveforms and the characterization of achievable rate-power (RP) region are

---

[1]In the literature, the rectifier is usually considered as a nonlinear device (usually a diode) followed by a low-pass filter. The diode is the main source of nonlinearity induced in the system.

[2]The linear model has for consequence that the RF-to-DC conversion efficiency of the energy harvester is constant and independent of the harvester's input waveform (power and shape) [4], [10].

studied on deterministic Additive White Gaussian Noise (AWGN) channels accounting for the rectenna's nonlinearity with a power splitter at the receiver. In single-carrier transmission, it is shown that modulation with Circular Symmetric Complex Gaussian (CSCG) input is beneficial to wireless power delivery compared to an unmodulated continuous wave. In multi-carrier transmission, however, it is shown that a non-zero mean Gaussian input distribution leads to an enlarged RP region compared to a CSCG input distribution. This highlights that the choice of a suitable input distribution (and therefore modulation and waveform) for SWIPT is affected by the rectifier nonlinearity and motivates the study of the capacity of AWGN channels under nonlinear power constraints.

The capacity of complex and real, discrete-time memoryless AWGN channels has been investigated in the literature under various constraints, extensively. The most classical one is the channel input average power constraint, under which the optimal input is demonstrated to be Gaussian distributed [13]. It seems that the linear AWGN channel subject to transmit average power constraint is an exception and under many other constraints, the optimal input leads to discrete inputs. To mention a few, Smith in [14] considered a real AWGN channel with average power and amplitude constrained inputs, where he established that the optimal capacity achieving input distribution is discrete with a finite number of mass points. Similar results were reported in [15] for complex AWGN channels with average and peak-power constraint and in [16] for complex Rayleigh-fading channel when no channel state information (CSI) is assumed either at the receiver or the transmitter. As a more general result, in [17] a real channel is considered in which sufficient conditions for the additive noise are provided such that the support of the optimal bounded input has a finite number of mass points. In [18], real AWGN channels with nonlinear inputs are considered subject to multiple types of constraints such as the even-moment and/or compact-support constraints under which the optimal input is proved to be discrete with a finite number of mass points in the vast majority of the cases.

Nonlinearity has appeared in a number of applications including intermodulation distortion [19], optical channels [20], [21], magnetic recording [22], power amplifiers [23], and more recently in wireless energy harvesters (for WPT and SWIPT). In all those applications, the characterization of the capacity or capacity-achieving input distributions in nonlinear channels remains an open and challenging problem in general settings.

A survey of the literature reveals that almost all models considered for AWGN channels are not inclusive of the inevitable nonlinearities, such as fibre optic channels, power amplifiers or energy harvesters. The lack of fundamental results in the literature relating to nonlinear models is becoming more sensible due to the growth of applications involving devices with nonlinear responses. The typical and straightforward

approaches to tackle such problems are either considering linearized models or obtaining approximations and lower bounds on capacity [24]. As one of the novel works in the information theory literature, in [18], the authors consider a real AWGN channel with their focus on nonlinear channel inputs and different types of transmit power constraints.

Leveraging the aforementioned observations, we provide a step closer at identifying the fundamental limits of SWIPT structures taking into account the nonlinearities of the power harvester, i.e., rectenna. In this paper, we study a deterministic, complex and discrete time memoryless AWGN channel under the transmit average power and amplitude constraints as well as a constraint on the linear combination of even-moment statistics of the channel input. The contributions of this paper are listed below.

- First, we show that the capacity of an AWGN channel under a transmit average power constraint and receiver delivered power constraint is the same as the capacity of an AWGN channel. However, depending on the two constraints, the capacity can be either achieved using a unique CSCG input or approached arbitrarily (irrespectively of the delivered power constraint) using time sharing between signals with high information content and signals with high energy content.

- Second, we show that under an input amplitude constraint and receiver delivered power constraint, similarly to the results reported in [14]–[16] and [18], the capacity achieving input distribution is discrete in amplitude with a finite number of mass-points and with a uniformly distributed independent phase. The system model studied in this paper focuses on the nonlinearities at the receiver (over complex AWGN channels) and indeed can be considered as a reciprocal of [18], where the main focus was on the nonlinearities at the transmitter (transmit nonlinear constraints as well as nonlinear channel inputs over real AWGN channels).

- Third, as an application of the obtained results, we consider SWIPT over a complex AWGN channel, where the receiver is equipped with a rectenna in order to harvest and convert RF power into DC power. Taking the advantage of the small-signal approximation for rectenna's nonlinear output introduced in [7], [11], we obtain the general form of the delivered power for independent and identically distributed (iid) complex inputs in terms of system baseband parameters. Assuming that the receiver jointly extracts information and harvests power from the received RF signal,[3] it is shown that the delivered power at the receiver is dependent on the even-moment statistics of the channel input.

[3]We note that, leveraging the results in thermodynamics of computing, it is demonstrated that energy need not be dissipated in the decoding process. This is due to the reason that to perform a mathematical work, energy is not required [25, Ch. 5]. In particular, decoders that are reversible computational devices would not dissipate any energy [26] and electronic circuits that are almost thermodynamically reversible have been built [27]. Motivated by this, we also assume that at the receiver, the decoder is able to jointly harvest power and extract information from the received RF signal.

Defining RP region for the considered application, we obtain two inner bounds for the RP region. The first inner bound is based on merely iid complex Gaussian inputs, where we show that the optimal complex Gaussian inputs are zero mean. We also recognize a tradeoff between transmitted information and transferred power resulting from asymmetric power allocations between Inphase and Quadrature subchannels. The second inner bound is based on convexifying the optimization probability space and obtaining the necessary and sufficient condition for optimality for the convexified optimization space. Using numerical programming, it is observed that the Numerically Obtained Input (NOI) distributions outperform their Gaussian counterparts.

- Fourth, the analysis provides new engineering guidelines and refreshing views on the crucial role played by nonlinearity in SWIPT design. First, in contrast with the conventional linear model of the energy harvester for which CSCG inputs are capacity achieving under average power constraints [2], [3], CSCG inputs cannot achieve the optimal RP region boundaries in the presence of nonlinearity. Second, the energy harvester nonlinearity enlarges the RP region. Hence, in contrast with other systems subject to nonlinear responses, where nonlinearity is compensated (e.g. [20]), the nonlinearity in SWIPT is exploitable in the signal and system design and is beneficial to the system performance. Third, in contrast with the linear model for which time sharing between power and information transmission is suboptimal [4], time sharing between distributions with high amount of information, e.g. CSCG inputs, and distributions with high amount of power, reminiscent of flash signaling[4] and exhibiting a low probability of high amplitude signals, is sufficient to approach the capacity in the presence of nonlinearity. Fourth, the efficacy of the derived and optimized input distributions to boost the harvested DC power is validated and confirmed through realistic circuit simulations. This sheds light on a new form of signal design for WPT relying on (energy) modulation for single-carrier transmission, as an alternative to the multi-carrier (energy) waveform approach of [7].

- Fifth, as an independent result, we note that in analyzing complex AWGN channels, Bessel modified function of first kind of order zero appears frequently. Due to the lack of explicit expressions for Bessel functions in general, it is sometimes hard to analyze such channels. Accordingly, we obtain a tight upper bound on the Bessel modified function of first kind of order zero, which might also come useful in future applications and analysis.

Organization: In Section II, we introduce the system model and define the channel capacity problem studied here. In Section III, we introduce the main results of the paper. A SWIPT problem is considered in

---

[4]In general, flash signaling is the mixture of a probability distribution that asymptotically concentrates all its mass at 0 and a probability distribution that migrates to infinity; the weight of the latter vanishes sufficiently fast to satisfy the vanishing power constraint [28].

Section IV as an application of the main results introduced in Section III. In Section IV-A, the delivered power for the considered SWIPT problem is obtained in terms of channel baseband parameters for iid channel inputs accounting for small-signal approximations of rectenna. Defining the RP region in Section IV-B, an inner bound on the RP region based on complex Gaussian distributed inputs and an inner bound on the RP region based on the results developed in Section III are introduced in Section IV-B1 and Section IV-B2, respectively. In Section V, numerical results are illustrated in order to clarify the inner bounds obtained for the RP region. In Section VI, some problems are posed as potential future research directions. We conclude the paper in Section VII and the proofs for some of the results are provided in the Appendices at the end of the paper.

Notations: Throughout this paper, the standard CSCG distribution is denoted by $\mathcal{CN}(0, 1)$. Complex conjugate of a complex number c is denoted by $\bar{c}$. For a random process $X(t)$, the corresponding random variable at time index k is represented by $\boldsymbol{x}_k$. The support of the random variable $\boldsymbol{x}_k$ is denoted as $\text{supp}\{\boldsymbol{x}_k\}$. $\boldsymbol{x}_r$ and $\boldsymbol{x}_i$ denote the real and imaginary parts of the complex random variable $\boldsymbol{x}$, respectively. $\text{Re}\{\cdot\}$ and $\text{Im}\{\cdot\}$ are real and imaginary operators, respectively. We use the notations $\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$ and $s_l = \text{sinc}(l + 1/2)$ for integer l. $F_{\boldsymbol{x}}(x)$ and $f_{\boldsymbol{x}}(x)$ denote, respectively, the cumulative distribution function (cdf) and the probability density function (pdf) of the random variable $\boldsymbol{x}$. For the random process $X(t)$, the expectation over statistical randomness $\mathbb{E}[\cdot]$ and averaging over time $\mathcal{E}[\cdot]$ is defined as

$$\mathbb{E}[X(t)] = \int_{-\infty}^{\infty} x(t) dF_{\boldsymbol{X}(t)}(x), \tag{1}$$

$$\mathcal{E}[X(t)] = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} X(t) dt, \tag{2}$$

respectively. $\Phi(\cdot, \cdot; \cdot)$ denotes the confluent hypergeometric function defined as in [29, Section 9.21]. We define the kernel $K(R, r)$ as

$$K(R, r) \triangleq R e^{-\frac{R^2 + r^2}{2}} I_0(rR), \tag{3}$$

where $I_0(x) = 1/\pi \int_0^\pi e^{x \cos(\theta)} d\theta$ is the modified Bessel function of the first kind and order zero. The error function is defined as $\text{erf}(x) = 2/\sqrt{\pi} \int_0^x e^{-t^2} dt$. The Heaviside step function is denoted by $U(x)$.

II. SYSTEM MODEL, PROBLEM DEFINITION AND PRELIMINARIES

Consider the following complex representation of a discrete-time AWGN channel,

$$\boldsymbol{y}_k = \boldsymbol{x}_k + \boldsymbol{n}_k, \tag{4}$$

where $\{\boldsymbol{y}_k\}$, $\{\boldsymbol{x}_k\}$ and $\{\boldsymbol{n}_k\}$ represent the sequences of complex-valued samples of the channel output, input and AWGN, respectively, and k is the discrete-time index. The real and imaginary parts of the signal

$\{\boldsymbol{y}_k\}$ indicate the Inphase and Quadrature components, respectively. The noise samples $\{\boldsymbol{n}_k\}$ are assumed to be CSCG distributed as $\mathcal{CN}(0,2)$, i.e., $\mathbb{E}[\text{Re}\{\boldsymbol{n}_k\}^2] = \mathbb{E}[\text{Im}\{\boldsymbol{n}_k\}^2] = 1$ and $\mathbb{E}[\text{Re}\{\boldsymbol{n}_k\}\text{Im}\{\boldsymbol{n}_k\}] = 0$.

We are interested in the capacity of the channel in (4) with input samples subject to

$$
\begin{cases}
\mathbb{E}[|\boldsymbol{x}_k|^2] \leq P_a \\
P_d \leq \mathbb{E}[g(|\boldsymbol{x}_k|)] \quad , \\
|\boldsymbol{x}_k| \leq r_p
\end{cases}
\tag{5}
$$

for all k, where throughout the paper $P_a < \infty$, $P_d < \infty$ and $r_p \leq \infty$ are interpreted as the transmitter maximum allowable average power, minimum delivered power and channel input amplitude constraints, respectively. $g(\cdot)$ is assumed to be a continuous positive function having the form of

$$
g(r) = \sum_{i=0}^{n} \alpha_i r^{2i}, \;\; r \geq 0,
\tag{6}
$$

where $n \geq 2$ is an arbitrary integer. Note that since $g(r)$ is assumed to be a positive function, we have $\alpha_n > 0$, and hence, $\lim_{r \to \infty} g(r) = \infty$.

Remark 1. *The scenario $g(r) = \alpha_0 + \alpha_1 r^2$ is not considered in this paper, as the capacity problem in (5) boils down to either [13] (when $r_p = \infty$), where a CSCG distribution is optimal, or [15] (when $r_p < \infty$), where optimal distribution is discrete with a finite number of mass points. Accordingly, we are interested in $g(r)$ with $\alpha_i \neq 0$ for at least one of $i = 2, \ldots, n$.*

The capacity of a discrete-time complex AWGN channel [30, Chapter 7] is therefore given by

$$
C(P_a, P_d, r_p) = \sup_{f_{\boldsymbol{x}}(x)} \; I(\boldsymbol{x}; \boldsymbol{y})
\tag{7}
$$

$$
\text{s.t.} \quad
\begin{cases}
\mathbb{E}[|\boldsymbol{x}|^2] \leq P_a, \\
P_d \leq \mathbb{E}[g(|\boldsymbol{x}|)], \\
|\boldsymbol{x}| \leq r_p,
\end{cases}
$$

By expressing $I(\boldsymbol{x}; \boldsymbol{y})$ in terms of differential entropies, i.e., $I(\boldsymbol{x}; \boldsymbol{y}) = h(\boldsymbol{y}) - \ln 2\pi e$, (7) boils down to the supremization of differential entropy $h(\boldsymbol{y})$. Using the polar coordinates[5] $\boldsymbol{x} = \boldsymbol{r}e^{i\boldsymbol{\theta}}$ and $\boldsymbol{y} = \boldsymbol{R}e^{i\boldsymbol{\phi}}$ ($r, R \geq 0$ and $\boldsymbol{\theta}, \boldsymbol{\phi} \in [-\pi, \pi)$) and following the same steps in [15, eq. 5 to eq. 12], we have

$$
h(\boldsymbol{y}) \leq - \int_{0}^{\infty} f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}}) \ln \frac{f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}})}{R} dR + \ln 2\pi,
\tag{8}
$$

---

[5]The polar representation simplifies the problem, since the constraints are circular.

where $f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}})$ is the pdf of $\boldsymbol{R}$ induced by $F_{\boldsymbol{r}}$ and is given by

$$f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}}) = \int_0^{r_p} K(R, r) dF_{\boldsymbol{r}}(r). \tag{9}$$

Note that by selecting $\boldsymbol{r}$ and $\boldsymbol{\theta}$ independent with uniformly distributed $\boldsymbol{\theta}$ over $[-\pi, \pi)^6$, (8) holds with equality and we have

$$f_{\boldsymbol{R},\boldsymbol{\phi}}(R, \phi) = \frac{1}{2\pi} f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}}). \tag{10}$$

Therefore, the optimization problem in (7) is reduced to the following problem

$$C(P_a, P_d, r_p) = \sup_{F_{\boldsymbol{r}} \in \Omega_1 \cap \Omega_2} H(F_{\boldsymbol{r}}) - \ln e, \tag{11}$$

where $F_{\boldsymbol{r}}(0^-) = 0$, $F_{\boldsymbol{r}}(r_p) = 1$ and $H(F_{\boldsymbol{r}})$, Ω_1 and Ω_2 are given as

$$H(F_{\boldsymbol{r}}) \triangleq - \int_0^{\infty} f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}}) \ln \frac{f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}})}{R} dR, \tag{12}$$

and

$$\Omega_1 = \left\{ F_{\boldsymbol{r}} : \int_0^{r_p} r^2 dF_{\boldsymbol{r}}(r) \le P_a \right\}, \tag{13a}$$

$$\Omega_2 = \left\{ F_{\boldsymbol{r}} : P_d \le \int_0^{r_p} g(r) dF_{\boldsymbol{r}}(r) \right\}. \tag{13b}$$

III. MAIN RESULTS

In this section, we provide the main results of this paper. In the following, we first characterize the capacity in (11) when the channel input amplitude constraint is $r_p = \infty$. In the next theorem, we study the capacity problem in (11), when $r_p < \infty$. We accordingly, derive the necessary and sufficient condition for the optimal distributions achieving the capacity.

Theorem 1. *The capacity of the channel in (4) for $r_p = \infty$, i.e., $C(P_a, P_d, \infty)$ is characterized as*

$$C(P_a, P_d, \infty) = \ln \left(1 + \frac{P_a}{2} \right). \tag{14}$$

Let $P_G = 1/P_a \int_0^{\infty} rg(r) e^{-\frac{r^2}{2P_a}} dr$ be the delivered power corresponding to an input distributed as $\boldsymbol{x} \sim \mathcal{CN}(0, P_a)$. If $P_d \le P_G$, (14) is attained by a unique input, distributed as $\boldsymbol{x} \sim \mathcal{CN}(0, P_a)$, and if $P_d > P_G$, (14) is not attained.

---

[6]Note that this causes no loss of optimality, since the constraints are circularly symmetric.
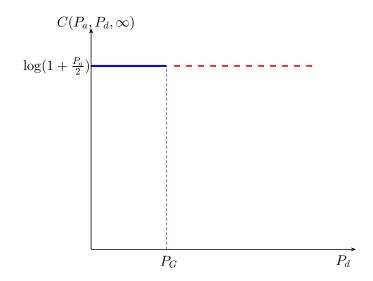
Figure 1: The capacity $C(P_a, P_d, \infty)$ of an AWGN channel. The solid blue line is achievable by a unique input $\boldsymbol{x} \sim \mathcal{CN}(0, P_a)$, however, the red dashed line can be approached.

Proof: See Appendix B.

From Theorem 1, it is verified that for $n \geq 2$ in (6), the capacity of an AWGN channel in (4) for $r_p = \infty$ is independent of the value of the delivered power constraint, i.e., P_d. That is, given P_a, the capacity $C(P_a, P_d, \infty)$ is constant with P_d. This is represented in Figure 1, where the solid line illustrates the capacity $C(P_a, P_d, \infty)$ achievable by $\boldsymbol{x} \sim \mathcal{CN}(0, P_a)$, and the dashed line illustrates the capacity $C(P_a, P_d, \infty)$ that can be approached arbitrarily using time sharing[7] between distributions with high amount of information and distributions with high amount of power (see Appendix B for construction of such inputs).

Note that, the result of Theorem 1 is due to the fact that the function $g(r)$ is of the order of at least 4. In Section IV, we show that accounting for the nonlinearity of the rectifier at the receiver, the delivered power is dependent on higher order moment statistics of the channel input $\boldsymbol{x}$. This, accordingly, explains why nonlinearity[8] is actually beneficial to system performance in contrast with the linear scenario (,i.e., $n = 1$ in (6)).

Theorem 2. *The optimal distribution denoted by $F_{\boldsymbol{r}^\circ}$ achieving the capacity $C(P_a, P_d, r_p)$ for $r_p < \infty$,*

---

[7]Given two distributions, one having high information (e.g. CSCG input distribution) and the other having high power (e.g. flash signalling [28]), power splitting at the receiver is always better than time sharing [4]. The main use of time sharing in our results is that from approaching the capacity point of view, time sharing is sufficient. We also note that, in a practical receiver, the information decoder and the energy harvester are separate. Accordingly, in practical applications, power splitting may still be preferred.

[8]We note that, in practice, nonlinearity of the energy harvester (rectenna) occurs in the low average RF input power regime..

is unique and its corresponding set of points of increase[9] *is finite (the cardinality of the support of the random variable r^o is finite, i.e., $|supp\{r^o\}| < \infty$). Furthermore, F_{r^o} is optimal if and only if there exist unique parameters $\lambda \geq 0$ and $\mu \geq 0$ for which*

$$h(r; F_{r^o}) - \lambda r^2 + \mu g(r) - K = 0, \ \forall r \in supp\{r^o\}, \tag{15a}$$

$$h(r; F_{r^o}) - \lambda r^2 + \mu g(r) - K \leq 0, \ \forall r \in [0, r_p], \tag{15b}$$

where $K \triangleq H(F_{r^o}) - \lambda P_a + \mu P_d$ and

$$h(r; F_{r^o}) = - \int\limits_0^\infty K(R, r) \ln \frac{f_R(R; F_{r^o})}{R} dR. \tag{16}$$

Proof: See Appendix C.

Note that the results in (15) are important in the sense that they can be utilized to obtain the optimal distributions using numerical programming. In [31], the capacity of a real AWGN channel is studied with $g(r) = I_0(r)$. It can be easily verified that for both real and complex AWGN channels the obtained results (uniqueness and finite cardinality of the optimal input distribution) in [31] and here in Theorem 2 remain valid if the function $g(r)$ grows faster than r^2, i.e., $r^2 = \mathcal{O}(g(r))$[10].

Remark 2. *Rewriting the KKT condition for the inequality in (15), we get*

$$0 \leq \mu \leq \frac{K + 2 + \lambda r^2}{g(r)}, \ r \in [0, r_p], \tag{17}$$

where we used the inequality $h(r; F_r) \geq -2$ for any $F_r \in \Omega_1 \cap \Omega_2$ (see (70) in Appendix C). We note that, since by definition the function $g(r)$ grows faster than r^2, we have $\mu \to 0$ as $r_p \to \infty$. The intuition behind this is as follows. μ can be considered as the opposite sign of $\partial C(P_a, P_d, r_p)/\partial r_p$. As r_p increases, $C(P_a, P_d, r_p)$ approaches $C(P_a, P_d, \infty)$. From Theorem 1, we already know that capacity $C(P_a, P_d, \infty)$ is unchanged for any $P_d < \infty$. Therefore, $\partial C(P_a, P_d, r_p)/\partial r_p \to 0$, and accordingly, $\mu \to 0$ as r_p increases. In other words, the dependency of the capacity on r_p reduces as r_p grows large.

Remark 3. *Though the primary focus of this paper is on point-to-point scenario with co-located receivers, we also note the results are valid for the scenario where the energy receiver is separated from the information receiver.*

---

[9]x is said to be a point of increase of F_x if and only if $\Pr(x - \eta < x < x + \eta) > 0$ for all $\eta > 0$.

[10]By definition, given two functions $f(\cdot)$ and $g(\cdot)$, we write $f(x) = \mathcal{O}(g(x))$ if and only if there exist two positive scalars, $c > 0$, $x_0 > 0$, such that $|f(x)| \leq c|g(x)|, \forall x > x_0$.

Remark 4. *In [31, Corollary 2], it is stated that for a real AWGN channel and $g(r) = I_0(r)$, when $r_p \to \infty$ and P_d is greater than the feasible delivered power corresponding to Gaussian input, the capacity is still achievable and the corresponding input distribution is discrete with a finite number of mass points. We note that, this claim cannot hold, since as in Theorem 1, the capacity is not achievable, however, it can be approached arbitrarily (See Appendix B, for construction of such distributions approaching capacity when $r_p = \infty$.).*

IV. APPLICATION

As an application of the results in Section III, in this section, we consider the channel in (4), under a scenario where the receiver is equipped with a nonlinear energy harvester. In the following, we first explain the transmission process. Next, we obtain a baseband equivalent for the harvested power at the receiver. Later, we define the rate-power region, and obtain two inner bounds on the rate-power region.

Transmitter: The transmitted process $X(t)$ is produced as

$$X(t) = \sum_k \boldsymbol{x}_k \text{sinc}(f_w t - k), \tag{18}$$

where $\boldsymbol{x}_k$ is an information-power symbol at time index k, modelled as a random variable, which is produced in an iid fashion. Next, the process $X(t)$ is upconverted to the carrier frequency f_c and is sent over the channel.

Receiver: The filtered received RF waveform at the receiver is modelled as

$$Y_{\text{rf}}(t) = \sqrt{2} \text{Re} \left\{ Y(t) e^{j2\pi f_c t} \right\}, \tag{19}$$

where $Y(t)$ is the baseband equivalent of the channel output with bandwidth $[-f_w/2, f_w/2]$. In order to have a narrowband transmission, we assume that $f_c \gg 2f_w$.

Power: At the receiver, the power of the RF signal $Y_{\text{rf}}(t)$ is captured via the rectenna. Leveraging the small-signal approximation for rectenna's output introduced in [7], [11],[11] the delivered power, denoted by P_{del} is modelled as[12]

$$P_{\text{del}} = \mathbb{E}\mathcal{E}[k_2 Y_{\text{rf}}(t)^2 + k_4 Y_{\text{rf}}(t)^4], \tag{20}$$

---

[11]According to [7], due to the presence of a diode in rectenna's structure, its output current is an exponential function, which is approximated by expanding its Taylor series. The approximation used here, is the fourth moment truncation of Taylor series, in which the first and third moments are zero with respect to the time averaging. Discussions on the assumptions and validity of this model can be found in [7].

[12]According to [7], rectenna's output in (20) is in the form of current with unit Ampere. However, since power is proportional to current, with abuse of notation, we refer to the term in (20) as power.

where k_2 and k_4 are constants. Note that, in the linear model for the delivered power P_{del}, in (20), we have only the second moment of the received RF signal $Y_{\text{rf}}(t)$, where the optimal input is shown to be a CSCG distribution [13].

Information: The signal $Y_{\text{rf}}(t)$ is downconverted producing the baseband signal $Y(t)$ given as[13]

$$Y(t) = X(t) + W(t). \tag{21}$$

Next, $Y(t)$ is sampled with a sampling frequency f_w producing $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}$ as in (4).[14]

A. Delivered power in the baseband

From a communications system design point of view, it is most preferable to have baseband equivalent representation of the system. Henceforth, in the following Proposition, we derive the delivered power P_{del} at the receiver (see (20)) in terms of the system baseband parameters.

Lemma 1. *Assuming the channel input distributions are iid, the delivered power P_{del} at the receiver can be expressed as*

$$P_{del} = \alpha(Q + \tilde{Q}) + \beta P + \gamma, \tag{22}$$

where $\tilde{Q}$ is given by

$$\tilde{Q} = \frac{1}{3}\big(Q_r + Q_i + 2(\mu_r T_r + \mu_i T_i)$$
$$+ 6P_r P_i + 6P_r(P_r - \mu_r^2) + 6P_i(P_i - \mu_i^2)\big), \tag{23}$$

and the parameters α, β and γ are given as

$$\alpha = \frac{3k_4}{2}, \tag{24}$$

$$\beta = k_2 + 48k_4, \tag{25}$$

$$\gamma = 4k_2 + 96k_4, \tag{26}$$

and $Q = \mathbb{E}[|\boldsymbol{x}|^4]$, $T = \mathbb{E}[|\boldsymbol{x}|^3]$, $P = \mathbb{E}[|\boldsymbol{x}|^2]$, $\mu = \mathbb{E}[\boldsymbol{x}]$. Similarly, $Q_r = \mathbb{E}[\boldsymbol{x}_r^4]$, $T_r = \mathbb{E}[\boldsymbol{x}_r^3]$, $P_r = \mathbb{E}[\boldsymbol{x}_r^2]$, $\mu_r = \mathbb{E}[\boldsymbol{x}_r]$ and $Q_i = \mathbb{E}[\boldsymbol{x}_i^4]$, $T_i = \mathbb{E}[\boldsymbol{x}_i^3]$, $P_i = \mathbb{E}[\boldsymbol{x}_i^2]$, $\mu_i = \mathbb{E}[\boldsymbol{x}_i]$.

Proof: See Appendix K.

---

[13]We model the baseband equivalent channel impulse response as $H(\tau, t) = \sum_i \delta(\tau) + W(t)$, where the delay and the gain of the channel are assumed to be 0 and 1, respectively.

[14]Due to the assumption of iid channel inputs and discrete memoryless channel, we neglect the time index k.

Remark 5. *We note that, obtaining a closed form expression for the delivered power P_{del} at the receiver, when the channel inputs are not iid is cumbersome. This is due to the fact that the fourth moment of the received RF signal $Y_{rf}(t)$ creates dependencies of the statistics of the present channel input on the statistics of the channel inputs on the other time indices (see e.g., eq. (158) and eq. (154) in Appendix K).*

B. Rate-Power (RP) region

We define the RP region as the convex hull of the following union of regions

$$\mathcal{R}(P_a, r_p) = \bigcup_{P_d} \{(R, P) : R < C_{\text{SWIPT}}(P_a, P_d, r_p), P \leq P_d\}, \tag{27}$$

where $C_{\text{SWIPT}}(P_a, P_d, r_p)$ is defined similarly to (7) as

$$C_{\text{SWIPT}}(P_a, P_d, r_p) = \sup_{f_{\boldsymbol{x}}(x):} \quad I(\boldsymbol{x}; \boldsymbol{y}) \qquad \begin{cases} \mathbb{E}[|\boldsymbol{x}|^2] \leq P_a, \\ P_d \leq P_{\text{del}}, \\ |\boldsymbol{x}| \leq r_p, \end{cases} \tag{28}$$

and P_{del} is given in (22).

In the following, we consider two different lower bounds on the RP region defined in (27). In the first approach, we assume that the inputs are Gaussian distributed, where it is shown that the optimal Gaussian inputs are zero mean. In the second, we obtain an inner bound on the harvested power in (22) by considering a convex subset of optimization probability space, and accordingly, apply the result of Theorem 2.

1) Complex Gaussian Inputs: Assuming that the inputs are Gaussian distributed, we show that for the considered scenario, there is a tradeoff between the rate of the transmitted information, namely $I(\boldsymbol{x}; \boldsymbol{y})$ and delivered power P_{del} at the receiver, and accordingly, we characterize the tradeoff.

Lemma 2. *If a channel input distribution $f_{\boldsymbol{x}}(x)$ is complex Gaussian, the supremum in (28) is achieved by zero mean inputs, i.e., $\boldsymbol{x}_r \sim \mathcal{N}(0, P_r)$, and $\boldsymbol{x}_i \sim \mathcal{N}(0, P_i)$, where $P_r + P_i = P_a$. Furthermore, let $P_{del,max} = 3\alpha P_a^2 + 2\beta P_a + \gamma$ and $P_{del,min} = 2\alpha P_a^2 + 2\beta P_a + \gamma$ be the maximum and minimum delivered power at the receiver, respectively. If $P_d > P_{del,max}$, the solution does not exist. If $P_d = P_{del,max}$, the maximum in (28) is attained by $P_i = 0, P_r = P_a$ or $P_i = P_a, P_r = 0$. If $P_{del,min} < P_d < P_{del,max}$, the optimal power allocation that attains the maximum in (28) is given by P_i^* and $P_r^* = P_a - P_i^*$, where P_i^* is chosen, such that the following equation is satisfied*

$$2\alpha(4P_i^{*2} + 3P_a^2 - 8P_a P_i^*) + 2\beta P_a + \gamma = P_d. \tag{29}$$

For $P_d \leq P_{del,min}$, the optimal power allocation is attained by $P_i^* = P_r^* = P_a/2$ and the delivered power is still $P_{del,min}$.

Proof: See Appendix L.

We note that the tradeoff between information and power for Gaussian inputs, results from the asymmetric power allocation between Inphase and Quadrature subchannels. We have illustrated the RP region corresponding to Gaussian inputs in Section V.

Remark 6. *From (22), it is seen that the delivered power P_{del} at the receiver depends on the second moment statistics P_r, P_i, as well as the fourth moment statistics Q_r, Q_i of the channel input $\boldsymbol{x}$. This is due to the presence of the fourth moment of the received RF signal in modelling the rectenna's output. From Lemma 2, it is seen that the maximum rate corresponding to $P_d = P_{del,max}$ is when the available power at the transmitter is fully allocated to one of the real or imaginary dimensions. This is because allocating power to one dimension, leads to a higher fourth moment statistic. On the other hand, the maximum rate corresponding to $P_d = P_{dc,min}$ is when the available power is equally distributed between the real and the imaginary dimensions. Note that as also mentioned in Remark 1, there is no tradeoff when the linear model is considered for the delivered power ,i.e., $n < 2$ in (6).*

2) Convexified optimization probability space: In this section, we consider an inner bound on the RP region defined in (27), by considering a convex subset of the optimization probability space in (28). Note that the delivered power at the receiver in (20) can be lower bounded as below

$$P_{\text{del}} = k_2 \mathbb{E}\left[|\boldsymbol{y}_k|^2\right] + \frac{3k_4}{2}\left(\mathbb{E}[|\boldsymbol{s}_{2k+1}|^2] + \mathbb{E}[|\boldsymbol{s}_{2k}|^2]\right) \tag{30}$$

$$> k_2 \mathbb{E}\left[|\boldsymbol{y}_k|^2\right] + \frac{3k_4}{2}\mathbb{E}[|\boldsymbol{s}_{2k}|^2] \tag{31}$$

$$= k_2 \mathbb{E}\left[|\boldsymbol{y}_k|^2\right] + \frac{3k_4}{2}\mathbb{E}[|\boldsymbol{y}_k|^4] \tag{32}$$

$$= \frac{3k_4}{2}\mathbb{E}\left[|\boldsymbol{x}_k|^4\right] + (k_2 + 24k_4)\mathbb{E}\left[|\boldsymbol{x}_k|^2\right] + 4k_2 + 48k_4 \tag{33}$$

$$= \mathbb{E}[g_{\text{NL}}(\boldsymbol{r})], \tag{34}$$

where (30) is due to (138) and (146) (see Appendix K for the definition of $\boldsymbol{s}_{2k+1}$ and $\boldsymbol{s}_{2k}$). (32) is due to (147). In (34), we have $\boldsymbol{r} = |\boldsymbol{x}|$ and $g_{\text{NL}}(r)$ is given as

$$g_{\text{NL}}(r) = \frac{3k_4}{2}r^4 + (k_2 + 24k_4)r^2 + 4k_2 + 48k_4. \tag{35}$$

By $g_{\text{NL}}(r)$ in hand and noting that $I(\boldsymbol{x}; \boldsymbol{y}) = H(F_{\boldsymbol{r}}) - 1$, (28) can be written as

$$C_{\text{IB}}(P_a, P_d, r_p) = \sup \quad H(F_{\boldsymbol{r}}) - 1$$

$$F_{\boldsymbol{r}} : \begin{cases} \mathbb{E}[\boldsymbol{r}^2] \leq P_a, \\ P_d \leq \mathbb{E}[g_{\text{NL}}(\boldsymbol{r})], \\ \boldsymbol{r} \leq r_p. \end{cases} \tag{36}$$

The inner bound for the RP region in (27) is obtained by finding the corresponding delivered power $\mathbb{E}[g_{\text{NL}}(\boldsymbol{r})]$ and transmitted information $I(\boldsymbol{x}; \boldsymbol{y})$ of the optimal solutions of the problem (36). We illustrate the related results in Section V.

V. NUMERICAL RESULTS

In this section, we first illustrate through numerical evaluations the RP regions and highlight the benefits of nonlinear energy harvesting. We then evaluate through realistic circuit simulations the impact of various input distributions on the harvested DC power in WPT and contrast with the analytical results.

A. Numerical Evaluations of SWIPT RP Regions

In this section, we provide some numerical illustrations of the two inner bounds (see Section IV-B1 and IV-B2) for the RP region defined in Section IV-B. In the following, we first summarize the steps in obtaining the bounds, and next, we illustrate the obtained numerical results.

Complex Gaussian inputs: To obtain the RP region corresponding to Gaussian inputs, we use (29). Note that when symmetric power allocation is used between the real and imaginary subchannels, i.e., $\mathbb{E}[\boldsymbol{x}_i^2] = \mathbb{E}[\boldsymbol{x}_r^2] = P_a/2$, the delivered power is $P_{\text{del,min}}$ with the transmitted information $\ln(1 + P_a/2)$. We gradually increase P_d ($P_d \geq P_{\text{del,min}}$) and using the fact that the average power constraint is satisfied with equality (see Lemma 1) and using (29), the optimal power allocations for Inphase and Quadrature channels are obtained. We continue increasing P_d until allocated power for one of the subchannels gets zero. At this point, the delivered power is equal to $P_{\text{del,max}}$ and the transmitted information is $1/2 \ln(1 + P_a)$.

Inputs obtained by convexifying optimization probability space: To obtain the RP region corresponding to the distributions obtained by solving (36), we resort to numerical programming. Accordingly, we solve the optimization problem in (36) using the interior-point algorithm implemented by the fmincon function in MATLAB software. Note that, since we already know that the optimal distribution is discrete with a finite number of mass points, the numerical optimization is over the position, the probabilities and the number of the mass points. Hence, there are $2m$ parameters to be optimized, where m is the number of the mass points. We aim at calculating the capacity $I(\boldsymbol{x}; \boldsymbol{y})$ in (36) under given an average power and an amplitude constraints and for different values of the delivered power constraint. As a result, we consider the following unconstraint optimization problem

$$H(F_{\boldsymbol{r}}) - \lambda \mathbb{E}[\boldsymbol{r}^2] + \mu \mathbb{E}[g_{\text{NL}}(\boldsymbol{r})], \ 0 \leq \boldsymbol{r} \leq r_p, \ \lambda, \mu \geq 0. \tag{37}$$

In the following, the different steps of the optimization are summarized:

1) Fix the average power constraint. Set $P_d = P_{\text{del,min}} + \delta$, where δ is the step size (Note that for $P_d \leq P_{\text{del,min}}$ and $r_p = \infty$, Gaussian inputs are optimal [13] and for $P_d \leq P_{\text{del,min}}$ and $r_p < \infty$, the

optimal distributions for the input amplitude $\boldsymbol{r}$ are discrete with a finite number of mass points [15]). Set $m = 1$.

2) Utilizing interior-point algorithm, minimize the objective function in (37) initialized by a random guess.

3) Once the optimal positions and their respective probabilities are found, the answer is validated by checking the average power constraint and the necessary and sufficient KKT conditions in (15). If the conditions are not satisfied, the initial guess is changed. We continue changing the initial guess for a large number of times.

4) If the KKT conditions are not satisfied, the number of mass points is increased by one. We continue from stage 1 to 4 until at some values of m, KKT conditions are met.

5) Obtain the delivered power corresponding to the optimal solution.

Note that despite the fact that the problem is concave with respect to probability laws, however, for a given number of mass points m, the problem is not concave and the obtained solution is not guaranteed to be a global one.

Illustration of the numerical results: In Figure 2, simulation results for the transmitted information in terms of mutual information $I(\boldsymbol{x}; \boldsymbol{y})$ and harvested power in terms of the expectation $\mathbb{E}[g_{\mathrm{NL}}(|\boldsymbol{x}|)]$ are illustrated for an average power constraint $P_a = 5$ and $g_{\mathrm{NL}}(r) = 0.01(r^4 + r^2 + 1)$.[15] The horizontal solid line related to $C_{\mathrm{IB}}(5, P_d, \infty)$ corresponds to the AWGN channel capacity under an average power constraint $P_a = 5$ achieved by only a CSCG distribution. The horizontal dashed line related to $C_{\mathrm{IB}}(5, P_d, \infty)$ corresponds to the capacity under an average power constraint $P_a = 5$, which is not achievable, however, can be approached arbitrarily (see Theorem 1). $C_{\mathrm{IB}}(5, P_d, 4)$, $C_{\mathrm{IB}}(5, P_d, 5)$ and $C_{\mathrm{IB}}(5, P_d, 6)$ correspond to the optimal solution in (36) for $r_p = 4,\ 5$ and 6, respectively. The RP region obtained from Gaussian inputs is denoted by Gaussian Asymmetric Power Allocation (GAPA). The distributions obtained numerically by convexifying the probability optimization space are denoted as numerically obtained input (NOI) distributions. As it is observed from Figure 2, NOI distributions yield significantly larger RP region compared to the region corresponding to GAPA. It is also observed that by increasing the amplitude constraint r_p, the RP region tends to the RP region corresponding to $r_p = \infty$. This observation is inline with Remark 2, that increasing r_p, reduces the dependency of the capacity on r_p. Note that given the value of r_p, the amount of harvested power at the receiver is limited. This is the reason for the vertical lines corresponding to $C_{\mathrm{IB}}(5, P_d, 4)$, $C_{\mathrm{IB}}(5, P_d, 5)$ and $C_{\mathrm{IB}}(5, P_d, 6)$.

---

[15]We chose the coefficients in (35) such that the numerical results are readable, however, the baseline of the results remain valid for the realistic values of the coefficients in (35).
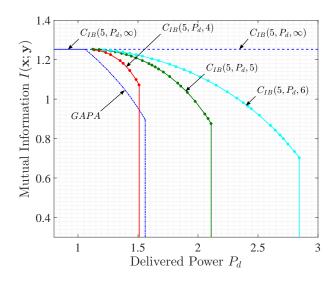
Figure 2: Mutual information $I(\boldsymbol{x}; \boldsymbol{y})$ corresponding to the complex Gaussian inputs (denoted by GAPA). Mutual information $I(\boldsymbol{x}; \boldsymbol{y})$ corresponding to the optimal solutions of (36) with respect to different values of the minimum delivered power constraint P_d with amplitude constraints $r_p = 4$, 5, 6 and $r_p = \infty$. Average power constraint is $P_a = 5$.

In Figures 3, 4 and 5 the position of the mass points $\boldsymbol{r} = |\boldsymbol{x}|$ corresponding to $C_{\text{IB}}(5, P_d, 4)$, $C_{\text{IB}}(5, P_d, 5)$ and $C_{\text{IB}}(5, P_d, 6)$ are illustrated, respectively, with respect to different delivered power constraints P_d. It is observed that by increasing the delivered power constraint P_d at the receiver, the number of mass points decreases. Also, as it is seen from the figures, one of the mass points is always equal to r_p.

In Figure 6, the information rate $I(\boldsymbol{x}; \boldsymbol{y})$ and delivered power P_{del} for complex Gaussian inputs is shown versus the inphase subchannel power allocation P_i ($P_r = P_a - P_i$). In line with the previous results, it is observed that (unlike the linear model for the energy harvester), under nonlinear model for the energy harvester, the information rate and delivered power are maximized and minimized, respectively, for $P_i = P_r = \frac{P_a}{2}$. Alternatively the information rate and delivered power are minimized and maximized, respectively when $P_i = 0, P_r = P_a$ or $P_i = P_a, P_r = 0$.

Finally, we note that the algorithm used for finding NOI distributions is extremely sensitive on the first guess as the number of mass points m increases. This is due to the fact that the optimization of the capacity given that the number of mass points m is fixed, is not a concave function. This, accordingly, makes the problem computationally demanding with m.
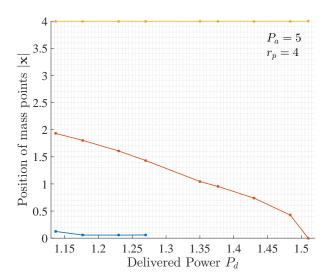
Figure 3: The position of the optimal mass points for $C_{\text{IB}}(5, P_d, 4)$ versus different values of the minimum delivered power P_d constraint.
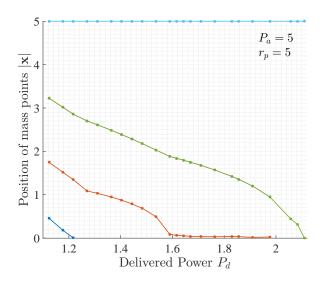

Figure 4: The position of the optimal mass points for $C_{\text{IB}}(5, P_d, 5)$ versus different values of the minimum delivered power P_d constraint.

B. Realistic Circuit Simulations for WPT

In order to assess and validate the analysis and the benefits of flash signalling[16] and asymmetric Gaussian distribution (from WPT perspective only), we designed, optimized and simulated the rectenna circuit of Figure 7. We used a conventional single series rectifier circuit that consists of a rectifying diode, impedance

---

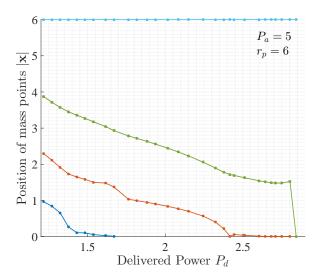[16]For flash signalling, we use the distributions introduced in (56) for different values of the parameter l.

Figure 5: The position of the optimal mass points for $C_{\text{IB}}(5, P_d, 6)$ versus different values of the minimum delivered power P_d constraint.
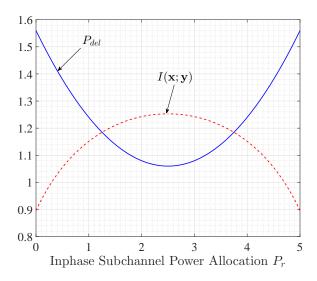
Figure 6: Mutual information $I(\boldsymbol{x}; \boldsymbol{y})$ (Red dashed line) and delivered power P_{del} (blue solid line) corresponding to the complex Gaussian inputs with asymmetric power allocation. The transmitted information rate is maximized for $P_i = P_r = \frac{P_a}{2}$ and delivered power is maximized when $P_i = 0, P_r = P_a$ or $P_i = P_a, P_r = 0$.

matching circuit, and low pass filter. The Schottky diode Skyworks SMS7630 is chosen for the rectifying diode because it requires low biasing voltage level, which is suitable for low power rectifier. The impedance matching and low pass filter circuits are designed for an Inphase 4-tone multisine input signal centered around 2.45GHz with an average power of -20dBm and with 2.5MHz inter-carrier frequency spacing. The load impedance $R2$ is chosen as $10\text{K}\Omega$ in order to reach maximum RF-to-DC conversion efficiency with the 4-tone multisine waveform. The matching network capacitor $C1$, inductor $L1$ and output capacitor $C2$ values are optimized (using an iterative process) to maximize the output DC power under a given load impedance and for the given multisine input waveform at -20dBm RF input power. The chosen values are given by 0.4pF for $C1$, 8.8nH for $L1$, and 1nF for $C2$. The antenna impedance is set as $\text{R1} = 50\Omega$ and the voltage source V1 is expressed as $\text{V1} = 2Y_{\text{rf}}(t)\sqrt{R1}$.

In Table I, the measured delivered DC power is shown for four types of channel input, namely, continuous wave (CW)[17], Complex Gaussian (CG), Real Gaussian (RG) and inputs of (56) for different values of parameter l is shown. A first observation is to note that the second moment (i.e., average input power) of the input distribution is the same for all distributions, though a significant range of harvested DC power is observed. This is due to the rectenna nonlinearity that favors distributions with a large fourth moment. Indeed, the fourth moment increases proportionally to 1, 2, 3 and l^2 for the CW, CG, RG and flash signaling (with l), respectively. This shows that the nonlinearity model through a polynomial expansion with a second and fourth order terms as in (20) predicts the dependency of the rectenna nonlinearity on the input signal quite accurately, and confirm observations made in [7], [8], [11]. Recall that the linear model of the rectifier would not capture this dependency since it only accounts for the second order term in (20) [7], [11]. A second observation is the significantly larger power delivered with inputs in (20) compared to other schemes. Specifically, the maximum delivered power occurs at $l = 4$. The reason that the delivered power decreases for $l > 4$ is due to the finite RC constant in the low pass filter of the rectenna.

VI. DISCUSSION AND FUTURE WORKS

In the following, we discuss about a number of interesting research avenues that can be considered in the future.

- Note that the delivered power in (22), contains odd moments of the channel input $\boldsymbol{x}$. Accordingly, for the problem considered in (7), it is interesting to find optimal input distributions when the function $g(r)$ (we recall that $g(r)$ models the baseband representation) contains odd powers of the argument.

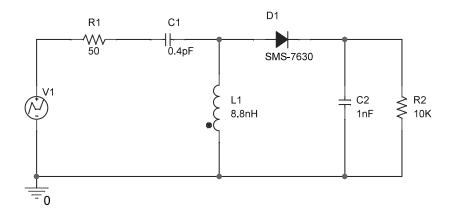[17]A single tone with frequency 2.45GHz

Figure 7: Conventional single series rectifier circuit consisting of a rectifying diode, impedance matching circuit, and low pass filter.

Transmission type	Delivered DC Power (μW)
CW	1.0959
CG	1.5296
RG	1.7547
l=2	2.6899
l=3	3.4262
l=4	3.4884
l=5	3.2965

Table I: Conventional single series rectifier circuit consisting of a rectifying diode, impedance matching circuit, and low pass filter.

- The practical power harvesters exhibit nonlinear behaviors since their efficiency becomes different (not constant) when the received RF power level changes. Specifically, the efficiency is very small in low RF power level (due to the turn-on voltage of the diode), is large in the middle RF power level, and is again very small in the high RF power level (due to the reverse breakdown of the diode). In order to capture this behaviour, the function $g(r)$ should not tend to infinity when $r \to \infty$. Accordingly, finding optimal inputs for bounded $g(r)$ is of interest.

- The problem considered in (7), is indeed an optimization over circular symmetric solutions. However, in practical SWIPT problems, harvesters are also phase dependent and circuit simulations reveal that phase variations in the channel input can also affect the delivered power at the receiver significantly [32]. Hence, it is interesting to develop a systematic approach in order to capture the effect of phase variations as well.

- Note that the harvester's input is the RF signal $Y_{\text{rf}}(t)$ (see (20)), and therefore, in the baseband

representation (for nonlinear harvesters), it appears that we have higher order moment statistics of the baseband equivalent of the channel output, i.e., $Y(t)$ (see (142) in Appendix K). Accordingly, to represent the signal perfectly in terms of its samples, we require to consider more values of the baseband channel output $Y(t)$ between any consecutive information samples (see (143) in Appendix K). If unlike the assumption of this paper, we assume that the samples possess a level of correlation with each other, then the problem gets cumbersome to approach. However, it seems to the authors that from a power harvesting point of view, correlation among different samples is good, in opposition to information transmission. Hence, it is also interesting to consider even very simple achievable schemes which utilize the effect of correlation.

- Finally, we note that the results presented here can be extended to vector Gaussian channels with bounded inputs [33] and Gaussian multiple access channels [34], utilizing the similar tools presented therein.

- There might be interesting connections to make with other systems subject to nonlinear responses. In optical communications, for instance, the nonlinearity is commonly compensated and transmission is performed using constellations approximating the zero-mean Gaussian distribution optimum for AWGN channels (e.g. ring constellations) [20]. The information theoretic limits of optical channels are studied by modelling the nonlinear optical communication channel as a linear channel with a multiplicative noise or using a finite-memory model with additive noise [20], [21]. On the contrary, in SWIPT, the diode nonlinearity is exploited in the signal design and in the characterization of the RP region, therefore leading to non-zero mean Gaussian inputs and enlarged region compared to that obtained with zero-mean inputs.

VII. CONCLUSIONS

In this paper, we studied the capacity of a complex AWGN channel under transmit average power, amplitude and receiver delivered power constraints. We focused on nonlinear delivered power constraints at the receiver. We showed that under an average power constraint and for any given delivered power constraint, the capacity of an AWGN channel can be either achieved or approached arbitrarily. In line with the similar results in the literature, we showed that including the amplitude constraint causes the optimal inputs to be discrete with a finite number of mass points. As an application of the presented results, we considered SWIPT over a complex AWGN channel in the presence of a nonlinear power harvester at the receiver. Defining the RP region, we provided two inner bounds for the RP region. Considering general complex Gaussian inputs as the first inner bound, we showed that the optimal Gaussian inputs are zero mean. A tradeoff between the transmitted information and harvested power is recognized by allocating the

power budget asymmetrically between the real and imaginary subchannels. Obtaining a convexified subset of optimization probability space, we utilized the obtained results in this paper to derive the second inner bound. Numerical results reveal that there are significant improvements in the second inner bound with respect to the first inner bound corresponding to complex Gaussian inputs.

VIII. ACKNOWLEDGMENT

APPENDIX A

LEMMAS

In this appendix, we provide the lemmas required to prove Theorems 1 and 2.

Lemma 3. *In the Levy's metric, the space $\Omega_1 \cap \Omega_2$ is convex, however, compact only if $r_p < \infty$.*

Proof: The proof is obtained by following exactly the same approach used in [14]. In the following, we bring a counterexample which proves that the space $\Omega_1 \cap \Omega_2$ for $r_p = \infty$ is not compact. For simplicity, assume $g(r) = r^4$ (the following argument can be extended to the general definition of $g(r)$ in (6)) and consider the following sequence of probability distributions

$$F_{\boldsymbol{r},l}(r) = \begin{cases} 0 & r < 0, \\ 1 - \frac{1}{l^4} & 0 \leq r < \sqrt[4]{P_d}l, \quad l = 0, 1, \dots. \\ 1 & r \geq \sqrt[4]{P_d}l, \end{cases} \tag{38}$$

It can be verified that $\mathbb{E}[r^4] = P_d$ and for integer $l \geq \sqrt[4]{P_d/P_a^2}$ we have $\mathbb{E}[r^2] \leq P_a$. However, the limiting distribution (when $l \to \infty$) is $F_{\boldsymbol{r}}^*(r) = U(r)$ does not satisfy the second constraint, i.e., $\mathbb{E}[r^4] = 0$. This establishes that the space $\Omega_1 \cap \Omega_2$ for $P_d, P_a < \infty$ and $r_P = \infty$, is not compact[18].

Lemma 4. *For all $x \geq 0$ we have*

$$I_0(x) < \min_{0 \leq a < 1} e^x \left(\frac{\hat{a}(1 - e^{-2ax})}{\pi x} + \frac{erf\left(\sqrt{2ax}\right)}{\sqrt{2\pi x}} + e^{-2ax} \right), \tag{39}$$

where $\hat{a} = \frac{\frac{1}{\sqrt{1-a}} - 1}{2\sqrt{a}}$.

Proof: See Appendix D.

---

[18]Note that compactness is a sufficient condition for continuous functions to achieve their supremum or infimum, however, not necessary.

Remark 7. *From (39), it can be easily verified that*

$$\lim_{x \to 0} e^x \left(\frac{\hat{a}(1 - e^{-2ax})}{\pi x} + \frac{erf\left(\sqrt{2ax}\right)}{\sqrt{2\pi x}} + e^{-2ax} \right) = 1 + \frac{\sqrt{a}}{\pi} + \frac{\sqrt{a}}{\pi\sqrt{1 - a}}. \tag{40}$$

We can also obtain a looser upper bound as below. Substituting $a = 1/2$ in (39) and noting that $erf(x) \leq 1$ and $1 - e^{-x} \leq \sqrt{\pi x}$ we have

$$I_0(x) < \frac{e^x}{\sqrt{\pi x}} + 1. \tag{41}$$

It can be easily verified that $\sqrt{\pi x} < e^x(\sqrt{\pi} - 1)$. Using this inequality, we can further upper bound (41) as

$$I_0(x) < \frac{e^x}{\sqrt{x}}. \tag{42}$$

Lemma 5. *The following integral will come useful in the proof of Theorems 1 and 2.*

$$\int_0^\infty R^b K(R, r) dR = 2^{\frac{b}{2}} \Gamma\left(\frac{b}{2} + 1\right) e^{-\frac{r^2}{2}} \Phi\left(\frac{b}{2} + 1, 1; \frac{r^2}{2}\right), \text{ for } 0 \leq r < \infty, \ b > -2. \tag{43}$$

Proof: This can be verified by the transform $t = u^2/2$ and [29, MI 45]. $\qquad\square$

Lemma 6. *In the following integral transform*

$$\int_0^\infty K(R, r) G(R) dR = g(r), \tag{44}$$

where $g(r)$ is defined in (6), $G(R)$ has the following form

$$G(R) = \sum_{i=0}^n c_i R^{2i}, \tag{45}$$

where $c_i, \ i = 0, \ldots, n$ are coefficients determined uniquely.

Proof: See Appendix E.

The following Lemma is indeed a generalization of [18, Theorem 13] to complex channels.

Lemma 7. *Let $\boldsymbol{n} = \boldsymbol{r_n} e^{j\boldsymbol{\theta_n}}$ be a CSCG random variable of variance 2, and let $\boldsymbol{x}$ be a complex random variable that is independent of $\boldsymbol{n}$. The PDF of the random variable $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n} = \boldsymbol{R} e^{j\boldsymbol{\theta}}$ is such that*

$$f_{\boldsymbol{y}}(y) \neq \mathcal{O}\left(e^{-AR^2}\right), \ \forall A > \frac{1}{2}. \tag{46}$$

Proof: See Appendix F.

Lemma 8. *$f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}}), \ R \geq 0, \ F_{\boldsymbol{r}} \in \Omega_1 \cap \Omega_2$ is bounded and continuous in both of its arguments.*

Proof: Continuity of $K(R, r)$ follows by the continuity of $I_0(rR)$. Noting that

$$K(0, r) = K(\infty, r) = K(R, \infty) = 0, \tag{47}$$

$$K(R, 0) = Re^{-\frac{R^2}{2}} < \infty, \tag{48}$$

$$K(\infty, \infty) < \sqrt{\frac{R}{r}} e^{-\frac{(R-r)^2}{2}} < \infty, \tag{49}$$

where (49) is due to (42). Therefore the function $K(R, r)$ is bounded.

Using (42), it can be easily verified that

$$K(R, r) < \sqrt{\frac{R}{r}} e^{-\frac{(R-r)^2}{2}} \leq 1. \tag{50}$$

Note that the first inequality in (50) is strict. Accordingly, to avoid extra notation and for brevity, we will use 1 as an upper bound for $K(R, r)$ when needed. Continuity of $f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}})$ is obtained by following the same steps as in [17, Lemma 3]. From (50) and $K(R, r) > 0$ it can also be easily verified that

$$0 < f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}}) < 1, \ \ R > 0. \tag{51}$$

$\square$

Lemma 9. $f_{\boldsymbol{R}}(R; F_{\boldsymbol{r},n}) \ln f_{\boldsymbol{R}}(R; F_{\boldsymbol{r},n})$ *for* $R \geq 0$, $F_{\boldsymbol{r},n} \in \Omega_1 \cap \Omega_2$ *is dominated by the following absolutely integrable function*

$$g(R) = \begin{cases} 4 & R \leq 2 \\ \frac{c}{R^{\frac{3}{2}}} & R > 2 \end{cases}, \tag{52}$$

where $c = 4(128 + 4P_a)^{\frac{3}{4}}$.

Proof: See Appendix G.

Lemma 10. *For every* $F_{\boldsymbol{r}} \in \Omega_1 \cap \Omega_2$, $H(F_{\boldsymbol{r}})$ *exists, and is continuous, strictly concave and weakly differentiable.*

Proof: See Appendix H.

APPENDIX B
PROOF OF THEOREM 1

It is easy to verify that for a given average power constraint P_a, capacity $C(P_a, P_d, \infty)$ is a non-increasing function with P_d. Therefore, we have

$$C(P_a, 0, \infty) \geq C(P_a, P_d, \infty). \tag{53}$$

Note that $C(P_a, 0, \infty) = \ln(1 + P_a/2)$ and is achieved by a unique CSCG input distribution as $\boldsymbol{x} \sim \mathcal{CN}(0, P_a)$ (with its amplitude $\boldsymbol{r}$ distributed as Rayleigh distribution according to the CDF $F_{\boldsymbol{r}_R}(r) = 1 - e^{-\frac{r^2}{2P_a}}$).[19] The uniqueness of the input can be verified from [15, appendix II]. The delivered power corresponding to $\boldsymbol{x} \sim \mathcal{CN}(0, P_a)$ is obtained as

$$P_G = \frac{1}{P_a} \int\limits_0^\infty r g(r) e^{-\frac{r^2}{2P_a}} \, dr. \tag{54}$$

Hence, we have

$$C(P_a, 0, \infty) = C(P_a, P_d, \infty), \ \ P_d \le P_G. \tag{55}$$

Since $\boldsymbol{x} \sim \mathcal{CN}(0, P_a)$ is the only distribution achieving the capacity $C(P_a, 0, \infty)$, therefore, $C(P_a, 0, \infty)$ is not achieved for $P_d > P_G$[20]. In what follows, we show that, (55) holds for $P_d > P_G$. In other words, when $P_d > P_G$, any rate lower than $C(P_a, 0, \infty)$ can be achieved by a distribution whose corresponding delivered power is greater than P_d. Consider the following sequence of distribution function

$$F_{\boldsymbol{r}_l}(r) = \begin{cases} 0 & r < 0 \\ 1 - \frac{1}{l^2} & 0 \le r < \sqrt{P_a}l \\ 1 & r \ge \sqrt{P_a}l \end{cases}, \ l = 2, 3, \dots. \tag{56}$$

It is easy to verify that $F_{\boldsymbol{r}_l}(r), \ l = 2, \dots$ satisfy $\mathbb{E}_{F_{\boldsymbol{r}_l}}[\boldsymbol{r}_l^2] = P_a$, hence, satisfying the average power constraint. Also, for the delivered power constraint we have

$$P_{d,l} \triangleq \mathbb{E}_{F_{\boldsymbol{r}_l}}[g(\boldsymbol{r}_l)] = \alpha_0 + \alpha_1 P_a + \sum_{i=2}^n \alpha_i P_a^i l^{2i-2}. \tag{57}$$

Since $n \ge 2$ by construction, it is guaranteed that there exists an integer number L, such that for $l > L$, $P_{d,l} \ge P_d$ (note that $P_{d,l} \to \infty$ as $l \to \infty$). Due to Lemma 3, time sharing is valid in our system model. Hence, we can construct a complex input with its phase uniformly distributed over $[-\pi, \pi)$ and its amplitude distributed according to the following CDF

$$F_{\boldsymbol{r}_{ts}}(r) = (1 - \tau) F_{\boldsymbol{r}_R}(r) + \tau F_{\boldsymbol{r}_l}(r), \ \ \tau \in (0, 1), \ l > L, \tag{58}$$

where the subscript ts in $F_{\boldsymbol{r}_{ts}}$ stands for time-sharing. By choosing $\tau = (P_d - P_G)/(P_{d,l} - P_G)$, we have $0 < \tau < 1$ and the constraints

$$\begin{cases} \mathbb{E}_{F_{\boldsymbol{r}_{ts}}}[\boldsymbol{r}_{ts}^2] = P_a, \\ \mathbb{E}_{F_{\boldsymbol{r}_{ts}}}[g(\boldsymbol{r}_{ts})] \ge P_d. \end{cases} \tag{59}$$

---

[19]The subscript R stands for the Rayleigh distribution.

[20]Note that although the probability space is not compact (a sufficient condition for achieving the supremum or infimum), here, supremum of the capacity is not attained due to the contradiction in uniqueness of the achievable input.

are both satisfied. On the other hand, due to strict concavity of the entropy $H(F_{\boldsymbol{r}})$ (see Lemma 10), we have

$$H(F_{\boldsymbol{r}_{ts}}) > (1 - \tau)H(F_{\boldsymbol{r}_R}) + \tau H(F_{\boldsymbol{r}_l}), \ \ \tau \in (0,1), \ \ l > L. \tag{60}$$

For a given P_d, we can increase l arbitrarily. Therefore τ can be made arbitrarily close to zero by letting $l \to \infty$. Rewriting (61), we have

$$H(F_{\boldsymbol{r}_R}) > H(F_{\boldsymbol{r}_{ts}}) > (1 - \tau)H(F_{\boldsymbol{r}_R}) + \tau H(F_{\boldsymbol{r}_l}), \tag{61}$$

where by letting τ tend to zero (equivalently letting $P_{d,l} \to \infty$) the result of Theorem 1 is concluded. We note that, there is no distribution achieving the supremum.

APPENDIX C

PROOF OF THEOREM 2

The main steps of the proof of Theorem 2 are parallel to those provided in [14]–[18]. However, the problem at hand is different mainly because of the constraints in (5).[21] Therefore, we provide the details for the different arguments and briefly mention (for brevity) the straightforward outcomes of [14]–[18].

A. Proof of Theorem 2

Since the set $\Omega_1 \cap \Omega_2$ is compact for $r_p < \infty$ (see Lemma 3) and $H(F_{\boldsymbol{r}})$ is continuous (see Lemma 9), it is verified that the supremum in (7) is achieved and therefore it can be replaced by maximum. Due to convexity of the set $\Omega_1 \cap \Omega_2$ (see Lemma 3) and strict concavity of $H(F_{\boldsymbol{r}})$ (see Lemma 9), it is concluded that the maximum is achieved by a unique $F_{\boldsymbol{r}^o} \in \Omega_1 \cap \Omega_2$. It is verified from Lemmas 3 and 9 that the conditions of the Lagrangian theorem [35, Section 8.3] are met. By writing the Lagrangian we have

$$L(F_{\boldsymbol{r}}, \lambda, \mu) = \int_0^{r_p} h(r; F_{\boldsymbol{r}}) - \lambda(r^2 - P_a) + \mu(g(r) - P_d)dF_{\boldsymbol{r}}(r), \tag{62}$$

where $\lambda \geq 0$, $\mu \geq 0$ are Lagrange multipliers and $h(r; F_{\boldsymbol{r}})$ is defined in (16). By weak differentiability of $H(F_{\boldsymbol{r}})$ (see Lemma 9) and the linear constraints in (13), the weak derivative [35, Section 7.4] of (62) with respect to $F_{\boldsymbol{r}^o}$ reads as

$$L'_{F_{\boldsymbol{r}^o}}(F_{\boldsymbol{r}}, \lambda, \mu) = \int_0^{r_p} h(r; F_{\boldsymbol{r}^o}) - \lambda r^2 + \mu g(r) - K dF_{\boldsymbol{r}}(r), \tag{63}$$

---

[21]The optimization problem we consider in this paper, essentially differs from [15], in the sense of the constraints. More specifically, due to the fact that the amplitude constraint can also take the infinite value, i.e., $r_p = \infty$, a different approach than [15] is required to prove the results.

where $K \triangleq H(F_{\boldsymbol{r}^o}) - \lambda P_a + \mu P_d$. From Lagrangian theory, we obtain that in order for a distribution $F_{\boldsymbol{r}^o}$ to be optimal (achieving the maximum), it is necessary and sufficient to

$$L'_{F_{\boldsymbol{r}^o}}(F_{\boldsymbol{r}}, \lambda, \mu) \leq 0, \quad \forall F_{\boldsymbol{r}} \in \Omega_1 \cap \Omega_2. \tag{64}$$

Following the same approach in [14]–[18], it is verified that (64) is equivalent to

$$\begin{cases} h(r; F_{\boldsymbol{r}^o}) - \lambda r^2 + \mu g(r) & = K, \quad r \in \text{supp}\{\boldsymbol{r}^o\} \\ h(r; F_{\boldsymbol{r}^o}) - \lambda r^2 + \mu g(r) & \leq K, \quad r \in [0, r_p]. \end{cases} \tag{65}$$

Assume that the optimal input $\boldsymbol{r}^o$ contains at least one limit point in its support. This case occurs if support of $\boldsymbol{r}^o$ contains an interval or it is discrete with an infinite number of mass points[22]. Extending the equation in (65) to the complex domain, we have

$$h(z; F_{\boldsymbol{r}^o}) = \lambda z^2 - \mu g(z) + K, \quad z \in \text{Re}(z) > 0. \tag{66}$$

$h(z; F_{\boldsymbol{r}^o})$ is analytic due to analyticity of $K(R, z)$ (see (16)) on the domain defined by $\text{Re}(z) > 0$. (66) holds if z is the support of $\boldsymbol{r}^o$ on $[0, r_p]$ (due to (65)). Hence, by the identity theorem, we have $h(z; F_{\boldsymbol{r}^o}) = \lambda z^2 - \mu g(z) + K$ over the whole domain $\text{Re}(z) > 0$ if $z \in \text{supp}\{\boldsymbol{r}^o\}$ is a limit point. In the following, we examine (66) for different range of values for $\lambda \geq 0$ and $\mu \in \mathbb{R}$.

- ($\lambda = \mu = 0$): Expanding $h(r; F_{\boldsymbol{r}^o})$ from (16), the KKT equality condition in (65) reads as

$$\int_0^\infty K(R, r) \ln \frac{R}{f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}^o})} dR = H(F_{\boldsymbol{r}}). \tag{67}$$

Noting that the integral transform in (67) is invertible, i.e., the solution is unique (see Appendix J), we have

$$f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}^o}) = R e^{-H(F_{\boldsymbol{r}})}, \tag{68}$$

which can be easily verified that is not a legitimate pdf.

- ($\lambda > 0, \mu = 0$): In this case the problem at hand is reduced to the capacity of an AWGN channel under average power and amplitude constraints. In [15], it shown that the optimal inputs for this setup are discrete with a finite number of mass points.

- ($\lambda \geq 0, \mu \neq 0$): By expanding $h(r; F_{\boldsymbol{r}^o})$ from (16), we have

$$\int_0^\infty K(R, r) \ln \frac{R}{f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}^o})} dR = \int_0^\infty K(R, r) \ln R dR - \int_0^\infty K(R, r) \ln f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}^o}) dR \tag{69}$$

---

[22]The existence of a limit point in this case follows by Bolzano-Weierstrass theorem.

$$> \int_0^1 \ln R \, dR - \int_0^\infty f_{\mathbf{R}}(R; F_{\mathbf{r}^o})) dR > -2, \tag{70}$$

where the first inequality (70) is due to (50) and $\ln x < x$.

Assuming $\mu \neq 0$, from Lemma 6, it can be easily verified that $f_{\mathbf{R}}(R)$ is in the form of

$$f_{\mathbf{R}}(R) = R \exp\left\{\sum_{i=1}^n c_i R^{2i}\right\} \tag{71}$$

It clear that the resulting distribution for the channel output is not a legitimate distribution due the presence of terms R^{2i} in the exponent.

Therefore, the only possibility for the optimal amplitude $\mathbf{r}^o$ is to be discrete with a finite number of mass points. We note that, the channel input is indeed continuous due to the uniformly distributed phase.

$\square$

APPENDIX D

PROOF OF LEMMA 4

Rewriting the function $I_0(x)$, we have

$$I_0(x) = \frac{1}{\pi} \int_0^\pi e^x \cos(t) dt \tag{72}$$

$$= \frac{e^x}{\pi\sqrt{x}} \int_0^{2\sqrt{x}} \frac{e^{-\frac{u^2}{2}}}{\sqrt{1 - \frac{u^2}{4x}}} du \tag{73}$$

$$= \frac{e^x}{\pi\sqrt{x}} \int_0^{2\sqrt{ax}} \frac{e^{-\frac{u^2}{2}}}{\sqrt{1 - \frac{u^2}{4x}}} du + \frac{e^x}{\pi} \int_a^1 \frac{e^{-2xt}}{\sqrt{t(1-t)}} dt \tag{74}$$

$$< \frac{e^x}{\pi\sqrt{x}} \int_0^{2\sqrt{ax}} \left(\frac{\hat{a}u}{\sqrt{x}} + 1\right) e^{-\frac{u^2}{2}} du + \frac{e^x}{\pi} \int_a^1 \frac{e^{-2xt}}{\sqrt{(t-a)(1-t)}} dt, \tag{75}$$

where (72) is the definition, in (73), we used the transformation $u = 2\sqrt{x}\sin(t/2)$, in (74), $0 < a < 1$ and in the last term of (74), we used the transformation $u^2/2 = t$. In (75), we used the inequalities $1/\sqrt{1 - u^2/4x} < \hat{a}u/\sqrt{x} + 1$, $0 \leq u \leq 2\sqrt{ax}$, $\hat{a} = (1/\sqrt{1-a} - 1)/(2\sqrt{a})$[23] and $\sqrt{t} > \sqrt{t-a}$, $t \geq a$, for the first and second terms, respectively. The first integral in (75) is the error function. From [29, ET I 139(23)], the second integral in (75) can be obtained as

$$\int_a^1 \frac{e^{-2xt}}{\sqrt{(t-a)(1-t)}} dt = \pi e^{-2x} \Phi(1/2, 1; 2x(1-a)) < \pi e^{-2ax}. \tag{76}$$

[23]This can be easily verified by noting that the function $f(u) = \frac{1}{\sqrt{1-u^2/4x}} - (\hat{a}u + 1)$ is concave and $f(0) = f(2\sqrt{ax}) = 0$.

The inequality in (76) can be easily verified from the definition of $\Phi(\cdot, \cdot; \cdot)$, that is, we have

$$\Phi(1/2, 1; 2x(1-a)) < \Phi(1, 1; 2x(1-a)) = e^{2x(1-a)}, \tag{77}$$

where the equality in (77) is due to [29, MO 15]. Hence, the term in (75) can be further upper bounded by (76) as follows

$$I_0(x) < e^x \left(\frac{\hat{a}(1 - e^{-2ax})}{\pi x} + \frac{\text{erf}\left(\sqrt{2ax}\right)}{\sqrt{2\pi x}} + e^{-2ax} \right). \tag{78}$$

Since (78) is valid for any $a \in [0, 1)$, therefore, the result of the lemma is concluded.

$\square$

APPENDIX E

PROOF OF LEMMA 6

By substituting $G(R)$ in (44), and using the result of Lemma 5, we have

$$\sum_{i=0}^{n} c_i \int_{0}^{\infty} R^{2i} K(R, r) dR = \sum_{i=0}^{n} c_i 2^i i! e^{-\frac{r^2}{2}} \Phi\left(i + 1, 1; \frac{r^2}{2}\right). \tag{79}$$

The function $\Phi\left(i + 1, 1; \frac{r^2}{2}\right)$ can be easily found for integer values of i using the following two properties of *Confluent Hypergeometric functions* (see [29, MO 15, MO 112])

$$\Phi(i, i; x) = e^x, \ \ i = 1, 2, \ldots, \tag{80a}$$

$$\Phi(a + 1, b; x) = \frac{x}{b}\Phi(a + 1, b + 1; x) + \Phi(a, b; x). \tag{80b}$$

Denoting $e^{-\frac{r^2}{2}}\Phi\left(i, k; \frac{r^2}{2}\right) \triangleq \Phi_r(i, k)$ for $i = 1, 2, \ldots$ and $k = 0, ., i - 1$, we have

$$\Phi_r(i, i) = 1, \tag{81a}$$

$$\Phi_r(i, i - 1) = \frac{r^2}{2(i - 1)}\Phi_r(i, i) + \Phi_r(i - 1, i - 1) \tag{81b}$$

$$= \frac{r^2}{2(i - 1)} + 1, \tag{81c}$$

$$\Phi_r(i, i - 2) = \frac{r^2}{2(i - 2)}\Phi_r(i, i - 1) + \Phi_r(i - 1, i - 2) \tag{81d}$$

$$= \frac{r^2}{2(i - 2)}\left(\frac{r^2}{2(i - 1)} + 1\right) + \frac{r^2}{2(i - 2)} + 1, \tag{81e}$$

$$\vdots \tag{81f}$$

$$\Phi_r(i, k) = \frac{r^2}{2k}\Phi_r(i, k + 1) + \Phi_r(i - 1, k), \tag{81g}$$

$$\vdots \tag{81h}$$

$$\Phi_r(i, 2) = \frac{r^2}{4} \Phi_r(i, 3) + \Phi_r(i - 1, 2), \tag{81i}$$

$$\Phi_r(i, 1) = \frac{r^2}{2} \Phi_r(i, 2) + \Phi_r(i - 1, 1). \tag{81j}$$

Note that for example in (81g), both $\Phi_r(i, k + 1)$, $\Phi_r(i - 1, k)$ can be obtained from the previous stage. Also, it is verified that $\Phi_r(i, k)$ is a polynomial of degree $2(i - k)$, $1 \leq k \leq i$, i.e., the degree of the polynomial depends on the difference of the arguments i, k. Therefore, $\Phi_r(i, 1)$ is a polynomial of degree $2(i - 1)$.

Using the aforementioned approach, in the following, we have calculated $\Phi_r(i, 1)$ for $i = 2, \ldots, 6$

$$\Phi_r(2, 1) = \frac{r^2}{2} + 1, \tag{82a}$$

$$\Phi_r(3, 1) = \frac{r^4}{8} + r^2 + 1, \tag{82b}$$

$$\Phi_r(4, 1) = \frac{r^6}{48} + \frac{3r^4}{8} + \frac{3r^2}{2} + 1, \tag{82c}$$

$$\Phi_r(5, 1) = \frac{r^8}{384} + \frac{r^6}{12} + \frac{3r^4}{4} + 2r^2 + 1, \tag{82d}$$

$$\Phi_r(6, 1) = \frac{r^{10}}{3840} + \frac{5r^8}{384} + \frac{5r^6}{24} + \frac{5r^4}{4} + \frac{5^2}{2} + 1. \tag{82e}$$

Therefore, c_is can be simply found by comparing the RHS of (79) with $g(r)$. Uniqueness of the coefficients c_i is guaranteed by the fact that the integral transform in (44) is invertible (see Appendix J).

$\square$

APPENDIX F
PROOF OF LEMMA 7

By calculating the characteristic function of the complex random variable $\boldsymbol{y}$, we have

$$|M_{\boldsymbol{y}}(z = r_z e^{j\theta_z})| = |\mathbb{E}[e^{j\mathrm{Re}(z^*\boldsymbol{y})}]| \tag{83}$$

$$= |\mathbb{E}[e^{j\mathrm{Re}(z^*\boldsymbol{x})}]| \cdot |\mathbb{E}e^{j\mathrm{Re}(z^*\boldsymbol{n})}]| \tag{84}$$

$$\leq |\mathbb{E}e^{j\mathrm{Re}(z^*\boldsymbol{n})}]| \tag{85}$$

$$= \left| \int_0^\infty \int_0^{2\pi} \frac{r_n}{2\pi} e^{-\frac{r_n^2}{2}} e^{jr_n r_z \cos(\theta_n - \theta_z)} dr_n d\theta_n \right| \tag{86}$$

$$= \left| \int_0^\infty r_n e^{-\frac{r_n^2}{2}} I_0(jr_n r_z) dr_n \right| \tag{87}$$

$$= e^{-\frac{r_z^2}{2}}, \tag{88}$$

where (88) is due to the transform $t = \frac{r_n^2}{2}$ and [29, ET I 197(20)a].

Continuity of $\boldsymbol{y}$ is verified due to continuity of the complex Gaussian noise $\boldsymbol{n}$. From Lemma 8, existence of the pdf of $\boldsymbol{y}$ is guaranteed. Hence, the result of the lemma is proved by Hardy's theorem (see [36]) and (88) and noting that any pdf in the form of $f_{\boldsymbol{y}}(y) = \mathcal{O}(e^{-A|y|^2})$, $A > 1/2$ is identically zero, i.e., $f_{\boldsymbol{y}}(y) = 0$, which is not a legitimate pdf. $\qquad\square$

APPENDIX G

PROOF OF LEMMA 9

Solving $\frac{\partial K(R,r)}{\partial r} = 0$, we have

$$r^* = R\frac{I_0'(rR)}{I_0(rR)} = R\frac{I_1(rR)}{I_0(rR)}, \tag{89}$$

where the second equality in (89) is due to the equality $I_0'(x) = I_1(x)$. Using the inequalities $\frac{I_0'(x)}{I_0(x)} < 1$ and $\frac{I_1(x)}{I_0(x)} \geq \frac{x}{x+1}$ from [37], we have

$$R - \frac{1}{R} \leq r^* < R. \tag{90}$$

Note that for $R > 2$ we have $r^* > R/2$. Rewriting $f_{\boldsymbol{R}}(R, F_{\boldsymbol{r},n})$ for $R > 2$, we have

$$f_{\boldsymbol{R}}(R, F_{\boldsymbol{r},n}) = \int_0^\infty K(R,r)dF_{\boldsymbol{r}}(r) \tag{91}$$

$$= \int_0^{\frac{R}{2}} K(R,r)dF_{\boldsymbol{r}}(r) + \int_{\frac{R}{2}}^\infty K(R,r)dF_{\boldsymbol{r}}(r) \tag{92}$$

$$< K\left(R, R/2\right)\Pr\left(\boldsymbol{r} \leq \frac{R}{2}\right) + K\left(R, r^*\right)\Pr\left(\boldsymbol{r} > \frac{R}{2}\right) \tag{93}$$

$$< K\left(R, R/2\right) + \frac{4P_a}{R^2} \tag{94}$$

$$= Re^{-\frac{5R^2}{8}}I_0(R^2/2) + \frac{4P_a}{R^2} \tag{95}$$

$$< Re^{-\frac{R^2}{8}} + \frac{4P_a}{R^2} \tag{96}$$

$$< \frac{128R}{R^4} + \frac{4P_a}{R^2} \tag{97}$$

$$< \frac{128}{R^2} + \frac{4P_a}{R^2} \tag{98}$$

$$= \frac{128 + 4P_a}{R^2}, \tag{99}$$

where (94) is due to the Markov's inequality and (50). (96) is due to $I_0(x) < e^x$. (97) is due to $e^{-x} \leq \frac{k!}{x^k}$ for any nonnegative integer k (here $k = 2$).

Finally, from (51) and the inequality $|x \ln x| < 4x^{\frac{3}{4}}$, $0 \le x \le 1$ we have

$$|f_{\boldsymbol{R}}(R, F_{\boldsymbol{r},n}) \ln f_{\boldsymbol{R}}(R, F_{\boldsymbol{r},n})| < 4 f_{\boldsymbol{R}}(R, F_{\boldsymbol{r},n})^{\frac{3}{4}} \tag{100}$$

$$< g(R) = \begin{cases} 4 & R \le 2 \\ \frac{c}{R^{\frac{3}{2}}} & R > 2 \end{cases}, \tag{101}$$

where $c = 4(128 + 4P_a)^{\frac{3}{4}}$. It is easy to verify that $g(R)$ is integrable.

$\square$

APPENDIX H

PROOF OF LEMMA 10

1) Existence: Rewriting $|H(F_{\boldsymbol{r}})|$ in (12), we have

$$|H(F_{\boldsymbol{r}})| \le \int_0^\infty f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}}) \ln \frac{1}{f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}})} dR + \int_0^\infty f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}}) |\ln R| dR. \tag{102}$$

The first term in the RHS of (102) is the entropy of the random variable $\boldsymbol{R}$, which exists and is finite due to Lemma 9 and is always positive due to (51). For the second term in the RHS of (102) and for any $F_{\boldsymbol{r}} \in \Omega_1 \cap \Omega_2$ we have

$$\int_0^\infty f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}}) |\ln R| dR = \int_0^1 f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}}) |\ln R| dR + \int_1^\infty f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}}) \ln R dR. \tag{103}$$

The first term in (103) is bounded by noting that $\int_0^1 f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}}) \ln R dR < 0$ and due to

$$\int_0^1 f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}}) \ln R dR > \int_0^1 \ln R dR = -1, \tag{104}$$

where the inequality in (104) is due to (51). The second term in (103) is bounded due to the inequality $\ln x < \sqrt{x}$ and the following lemma

Lemma 11. *The expectation $\mathbb{E}[\sqrt{\boldsymbol{R}}]$ for any $F_{\boldsymbol{r}} \in \Omega_1 \cap \Omega_2$ exists and is bounded.*

Proof: See Appendix I.

Existence of (103) validates existence of $H(F_{\boldsymbol{r}})$ and this concludes the proof.

2) Continuity: Let $F_{\boldsymbol{r},n} \xrightarrow{w} F$. Using the weak topology, the continuity of $H(F_{\boldsymbol{r}})$ is equivalent to

$$F_{\boldsymbol{r},n} \xrightarrow{w} F \implies H(F_{\boldsymbol{r},n}) \to H(F_{\boldsymbol{r}}). \tag{105}$$

Therefore, we have

$$\lim_n H(F_{\boldsymbol{r},n}) = -\lim_n \int_0^\infty f_{\boldsymbol{R}}(R, F_{\boldsymbol{r},n}) \ln \frac{f_{\boldsymbol{R}}(R, F_{\boldsymbol{r},n})}{R} dR \tag{106}$$

$$= -\int_0^\infty \lim_n f_{\boldsymbol{R}}(R, F_{\boldsymbol{r},n}) \ln \frac{f_{\boldsymbol{R}}(R, F_{\boldsymbol{r},n})}{R} dR \tag{107}$$

$$= \int_0^\infty f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}}) \ln \frac{f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}})}{R} dR \tag{108}$$

$$= H(F_{\boldsymbol{r}}), \tag{109}$$

where (106) and (109) are definitions. (107) is due to Lebesgue Dominated Convergence Theorem and absolute integrability of the integrand in (106) due to Lemma 9. (108) is due to continuity of $x \ln x$.

3) Strict concavity: Concavity follows by noting that in (102), the first term is the entropy function and therefore concave with respect to the distribution function $f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}})$, and the second term is a linear function of $f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}})$. Strict concavity follows by noting that the transform

$$f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}}) = \int_0^\infty K(R, r) dF_{\boldsymbol{r}}(r), \tag{110}$$

is invertible (for the proof see [15, Appendix II]).

3) Weak differentiability: The proof for weak differentiability is the same as [15, Proposition 4] or [16, AppendixII.B]. For brevity we avoid the details and conclude the proof by providing the final result of applying weak derivative over (102), which is given as

$$H'_{F_{\boldsymbol{r}^o}}(F_{\boldsymbol{r}}) = \lim_{\theta \to 0} \frac{H((1-\theta)F_{\boldsymbol{r}}^0 + \theta F_{\boldsymbol{r}}) - H(F_{\boldsymbol{r}}^0)}{\theta}, \ \theta \in [0,1] \tag{111}$$

$$= \int_0^\infty h(r; F_{\boldsymbol{r}^o}) dF_{\boldsymbol{r}} - H(F_{\boldsymbol{r}}^0), \tag{112}$$

where $h(r; F_{\boldsymbol{r}^o})$ is defined as in (16).

We conclude the proof by noting that the integral transform in (16) is invertible (see Appendix J).

APPENDIX I

PROOF OF LEMMA 11

For $\mathbb{E}[\boldsymbol{R}^\alpha]$ we have

$$\mathbb{E}[\boldsymbol{R}^\alpha] = \int_0^\infty R^\alpha f_{\boldsymbol{R}}(R; F_{\boldsymbol{r}}) dR \tag{113}$$

$$= \int_0^\infty \int_0^\infty R^\alpha K(R, r) dF_{\boldsymbol{r}}(r) dR \tag{114}$$

$$= \int_0^2 \int_0^\infty R^\alpha K(R, r) dF_{\boldsymbol{r}}(r) dR + \int_2^\infty \int_0^1 R^\alpha K(R, r) dF_{\boldsymbol{r}}(r) dR + \int_2^\infty \int_1^\infty R^\alpha K(R, r) dF_{\boldsymbol{r}}(r) dR, \tag{115}$$

where we have divided the integrals due to the similar reason explained in Appendix (G) (see equation (90)).

For the first integral in the RHS of (115) we have

$$\int\limits_{0}^{2}\int\limits_{0}^{\infty} R^{\alpha}K(R,r)dF_{\boldsymbol{r}}(r)dR < \int\limits_{0}^{2} R^{\alpha}dR = \frac{2^{1+\alpha}}{1+\alpha} < \infty, \ \alpha \geq 0, \tag{116}$$

where the inequality is due to (50). For the second integral in the RHS of (115) we have

$$\int\limits_{2}^{\infty}\int\limits_{0}^{1} R^{\alpha}K(R,r)dF_{\boldsymbol{r}}(r)dR < \int\limits_{2}^{\infty}\int\limits_{0}^{1} R^{\alpha}K(R,1)dF_{\boldsymbol{r}}(r)dR \tag{117}$$

$$< \int\limits_{0}^{\infty} R^{1+\alpha}e^{-\frac{(R-1)^2}{2}} < \infty, \ \alpha \geq 0, \tag{118}$$

where in (117) we used $K(R,r) \leq K(R,1)$ for $R \geq 2, r \leq 1$ due to (90). In (118) we used the inequality $I_0(x) < e^x$.

Note that for $R \geq 2$, it is easy to verify that $1 + R/4 \leq R - 1/R$. This along with (90), guarantee that

$$K(R,r) \leq K(R, 1 + R/4), \quad R \geq 2, \ r \leq 1 + R/4. \tag{119}$$

Therefore, for the third integral in the RHS of (115) we have

$$\int\limits_{2}^{\infty}\int\limits_{1}^{\infty} R^{\alpha}K(R,r)dF_{\boldsymbol{r}}(r)dR < \int\limits_{2}^{\infty}\int\limits_{1}^{1+\frac{R}{4}} R^{\alpha}K(R,r)dF_{\boldsymbol{r}}(r)dR + \int\limits_{2}^{\infty}\int\limits_{1+\frac{R}{4}}^{\infty} R^{\alpha}K(R,r)dF_{\boldsymbol{r}}(r)dR \tag{120}$$

$$< \int\limits_{2}^{\infty} R^{\alpha}K(R, 1 + R/4)dR + \int\limits_{2}^{\infty} R^{\alpha}\mathrm{Pr}(r > 1 + R/4)dR \tag{121}$$

$$< \int\limits_{2}^{\infty} \frac{R^{\alpha+\frac{1}{2}}}{1 + \frac{R}{4}}e^{-\frac{(3R-4)^2}{32}} dR + \int\limits_{2}^{\infty} \frac{P_a R^{\alpha}}{(1 + R/4)^2} dR < \infty, \ 0 \leq \alpha < 1, \tag{122}$$

where (121) is due to (119) and (50). In (122) we used Markov's inequality. From (116), (118) and (122), it can be easily verified that $E[\boldsymbol{R}^{\alpha}]$ for $0 \leq \alpha < 1$ exists, which also concludes the result of Lemma 11.

$\square$

APPENDIX J

PROOF OF INVERTIBILITY OF THE INTEGRAL TRANSFORM

Consider the following transform

$$V(r) = \int\limits_{0}^{\infty} K(R,r)S(R)dR, \tag{123}$$

where $S(R)$ is allowed to be a polynomial with a finite degree in order to guarantee the existence of the transform.

To prove the invertibility, it is enough to show that $S(R) = 0$ if and only if $V(r) = 0$. It is easily verified that $S(R) = 0$ yields $V(r) = 0$. For the converse, assume $V(r) = 0$. By taking the second integral over r as below, we have

$$\int_0^\infty re^{-sr^2} \int_0^\infty K(R,r)S(R)dRdr = 0, \quad s \geq 0. \tag{124}$$

By changing the order of the integrals in (124) (This is validated by our assumption on $S(R)$ and due to Fubini's theorem), we have

$$\int_0^\infty \int_0^\infty re^{-sr^2} K(R,r)S(R)drdR = \int_0^\infty \frac{Re^{-\frac{R}{2}}S(R)}{2} \int_0^\infty e^{-t(s+\frac{1}{2})} I_0\left(R\sqrt{t}\right) drdR \tag{125}$$

$$= \frac{1}{1+2s} \int_0^\infty Re^{-R^2\left(\frac{1+s}{1+2s}\right)} S(R)dR = 0, \quad s \geq 0, \tag{126}$$

where (125) is obtained by expanding $K(R,r)$ and transformation $R^2 = t$. (126) is obtained using [29, ET I 197(20)a, MO 115, MO 15]. From (126) it is verified that (124) is valid only if $S(R) = 0$.

$\square$

APPENDIX K

PROOF OF LEMMA 1

The following series will be useful throughout the proof of Lemma 1.

Lemma 12. *We have the following series[24]:*

$$S_0 \triangleq \sum_l s_l^2 = 1, \tag{127}$$

$$S_1 \triangleq \sum_l \sum_{k:k\neq l} s_l s_k = 0, \tag{128}$$

$$S_2 \triangleq \sum_l \sum_{k:k\neq l} \sum_{\substack{d:d\neq l \\ d\neq k}} \sum_{\substack{m:m\neq l \\ m\neq d \\ m\neq k}} s_l s_k s_d s_m = 0, \tag{129}$$

$$S_3 \triangleq \sum_l \sum_{k:k\neq l} s_l^2 s_k^2 = \frac{2}{3}, \tag{130}$$

[24]The summations are from $-\infty$ to ∞. They are removed due to brevity.

$$S_4 \triangleq \sum_l \sum_{k:k\neq l} \sum_{\substack{d:d\neq l \\ d\neq k}} s_l^2 s_k s_d = -\frac{1}{3}, \tag{131}$$

$$S_5 \triangleq \sum_l s_l^4 = \frac{1}{3}, \tag{132}$$

$$S_6 \triangleq \sum_l \sum_{k:k\neq l} s_l^3 s_k = \frac{1}{6}. \tag{133}$$

Proof: See Appendix M.

Considering first the term $\mathbb{E}\mathcal{E}[Y_{\mathrm{rf}}(t)^2]$, we have

$$\mathbb{E}\mathcal{E}[Y_{\mathrm{rf}}(t)^2] = \frac{1}{2}\mathbb{E}\mathcal{E}\left[\left(Y(t)e^{j2\pi f_c t} + \overline{Y(t)}e^{-j2\pi f_c t}\right)^2\right] \tag{134}$$

$$= \mathbb{E}\mathcal{E}\left[|Y(t)|^2\right] \tag{135}$$

$$= \mathbb{E}\mathcal{E}\left[\sum_{n,k} \boldsymbol{y}_n \overline{\boldsymbol{y}_k}\mathrm{sinc}(f_w t - n)\mathrm{sinc}(f_w t - k)\right] \tag{136}$$

$$= \sum_{n,k} \mathbb{E}\left[\boldsymbol{y}_n \overline{\boldsymbol{y}_k}\right]\mathcal{E}\left[\mathrm{sinc}(f_w t - n)\mathrm{sinc}(f_w t - k)\right] \tag{137}$$

$$= \lim_{T\to\infty} \frac{1}{f_w T}\sum_k \mathbb{E}\left[|\boldsymbol{y}_k|^2\right] \tag{138}$$

$$= P + \sigma_n^2, \tag{139}$$

where (135) is because we have $\mathcal{E}\{Y(t)^2 e^{j4\pi f_c t}\} = \mathcal{E}\{\overline{Y(t)}^2 e^{-j4\pi f_c t}\} = 0$. (136) is due to the fact that the signal $Y(t)$ is bandlimited to f_w and we have

$$Y(t) = \sum_k \boldsymbol{y}_k \mathrm{sinc}(f_w t - k). \tag{140}$$

In (138), we used the equation

$$\mathcal{E}\left[\mathrm{sinc}(f_w t - n)\mathrm{sinc}(f_w t - k)\right] = \lim_{T\to\infty} \frac{1}{f_w T}\delta_{n-k}. \tag{141}$$

Considering the term $\mathbb{E}\mathcal{E}[Y_{\mathrm{rf}}(t)^4]$, similarly, we have

$$\mathbb{E}\mathcal{E}[Y_{\mathrm{rf}}(t)^4] = \frac{3}{2}\mathbb{E}\mathcal{E}\left[|Y(t)|^4\right]. \tag{142}$$

Note that, the signal $|Y(t)|^2$ is real with bandwidth $[-f_w, f_w]$. Hence, it can be represented by its samples taken each $t = 1/2f_w$ seconds. Therefore, we have

$$|Y(t)|^2 = \sum_k \boldsymbol{s}_k \mathrm{sinc}(2f_w t - k), \tag{143}$$

where $\boldsymbol{s}_k \triangleq |Y(k/2f_w)|^2$. Accordingly, (142) reads as

$$\frac{3}{2}\mathbb{E}\mathcal{E}\left[|Y(t)|^4\right] = \lim_{T\to\infty} \frac{3}{2f_w}\sum_k \mathbb{E}[|\boldsymbol{s}_k|^2] \tag{144}$$

$$= \lim_{T \to \infty} \frac{3}{2Tf_w} \sum_k \mathbb{E}[|\boldsymbol{s}_{2k+1}|^2] + \frac{3}{2Tf_w} \sum_k \mathbb{E}[|\boldsymbol{s}_{2k}|^2] \tag{145}$$

$$= \frac{3}{2} \left(\mathbb{E}[|\boldsymbol{s}_{2k+1}|^2] + \mathbb{E}[|\boldsymbol{s}_{2k}|^2] \right). \tag{146}$$

Note that $\boldsymbol{s}_{2k} = |Y(2k/2f_w)|^2 = |\boldsymbol{y}_k|^2$. Hence, $\mathbb{E}[|\boldsymbol{s}_{2k}|^2]$ in (145) reads

$$\mathbb{E}[|\boldsymbol{s}_{2k}|^2] = \mathbb{E}[|\boldsymbol{y}_k|^4] \tag{147}$$

$$= \mathbb{E}[|(\boldsymbol{x} + \boldsymbol{n})(\overline{\boldsymbol{x}} + \overline{\boldsymbol{n}})|^2]$$

$$= Q + 16P + 32. \tag{148}$$

To calculate the term $\mathbb{E}[|\boldsymbol{s}_{2k+1}|^2]$ in (145), we note that the channel's baseband equivalent signal $Y(t)$ can be written as

$$Y(t) = \sum_n \boldsymbol{x}_n \mathrm{sinc}(f_w t - n) + W(t), \tag{149}$$

Substituting $t = (2k+1)/f_w$ we have

$$\tilde{\boldsymbol{y}}_k \triangleq Y(t)|_{t=\frac{2k+1}{2f_w}} \tag{150}$$

$$= \tilde{\boldsymbol{x}} + \tilde{\boldsymbol{n}}. \tag{151}$$

where $\tilde{\boldsymbol{x}} \triangleq \sum_{n=-\infty}^{\infty} \boldsymbol{x}_n s_{k-n}$ and $\tilde{\boldsymbol{n}} \triangleq W((2k+1)/2f_w)$. Similarly to (148), we have

$$\mathbb{E}[|\boldsymbol{s}_{2k+1}|^2] = \mathbb{E}[|\tilde{\boldsymbol{y}}_k|^4] \tag{152}$$

$$= \tilde{Q} + 16\tilde{P} + 32, \tag{153}$$

where $\tilde{Q} = \mathbb{E}[|\tilde{\boldsymbol{x}}|^4]$, $\tilde{P} = \mathbb{E}[|\tilde{\boldsymbol{x}}|^2]$. For $\tilde{P}$, we have

$$\tilde{P} = \mathbb{E}\left[\sum_{n,m} \boldsymbol{x}_n \overline{\boldsymbol{x}_m} s_{k-n} s_{k-m} \right] \tag{154}$$

$$= \sum_{n,m:n=m} \mathbb{E}[|\boldsymbol{x}_n|^2] s_{k-n}^2 + \sum_{n,m:n \neq m} \mathbb{E}[\boldsymbol{x}_n] \mathbb{E}[\overline{\boldsymbol{x}_m}] s_{k-n} s_{k-m} \tag{155}$$

$$= S_0 P + S_1 |\mu|^2 \tag{156}$$

$$= P, \tag{157}$$

where in (155) we used the assumption that $\boldsymbol{x}_n$ is i.i.d. with respect to different values of n. For $\tilde{Q}$, we have

$$\tilde{Q} = \mathbb{E}\left[\sum_{l,k,d,m} \boldsymbol{x}_l \overline{\boldsymbol{x}_k} \boldsymbol{x}_d \overline{\boldsymbol{x}_m} s_{n-l} s_{n-k} s_{n-d} s_{n-m} \right]. \tag{158}$$

Accounting for the different cases for the possible values of l, k, d, m, we have

- If all the indices l, k, d, m are with different values, we have

$$\tilde{Q} = |\mu|^4 S_2. \tag{159}$$

- If $(l = k, \ d \neq k, \ d = m)$ or $(l = d, \ k \neq d, \ k = m)$, we have

$$\tilde{Q} = P^2 S_3. \tag{160}$$

- If $(l = m, \ k \neq m, \ k = d)$, we have

$$\tilde{Q} = |P'|^2 S_3. \tag{161}$$

- If $(l = k, \ d \neq m, \ d \neq k, \ m \neq k)$ or $(l = d, \ k \neq m, \ k \neq d, \ m \neq d)$ or $(k = m, \ l \neq d, \ l \neq m, \ d \neq m)$ or $(d = m, \ l \neq k, \ l \neq m, \ k \neq m)$, we have

$$\tilde{Q} = P|\mu|^2 S_4. \tag{162}$$

- If $(l = m, \ k \neq d, \ k \neq m, \ d \neq m)$, we have

$$\tilde{Q} = P'\overline{\mu}^2 S_4. \tag{163}$$

- If $(k = d, \ l \neq m, \ l \neq d, \ m \neq d)$, we have

$$\tilde{Q} = \overline{P'}\mu^2 S_4. \tag{164}$$

- If $l = k = d = m$, we have

$$\tilde{Q} = QS_5. \tag{165}$$

- If $l = k = d \neq m$ or $k = d = m \neq l$, we have

$$\tilde{Q} = \overline{T'}\mu S_6. \tag{166}$$

- If $l = d = m \neq k$ or $l = k = m \neq d$, we have

$$\tilde{Q} = T'\overline{\mu}S_6. \tag{167}$$

In the above expressions we define $P' \triangleq \mathbb{E}[\boldsymbol{x}^2]$, $T' \triangleq \mathbb{E}[|\boldsymbol{x}|^2\boldsymbol{x}]$. Hence, (158) reads

$$\tilde{Q} = |\mu|^4 S_2 + (2P^2 + |P'|^2)S_3 + (4P|\mu|^2 + P'\overline{\mu}^2 + \overline{P'}\mu^2)S_4 + QS_5 + 2(T'\overline{\mu} + \overline{T'}\mu)S_6$$

$$= \frac{1}{3}\left[Q + 4P(P - |\mu|^2) + 2(|P'|^2 - \text{Re}\{P'\overline{\mu}^2\}) + 2\text{Re}\{T'\overline{\mu}\} \right]. \tag{168}$$

Expanding the terms $|P'|^2 - \text{Re}\{P'\overline{\mu}^2\}$ and $\text{Re}\{T'\overline{\mu}\}$ in (168), we have

$$|P'|^2 - \text{Re}\{P'\overline{\mu}^2\} = (P_r - P_i)(P_r - P_i - (\mu_r^2 - \mu_i^2)), \tag{169}$$

$$\text{Re}\{T'\overline{\mu}\} = \mu_r(T_r + \mu_r P_i) + \mu_i(T_i + \mu_i P_r). \tag{170}$$

Noting that $Q = Q_i + Q_r + 2P_r P_i$ and substituting in (168) along with (169) and (170), after some manipulations $\tilde{Q}$ reads

$$\tilde{Q} = \frac{1}{3}\big(Q_r + Q_i + 2(\mu_r T_r + \mu_i T_i) + 6(P_r P_i + P_r(P_r - \mu_r^2) + P_i(P_i - \mu_i))\big). \tag{171}$$

Substituting (171), (157) in (153) and substituting the result along with (148) in (145), and adding with (139) yields the result of the Proposition.

APPENDIX L

PROOF OF LEMMA 2

Note that constraining the input distributions $f_{\boldsymbol{x}}(x)$ to those of non-zero mean Gaussian distributionS for each dimension, we have $\text{Re}\{\boldsymbol{x}\} \sim (\mu_r, \sigma_r^2)$ and $\text{Im}\{\boldsymbol{x}\} \sim (\mu_i, \sigma_i^2)$, where $\sigma_r^2 \triangleq P_r - \mu_r^2$ and $\sigma_i^2 \triangleq P_i - \mu_i^2$. Therefore, the rate maximization problem reads

$$\max_{\mu_r, \mu_i, P_r, P_i} \quad \frac{f_w}{2}\big(\ln(1 + a\sigma_r^2) + \ln(1 + a\sigma_i^2)\big)$$

$$\text{s.t.} \quad \begin{cases} P_r + P_i \le P_a \\ \alpha(Q + \tilde{Q}) + \beta P + \gamma \ge P_d \\ \sigma_r^2 \ge 0, \ \sigma_i^2 \ge 0 \end{cases} \tag{172}$$

where $a \triangleq 2/f_w \sigma_n^2$. Writing the K.K.T. conditions for the optimization problem in (172), we have

$$\lambda_1(P_r + P_i - P_a) = 0, \ \lambda_1 \ge 0 \tag{173}$$

$$\lambda_2(\alpha(Q + \tilde{Q}) + \beta P + \gamma - P_d) = 0, \ \lambda_2 \ge 0, \tag{174}$$

$$\zeta_r \sigma_r^2 = 0, \ \zeta_i \sigma_i^2 = 0, \ \zeta_r, \zeta_i \ge 0 \tag{175}$$

$$\zeta_r = \frac{-f_w a}{2(1 + a\sigma_r^2)} + \lambda_1 - \lambda_2(2\alpha(3P_r + P_i) + \beta), \tag{176}$$

$$\zeta_i = \frac{-f_w a}{2(1 + a\sigma_i^2)} + \lambda_1 - \lambda_2(2\alpha(3P_i + P_r) + \beta), \tag{177}$$

$$\frac{f_w a \mu_r}{1 + a\sigma_r^2} + 8\lambda_2 \alpha \mu_r^3 + 2\zeta_r \mu_r = 0, \tag{178}$$

$$\frac{f_w a \mu_i}{1 + a\sigma_i^2} + 8\lambda_2 \alpha \mu_i^3 + 2\zeta_i \mu_i = 0, \tag{179}$$

where in (176) to (179) we used the following

$$\frac{\partial Q}{\partial P_r} = \frac{\partial \tilde{Q}}{\partial P_r} = 6P_l + 2P_i, \tag{180}$$

$$\frac{\partial Q}{\partial P_i} = \frac{\partial \tilde{Q}}{\partial P_i} = 6P_i + 2P_l, \tag{181}$$

$$\frac{\partial Q}{\partial \mu_r} = \frac{\partial \tilde{Q}}{\partial \mu_r} = -8\mu_r^3, \tag{182}$$

$$\frac{\partial Q}{\partial \mu_i} = \frac{\partial \tilde{Q}}{\partial \mu_i} = -8\mu_i^3. \tag{183}$$

It can be easily verified from (173), (176) and (177) that when $\lambda_2 = 0$, the maximum is achieved when $\mu_r = \mu_i = 0$ and $P_r = P_i = \frac{P_a}{2}$, yielding $P_{\mathrm{del}} = 2\alpha P_a^2 + \beta P_a + \gamma$. For positive values of λ_2 from (176) it is verified that $\lambda_1 > 0$, which from (173) results that $P_r + P_i = P_a$. The condition $P_r + P_i = P_a$ reduces the number of variables P_i, P_r to one. Accordingly, since the rate (expansion of the mutual information accounting Gaussian input) is concave wrt $P_i \in [0, P_a]$ attaining its maximum and minimum at $P_i = P_a/2$ and $P_i = 0, \ P_a$, respectively and the delivered power P_{del} is convex wrt $P_i \in [0, P_a]$ attaining its maximum and minimum at $P_i = 0, \ P_a$ and $P_i = P_a/2$, respectively, the Proposition is proved.

APPENDIX M
PROOF OF LEMMA 12

In the following, if not mentioned, the summations are from $-\infty$ to ∞. We have

$$T_0 = \sum_l s_l \tag{184}$$

$$= \frac{1}{\pi} \sum_l \frac{(-1)^l}{\left(\frac{1}{2} + l\right)} \tag{185}$$

$$= \frac{2}{\pi} \left[\sum_{l=-\infty}^{-1} \frac{(-1)^l}{(2l+1)} + \sum_0^\infty \frac{(-1)^l}{(2l+1)} \right] \tag{186}$$

$$= \frac{2}{\pi} \left(\frac{\pi}{4} + \frac{\pi}{4} \right) = 1, \tag{187}$$

$$S_0 = \sum_l s_l^2 \tag{188}$$

$$= \sum_l \frac{(-1)^{2l}}{\pi^2 \left(\frac{1}{2} + l\right)^2} \tag{189}$$

$$= \frac{4}{\pi^2} \sum_l \frac{1}{(2l+1)^2} \tag{190}$$

$$= \frac{4}{\pi^2} \left[\sum_{l=-\infty}^{-1} \frac{1}{(2l+1)^2} + \sum_0^\infty \frac{1}{(2l+1)^2} \right] \tag{191}$$

$$= \frac{4}{\pi^2} \left(\frac{\pi^2}{8} + \frac{\pi^2}{8} \right) = 1, \tag{192}$$

$$T_1 = \sum_l s_l^3 \tag{193}$$

$$= \sum_l \frac{(-1)^{3l}}{\pi^3 \left(\frac{1}{2} + l\right)^3} \tag{194}$$

$$= \frac{8}{\pi^3} \left[\sum_{l=-\infty}^{-1} \frac{1}{(2l+1)^3} + \sum_0^\infty \frac{1}{(2l+1)^3} \right] \tag{195}$$

$$= \frac{8}{\pi^3} \left(\frac{\pi^3}{32} + \frac{\pi^3}{32} \right) = \frac{1}{2}, \tag{196}$$

$$S_5 = \sum_l s_l^4 \tag{197}$$

$$= \sum_l \frac{(-1)^{4l}}{\pi^4 \left(\frac{1}{2} + l\right)^4} \tag{198}$$

$$= \frac{16}{\pi^4} \sum_l \frac{1}{(2l+1)^4} \tag{199}$$

$$= \frac{16}{\pi^4} \left[\sum_{l=-\infty}^{-1} \frac{1}{(2l+1)^4} + \sum_0^\infty \frac{1}{(2l+1)^4} \right] \tag{200}$$

$$= \frac{16}{\pi^4} \left(\frac{\pi^4}{96} + \frac{\pi^4}{96} \right) = \frac{1}{3}, \tag{201}$$

$$S_1 = \sum_l \sum_{k, k \neq l} s_l s_k \tag{202}$$

$$= \sum_l s_l \left(\sum_k s_k - s_l \right) \tag{203}$$

$$= \left(\sum_l s_l \right)^2 - \sum_l s_l^2 \tag{204}$$

$$= 1 - 1 = 0, \tag{205}$$

$$S_3 = \sum_l \sum_{k, k \neq l} s_l^2 s_k^2 \tag{206}$$

$$= \sum_l s_l^2 \left(\sum_k s_k^2 - s_l^2 \right) \tag{207}$$

$$= \left(\sum_l s_l^2 \right)^2 - \sum_l s_l^4 \tag{208}$$

$$= 1 - \frac{1}{3} = \frac{2}{3}, \tag{209}$$

$$S_6 = \sum_l \sum_{k, k \neq l} s_l^3 s_k \tag{210}$$

$$= \sum_l s_l^3 \left(\sum_k s_k - s_l \right) \tag{211}$$

$$= \frac{1}{2} - \frac{1}{3} = \frac{1}{6}, \tag{212}$$

$$S_4 = \sum_l \sum_{k,k\neq l} \sum_{\substack{d,d\neq l \\ d\neq k}} s_l^2 s_k s_d \tag{213}$$

$$= \sum_l \sum_{k,k\neq l} s_l^2 s_k \left(\sum_d s_d - s_l - s_k \right) \tag{214}$$

$$= \sum_l s_l^2 \left((1-s_l) \sum_{k,k\neq l} s_k - \sum_{k,k\neq l} s_k^2 \right) \tag{215}$$

$$= \sum_l s_l^2 \left((1-s_l)^2 - (1-s_l^2) \right) \tag{216}$$

$$= \sum_l 2s_l^2 (s_l^2 - s_l) \tag{217}$$

$$= 2 \left(\frac{1}{3} - \frac{1}{2} \right) = -\frac{1}{3}, \tag{218}$$

$$S_2 = \sum_l \sum_{k,k\neq l} \sum_{\substack{d,d\neq l \\ d\neq k}} \sum_{\substack{m,m\neq d \\ m\neq l \\ m\neq k}} s_l s_k s_d s_m \tag{219}$$

$$= \sum_l \sum_{k,k\neq l} \sum_{\substack{d,d\neq l \\ d\neq k}} s_l s_k s_d (1 - s_d - s_l - s_k) \tag{220}$$

$$= \sum_l \sum_{k,k\neq l} s_l s_k \left((1 - s_l - s_k) \sum_{\substack{d,d\neq l \\ d\neq k}} s_d - \sum_{\substack{d,d\neq l \\ d\neq k}} s_d^2 \right) \tag{221}$$

$$= \sum_l \sum_{k,k\neq l} s_l s_k \left((1 - s_l - s_k)^2 - (1 - s_l^2 - s_k^2) \right) \tag{222}$$

$$= \sum_l s_l \left(2s_l(s_l - 1)(1 - s_l) + \sum_{k,k\neq l} 2s_k(s_k^2 + s_l s_k - s_k) \right) \tag{223}$$

$$= \sum_l s_l(-6s_l^3 + 6s_l^2 - 1)) \tag{224}$$

$$= -\frac{6}{3} + \frac{6}{3} - 1 = 0. \tag{225}$$

REFERENCES

[1] M. Varasteh, B. Rassouli, and B. Clerckx, "Wireless information and power transfer over an awgn channel: Nonlinearity and asymmetric gaussian signaling," in *IEEE Inf. Theory Workshop (ITW)*, Nov. 2017, pp. 181–185.

[2] L. R. Varshney, "Transporting information and energy simultaneously," in *IEEE Int. Sym. Inf. Theory*, Jul. 2008, pp. 1612–1616.

[3] P. Grover and A. Sahai, "Shannon meets tesla: Wireless information and power transfer," in *IEEE Int. Sym. Inf. Theory*, Jun. 2010, pp. 2363–2367.

[4] R. Zhang and C. Keong, "Mimo broadcasting for simultaneous wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 1989–2001, May 2013.

[5] 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.

[6] M. S. Trotter, J. D. Griffin, and G. D. Durgin, "Power-optimized waveforms for improving the range and reliability of rfid systems," in *IEEE Int. Conf. RFID*, Apr. 2009, pp. 80–87.

[7] B. Clerckx and E. Bayguzina, "Waveform design for wireless power transfer," *IEEE Trans. Signal Processing*, vol. 64, no. 23, pp. 6313–6328, Dec. 2016.

[8] ——, "Low-complexity adaptive multisine waveform design for wireless power transfer," *IEEE Antennas and Wireless Propagation Letters*, vol. 16, pp. 2207–2210, 2017.

[9] J. Kim, B. Clerckx, and P. D. Mitcheson, "Prototyping and experimentation of a closed-loop wireless power transmission with channel acquisition and waveform optimization," *CoRR*, vol. abs/1704.01799, 2017. [Online]. Available: http://arxiv.org/abs/1704.01799

[10] 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.

[11] B. Clerckx, "Wireless information and power transfer: Nonlinearity, waveform design, and rate-energy tradeoff," *IEEE Trans. Signal Processing*, vol. 66, no. 4, pp. 847–862, Feb. 2018.

[12] ——, "Waveform optimization for swipt with nonlinear energy harvester modeling," in *20th International ITG Workshop Smart Antennas (WSA)*, Mar. 2016, pp. 1–5.

[13] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, Jul. and Oct. 1948.

[14] J. G. Smith, "The information capacity of amplitude- and variance-constrained scalar Gaussian channels," *Inform. Control*, vol. 18, pp. 203–219, 1971.

[15] S. Shamai and I. Bar-David, "The capacity of average and peak-power-limited quadrature gaussian channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 1060–1071, Jul. 1995.

[16] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless rayleigh-fading channels," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.

[17] A. Tchamkerten, "On the discreteness of capacity-achieving distributions," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2773–2778, Nov. 2004.

[18] 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.

[19] J. Pedro and N. Carvalho, *Intermodulation Distortion in Microwave and Wireless Circuits*. 1st ed. Norwood, MA: Artech House, 2003.

[20] R. J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *Journal of Lightwave Technology*, vol. 28, no. 4, pp. 662–701, Feb. 2010.

[21] E. Agrell, A. Alvarado, G. Durisi, and M. Karlsson, "Capacity of a nonlinear optical channel with finite memory," *Journal of Lightwave Technology*, vol. 32, no. 16, pp. 2862–2876, Aug. 2014.

[22] E. Biglieri, E. Chiaberto, G. P. Maccone, and E. Viterbo, "Compensation of nonlinearities in high-density magnetic recording channels," *IEEE Trans. Magnetics*, vol. 30, no. 6, pp. 5079–5086, Nov. 1994.

[23] P. Banelli and S. Cacopardi, "Theoretical analysis and performance of ofdm signals in nonlinear awgn channels," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 430–441, Mar. 2000.

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

[25] R. P. Feynman, *Feynman Lectures on Computation*, J. G. Hey and R. W. Allen, Eds. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1998.

[26] R. Landauer, *Computation, Measurement, Communication and Energy Dissipation*. IBM Thomas J. Watson Research Division, 1987.

[27] M. P. Frank, *Reversibility for efficient computing*. Ph.D. dissertation, Massachusetts Institute of Technology, 1999.

[28] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[29] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 7th ed.   Elsevier/Academic Press, Amsterdam, 2007.

[30] R. G. Gallager, *Information Theory and Reliable Communication*.   New York, NY, USA: John Wiley & Sons, Inc., 1968.

[31] R. Morsi, V. Jamali, D. K. Ng, and R. Schober, "On the capacity of swipt systems with a nonlinear energy harvesting circuit," *CoRR*, vol. abs/1711.01082, 2017. [Online]. Available: http://arxiv.org/abs/1711.01082

[32] E. Bayguzina and B. Clerckx, "Modulation design for wireless information and power transfer with nonlinear energy harvester modeling," *CoRR*, vol. abs/1802.06512, 2018. [Online]. Available: http://arxiv.org/abs/1802.06512

[33] B. Rassouli and B. Clerckx, "On the capacity of vector gaussian channels with bounded inputs," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 6884–6903, Dec. 2016.

[34] B. Rassouli, M. Varasteh, and D. Gündüz, "Gaussian multiple access channels with one-bit quantizer at the receiver," *CoRR*, vol. abs/1703.02324, 2017. [Online]. Available: http://arxiv.org/abs/1703.02324

[35] D. Luenberger, *Optimization by Vector Space Methods*.   New York: John Wiley & Sons, Inc, 1969.

[36] E. M. Stein and R. Shakarchi, *Complex Analysis*.   Princeton, NJ: Princeton Univ. Pres, 2003.

[37] A. Laforgia and P. Natalini, "Some inequalities for modified bessel functions," *Journal of Inequalities and Applications*, vol. 2010, no. 1, p. 253035, Jan. 2010.