# Sum-Rate Maximization for Energy Harvesting Nodes With a Generalized Power Consumption Model

Maria Gregori, Miquel Payaró, *Senior Member, IEEE*, and Daniel P. Palomar, *Fellow, IEEE*

***Abstract*— This paper considers a network of energy harvesting wireless nodes transmitting simultaneously in a Gaussian interference channel and investigates a distributed power allocation algorithm that maximizes the sum-rate. The power consumption model is based on a series of step functions that allow to model, among others, radio frequency circuits being on/off and the startup power consumption of the transmitter. After showing that the sum-rate maximization problem is nonsmooth, nonconvex, and NP-hard, the Iterative Smooth and Convex approximation Algorithm (ISCA) is proposed, which successively approximates the step functions by proper smooth functions to obtain a sequence of smooth nonconvex problems that can be solved by means of the successive convex approximation method. It is demonstrated that the ISCA distributedly converges to a stationary solution of the sum-rate maximization problem. For the particular case of point to point communications, the numerical results show that the ISCA is able to avoid bad stationary solutions, performing close to the globally optimal solution. The performance of the ISCA is also evaluated in the interference channel and with real solar energy harvesting data.**

***Index Terms*— Energy harvesting, Gaussian interference channel, circuitry power consumption, step functions, nonconvex optimization, sum-rate maximization.**

## I. Introduction

ENERGY Harvesting Wireless Nodes (EHWNs) are battery operated devices that exploit current energy harvesting technologies, e.g., a solar panel or a piezoelectric generator, to recharge their batteries. Since the output powers provided by energy harvesters are generally low, EHWNs are critically affected by their energy availability and must adopt energy saving policies. For example, reducing the transmission range by implementing multiple hops or switching "off" the device when the transmission conditions are not adequate. Accordingly, Energy Harvesting Wireless Nodes (EHWNs) generally operate at low transmission powers that might be comparable to other sinks of power consumption at the transmitter, which include, among others, the consumption of the radio frequency circuits being on and the startup power consumption of the transmitter [1].

In this context, efficient transmission strategies must be designed taking into account the limited energy availability of EHWNs. Transmission strategies can be classified into two well defined categories, namely, offline and online. The *offline* strategies assume that the transmitter has full knowledge of the energy harvesting process and channel state, which is a realistic assumption when the channel is static and the energy source is controllable or predictable (e.g., in wireless power transmission scenarios or with solar panels). Contrarily, *online* transmission strategies consider only causal knowledge of these processes at the transmitter. Although the offline assumption is, in some cases, idealistic, it has been broadly used in the literature (e.g., [2]–[5], and references therein) because it can be used as a benchmark for the later design and evaluation of online transmission strategies.

Traditional power allocation polices, e.g., the famous classical waterfilling [6] for non-harvesting devices and its recent (offline) generalization for EHWNs, named directional waterfilling [2], assume that the transmission power is the unique energy sink at the transmitter. This is a reasonable assumption when the transmission range is large (because the radiated power dominates over other energy sinks), but it no longer holds when the transmission range is short, e.g., as occurs in energy efficient network topologies. In this context, several authors have recently considered a more realistic power consumption model [1], which accounts for the cost of having the transceiver "on", $\alpha_t$. Accordingly, the total consumed power is $p + \alpha_t \mathcal{H}(p)$, where $p$ denotes the radiated power and $\mathcal{H}(x)$ is the *unit step function* defined for $x \in \mathbb{R}_+$ as[1]

$$\mathcal{H}(x) = \begin{cases} 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases} \tag{1}$$

By using this power consumption model, the works

Manuscript received August 6, 2015; revised December 17, 2015 and March 16, 2016; accepted April 5, 2016. Date of publication April 20, 2016; date of current version August 10, 2016. This work was supported in part by the ERA-NET CHIST-ERA Program within the E-CROPs Project through the Spanish Government under Grant (PCIN-2013-027), in part by the Agència de Gestió d'Ajuts Universitaris i de Recerca through the Catalan Government under Grant 2014 SGR 1551 and Grant 2014 SGR 1567, in part by the Consejo Superior de Investigaciones Científicas through the AEThER Project under Grant TEC2014-58341-C4-4-R and through the INTENSYV Project under Grant TEC2013-44591-P, and in part by the Research Grants Council, University Grants Committee under Grant 16207814. The associate editor coordinating the review of this paper and approving it for publication was M. R. Nakhai.

M. Gregori and M. Payaró are with the Centre Tecnològic de Telecomunicacions de Catalunya, 08860, Castelldefels, Barcelona, Spain (e-mails: maria.gregori@cttc.es; mpayaro@cttc.es).

D. P. Palomar is with The Hong Kong University of Science and Technology, Hong Kong (e-mail: palomar@ust.hk).

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

Digital Object Identifier 10.1109/TWC.2016.2556658

[1] **Notation:** $\mathbb{R}_+$ denotes the set of nonnegative real numbers. Vectors and vector valued functions are denoted by lower case boldface letters, i.e., $\mathbf{v}$ and $\boldsymbol{\phi}(\mathbf{v})$, respectively. $(\mathbf{v}_u)_{u=1}^U$ defines a column vector obtained by stacking the column vectors $\mathbf{v}_1, \ldots, \mathbf{v}_U$ and $[\mathbf{v}]_k$ returns the $k$-th element of the vector $\mathbf{v}$. Symbol $\preceq$ ($\succeq$) denotes the component-wise "smaller (greater) than or equal to" inequality. Finally, $[x]_a^b \triangleq \max\{a, \min\{x, b\}\}$.

in [4] and [7]–[14] derived transmission strategies in different point to point scenarios. The power allocation that maximizes the mutual information for a non-harvesting node was derived in [7] and [8]. Optimal policies for an EHWN operating in a static and fading continuous time channel were derived in [9] and [4], respectively. In contrast, [12]–[14] studied the optimal offline resource allocation for an EHWN operating in a discrete time channel and showed that the step functions can be replaced by additional optimization variables, namely, *indicator variables*, which must belong to the set $\{0, 1\}$. Integer relaxation was then used to obtain an upper bound on the achievable mutual information and a feasible solution that tends to the optimal when the number of streams and channel accesses grow without bound [14].

To the best of our knowledge, few works have considered the interference channel for EHWNs. The two-user Gaussian interference channel was studied in [11], [15], and [16] from an information theory perspective. The transmission scheme that maximizes the sum-rate was derived in [15] by using the time-sharing argument, while [16] explored energy cooperation among transmitters to enlarge the rate region. Among these works, only [11] considered the energy consumed in the circuitry when the transmitter is "on". In contrast to the aforementioned works, [17] considered an arbitrary number of transmitter-receiver pairs with Gaussian distributed signals and derived the Nash equilibrium of the game that is obtained when each user aims at maximizing its own rate.

Similarly, in this paper, we consider $T$ energy harvesting transmitters that simultaneously transmit Gaussian distributed signals to their respective receivers. In contrast to [17], where the objective of each transmitter-receiver pair is to maximize its own rate, the objective of this paper is to devise a distributed power allocation algorithm that aims at maximizing the network sum-rate. Additionally, we consider a general power consumption model that can carefully account for the different sources of power consumption in each transmitter, such as radio frequency circuits being on/off and the startup power consumption associated to off-on transitions. Mathematically, this power consumption model contains additions and products of step functions and, to the best of our knowledge, has not been considered before in the literature. In this context, the major contributions of this paper are:

- Formulating the sum-rate maximization problem in the interference channel by considering (i) energy harvesting capabilities at the transmitter nodes, and (ii) a generalized power consumption model that accepts products of step functions. This generic formulation allows for the first time to model certain energy sinks (e.g., the startup power consumption). Hence, by addressing this problem, the contributions of this paper are two-fold. First, it addresses an open problem for energy harvesting devices in the interference channel. Second, it expands the existing algorithms in the literature by refining the energy consumption model at the transmitters; for example, the works [4], [7], [8], and [12]–[14] in point to point links (both for harvesting and non-harvesting nodes) as well as [18] in the Gaussian interference channel for non-harvesting transmitters.
- Proposing a distributed algorithm, namely, the Iterative Smooth and Convex approximation Algorithm (ISCA), that addresses the nonsmooth and nonconvex sum-rate maximization problem. The ISCA successively approximates the step functions by smooth functions in order to derive a smooth nonconvex optimization problem that can be solved by the Successive Convex Approximation (SCA) method [19]. The main advantages of the proposed approach with respect to the use of indicator variables with integer relaxation are: (i) it applies to nonconvex objective functions; (ii) it accepts products of step functions (note that if the step functions were replaced by additional indicator variables as in [14], the resulting relaxed problem would still be nonconvex due to products of optimization variables); and (iii) the problem can be solved in a distributed way under very mild assumptions. We believe that the ISCA may shed light on solving other problems that contain step functions.
- Proving that the ISCA converges to a stationary solution of the original nonsmooth nonconvex sum-rate maximization problem.
- Evaluating the performance of the ISCA numerically for two different scenarios. First, we consider a point to point link, and show that the stationary point obtained with the ISCA achieves almost the same sum-rate than the asymptotically optimal solution obtained in [14]. Second, we consider multiple transmitters simultaneously operating in an interference channel, and evaluate the performance of the ISCA when solar energy is harvested.

The remainder of the paper is structured as follows. In Section II, the sum-rate maximization problem for a network of EHWNs is formulated. In Section III, the smooth and convex approximations of the step functions are given. In Section IV, the ISCA is presented for a general power consumption model, which is particularized in Section V to the power consumption model used in [14]. The performance of the ISCA in terms of achieved rate and computational complexity is numerically evaluated in Section VI. Finally, the paper is concluded in Section VII.

## II. System Model and Problem Formulation

We consider a Gaussian interference channel composed of $T$ transmitter and receiver pairs sharing the same band over single-input single-output frequency-selective links composed of $K$ parallel subcarriers. Transmission takes place during $N$ time slots of duration $T_s$, where, at the $n$-th slot, the channel power gain from transmitter $t$, $t = 1, \ldots, T$, to receiver $r$, $r = 1, \ldots, T$, at the $k$-th subcarrier is denoted by $h_{tr}(k, n)$. We do not consider interference cancellation techniques in order to avoid the need of having a centralized control or coordination in the network and, accordingly, we treat the multiuser interference as additive noise. Thus, assuming that Gaussian signaling is used, the rate of user $t$ depends on its radiated power, $\mathbf{p}_t \in \mathbb{R}_+^{KN}$, $\mathbf{p}_t = (\mathbf{p}_{tn})_{n=1}^N, \mathbf{p}_{tn} = (p_t(k, n))_{k=1}^K$ and on the transmission power of all the other transmitters,

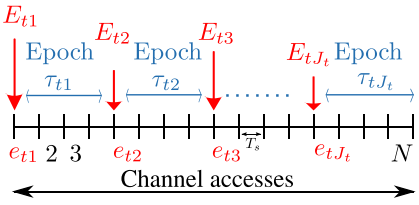



Fig. 1. Representation of the energy harvesting process at the $t$-th transmitter.

i.e., $\mathbf{p}_{-t} = (\mathbf{p}_{t'})_{t'\neq t=1}^{T}$, as

$$r_t(\mathbf{p}_t, \mathbf{p}_{-t}) = \sum_{n=1}^{N}\sum_{k=1}^{K} \log\left(1 + \frac{p_t(k,n)h_{tt}(k,n)}{\sigma_t^2(k) + \sum\limits_{t'\neq t} p_{t'}(k,n)h_{t't}(k,n)}\right), \quad (2)$$

where $\sigma_t^2(k)$ denotes the noise power at the $t$-th receiver and $k$-th subcarrier.[2]

We consider that transmitters can harvest energy from the environment to recharge their batteries. As it is commonly done in the literature [2], we characterize the energy harvesting process with a packetized model. Thus, a packet of energy containing $E_{tj}$ Joules is harvested at the beginning of the $e_{tj}$-th channel access,[3] $j = 1, \ldots, J_t$, with $J_t$ denoting the total number of harvested energy packets at transmitter $t$. The initial battery level is modeled as the first harvested packet $E_{t1}$ at $e_{t1} = 1$ and the battery capacity is assumed to be infinite. We use the term *epoch* $\tau_{tj}$ to denote the set of channel accesses between two consecutive energy arrivals, i.e., $\tau_{tj} = \{e_{tj}, e_{tj}+1, \ldots, e_{t(j+1)}-1\}$ with $e_{t(J_t+1)} = N+1$ so that the last epoch is well defined. A temporal representation is given in Fig. 1. Note that this packetized model can capture any continuous energy harvesting profile by considering that all the energy harvested within a certain slot is aggregated in an energy packet that arrives at the beginning of the subsequent slot.

The transmission strategy $\mathbf{p}_t$ must satisfy the *energy causality constraints*, which impose that the battery level must be nonnegative or, equivalently, that the energy cumulatively expended by the end of the $\ell$-th epoch, $\ell = 1, \ldots, J_t$, is not greater than the energy cumulatively harvested, i.e.,

$$[\mathbf{B}_t(\mathbf{p}_t)]_\ell \triangleq \underbrace{\sum_{j=1}^{\ell} E_{tj}}_{\text{Harvested energy}} - \underbrace{T_s \sum_{j=1}^{\ell}\sum_{n\in\tau_{tj}} \mathcal{C}_{tn}(\mathbf{p}_t)}_{\text{Expended energy}} \geq 0, \quad \forall \ell, t, \quad (3)$$

where $\mathcal{C}_{tn}(\mathbf{p}_t)$ denotes the power consumption model at transmitter $t$ and slot $n$.

Since EHWNs operate at low energy levels, the power consumption model must account not only for the transmission radiated power but also for the other energy sinks. In [14], the power consumption of transmitter $t$ at slot $n$ was modeled as

$$\mathcal{C}_{tn}^1(\mathbf{p}_{tn}) = \underbrace{\left(\sum_{k=1}^{K} p_t(k,n)\right)}_{\text{Transmission power}} + \underbrace{\alpha_t \mathcal{H}\left(\sum_{k=1}^{K} p_t(k,n)\right)}_{\text{Power consumption per active slot}} + \sum_{k=1}^{K} \underbrace{\beta_t \mathcal{H}(p_t(k,n))}_{\text{Power consumption per active stream}}, \quad (4)$$

were the constant $\alpha_t \geq 0$ models the energy consumption associated to the different components of the radio frequency chain when the transceiver is "on"; and $\beta_t \geq 0$ accounts for the additional cost of activating a certain subcarrier. As a result of considering these energy sinks, the optimal transmission strategy alternates between "off" and "on" cycles; however, due to the startup time of the transceiver, "off-on" transitions also entail energy consumption [20] that can be accounted by refining the previous model as

$$\mathcal{C}_{tn}^2(\mathbf{p}_{t(n-1)}, \mathbf{p}_{tn}) = \mathcal{C}_{tn}^1(\mathbf{p}_{tn}) + \underbrace{\gamma_t\left(1 - \mathcal{H}\left(\sum_{k=1}^{K} p_t(k,n-1)\right)\right)\mathcal{H}\left(\sum_{k=1}^{K} p_t(k,n)\right)}_{\text{Startup power consumption}}, \quad (5)$$

where $\gamma_t \geq 0$ denotes the startup power consumption at transmitter $t$; $(1 - \mathcal{H}(\sum_{k=1}^{K} p_t(k,n-1)))$ is one when the channel access $n-1$ is "off", $\mathcal{H}(\sum_{k=1}^{K} p_t(k,n))$ is one when the $n$-th channel access is "on", and their product takes value one when an "off-on" transition occurs.

In this work we consider a general power consumption model of the form:

$$\mathcal{C}_{tn}(\mathbf{p}_t) = \underbrace{\left(\sum_{k=1}^{K} p_t(k,n)\right)}_{\text{Transmission power}} + \underbrace{\sum_{s=1}^{S_t} w_{ts} \prod_{q=1}^{Q_{ts}} \mathcal{H}(\phi_{tsq}(\mathbf{p}_t))}_{\text{Remaining power sinks}}, \quad (6)$$

where $w_{ts} \neq 0$ is a given weight; $S_t \in \mathbb{N}$ is the number of summands containing step functions; $Q_{ts} \in \mathbb{N}$ stands for the number of factors of the $s$-th summand; and $\phi_{tsq} : \mathbb{R}_+^{KN} \to \mathbb{R}_+$ must be concave, Lipschitz continuous, and continuously differentiable. Note that $\mathcal{C}_{tn}^1$ and $\mathcal{C}_{tn}^2$ are particular cases of $\mathcal{C}_{tn}$ in (6); the associated weights and inner functions, $\phi_{tsq}(\mathbf{p}_t)$, are given in Table I, where $\mathcal{C}_{tn}^1$ has $S_t = K+1$ summands, and $\mathcal{C}_{tn}^2$ has $S_t = K+3$. Observe that a different power consumption model can be used for each of the network nodes, which can be specified when the network

TABLE I
WEIGHTS AND INNER FUNCTIONS OF $\mathcal{C}_{tn}^1$ AND $\mathcal{C}_{tn}^2$

| $s$ | $1 \ldots K$ | $K+1$ | $K+2$ | $K+3$ |
|---|---|---|---|---|
| $w_{ts}$ | $\beta_t$ | $\alpha_t$ | $\gamma_t$ | $-\gamma_t$ |
| $\phi_{ts1}$ | $p_t(s,n)$ | $\sum_{k=1}^{K} p_t(k,n)$ | $\sum_{k=1}^{K} p_t(k,n)$ | $\sum_{k=1}^{K} p_t(k,n-1)$ |
| $\phi_{ts2}$ | - | - | - | $\sum_{k=1}^{K} p_t(k,n)$ |
| | $\mathcal{C}_{tn}^1$ | | $\mathcal{C}_{tn}^2$ | |

[2]We consider that each channel varies sufficiently slowly, so that the information theoretical results are meaningful.

[3]Since the transmission strategy can only be changed in a channel use basis, we consider that, independently of the energy packet arrival instant, it becomes available for the transmitter at the beginning of the next channel use.

is being deployed and one knows the different energy sinks of each node.

Our objective is to design a distributed offline power allocation strategy that maximizes the sum-rate of a network of EHWNs transmitting simultaneously in the Gaussian interference channel. As mentioned in the introduction, offline strategies assume non-causal knowledge of the harvested energy at the transmitters, which is realistic when the energy source is predictable. Note that if the offline power allocation can be computed distributedly by the network nodes, then it can be updated when a substantial change in the energy harvesting prediction is observed. Additionally, for scenarios where energy cannot be predicted, the offline solution can be used as a benchmark to evaluate online policies. The offline sum-rate maximization problem is

$$(\hat{P}): \max_{\mathbf{p}} \sum_{t=1}^{T} r_t(\mathbf{p}_t, \mathbf{p}_{-t}) \tag{7a}$$

$$\text{s.t.} \quad \left.\begin{array}{ll} \mathbf{B}_t(\mathbf{p}_t) \succeq \mathbf{0}_{J_t}, & \forall t = 1, \ldots, T \\ \mathbf{p}_t \in \mathcal{P}_t, & \forall t = 1, \ldots, T \end{array}\right\} \triangleq \hat{\mathcal{P}} \tag{7b}$$

$$\tag{7c}$$

where $\mathbf{p} \triangleq (\mathbf{p}_t)_{t=1}^{T}$, and $\mathcal{P}_t = \{\mathbf{p}_t : \mathbf{p}_t \succeq \mathbf{0}_{KN}, \mathbf{p}_t \preceq \mathbf{p}_t^{\max}\}$ with $\mathbf{p}_t^{\max} \triangleq ((p_t^{\max}(k,n))_{k=1}^{K})_{n=1}^{N}$ limits the maximum transmit power. The classical power constraint does not appear directly in the formulation (as it happens in non-harvesting systems with a fixed energy budget) but in the so-called energy causality constraints (7b) since energy is created and consumed over time.

The problem in (7) has the following major difficulties: first, it is nonsmooth, nonconvex, and NP-hard, which was shown in [21] for a simpler scenario; and second, it is key that the solution can be computed distributedly by the network nodes to adjust their strategies when strong variations of the energy harvesting profile are observed. In this context, we propose the ISCA that is able to distributedly compute a stationary solution of the sum-rate maximization problem in (7). The details of the ISCA are presented in Section IV. By now, it is important to know that the ISCA is composed of two loops: the outer loop performs a Successive Smooth Approximation (SSA) of the step functions so that, at each iteration, a smooth nonconvex problem that approximates (7) is derived; then, the inner loop solves this smooth nonconvex problem by means of the SCA algorithm proposed in [19].

## III. Approximations of the Step Function

The objective of this section is twofold: (i) to design a smooth approximation of the unit step function in (1), which is used in the outer loop of the ISCA; and (ii) to derive a convex approximation of the smooth approximation in (i) that can be handled by the SCA algorithm in the inner loop. The later approximation, (ii), has to satisfy some tight technical requirements in order to guarantee convergence, as listed in [19, Assumption 3], that intrinsically couple the design of the approximations in (i) and (ii) because depending on the chosen smooth approximation, it might be either easy or extremely difficult to later find an accurate convex approximation. In this context, we can easily derive a convex approximation if the smooth approximation is: (C1) differentiable and (C2) decomposable as the summation of concave and convex functions.

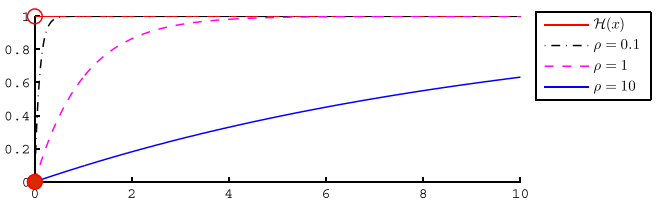



Fig. 2. Representation of $\mathcal{H}_\rho(x)$ in (8) for different values of the approximation control parameter $\rho$.

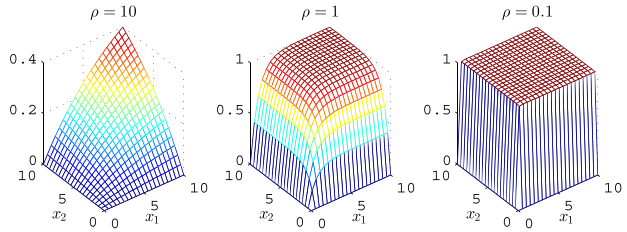



Fig. 3. Representation of the smooth approximation of $\mathcal{H}(x_1)\mathcal{H}(x_2)$, i.e., $\mathcal{H}_\rho(x_1, x_2)$ in (9) for different values of the approximation control parameter $\rho$.

### A. Smooth Approximation of the Step Function

In this section, we present a smooth approximation of the step function that satisfies (C1)-(C2).

*1) Single Step Function:* We approximate $\mathcal{H}(x)$ in (1) with the function $\mathcal{H}_\rho : \mathbb{R}_+ \to [0, 1]$, defined as

$$\mathcal{H}_\rho(x) = 1 - e^{-\frac{x}{\rho}}, \tag{8}$$

where $\rho > 0$ is a parameter that controls how good the approximation is (the smaller the value of $\rho$ the better the approximation) as illustrated in Fig. 2. Additionally, it can be easily shown that $\lim_{\rho \to 0} \mathcal{H}_\rho(x) = \mathcal{H}(x)$.

*2) Product of Step Functions:* In practice, it is also possible to encounter products of step functions as happens with the startup power consumption in (5). For illustrative reasons, we first consider a single product of step functions, i.e., $\mathcal{H}(x_1)\mathcal{H}(x_2)$, and later, in Lemma 1, we present a smooth approximation of higher order products. We approximate the product of step functions $\mathcal{H}(x_1)\mathcal{H}(x_2)$ with the function $\mathcal{H}_\rho : \mathbb{R}_+^2 \to [0, 1]$, defined as

$$\mathcal{H}_\rho(x_1, x_2) = \mathcal{H}_\rho(x_1)\mathcal{H}_\rho(x_2) = \underbrace{1 + e^{-\frac{x_1+x_2}{\rho}}}_{Convex} \underbrace{-e^{-\frac{x_1}{\rho}} - e^{-\frac{x_2}{\rho}}}_{Concave}. \tag{9}$$

This approximation is depicted in Fig. 3, where it is observed that the approximation improves when the control parameter $\rho$ is reduced. This is clearly observed for $\rho = 0.1$, where the approximation of the product of step functions takes value 0 at the axes and close to 1 elsewhere.[4]

[4]Note that, for compactness in the notation, we use $\mathcal{H}_\rho$ to denote both the smooth approximation of the single step and the product of step functions. Throughout the paper, we distinguish between them by the dimension of the argument.

*Lemma 1: Given a set of variables $x_q \in \mathbb{R}_+$, $q = 1, \ldots, Q$, and $\mathbf{x} = [x_1, \ldots, x_Q]^\mathsf{T}$, then the product of $Q$ step functions, $\prod_{q=1}^{Q} \mathcal{H}(x_q)$, can be approximated by the differentiable function $\mathcal{H}_\rho : \mathbb{R}_+^Q \to [0, 1]$, defined as*

$$\mathcal{H}_\rho(\mathbf{x}) = \underbrace{1 + \sum_{i \in \mathcal{E}} \sum_{0 < j_1 \cdots < j_i \leq Q} e^{-\frac{\sum_{k=1}^{i} x_{j_k}}{\rho}}}_{Convex} \underbrace{- \sum_{i \in \mathcal{O}} \sum_{0 < j_1 \cdots < j_i \leq Q} e^{-\frac{\sum_{k=1}^{i} x_{j_k}}{\rho}}}_{Concave}, \quad (10)$$

*where $\mathcal{E}$ and $\mathcal{O}$ are a partition of the set $\{1, \ldots, Q\}$ that take the even and odd elements, respectively; and $\rho > 0$ is the parameter that controls the approximation. Additionally, $\lim_{\rho \to 0} \mathcal{H}_\rho(\mathbf{x}) = \prod_{q=1}^{Q} \mathcal{H}(x_q)$.*

*Proof:* The function $\mathcal{H}_\rho(\mathbf{x})$ in (10) is obtained simply by expanding the product $\prod_{q=1}^{Q} \mathcal{H}_\rho(x_q)$. Since $e^{-x/\rho}$ is differentiable so it is $\mathcal{H}_\rho(\mathbf{x})$. The concavity and convexity of the different terms follows by noting that $e^{-x/\rho}$ is a convex function. Finally, $\lim_{\rho \to 0} \mathcal{H}_\rho(\mathbf{x}) = \prod_{q=1}^{Q} \lim_{\rho \to 0} \mathcal{H}_\rho(x_q) = \prod_{q=1}^{Q} \mathcal{H}(x_q)$. ∎

Note that in the inner summations the values of $j_i$ take all the possible combinations of $i$ elements from the set $\{1, \ldots, Q\}$; accordingly, each of these sums contains $\binom{Q}{i}$ terms.

### B. Convex Approximation of the Smooth Step Function

In this section, we derive a convex approximation of the smooth step function, $\mathcal{H}_\rho(\boldsymbol{\phi}(\mathbf{x}))$, around the point $\mathbf{x}_0$ for a given transformation $\boldsymbol{\phi}$ (whose component functions are all concave), which is denoted as $\breve{\mathcal{H}}_\rho(\mathbf{x}; \mathbf{x}_0, \boldsymbol{\phi})$.

*1) Convex Approximation of the Single Step Function:* We first consider the convex approximation of the single step function in (8), $\mathcal{H}_\rho(\phi(\mathbf{x}))$, which, from the rules of function composition, is a concave function [22]. Thanks to this concavity, it is easy to show that its linearization at the point $\mathbf{x}_0$, i.e., $\breve{\mathcal{H}}_\rho(\mathbf{x}; \mathbf{x}_0, \phi) = 1 + \xi_\rho(\mathbf{x}; \mathbf{x}_0, \phi)$, is a convex function that satisfies the requirements in [19, Assumption 3], which for completeness are given in Appendix A. We have defined $\xi_\rho(\mathbf{x}; \mathbf{x}_0, \phi)$ as the linearization of the concave term, $-e^{\frac{-\phi(\mathbf{x})}{\rho}}$, around the point $\mathbf{x}_0$, i.e.,

$$\xi_\rho(\mathbf{x}; \mathbf{x}_0, \phi) \triangleq \frac{\nabla_{\mathbf{x}} \phi(\mathbf{x}_0)}{\rho} e^{-\frac{\phi(\mathbf{x}_0)}{\rho}} (\mathbf{x} - \mathbf{x}_0) - e^{-\frac{\phi(\mathbf{x}_0)}{\rho}}. \quad (11)$$

*2) Convex Approximation of Products of Step Functions:* Similarly, by linearizing the concave terms of the smooth product of step functions, we can obtain a convex approximation of $\mathcal{H}_\rho(\boldsymbol{\phi}(\mathbf{x}))$ that satisfies the requirements in [19, Assumption 3].

*Lemma 2: Let $\boldsymbol{\phi} \triangleq [\phi_1, \ldots, \phi_Q]^\mathsf{T}$ with $\phi_q$ being concave, Lipschitz continuous, and continuously differentiable, then the function*

$$\breve{\mathcal{H}}_\rho(\mathbf{x}; \mathbf{x}_0, \boldsymbol{\phi}) = \underbrace{1 + \sum_{i \in \mathcal{E}} \sum_{0 < j_1 \cdots < j_i \leq Q} e^{-\frac{\sum_{k=1}^{i} \phi_{j_k}(\mathbf{x})}{\rho}}}_{Convex} + \underbrace{\sum_{i \in \mathcal{O}} \sum_{0 < j_1 \cdots < j_i \leq Q} \xi_\rho\left(\mathbf{x}; \mathbf{x}_0, \sum_{k=1}^{i} \phi_{j_k}\right)}_{Linear} \quad (12)$$

*is a convex approximation of $\mathcal{H}_\rho(\boldsymbol{\phi}(\mathbf{x}))$ around the point $\mathbf{x}_0$ that satisfies the required conditions in [19, Assumption 3], where $\xi_\rho$ is given in* (11).

*Proof:* See Appendix A. ∎

*Remark 1:* Using smooth approximations of step functions to deal with its discontinuity is not a new concept in the literature; it has been used before to approximate the cardinality operator (or $\ell^0$ norm), e.g., see [23] and [24]. However, to the best of our knowledge this is the first paper that derives an approximation for the product of step functions that takes into account the SCA requirements in [19, Assumption 3]. Note that the smooth approximation is not unique, for instance a logarithmic approximation could be used (e.g, $\mathcal{H}_\rho(x) = \mathsf{log}(1 + x/\rho)/\mathsf{log}(1 + x_{MAX}/\rho)$ with $x_{MAX}$ denoting the maximum possible value of the argument $x$). The key properties of the proposed approximation are: (i) it easily generalizes to products of step functions (products of logarithms are no longer concave and the approximation would be difficult to convexify); and (ii) the proposed smooth approximation is more accurate than the logarithmic approximation under the same value of $\rho$ (the logarithmic approximation above needs to reduce much more parameter $\rho$ to be accurate, i.e., $\rho \ll 10^{-20}$, which leads to numerical problems due to the finite precision of the solvers).

## IV. The Iterative Smooth and Convex Approximation Algorithm

In this section, we propose the ISCA that is composed of two loops as shown in Fig. 4. The outer loop indexed by $\varsigma$, performs a SSA of (7), deriving, at each iteration, a nonconvex smooth problem $(\tilde{P}^\varsigma)$, obtained by using the approximation of the step functions in Lemma 1 given the approximation control parameter $\rho^\varsigma$. Initially, we set $\rho^1 \gg 0$ so that the resulting problem is smooth (e.g., $\rho^1 = 5$), reducing the impact of the ISCA initial point, $\mathbf{p}^1$, over the final stationary solution to (7), $\hat{\mathbf{p}}$. Then, the inner loop, indexed by $\nu$, uses the SCA in [19] to determine a stationary solution, $\tilde{\mathbf{p}}^\varsigma$, of $(\tilde{P}^\varsigma)$. Next, a termination condition is checked: if it is satisfied, which implies that the approximation of the step functions is tight enough, the ISCA concludes that a good solution of the original problem (7) is $\hat{\mathbf{p}} = \tilde{\mathbf{p}}^\varsigma$; otherwise, a new outer iteration starts by reducing the approximation control parameter, which improves the approximation of the step functions. At each iteration, we use a warm start, i.e., a feasible initial point for the inner loop is obtained that resembles the stationary solution of the previous outer iteration. In the following sections, we present more details on the ISCA.

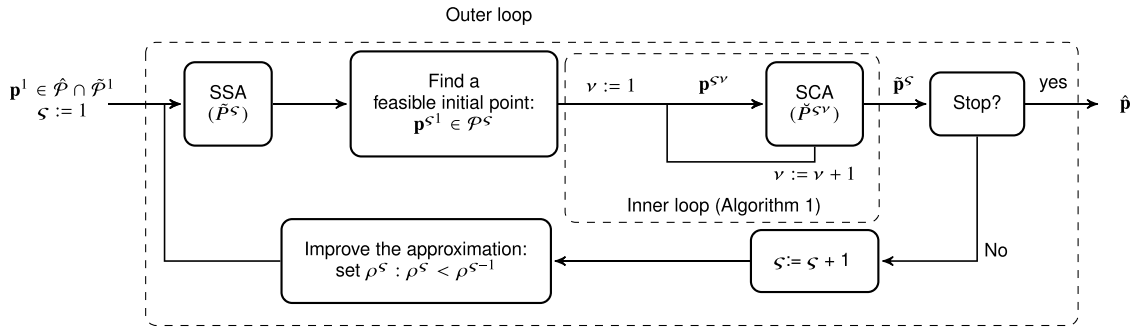



Fig. 4. Block diagram of the ISCA.

### A. The Outer Loop: Successive Smooth Approximation of the Step Functions

In this section, we formulate the smooth problem, $(\tilde{P}^\varsigma)$, at the $\varsigma$-th outer loop iteration. This problem, in spite of being nonconvex, has an inherently high degree of convexity, which is exploited by the inner loop:

$$(\tilde{P}^\varsigma): \max_{\mathbf{p}} \sum_{t=1}^{T} r_t(\mathbf{p}_t, \mathbf{p}_{-t}) \tag{13a}$$

$$\text{s.t.} \left.\begin{array}{ll} \mathbf{B}_t^\varsigma(\mathbf{p}_t) \succeq \mathbf{0}_{J_t}, & \forall t = 1, \ldots, T \\ \mathbf{p}_t \in \mathcal{P}_t, & \forall t = 1, \ldots, T \end{array}\right\} \triangleq \tilde{\mathcal{P}}^\varsigma \tag{13b, 13c}$$

where

$$[\mathbf{B}_t^\varsigma(\mathbf{p}_t)]_\ell \triangleq \sum_{j=1}^{\ell} \left[ E_{tj} - T_s \sum_{n \in \tau_{tj}} \left( \left( \sum_{k=1}^{K} p_t(k,n) \right) + \sum_{s=1}^{S_t} w_{ts} \mathcal{H}_{\rho^\varsigma}\left(\boldsymbol{\phi}_{ts}(\mathbf{p}_t)\right) \right) \right], \quad \ell = 1, \ldots, J_t;$$

$\mathcal{H}_\rho$ is given in (10); and $\boldsymbol{\phi}_{ts}(\mathbf{p}_t)$ is a vector function defined as $\boldsymbol{\phi}_{ts}(\mathbf{p}_t) = [\phi_{ts1}(\mathbf{p}_t), \ldots, \phi_{tsQ_{ts}}(\mathbf{p}_t)]^{\mathsf{T}}$.

### B. The Inner Loop: Nonconvex Optimization of Smooth Problems With SCA

Among the algorithms that converge to stationary solutions of smooth nonconvex problems (e.g., gradient-based descend schemes [25], SCA algorithms [18], [19], [26], feasible sequential quadratic programming [27], parallel variable distribution [28], etc.), we have selected the algorithm in [19] for the inner loop because it has the following main advantages: (i) it accepts nonconvex constraints; (ii) it exploits any degree of convexity present in the problem, which results in a much faster convergence; (iii) it can be solved in a distributed way under very mild assumptions; and (iv) it includes as special cases SCA-based algorithms, such as (proximal) gradient or Newton type method, block coordinate (parallel) descent schemes and difference of convex functions methods.

The algorithm proposed in [19] is based on SCA and consists on solving a sequence of strongly convex inner approximations of the nonconvex smooth problem. Under some structural assumptions, the algorithm converges to a stationary solution. These assumptions enforce a specific structure of: (i) the original nonconvex smooth problem [19, Assumption 1]; (ii) the convex approximation of the objective function [19, Assumption 2]; and (iii) the convex approximation of the constraints [19, Assumption 3]. In order to use the algorithm in [19] in our inner loop, we need to satisfy these structural requirements.

It can be easily shown that the smooth problem in (13) satisfies the structural requirements in [19, Assumption 1]. Since the objective function is nonconvex, we need to derive a proper convex approximation. To do so, we exploit the "partial" concavity of the rate of a certain user, $r_t(\mathbf{p}_t, \mathbf{p}_{-t})$, with respect to its own transmission power $\mathbf{p}_t$. Hence, we approximate the objective function in (13) around the current iterate $\mathbf{p}^{\varsigma\nu} = ((p_t^{\varsigma\nu}(k,n))_{k=1}^K)_{n=1}^N$ as $\sum_{t=1}^T \breve{r}_t(\mathbf{p}_t; \mathbf{p}^{\varsigma\nu})$, where $\breve{r}_t(\mathbf{p}_t; \mathbf{p}^{\varsigma\nu}) = r_t(\mathbf{p}_t, \mathbf{p}_{-t}^{\varsigma\nu}) + \boldsymbol{\pi}_t^{\varsigma\nu\mathsf{T}}(\mathbf{p}_t - \mathbf{p}_t^{\varsigma\nu}) - \frac{b_t}{2}||\mathbf{p}_t - \mathbf{p}_t^{\varsigma\nu}||^2$. The term $\boldsymbol{\pi}_t^{\varsigma\nu}$ linearizes the rate functions of the users $t' \neq t$ with respect to $\mathbf{p}_t$, i.e., $\boldsymbol{\pi}_t^{\varsigma\nu} \triangleq ((\pi_{tkn}^{\varsigma\nu})_{k=1}^K)_{n=1}^N = \sum_{t' \neq t} \nabla_{\mathbf{p}_t} r_{t'}(\mathbf{p}_{t'}, \mathbf{p}_{-t'})\big|_{\mathbf{p}^{\varsigma\nu}}$ with

$$\pi_{tkn}^{\varsigma\nu} = \sum_{t' \neq t} \frac{-SNR_{t'}^{\varsigma\nu}(k,n) h_{tt'}(k,n)}{MUI_{t'}^{\varsigma\nu}(k,n)(1 + SNR_{t'}^{\varsigma\nu}(k,n))}; \tag{14}$$

$SNR_t^{\varsigma\nu}(k,n) \triangleq \frac{h_{tt}(k,n) p_t^{\varsigma\nu}(k,n)}{MUI_t^{\varsigma\nu}(k,n)}$ and $MUI_t^{\varsigma\nu}(k,n) \triangleq \sigma_t^2(k) + \sum_{t' \neq t} p_{t'}^{\varsigma\nu}(k,n) h_{t't}(k,n)$ are the signal to interference plus noise ratio and the multiuser interference-plus-noise power experienced by user $t$ given the power profile $\mathbf{p}^{\varsigma\nu}$. The term $\frac{b_t}{2}||\mathbf{p}_t - \mathbf{p}_t^{\varsigma\nu}||^2$ with $b_t \geq 0$ is a proximal regularization term that relaxes the convergence conditions of the inner loop algorithm and enhances the convergence speed [18].

Accordingly, the strongly convex problem that has to be solved in the $\nu$-th inner loop iteration, which approximates the smooth problem $(\tilde{P}^\varsigma)$ around the current iterate, $\mathbf{p}^{\varsigma\nu}$, is

$$(\breve{P}^{\varsigma\nu}): \max_{\mathbf{p}} \sum_{t=1}^{T} \breve{r}_t(\mathbf{p}_t; \mathbf{p}^{\varsigma\nu}) \tag{15a}$$

$$\text{s.t.} \left.\begin{array}{ll} \breve{\mathbf{B}}_t^{\varsigma\nu}(\mathbf{p}_t; \mathbf{p}_t^{\varsigma\nu}) \succeq \mathbf{0}_{J_t}, & \forall t \\ \mathbf{p}_t \in \mathcal{P}_t, & \forall t \end{array}\right\} \triangleq \breve{\mathcal{P}}^{\varsigma\nu} \tag{15b, 15c}$$

**Algorithm 1** The Inner Loop: SCA of $(\tilde{P}^{\varsigma})$

**Input:** $\mathbf{p}^{\varsigma 1} \in \tilde{\mathcal{P}}^{\varsigma}$, $a^{\varsigma \nu} > 0$.

**Initialization**: Set $\nu := 1$.

**Step 1**: If a termination condition is satisfied: STOP.

**Step 2**: For every user $t \in [1, T]$, find $\breve{\mathbf{p}}_t^{\varsigma \nu}$ that is the unique optimal solution of the strongly convex problem $(\breve{P}_t^{\varsigma \nu})$.

**Step 3**: Update the iterate: $\mathbf{p}_t^{\varsigma (\nu+1)} = \mathbf{p}_t^{\varsigma \nu} + a^{\varsigma \nu}(\breve{\mathbf{p}}_t^{\varsigma \nu} - \mathbf{p}_t^{\varsigma \nu}), \forall t$.

**Step 4**: $\nu := \nu + 1$ and go to Step 1;

with

$$[\breve{\mathbf{B}}_t^{\varsigma \nu}(\mathbf{p}_t; \mathbf{p}_t^{\varsigma \nu})]_{\ell} \triangleq \sum_{j=1}^{\ell} \Bigg[ E_{tj} - T_s \sum_{n \in \tau_{tj}} \Bigg( \Big( \sum_{k=1}^{K} p_t(k,n) \Big) + \sum_{s \in \mathcal{S}_t^+} w_{ts} \breve{\mathcal{H}}_{\rho^{\varsigma}}(\mathbf{p}_t; \mathbf{p}_t^{\varsigma \nu}, \boldsymbol{\phi}_{ts}) + \sum_{s \in \mathcal{S}_t^-} w_{ts} \breve{\mathcal{H}}^{-}_{\rho^{\varsigma}}(\mathbf{p}_t; \mathbf{p}_t^{\varsigma \nu}, \boldsymbol{\phi}_{ts}) \Bigg) \Bigg], \tag{16}$$

where $\mathcal{S}_t^+ = \{s \in \{1, \ldots, S_t\} : w_{ts} > 0\}$ and $\mathcal{S}_t^- = \{s \in \{1, \ldots, S_t\} : w_{ts} < 0\}$; $\breve{\mathcal{H}}_{\rho}$ is given in Lemma 2; and $\breve{\mathcal{H}}_{\rho}^{-}$ is defined as $\breve{\mathcal{H}}_{\rho}$ but swapping the odd and even sets. We have defined $\breve{\mathcal{H}}_{\rho}^{-}$ because the negative weights invert the concavity or convexity of the terms of the smooth step function.

Additionally, since the objective function and constraints of the different transmitters are decoupled, the problem decouples into $T$ subproblems one for each transmitter-receiver pair, which leads to a distributed resource allocation strategy that requires very limited feedback as presented later. Accordingly, each transmitter must solve the following problem at each inner loop iteration:

$$(\breve{P}_t^{\varsigma \nu}) : \max_{\mathbf{p}_t} \ \breve{r}_t(\mathbf{p}_t; \mathbf{p}^{\varsigma \nu}) \tag{17a}$$

$$\text{s.t.} \ \left. \begin{array}{c} \breve{\mathbf{B}}_t^{\varsigma \nu}(\mathbf{p}_t; \mathbf{p}_t^{\varsigma \nu}) \succeq \mathbf{0}_{J_t}, \\ \mathbf{p}_t \in \mathcal{P}_t. \end{array} \right\} \triangleq \breve{\mathcal{P}}_t^{\varsigma \nu} \tag{17b, 17c}$$

Since $(\breve{P}_t^{\varsigma \nu})$ is a strongly convex problem, its unique solution, $\breve{\mathbf{p}}_t^{\varsigma \nu}$, can be easily determined by classical convex optimization algorithms, e.g., interior point methods [22]. However, since the solution to $(\breve{P}_t^{\varsigma \nu})$ has to be computed at each inner loop iteration, it is key to derive (if possible) a closed form solution in order to reduce the computational complexity of the ISCA. Section V derives an efficient solution for the power consumption model $\mathcal{C}_{tn}^1$ in (4).

The SCA-based inner loop algorithm is presented in Algorithm 1 [19]. The algorithm uses the unique optimal solution to $(\breve{P}_t^{\varsigma \nu})$, $\breve{\mathbf{p}}_t^{\varsigma \nu}$, to determine the initial point of the following iteration, $\mathbf{p}_t^{\varsigma (\nu+1)}$, which is computed as a convex combination of $\breve{\mathbf{p}}_t^{\varsigma \nu}$ and the previous iterate $\mathbf{p}_t^{\varsigma \nu}$.

*Theorem 1 [19]: Given the smooth nonconvex problem $(\tilde{P}^{\varsigma})$, suppose that one of the two following conditions holds:*

a) *The step size $a^{\varsigma \nu}$ is such that $0 < \inf_{\nu} a^{\varsigma \nu} \leq \sup_{\nu} a^{\varsigma \nu} \leq a^{\max} \leq 1$ and $2c_{\breve{r}} \geq a^{\max} L_{\nabla r}$, where $c_{\breve{r}}$ is the constant of uniform strong convexity of $\sum_{t=1}^{T} \breve{r}_t(\mathbf{p}_t; \mathbf{p}^{\varsigma \nu})$ and $L_{\nabla r}$ is Lipschitz continuity constant of $\sum_{t=1}^{T} \nabla_{\mathbf{p}} r_t(\mathbf{p}_t, \mathbf{p}_{-t})$.*

b) *(i) $\tilde{\mathcal{P}}^{\varsigma}$ is compact; (ii) $\breve{\mathbf{p}}^{\varsigma \nu}$ is regular for every possible initial point $\mathbf{p}^{\varsigma 1} \in \tilde{\mathcal{P}}^{\varsigma}$; and (iii) the step size $a^{\varsigma \nu}$ is such that $a^{\varsigma \nu} \in (0, 1]$, $a^{\varsigma \nu} \to 0$, and $\sum_{\nu} a^{\varsigma \nu} = +\infty$.*

*Then every regular limit point of $\{\breve{\mathbf{p}}^{\varsigma \nu}\}_{\nu=1}^{\infty}$ is a stationary solution of $(\tilde{P}^{\varsigma})$. Furthermore, none of such points is a local minimum.*

### C. Determining a Feasible Initial Point for the Inner Loop

The inner loop in Algorithm 1 requires a feasible initial point, i.e., $\mathbf{p}^{\varsigma 1} \in \tilde{\mathcal{P}}^{\varsigma}$. The stationary solution to $(\tilde{P}^{\varsigma - 1})$ cannot be directly used since, in most of the cases, is not feasible, i.e., $\tilde{\mathbf{p}}^{\varsigma - 1} \notin \tilde{\mathcal{P}}^{\varsigma}$. Finding the projection of $\tilde{\mathbf{p}}^{\varsigma - 1}$ to the nonconvex feasible set $\tilde{\mathcal{P}}^{\varsigma}$ would require to solve a nonconvex problem, which is not practical because we need something simple and fast. There are many heuristic approaches to find the initial feasible point; the simplest and most general option, which in practice works well, is to move from $\tilde{\mathbf{p}}^{\varsigma - 1}$ towards the ISCA initial point, $\mathbf{p}^1$, which is required to belong to $\hat{\mathcal{P}} \cap \tilde{\mathcal{P}}^1$. It can be easily shown that if $\mathbf{p}^1 \in \hat{\mathcal{P}} \cap \tilde{\mathcal{P}}^1$, then there exists a step length, $d^{\varsigma}$, such that an initial feasible point is obtained, i.e., $\mathbf{p}^{\varsigma 1} \triangleq \tilde{\mathbf{p}}^{\varsigma - 1} + d^{\varsigma}(\mathbf{p}^1 - \tilde{\mathbf{p}}^{\varsigma - 1}) \in \tilde{\mathcal{P}}^{\varsigma}$. Given the power consumption models $\mathcal{C}_{tn}^1$ and $\mathcal{C}_{tn}^2$ in (4) and (5), we can select $\mathbf{p}^1 = \mathbf{0}$.

Depending on the specific power consumption model one may find better ways to obtain the initial feasible point for the inner loop. For example, given the power consumption model $\mathcal{C}_{tn}^1$, it can be shown that the steepest descent direction of a given energy causality constraint is a descend direction of the remaining ones. Hence, the feasible initial point can be found by successively moving in the steepest descend direction of the unfulfilled energy causality constraints.

### D. Convergence of the ISCA and Distributed Implementation

Now, the details of all the building blocks of the ISCA have been introduced. The following lemma characterizes the relations between the feasible sets of the outer loop problems, $\tilde{\mathcal{P}}^{\varsigma}$, with respect to the feasible set of the original problem, $\hat{\mathcal{P}}$.

*Lemma 3: (a) The sequence of feasible sets of the smooth problems $\{\tilde{\mathcal{P}}^{\varsigma}\}_{\mathbb{N}}$ converges to $\hat{\mathcal{P}}$ in the Painlevé-Kuratowski sense [29], i.e., $\lim_{\varsigma \to \infty} \tilde{\mathcal{P}}^{\varsigma} \to \hat{\mathcal{P}}$. (b) If the step function weights are all positive, $w_{ts} > 0$, $\forall t, s$, then we have that $\hat{\mathcal{P}} \subseteq \tilde{\mathcal{P}}^{\infty} \subset \cdots \subset \tilde{\mathcal{P}}^{\varsigma+1} \subset \tilde{\mathcal{P}}^{\varsigma} \subset \cdots \subset \tilde{\mathcal{P}}^{1}$. (c) If the weights are all negative, $w_{ts} < 0$, $\forall t, s$, then $\tilde{\mathcal{P}}^{1} \subset \cdots \subset \tilde{\mathcal{P}}^{\varsigma} \subset \tilde{\mathcal{P}}^{\varsigma+1} \subset \ldots \tilde{\mathcal{P}}^{\infty} \subseteq \hat{\mathcal{P}}$.*

*Proof:* See Appendix B. ∎

Next, we analyze the convergence of the ISCA.

*Theorem 2: (a) Let $\tilde{\mathbf{p}}^{\star \varsigma}$ be a global solution of $(\tilde{P}^{\varsigma})$, then every cluster point of the sequence $\{\tilde{\mathbf{p}}^{\star \varsigma}\}$ converges to a globally optimal solution of $(\hat{P})$. (b) Every cluster point of the sequence $\{\tilde{\mathbf{p}}^{\varsigma}\}$ converges to a stationary solution of the problem $(\hat{P})$.*

*Proof:* See Appendix C. ∎

In general, we cannot obtain the globally optimal solution of $(\hat{P})$ as the problems $(\tilde{P}^{\varsigma})$ are nonconvex and we are not able to obtain their globally optimal solution, $\tilde{\mathbf{p}}^{\star\varsigma}$. However, from the previous theorem, we guarantee that the ISCA converges to a stationary solution of the original problem $(\hat{P})$ for any decreasing sequence of $\{\rho^{\varsigma}\}$. The rate at which the approximation parameter $\rho$ is reduced affects the ISCA performance and its computational complexity. If $\rho$ is reduced at a very fast rate in the outer loop, then the algorithm might converge to a worse stationary solution. Contrarily, if $\rho$ is reduced at a very slow rate, the algorithm generally converges to a better stationary solution, but the computational complexity of the algorithm increases due to the elevated number of outer loop iterations. In the numerical experiments, we have observed a wide range of sequences $\{\rho^{\varsigma}\}$ that achieve a good trade-off between performance and computational complexity.

In order to compute the solution in a distributed manner, which is meaningful when the channel is static in time and the energy harvesting process is predictable, the following signalling is required so that the remaining transmitters can compute the weights $\pi_{tkn}^{\varsigma\nu}$: (i) at each outer loop iteration, each transmitter has to broadcast the feasible initial point of the inner loop; and (ii) at each inner loop iteration, each transmitter $t$ solely has to broadcast $\nabla_{\mathbf{p}_{-t}} r_t(\mathbf{p}_t, \mathbf{p}_{-t}^{\varsigma\nu})$, which can be computed with the local measurements of the signal to interference plus noise ratio and the multiuser interference. Note that the energy causality constraints are not fulfilled until convergence of the ISCA. Accordingly, it is required that the nodes have a backup battery to be used for this transitory regime, which can be recharged with the harvested energy.

*Remark 2:* Note that if a different rate function is employed, the ISCA can be used by deriving a proper convex approximation of the objective function.

## V. The ISCA Algorithm for $\mathcal{C}_{tn}^1$ in (4)

In this section, we focus on the power consumption model $\mathcal{C}_{tn}^1$ in (4) and derive a closed form solution of the inner loop problem.

As it has been mentioned in Section IV, at each inner loop iteration, the $t$-th transmitter must solve (17) to obtain the update direction. Given the power consumption model $\mathcal{C}_{tn}^1$, we have

$$[\breve{\mathbf{B}}_t^{\varsigma\nu}(\mathbf{p}_t; \mathbf{p}_t^{\varsigma\nu})]_\ell = \sum_{j=1}^{\ell}\left[E_{tj} - \sum_{n\in\tau_{tj}}\left(\varepsilon_t^{\varsigma\nu}(n) + \sum_{k=1}^{K}\varphi_t^{\varsigma\nu}(k,n)p_t(k,n)\right)\right],$$

with

$$\varphi_t^{\varsigma\nu}(k,n) = T_s\left(1 + \frac{\alpha_t}{\rho^{\varsigma}}e^{\frac{-\sum_{k=1}^{K}p_t^{\varsigma\nu}(k,n)}{\rho^{\varsigma}}} + \frac{\beta_t}{\rho^{\varsigma}}e^{\frac{-p_t^{\varsigma\nu}(k,n)}{\rho^{\varsigma}}}\right)$$

$$\varepsilon_t^{\varsigma\nu}(n) = T_s\alpha_t\left(1 - \left(1 + \frac{\sum_{k=1}^{K}p_t^{\varsigma\nu}(k,n)}{\rho^{\varsigma}}\right)e^{\frac{-1}{\rho^{\varsigma}}\sum_{k=1}^{K}p_t^{\varsigma\nu}(k,n)}\right) + T_s\beta_t\sum_{k=1}^{K}\left(1 - \left(1 + \frac{p_t^{\varsigma\nu}(k,n)}{\rho^{\varsigma}}\right)e^{\frac{-1}{\rho^{\varsigma}}p_t^{\varsigma\nu}(k,n)}\right), \quad (18)$$

where the constants $\varphi_t^{\varsigma\nu}(k,n)$ and $\varepsilon_t^{\varsigma\nu}(n)$ are obtained after linearizing the step functions at the current iterate, $\mathbf{p}_t^{\varsigma\nu}$.

*Lemma 4: Given the power consumption model $\mathcal{C}_{tn}^1$, the optimal solution to* (17)*, $\breve{p}_t^{\varsigma\nu}(k,n)$, $n \in \tau_{tj}$, is obtained in closed form and is given in* (19) *at the bottom of this page, where $\gamma_t^{\varsigma\nu\star}(k,n) = -\pi_{tkn}^{\varsigma\nu} + \bar{\lambda}_{tj}^{\varsigma\nu\star}\varphi_t^{\varsigma\nu}(k,n)$, $\bar{\lambda}_{tj}^{\varsigma\nu\star} = \sum_{\ell=j}^{J_t}\lambda_{t\ell}^{\varsigma\nu\star}$ with $\{\lambda_{t\ell}^{\varsigma\nu\star}\}_{\ell=1}^{J_t}$ being the optimal Lagrange multipliers associated to the energy causality constraints in* (15b)*, which can be computed efficiently similarly to [3], [30, FSA]. Additionally, if we do not include the proximal regularization term, i.e., $b_t = 0$, we obtain the following iterative directional waterfilling like solution:*

$$\breve{p}_t^{\varsigma\nu}(k,n) = \left[\frac{1}{\gamma_t^{\varsigma\nu\star}(k,n)} - \frac{MUI_t^{\varsigma\nu}(k,n)}{h_{tt}(k,n)}\right]_0^{p_t^{\max}(k,n)}. \quad (20)$$

*Proof:* See Appendix D. ∎

From the expression in (20), we can get some intuition on the solution. First, if the water level, $\gamma_t^{\varsigma\nu\star}(k,n)^{-1}$, is smaller than $MUI_t^{\varsigma\nu}(k,n)/h_{tt}(k,n)$, then it is preferable to turn off the $(k,n)$-th subchannel. Second, the water level decreases with the interference produced to other users, which is quantified in the term $-\pi_{tkn}^{\varsigma\nu}$. This implies that the users will try to reduce the interference as much as possible to increase the sum-rate. Third, the water level depends on $\varphi_t^{\varsigma\nu}(k,n)$ in (18), i.e., the partial derivative of the smooth energy consumption with respect to $p_t(k,n)$ evaluated at the current iterate $p_t^{\varsigma\nu}(k,n)$. Accordingly, if the power of a certain subchannel is small, $p_t^{\varsigma\nu}(k,n) \to 0$, the derivative of the smooth step functions is large, and the water level is penalized; vice versa if the power is large the water level is rewarded. Note that these penalizations or rewards are weak at the initial ISCA iterations, because the approximation of the step functions is smooth, but they gain in importance as the ISCA iterations go by. Finally, the water level is a function of the Lagrange multipliers that depend on the energy availability of the node in a similar way than in the directional waterfilling solution [2].

$$\breve{p}_t^{\varsigma\nu}(k,n) = \left[\frac{1}{2}\left(p_t^{\varsigma\nu}(k,n) - \frac{MUI_t^{\varsigma\nu}(k,n)}{h_{tt}(k,n)}\right) - \frac{1}{2b_t}\left(\gamma_t^{\varsigma\nu\star}(k,n) - \sqrt{\left[\gamma_t^{\varsigma\nu}(k,n) - b_t\left(p_t^{\varsigma\nu}(k,n) + \frac{MUI_t^{\varsigma\nu}(k,n)}{h_{tt}(k,n)}\right)\right]^2 + 4b_t}\right)\right]_0^{p_t^{\max}(k,n)} \quad (19)$$

Note that the ISCA can be applied to a broad class of problems. In the following remarks, we use the ISCA to derive stationary solutions of power allocation problems that, to the best of our knowledge, have not been yet derived in the literature.

*Remark 3 (Transmission Power Only):* Consider the sum-rate maximization problem of a network of energy harvesting nodes, where the unique source of energy consumption is the transmission power ($\alpha_t = 0$, $\beta_t = 0$, $\gamma_t = 0$). Then, the inner loop of the ISCA (or the algorithm in [19]) can be used to determine distributedly a stationary solution, where the solution to the $\varsigma\nu$-th inner loop problem is

$$\breve{p}_t^{\varsigma\nu}(k,n) = \left[\frac{1}{-\pi_{tkn}^{\varsigma\nu} + \bar{\lambda}_{tj}^{\varsigma\nu\star} T_s} - \frac{MUI_t^{\varsigma\nu}(k,n)}{h_{tt}(k,n)}\right]_0^{p_t^{\max}(k,n)}, \quad n \in \tau_{tj}.$$

*Remark 4 (Power Consumption Model $\mathcal{C}_{tn}^2$):* Consider the problem of maximizing the sum-rate given the power consumption model $\mathcal{C}_{tn}^2$. Then, the ISCA determines distributedly a stationary solution, where the inner loop problem must be solved by numerical methods since it does not accept a closed form solution. Additionally, the power allocation strategy in a point to point link is obtained by particularizing $T = 1$, which implies that $\pi_{tkn}^{\varsigma\nu} = 0, \forall k, n$.

*Remark 5 (Non-Harvesting Nodes):* Finally, consider the sum-rate maximization problem of a network of non-harvesting nodes given any power consumption model of the form (6). Then, the ISCA distributedly determines a stationary power allocation policy, where the strongly convex problem that has to be solved at each inner loop iteration is (17) given that $J_t = 1, \forall t$, which imposes a sum-power constraint.

## VI. Results

In this section, we numerically evaluate the performance of the ISCA in terms of achieved rate and computational complexity of the algorithm. The most similar work is [14] that considered the problem of maximizing the mutual information in a point to point link with the power consumption model $\mathcal{C}_{tn}^1$. Accordingly, in order to have some benchmark with which to compare the performance of the ISCA, we first particularize the solution derived in the previous section to the case $T = 1$. The remaining system parameters have been set as follows. We have considered $N = 50$ channel accesses of duration $T_s = 20\ ms$ in which symbols are transmitted through $K = 2$ parallel streams. The power consumption constants are set to $\alpha_t = \gamma_t = 150\ mW$ and $\beta_t = 10\ mW$. A Rayleigh fading channel has been considered with unit mean channel power gain. The energy harvesting process is modeled as a compound Poisson process as done in [2], where the arrival instants follow a Poisson distribution with rate $\frac{1}{10}$ and the energy in the packets is drawn from a uniform distribution and normalized by the total harvested energy that varies along the $x$-axis of Figs. 5-7. The initial point of the ISCA is set to zero, and the approximation control parameter is $\rho^{\varsigma} = 0.5\rho^{\varsigma-1}$ with $\rho^1 = 5$. We have not used the proximal regularization term, $b_t = 0$, and the inner loop step size is $a^{\varsigma\nu} = a^{\varsigma(\nu-1)}(1 - 10^{-3}a^{\varsigma(\nu-1)})$ with $a^{\varsigma 0} = 1, \forall\varsigma$.

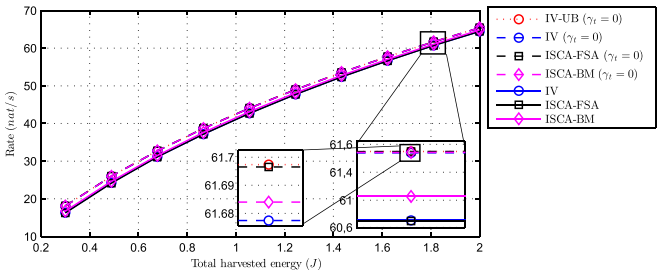



Fig. 5. Achieved rate versus total harvested energy.

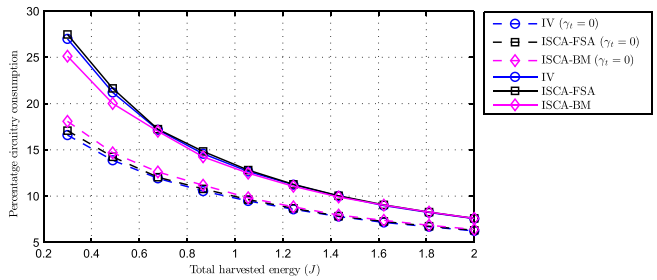



Fig. 6. Percentage of the total harvested energy expended in the circuitry.

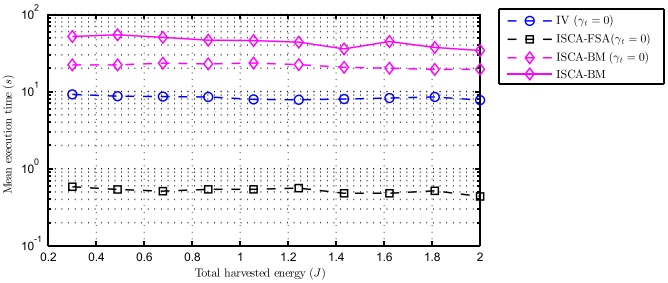



Fig. 7. Mean execution time versus total harvested energy.

We consider two classes of strategies. On the one hand, we consider strategies that use the power consumption model $\mathcal{C}_{tn}^1$ (i.e., strategies that disregard the startup power consumption, $\gamma_t = 0$), which can be based either on the use of indicator variables or on the use of the ISCA: (i) the indicator variables based strategies are an upper bound of the solution, *IV-UB* ($\gamma_t = 0$), and a feasible solution that performs close to the upper bound *IV* ($\gamma_t = 0$) and were derived in [14]; (ii) the ISCA based strategies are *ISCA-FSA* ($\gamma_t = 0$) that uses the closed form solution derived in Lemma 4, where the Lagrange multipliers are obtained by using the FSA algorithm in [3], and *ISCA-BM* ($\gamma_t = 0$) that solves (17) using the barrier method given that $\gamma_t = 0$. On the other hand, we consider strategies that account for the startup power consumption (i.e., strategies that use the power consumption model $\mathcal{C}_{tn}^2$), *ISCA-BM*, which solves (17) using the barrier method, and the strategies *IV* and *ISCA-FSA*, which scale *IV* ($\gamma_t = 0$), and *ISCA-FSA* ($\gamma_t = 0$) until the energy causality constraints with startup power consumption are satisfied.

In this setup, Fig. 5 shows the achieved rate versus total harvested energy. First, we observe that the stationary solution provided by the ISCA strategies is close to the global optimum

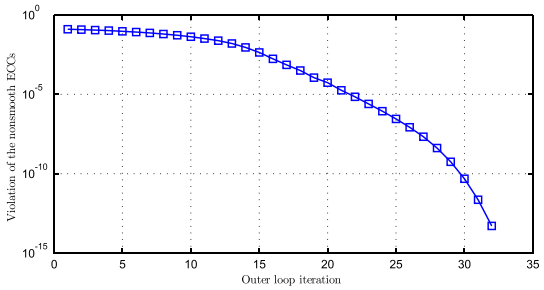



Fig. 8. Violation of the original energy causality constraints as a function of the ISCA outer iteration for the *ISCA-FSA* ($\gamma_t = 0$) power allocation strategy.

since the gap with the upper bound, *IV-UB* ($\gamma_t = 0$), is small. Second, the ISCA based solutions perform slightly better than the feasible strategy *IV* ($\gamma_t = 0$). Finally, as expected, the strategies that consider the startup power consumption achieve a lower sum-rate, where the stationary solution *ISCA-BM* performs better than the other strategies that consider the startup power consumption.

Fig. 6 shows the percentage of the total harvested energy that is expended in the circuitry. The percentage of energy spent in the circuitry is much higher at low harvested energies, where the cost for turning on a subchannel is a high fraction of the total available energy, and decreases in the high energy regime, where the transmission power in each subchannel increases. Additionally, at the high energy regime, the effect of disregarding the startup power consumption does not have a significant impact since most of the channel accesses are active and, accordingly, there are a few off-on transitions.

Fig. 7 evaluates the computational complexity of the different algorithms. It is observed that the worst performance is achieved by *ISCA-BM*, where most of the execution time is spent in the computation of the gradient and the Hessian required for the Newton method [22]; however, it also solves a more complex problem than the strategies that disregard the startup power consumption. Note that the performance of the ISCA with the barrier method improves when the startup power consumption is disregarded. Finally, it is important to mention that when a closed form solution of the inner loop problem is available, as happens with *ISCA-FSA*, the computational complexity of the ISCA is dramatically reduced outperforming the strategy based on indicator variables.

Fig. 8 shows the violation of the original nonsmooth energy causality constraints (7b) produced by the stationary solution of each smooth nonconvex problem, which is computed as the Euclidean norm of the nonfulfilled constraints. To obtain Fig. 8, we have reduced the approximation control parameter slightly slower to have more points, i.e., $\rho^{\varsigma} = 0.8\rho^{\varsigma-1}$ with $\rho^1 = 5$. As expected, when the outer loop iterations go by and the approximation control parameter is reduced, the violation of the original nonsmooth energy causality constraints is reduced.

Next, we evaluate the ISCA for the case of having $T = 3$ simultaneous transmitter-receiver pairs. Since the optimal solution is unknown for $T > 1$, the performance of the ISCA is evaluated with respect to the idealistic case of having multiple non-interfering point to point links with the upper bound

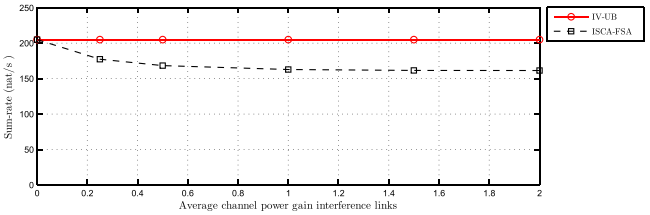



Fig. 9. Sum-rate achieved in the interference channel when $T = 3$ for different average power gains of the interference links.

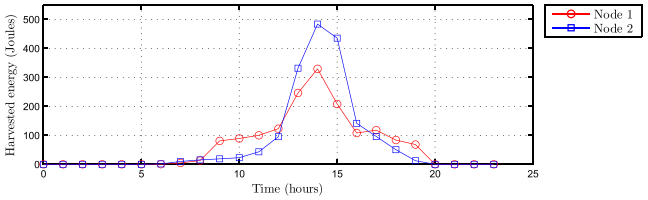



Fig. 10. Harvested energy using solar panels at the different hours of the day.

TABLE II
CHANNEL POWER GAINS

| $k = 1, \forall n$ | $t' = 1$ | $t' = 2$ | $k = 2, \forall n$ | $t' = 1$ | $t' = 2$ |
|---|---|---|---|---|---|
| $t = 1$ | 0.875 | 18.152 | $t = 1$ | 0.328 | 2.392 |
| $t = 2$ | 3.838 | 1.681 | $t = 2$ | 4.765 | 1.063 |

obtained with indicator variables IV-UB in [14]. As before, we consider a time window of 1 s where the total harvested energy of transmitters 1-3 is 2, 1.5 and 3 Joules, respectively. For the direct links, we have considered a Rayleigh fading channel with unit mean channel power gain. In this context, Fig. 9 evaluates the achieved rate for different average channel power gains of the interference links. It is observed that when the channel power gain of the interference links is close to zero, the performance of the ISCA is close to the idealistic case of having multiple non-interfering point to point links. However, as the channel power gain of the interference links increases, the sum-rate of the ISCA is reduced.

Finally, we evaluate the performance of the ISCA for realistic solar energy harvesting profiles. We consider $T = 2$ transmitters with a solar panel of 30% efficiency and dimensions of 45 cm $\times$ 45 cm and 63 cm $\times$ 63 cm. The data from the harvested energy is obtained from [31] and is depicted in Fig. 10. We consider that the continuous harvested energy is first stored in a super-capacitor and then transferred at every hour transition to the rechargeable battery to be used for transmission. The total transmission time is one day that is divided in transmission slots of 15 minutes. We consider $K = 2$ subcarriers. The channel power gains in each subcarrier, $h_{tt'}(k, n)$, are given in Table II and are assumed to be static in time. The power consumption constants are set to $\alpha_t = 100$ $\mu$W, $\beta_t = 10$ $\mu$W, and $\gamma_t = 0$. In this context, Fig. 11 depicts the transmission power obtained with the ISCA for each node and subcarrier. It is observed that both nodes transmit solely at subcarrier $k = 1$, where the direct link is better. At 7 am both nodes start harvesting energy.

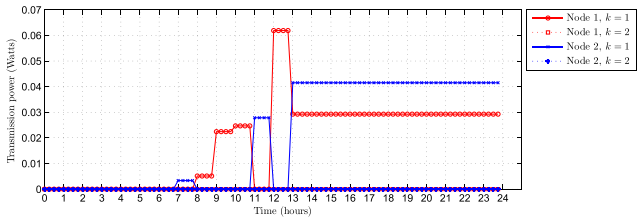



Fig. 11. Power allocation obtained for the ISCA algorithm under the solar energy harvesting depicted in Fig. 10.

Then, node 2 starts transmitting, while node 1 remains off to avoid generating interference. From 8 to 11, node 1 transmits and 2 switches off; this occurs because in this time interval node 1 harvests more energy. The nodes alternate transmission cycles in a time division fashion from 11 to 13. Finally, from 13 until midnight both nodes transmit simultaneously at a constant power.

## VII. Conclusion

In this paper, we have studied the offline sum-rate maximization problem of a Gaussian interference channel composed of EHWNs by considering a general power consumption model that is composed of step functions. We have proposed the ISCA, a distributed power allocation algorithm that is based on SSA of the step functions to derive a sequence of smooth nonconvex problems that can be solved by means of SCA. It has been shown that the ISCA converges to stationary solutions of the original optimization problem, which is nonsmooth, nonconvex, and NP-hard. The numerical results have first focused on point to point communications in order to have benchmarks to evaluate the performance of the ISCA. In this setup, we have shown that the ISCA is able to avoid bad stationary solutions performing close to the globally optimal solution and reduces the complexity of existing algorithms. The performance of the ISCA has been also evaluated using real solar energy harvesting traces. It has been observed that the power allocation obtained with the ISCA naturally tries to avoid interference to maximize the network sum-rate. In conclusion, the ISCA is a powerful offline power allocation algorithm able to compute stationary solutions in the Gaussian interference channel for a broad class of power consumption models, which was not possible with existing algorithms. In this context, the ISCA can be used to evaluate novel online policies that do not require non-causal knowledge of the harvested energy.

## Appendix

### A. Proof of Lemma 2

In the following lines, we show that $\breve{\mathcal{H}}_\rho(\mathbf{x};\mathbf{x}_0,\boldsymbol{\phi})$ satisfies all the requirements imposed from [19, Assumption 3] that are:

*Assumption 1: The function* $\breve{\mathcal{H}}_\rho(\cdot;\cdot,\boldsymbol{\phi}):\mathcal{X}\times\mathcal{Y}\to\mathbb{R}$ *must satisfy [19, Assumption 3] for all* $\boldsymbol{\phi}$: ***A1)*** $\breve{\mathcal{H}}_\rho(\cdot;\mathbf{x}_0,\boldsymbol{\phi})$ *is convex on* $\mathcal{X}$ *for all* $\mathbf{x}_0\in\mathcal{Y}$; ***A2)*** $\breve{\mathcal{H}}_\rho(\mathbf{x};\mathbf{x},\boldsymbol{\phi})=\mathcal{H}_\rho(\boldsymbol{\phi}(\mathbf{x}))$, *for all* $\mathbf{x}\in\mathcal{X}$; ***A3)*** $\mathcal{H}_\rho(\boldsymbol{\phi}(\mathbf{x}))\le\breve{\mathcal{H}}_\rho(\mathbf{x};\mathbf{x}_0,\boldsymbol{\phi})$ *for all* $\mathbf{x}\in\mathcal{X}$ *and* $\mathbf{x}_0\in\mathcal{Y}$; ***A4)*** $\breve{\mathcal{H}}_\rho(\cdot;\cdot,\boldsymbol{\phi})$ *is Lipschitz continuous on* $\mathcal{X}\times\mathcal{Y}$; ***A5)*** $\nabla_{\mathbf{x}}\breve{\mathcal{H}}_\rho(\mathbf{x}_0;\mathbf{x}_0,\boldsymbol{\phi})=\nabla_{\mathbf{x}}\mathcal{H}_\rho(\boldsymbol{\phi}(\mathbf{x}_0))$, *for all* $\mathbf{x}_0\in\mathcal{Y}$; ***A6)*** $\nabla_{\mathbf{x}}\breve{\mathcal{H}}_\rho(\cdot;\cdot,\boldsymbol{\phi})$ *is continuous on* $\mathcal{X}\times\mathcal{Y}$; *where* $\nabla_{\mathbf{x}}\breve{\mathcal{H}}_\rho(\mathbf{x}_0;\mathbf{x}_0,\boldsymbol{\phi})$ *denotes the partial gradient of* $\breve{\mathcal{H}}_\rho(\mathbf{x};\mathbf{x}_0,\boldsymbol{\phi})$ *with respect to* $\mathbf{x}$ *evaluated at* $(\mathbf{x}_0;\mathbf{x}_0,\boldsymbol{\phi})$.

First note that since the component functions of $\phi$ are all concave and $e^{\frac{-x}{\rho}}$ is convex and decreasing, the function $e^{-\frac{\sum_{k=1}^{i}\phi_{j_k}(\mathbf{x})}{\rho}}$ is convex [22], which proves the convexity of the terms in (12). Accordingly, A1 is satisfied because $\breve{\mathcal{H}}_\rho$ is the addition of convex and affine terms. Since $\breve{\mathcal{H}}_\rho(\cdot;\mathbf{x}_0,\boldsymbol{\phi})$ is obtained after linearizing the concave terms of $\mathcal{H}_\rho(\boldsymbol{\phi}(\cdot))$, it follows that $\breve{\mathcal{H}}_\rho(\cdot;\mathbf{x}_0,\boldsymbol{\phi})$ is a global over estimator that has the same value and gradient at $\mathbf{x}_0$. Hence, conditions A2, A3, and A5 are also satisfied. Finally, since $e^{\frac{-x}{\rho}}$ is Lipschitz continuous, $\breve{\mathcal{H}}_\rho(\cdot;\cdot,\boldsymbol{\phi})$ is also Lipschitz continuous.

### B. Proof of Lemma 3

Note that the difference between $\tilde{\mathcal{P}}^{\varsigma+1}$ and $\tilde{\mathcal{P}}^{\varsigma}$ is due to the reduction of the approximation control parameter ($\rho^{\varsigma+1}<\rho^{\varsigma}$) in the energy causality constraints. It can be easily shown that: (i) $\mathcal{H}_\rho$ is strictly decreasing in $\rho$ (for $x>0$) and (ii) $\mathcal{H}_\rho(\mathbf{x})\le\prod_{q=1}^{Q}\mathcal{H}(x_q)$, $\forall\mathbf{x}\in\mathbb{R}_+^Q$, $\rho>0$. Accordingly, when all the weights are positive, we have from (i) that the smooth energy causality constraints are tightened when the approximation control parameter is reduced, i.e., $\tilde{\mathcal{P}}^{\varsigma+1}\subset\tilde{\mathcal{P}}^{\varsigma},\forall\varsigma$. Additionally, from (ii), the energy causality constraints are relaxed when using the smooth approximation, we have that $\hat{\mathcal{P}}\subseteq\tilde{\mathcal{P}}^{\varsigma}$. This proves (b). The proof of (c) follows similarly by noting that when all the weights are negative the original energy causality constraints are tightened.

To prove (a), we define the sets $\tilde{\mathcal{P}}_+^{\varsigma}$ and $\tilde{\mathcal{P}}_-^{\varsigma}$ in (21) at the top of next page, where $\mathcal{S}_t^+$ and $\mathcal{S}_t^-$ are defined as in (16). Note that $\tilde{\mathcal{P}}_+^{\varsigma}$ approximates the positive step functions only and $\tilde{\mathcal{P}}_-^{\varsigma}$, the negative ones. Similarly than in the proofs of (b) and (c) it follows that $\tilde{\mathcal{P}}_+^{\varsigma+1}\subset\tilde{\mathcal{P}}_+^{\varsigma},\forall\varsigma$ and $\tilde{\mathcal{P}}_-^{\varsigma}\subset\tilde{\mathcal{P}}_-^{\varsigma+1},\forall\varsigma$. Additionally, following the same arguments, we have that $\tilde{\mathcal{P}}_-^{\varsigma}\subset\tilde{\mathcal{P}}^{\varsigma}\subset\tilde{\mathcal{P}}_+^{\varsigma}$. From [29, Exercise 4.3], we have that the limits of the sets $\tilde{\mathcal{P}}_+^{\varsigma}$ and $\tilde{\mathcal{P}}_-^{\varsigma}$ exists (the inner and outer limits are equal) and are $\tilde{\mathcal{P}}_+^{\varsigma}\to\hat{\mathcal{P}}$ and $\tilde{\mathcal{P}}_-^{\varsigma}\to\hat{\mathcal{P}}$. This leads to $\tilde{\mathcal{P}}^{\varsigma}\to\hat{\mathcal{P}}$, which proves (a).

### C. Proof of Theorem 2

Let us write the original optimization problem $(\hat{P})$ as $\min_{\mathbf{p}}\hat{f}(\mathbf{p})$ with $\hat{f}:\mathbb{R}^{NKT}\to\mathbb{R}\cup\{\infty\}$, $\hat{f}(\mathbf{p})=r(\mathbf{p})+I_{\hat{\mathcal{P}}}(\mathbf{p})$, where $r(\mathbf{p})=-\sum_{t=1}^{T}r_t(\mathbf{p}_t,\mathbf{p}_{-t})$ and $I_\Omega(x)$ is the indicator function of a given set $\Omega$ with $I_\Omega(x)=0$ if $x\in\Omega$ and $I_\Omega(x)=\infty$ otherwise. Similarly, we write the smooth problem, $(\tilde{P}^{\varsigma})$, as $\min_{\mathbf{p}}\tilde{f}^{\varsigma}(\mathbf{p})$ with $\tilde{f}^{\varsigma}(\mathbf{p})=r(\mathbf{p})+I_{\tilde{\mathcal{P}}^{\varsigma}}(\mathbf{p})$.

*Lemma 5:* $\tilde{f}^{\varsigma}(\mathbf{p})$ *epi-converges to* $\hat{f}(\mathbf{p})$, $\tilde{f}^{\varsigma}(\mathbf{p})\xrightarrow{e}\hat{f}(\mathbf{p})$ *(epi-convergence is defined in [29]).*

*Proof:* From the convergence of the feasible sets in Lemma 3 and [29, Proposition 7.4 f], we have that $I_{\tilde{\mathcal{P}}^{\varsigma}}\xrightarrow{e}I_{\hat{\mathcal{P}}}$. Given that $r(\mathbf{p})$ is continuous and finite, the lemma is readily proven by using [29, Exercise 7.8 a]. ■

$$\tilde{\mathcal{P}}_{+}^{\varsigma} \triangleq \left\{ \mathbf{p}_t \in \mathcal{P}_t : \sum_{j=1}^{\ell} \left[ E_{tj} - T_s \sum_{n \in \tau_{tj}} \left( \left( \sum_{k=1}^{K} p_t(k,n) \right) + \sum_{s \in \mathcal{S}_t^+} w_{ts} \mathcal{H}_{\rho^{\varsigma}}(\boldsymbol{\phi}_{ts}(\mathbf{p}_t)) \right. \right. \right.$$
$$\left. \left. \left. + \sum_{s \in \mathcal{S}_t^-} w_{ts} \prod_{q=1}^{Q_{ts}} \mathcal{H}(\phi_{tsq}(\mathbf{p}_t)) \right) \right] \geq 0, \forall \ell = 1, \ldots J_t, \forall t \right\}$$

$$\tilde{\mathcal{P}}_{-}^{\varsigma} \triangleq \left\{ \mathbf{p}_t \in \mathcal{P}_t : \sum_{j=1}^{\ell} \left[ E_{tj} - T_s \sum_{n \in \tau_{tj}} \left( \left( \sum_{k=1}^{K} p_t(k,n) \right) + \sum_{s \in \mathcal{S}_t^-} w_{ts} \mathcal{H}_{\rho^{\varsigma}}(\boldsymbol{\phi}_{ts}(\mathbf{p}_t)) \right. \right. \right.$$
$$\left. \left. \left. + \sum_{s \in \mathcal{S}_t^+} w_{ts} \prod_{q=1}^{Q_{ts}} \mathcal{H}(\phi_{tsq}(\mathbf{p}_t)) \right) \right] \geq 0, \forall \ell = 1, \ldots J_t, \forall t \right\} \tag{21}$$

$$\breve{p}_t^{\varsigma\nu}(k,n) = \begin{cases} 0 & \text{if } \gamma_t^{\varsigma\nu}(k,n) \geq \dfrac{h_{tt}(k,n)}{MUI_t^{\varsigma\nu}(k,n)} + b_t p_t^{\varsigma\nu}(k,n), \\ \bar{p}_t^{\varsigma\nu}(k,n) & \text{if } \dfrac{h_{tt}(k,n)}{h_{tt}(k,n)p_t^{\max}(k,n) + MUI_t^{\varsigma\nu}(k,n)} - b_t(p_t^{\max}(k,n) - p_t^{\varsigma\nu}(k,n)) \\ & \quad < \gamma_t^{\varsigma\nu}(k,n) < \dfrac{h_{tt}(k,n)}{MUI_t^{\varsigma\nu}(k,n)} + b_t p_t^{\varsigma\nu}(k,n) \\ p_t^{\max}(k,n) & \text{if } \gamma_t^{\varsigma\nu}(k,n) \leq \dfrac{h_{tt}(k,n)}{h_{tt}(k,n)p_t^{\max}(k,n) + MUI_t^{\varsigma\nu}(k,n)} - b_t(p_t^{\max}(k,n) - p_t^{\varsigma\nu}(k,n)) \end{cases} \tag{26}$$

Since the sets $\tilde{\mathcal{P}}^{\varsigma}$ and $\hat{\mathcal{P}}$ are closed and nonempty and $r(\mathbf{p})$ is continuous, it follows that the functions $\tilde{f}^{\varsigma}$ and $\hat{f}$ are proper, lower semi-continuous, and eventually level-bounded [29]. Then, the proof of the statement in (a) readily follows by using Lemma 5 and [29, Theorem 7.33].

Next, we prove the statement in (b). We will demonstrate that a given cluster point, $\tilde{\mathbf{p}}$, of the sequence $\{\tilde{\mathbf{p}}^{\varsigma}\}_{\mathbb{N}}$ converges to a stationary solution of $(\hat{P})$ and the same procedure can be applied to every cluster point of the sequence. Thus, there exists a suitable subsequence $\mathcal{N}' \subseteq \mathbb{N}$ such that $\lim_{\mathcal{N}' \in \varsigma \to \infty} \tilde{\mathbf{p}}^{\varsigma} = \tilde{\mathbf{p}}$.

First of all, note that if $\nabla r(\tilde{\mathbf{p}}) = \mathbf{0}$, then necessarily $\tilde{\mathbf{p}}$ is a stationary solution of $(\hat{P})$. Otherwise, if $\nabla r(\tilde{\mathbf{p}}) \neq \mathbf{0}$, then $\tilde{\mathbf{p}}$ is a stationary solution if and only if $\tilde{\mathbf{p}}$ is a local (or global) minimum of $\hat{f}$ in the boundary of the feasible set $\hat{\mathcal{P}}$. Thus, we need to show that when $\nabla r(\tilde{\mathbf{p}}) \neq \mathbf{0}$, then $\tilde{\mathbf{p}} \subset \operatorname{argmin}_{\mathbb{C}(\tilde{\mathbf{p}},\delta)} \hat{f} \triangleq \{\mathbf{p} \in \mathbb{C}(\tilde{\mathbf{p}},\delta) | \hat{f}(\mathbf{p}) = \inf_{\mathbb{C}(\tilde{\mathbf{p}},\delta)} \hat{f}\}$, where $\mathbb{C}(\tilde{\mathbf{p}},\delta)$ denotes a closed ball around the point $\tilde{\mathbf{p}}$ with radius $\delta > 0$. Similarly, let $\mathbb{O}(\tilde{\mathbf{p}},\delta)$ denote the open ball centered at $\tilde{\mathbf{p}}$ with radius $\delta$.

From the continuity of $\nabla r$, we know that there is a sufficiently large index $\bar{\varsigma} \in \mathcal{N}'$ such that $\nabla r(\tilde{\mathbf{p}}^{\varsigma}) \neq \mathbf{0}$ $\varsigma \geq \bar{\varsigma}$. Then, similarly as before, we know that the points $\tilde{\mathbf{p}}^{\varsigma}, \forall \varsigma \geq \bar{\varsigma}$, correspond to local minimums of $\tilde{f}^{\varsigma}$ at the boundary of $\tilde{\mathcal{P}}^{\varsigma}$, i.e., $\tilde{\mathbf{p}}^{\varsigma} \subset \operatorname{argmin}_{\mathbb{C}(\tilde{\mathbf{p}}^{\varsigma},\delta^{\varsigma})} \tilde{f}^{\varsigma}$, $\forall \varsigma \geq \bar{\varsigma}$. Then, there exist arbitrary small constants $\delta > 0$ and $\epsilon > 0$ and a sufficiently large index $\bar{\bar{\varsigma}} \in \mathcal{N}'$, $\bar{\bar{\varsigma}} \geq \bar{\varsigma}$, such that the points $\tilde{\mathbf{p}}^{\varsigma}$ satisfy that $\tilde{\mathbf{p}}^{\varsigma} \subset \operatorname{argmin}_{\mathbb{C}(\tilde{\mathbf{p}},\delta)} \tilde{f}^{\varsigma}$ and $\tilde{\mathbf{p}}^{\varsigma} \subset \operatorname{argmin}_{\mathbb{O}(\tilde{\mathbf{p}},\delta+\epsilon)} \tilde{f}^{\varsigma}$ for all $\varsigma \geq \bar{\bar{\varsigma}}$, or, equivalently,

$$\inf_{\mathbb{O}(\tilde{\mathbf{p}},\delta+\epsilon)} \tilde{f}^{\varsigma} = \inf_{\mathbb{C}(\tilde{\mathbf{p}},\delta)} \tilde{f}^{\varsigma} = \tilde{f}^{\varsigma}(\tilde{\mathbf{p}}^{\varsigma}), \quad \forall \varsigma \geq \bar{\bar{\varsigma}}. \tag{22}$$

The existence of $\delta$ and $\epsilon$ satisfying (22) follows from the differentiability of $r(p)$. Then, from [29, Proposition 7.29] and Lemma 5, we know that $\liminf_{\varsigma} (\inf_{\mathbb{C}(\tilde{\mathbf{p}},\delta)} \tilde{f}^{\varsigma}) \geq \inf_{\mathbb{C}(\tilde{\mathbf{p}},\delta)} \hat{f}$ and

$$\limsup_{\varsigma} \left( \inf_{\mathbb{O}(\tilde{\mathbf{p}},\delta+\epsilon)} \tilde{f}^{\varsigma} \right) = \limsup_{\varsigma} \left( \inf_{\mathbb{C}(\tilde{\mathbf{p}},\delta)} \tilde{f}^{\varsigma} \right) \leq \inf_{\mathbb{O}(\tilde{\mathbf{p}},\delta+\epsilon)} \hat{f} \leq \inf_{\mathbb{C}(\tilde{\mathbf{p}},\delta)} \hat{f}. \tag{23}$$

Combining these two results, we have that $\inf_{\mathbb{C}(\tilde{\mathbf{p}},\delta)} \tilde{f}^{\varsigma} \to \inf_{\mathbb{C}(\tilde{\mathbf{p}},\delta)} \hat{f}$.

Next, we show that $\tilde{\mathbf{p}}$ is a stationary solution of the original problem. Recall that $\tilde{\mathbf{p}}^{\varsigma} \subset \operatorname{argmin}_{\mathbb{C}(\tilde{\mathbf{p}},\delta)} \tilde{f}^{\varsigma}$, $\forall \varsigma \geq \bar{\bar{\varsigma}}$ and that these sets are nonempty by [29, Theorem 1.9]. Similarly, $\operatorname{argmin}_{\mathbb{C}(\tilde{\mathbf{p}},\delta)} \hat{f}$ is nonempty. Thus, we have

$$\hat{f}(\tilde{\mathbf{p}}) \leq \liminf_{\varsigma} \tilde{f}^{\varsigma}(\tilde{\mathbf{p}}^{\varsigma}) \leq \limsup_{\varsigma} \tilde{f}^{\varsigma}(\tilde{\mathbf{p}}^{\varsigma}) \tag{24}$$

$$= \limsup_{\varsigma} \inf_{\mathbb{C}(\tilde{\mathbf{p}},\delta)} \tilde{f}^{\varsigma} \leq \inf_{\mathbb{C}(\tilde{\mathbf{p}},\delta)} \hat{f}, \tag{25}$$

where the first inequality follows from the definition of epi-convergence [29, Proposition 7.2] and the last one from (23). Accordingly, we have that $\hat{f}(\tilde{\mathbf{p}}) = \inf_{\mathbb{C}(\tilde{\mathbf{p}},\delta)} \hat{f}$. Thus, whenever $\nabla r(\tilde{\mathbf{p}}) \neq \mathbf{0}$, we have that $\tilde{\mathbf{p}}$ is a local minimum of $\hat{f}$ on the boundary of the feasible set, i.e., $\tilde{\mathbf{p}} \subset \operatorname{argmin}_{\mathbb{C}(\tilde{\mathbf{p}},\delta)} \hat{f}$. Thus, $\tilde{\mathbf{p}}$ is a stationary solution of $(\hat{P})$.

### *D. Proof of Lemma 4*

We proof Lemma 4 by using the KKT sufficient optimality conditions. The Lagrangian of the problem in (17) is $\mathcal{L}_t^{\varsigma\nu}(\mathbf{p}_t, \boldsymbol{\lambda}_t^{\varsigma\nu}) = r_t(\mathbf{p}_t, \mathbf{p}_{-t}^{\varsigma\nu}) + \boldsymbol{\pi}_t^{\varsigma\nu\mathsf{T}}(\mathbf{p}_t - \mathbf{p}_t^{\varsigma\nu}) - \frac{b_t}{2}||\mathbf{p}_t - \mathbf{p}_t^{\varsigma\nu}||^2 + \boldsymbol{\lambda}_t^{\varsigma\nu\mathsf{T}} \breve{\mathbf{B}}_t^{\varsigma\nu}(\mathbf{p}_t; \mathbf{p}_t^{\varsigma\nu})$, where $\boldsymbol{\lambda}_t^{\varsigma\nu}$ are the Lagrange multipliers associated to the energy causality constraints in (17b). Taking the derivative of the Lagrangian with respect to $p_t(k,n)$, $n \in \tau_{tj}$, and equating to zero, we obtain $\breve{p}_t^{\varsigma\nu}(k,n)$ as given in (26) at the top of this page,

where $\gamma_t^{\varsigma\nu}(k,n) = -\pi_{tkn}^{\varsigma\nu} + \bar{\lambda}_{tj}^{\varsigma\nu}\varphi_t^{\varsigma\nu}(k,n)$ with $\bar{\lambda}_{tj}^{\varsigma\nu} = \sum_{\ell=j}^{J_t}\lambda_{t\ell}^{\varsigma\nu}$; and $\bar{p}_t^{\varsigma\nu}(k,n)$ is obtained as the solution of the following quadratic equation

$$\frac{h_{tt}(k,n)}{h_{tt}(k,n)\bar{p}_t^{\varsigma\nu}(k,n) + MUI_t^{\varsigma\nu}(k,n)} = \gamma_t^{\varsigma\nu}(k,n) + b_t(\bar{p}_t^{\varsigma\nu}(k,n) - p_t^{\varsigma\nu}(k,n)). \quad (27)$$

From [32, Lemma 35] (with $H_k := \frac{h_{tt}(k,n)}{MUI_t^{\varsigma\nu}(k,n)}$, $\tau := b_t$, $c_k := p_t^{\varsigma\nu}(k,n)$, $\tilde{\mu}_k := \gamma_t^{\varsigma\nu}(k,n)$), the previous equation has the following properties: (i) both roots are real, one root is always negative, and the other is nonnegative; (ii) both roots are decreasing in $\gamma_t^{\varsigma\nu}(k,n)$; and the nonnegative root is given by (19). Finally, note that if the proximal step is zero, $b_t = 0$, then (20) follows directly from the first order equation in (27).

## REFERENCES


[1] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 1089–1098, Aug. 2004.

[2] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with energy harvesting nodes in fading wireless channels: Optimal policies," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1732–1743, Sep. 2011.

[3] M. Gregori and M. Payaró, "On the precoder design of a wireless energy harvesting node in linear vector Gaussian channels with arbitrary input distribution," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 1868–1879, May 2013.

[4] O. Orhan, D. Gündüz, and E. Erkip, "Throughput maximization for an energy harvesting communication system with processing cost," in *Proc. IEEE Inf. Theory Workshop*, Sep. 2012, pp. 84–88.

[5] M. Gregori and M. Payaró, "Energy-efficient transmission for wireless energy harvesting nodes," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1244–1254, Mar. 2013.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[7] P. Youssef-Massaad, M. Medard, and L. Zheng, "Impact of processing energy on the capacity of wireless channels," in *Proc. Int. Symp. Inf. Theory Appl.*, Oct. 2004, pp. 1–6.

[8] P. Youssef-Massaad, L. Zheng, and M. Medard, "Bursty transmission and glue pouring: On wireless channels with overhead costs," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5188–5194, Dec. 2008.

[9] J. Xu and R. Zhang, "Throughput optimal policies for energy harvesting wireless transmitters with non-ideal circuit power," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 2, pp. 322–332, Feb. 2014.

[10] R. Rajesh, V. Sharma, and P. Viswanath, "Capacity of Gaussian channels with energy harvesting and processing cost," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2563–2575, May 2014.

[11] X. Liu and E. Erkip. (2013). "Energy-efficient communication over Gaussian interference networks with processing energy cost." [Online]. Available: arXiv:1301.1661

[12] M. Gregori and M. Payaró, "Throughput maximization for a wireless energy harvesting node considering the circuitry power consumption," in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2012, pp. 1–5.

[13] M. Gregori, A. Pascual-Iserte, and M. Payaró, "Mutual information maximization for a wireless energy harvesting node considering the circuitry power consumption," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2013, pp. 4238–4243.

[14] M. Gregori and M. Payaró, "On the optimal resource allocation for a wireless energy harvesting node considering the circuitry power consumption," *IEEE Trans. Wireless Commun.*, vol. 13, no. 11, pp. 5968–5984, Nov. 2014.

[15] K. Tutuncuoglu and A. Yener, "Sum-rate optimal power policies for energy harvesting transmitters in an interference channel," *J. Commun. Netw.*, vol. 14, no. 2, pp. 151–161, Apr. 2012.

[16] D. K. Shin, W. Choi, and D. I. Kim, "The two-user Gaussian interference channel with energy harvesting transmitters: Energy cooperation and achievable rate region," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4551–4564, Nov. 2015.

[17] M. Gregori and M. Payaró, "Multiuser communications with energy harvesting transmitters," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2014, pp. 5360–5365.

[18] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, "Decomposition by partial linearization: Parallel optimization of multi-agent systems," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 641–656, Feb. 2014.

[19] G. Scutari, F. Facchinei, L. Lampariello, and P. Song, "Parallel and distributed methods for nonconvex optimization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2014, pp. 840–844.

[20] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002.

[21] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 57–73, Feb. 2008.

[22] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[23] G. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed $\ell^0$ norm," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 289–301, Jan. 2009.

[24] X. Zheng, X. Sun, D. Li, and J. Sun, "Successive convex approximations to cardinality-constrained convex programs: A piecewise-linear DC approach," *Comput. Optim. Appl.*, vol. 59, no. 1, pp. 379–397, 2013.

[25] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Old Tappan, NJ, USA: Prentice-Hall, 1989.

[26] A. Beck, A. Ben-Tal, and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *J. Global Optim.*, vol. 47, no. 1, pp. 29–51, May 2010.

[27] C. T. Lawrence and A. L. Tits, "A computationally efficient feasible sequential quadratic programming algorithm," *SIAM J. Optim.*, vol. 11, no. 4, pp. 1092–1118, 2001.

[28] C. A. Sagastizábal and M. V. Solodov, "Parallel variable distribution for constrained optimization," *Comput. Optim. Appl.*, vol. 22, no. 1, pp. 111–131, 2002.

[29] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*. Berlin, Germany: Springer-Verlag, 1998, vol. 317.

[30] C. K. Ho and R. Zhang, "Optimal energy allocation for wireless communications with energy harvesting constraints," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4808–4818, Sep. 2012.

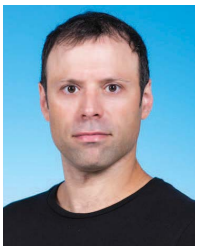
[31] M. Gorlatova, M. Zapas, E. Xu, M. Bahlke, I. Kymissis, and G. Zussman. (2011). *Crawdad Data Set Columbia/Enhants (v. 2011-04-07)*. [Online]. Available: http://crawdad.org/~crawdad/columbia/enhants/

[32] G. Scutari, F. Facchinei, J.-S. Pang, and D. P. Palomar, "Real and complex monotone communication games," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4197–4231, Jul. 2014.


**Maria Gregori** received the B.Sc., M.Sc., and Ph.D. degrees from the Universitat Politècnica de Catalunya (UPC), Barcelona, in 2009, 2011, and 2014, respectively, all in telecommunications. From 2008 to 2009, she was a Visiting Researcher with the Georgia Institute of Technology with a fellowship from the Vodafone Foundation. From 2009 to 2010, she was a Research Assistant with Intel–UPC. From 2010 to 2014, she pursued her Ph.D. studies with the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC). During her Ph.D., studies, she was a Visiting Researcher with the Hong Kong University of Science and Technology from 2013 to 2014. Since 2014, she has been a Researcher at CTTC. Her research interests include energy harvesting for wireless communication, content caching, and smart grid. She received predoctoral grants from the Generalitat de Catalunya and the CTTC fellowship program.

Dr. Gregori served as the Chair of the IEEEWCNC 2016 Workshop on Wireless Powered Communication Networks: From Theory to Industry Challenges. She participated in national and European projects, such as INTENSYV, E-CROPS, GRE3N, P2P-SmarTest, and NEWCOM#. She received the NEWCOM# 2013 Best Young Researcher's Paper Award and was a Designated Exemplary Reviewer of the IEEE WIRELESS COMMUNICATIONS LETTERS in 2013.

**Miquel Payaró** received the Electrical Engineering and Ph.D. degrees from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 2002 and 2007, respectively. He conducted his Ph.D. studies with the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Barcelona. He held a visiting research appointment with the University of New South Wales, Sydney, Australia, from 2005 to 2006. From 2007 to 2008, he was a Post-Doctoral Research Associate with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, under the supervision of Prof. D. P. Palomar. Since 2009, he has been a Research Associate at CTTC, where he led the Engineering Unit from 2011 to 2013. In 2013, he was appointed Head of the Communications Technologies Division. Since 2009, he has led and participated in several research contracts with the industry as well as in research projects from the Seventh Framework Program of the EC, such as BuNGee, BeFEMTO, or, more recently, Newcom#, where he led a Workpackage focused on experimental activities related to radio interfaces for fifth-generation (5G) systems. Since 2015, he has served as the Technical Manager on the H2020/5G-PPP project Flex5Gware, where flexible, reconfigurable HW/SW communication platforms for 5G mobile services and applications are studied and key building blocks are prototyped.



**Daniel P. Palomar** (S'99–M'03–SM'08–F'12) received the Electrical Engineering and Ph.D. degrees from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1998 and 2003, respectively.

He joined the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2006, where he is currently a Professor. Since 2013, he has been a fellow of the Institute for Advance Study at HKUST. He had previously held several research appointments, namely, at King's College London, London, U.K., Stanford University, Stanford, CA, USA, the Telecommunications Technological Center of Catalonia, Barcelona, Spain, the Royal Institute of Technology, Stockholm, Sweden, the University of Rome La Sapienza, Rome, Italy, and Princeton University, Princeton, NJ, USA. His current research interests include applications of convex optimization theory, game theory, and variational inequality theory to financial systems, big data systems, and communication systems.

He is a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING 2016 Special Issue on Financial Signal Processing and Machine Learning for Electronic Trading, and has been an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY and the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was a Guest Editor of the *IEEE Signal Processing Magazine* 2010 Special Issue on Convex Optimization for Signal Processing, the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2008 Special Issue on Game Theory in Communication Systems, and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2007 Special Issue on Optimization of MIMO Transceivers for Realistic Communication Networks.

Dr. Palomar was a recipient of a 2004/06 Fulbright Research Fellowship, the 2004 and 2015 (co-author) Young Author Best Paper Awards by the IEEE Signal Processing Society, the 2015-16 HKUST Excellence Research Award, the 2002/03 Best Ph.D. Prize in Information Technologies and Communications by UPC, the 2002/03 Rosina Ribalta First Prize for the Best Doctoral Thesis in Information Technologies and Communications by the Epson Foundation, and the 2004 Prize for the Best Doctoral Thesis in Advanced Mobile Communications by the Vodafone Foundation and COIT.