Robust Smart-Grid Powered Cooperative Multipoint Systems

Xin Wang, *Senior Member, IEEE*, Yu Zhang, *Student Member, IEEE*,
Georgios B. Giannakis, *Fellow, IEEE*, and Shuyan Hu

Abstract—A framework is introduced to integrate renewable energy sources (RES) and dynamic pricing capabilities of the smart grid into beamforming designs for coordinated multi-point (CoMP) downlink communication systems. To this end, novel models are put forth to account for harvesting, storage of nondispatchable RES, time-varying energy pricing, as well as stochastic wireless channels. Building on these models, robust energy management and transmit-beamforming designs are developed to minimize the worst-case energy cost subject to the worst-case user QoS guarantees for the CoMP downlink. Leveraging pertinent tools, this task is formulated as a convex problem. A Lagrange dual based subgradient iteration is then employed to find the desired optimal energy-management strategy and transmit-beamforming vectors. Numerical results are provided to demonstrate the merits of the proposed robust designs.

Index Terms—Smart grid, renewable energy, downlink beamforming, CoMP systems, robust optimization.

NOMENCLATURE

A. Indices, numbers, and sets

T, t	Number and index of time slots.
I, i	Number and index of distributed BSs.
K, k	Number and index of mobile users.
S, s	Number and index of sub-horizons.
j	Iteration index of the dual subgradient ascent.
$\mathcal{T}$	Set of the total scheduling horizon.
$\mathcal{T}_{i,s}$	Sub-horizon s for BS i.
$\mathcal{K}$	Set of mobile users.
$\mathcal{H}_k^t$	Channel uncertainty set of user k in time slot t.
$\mathcal{E}_i$	Uncertainty set of the energy harvested at BS i.
$\mathcal{E}_i^{\mathrm{p}}$, $\mathcal{E}_i^{\mathrm{e}}$	Polyhedral and ellipsoidal cases of $\mathcal{E}_i$.
ℓ	Iteration index of the Bundle method.

B. Constants

σ_k^2	Variance of channel additive noise.
ϵ_k^t	Radius of the channel uncertainty region for user k in slot t.
γ_k	Target signal-to-interference-plus-noise-ratio (SINR) of user k.
$\underline{E}_i^t$, $\overline{E}_i^t$	Lower and upper bounds of the renewable energy at BS i in slot t.
$E_{i,s}^{\min}$, $E_{i,s}^{\max}$	Lower and upper bounds of the renewable energy at BS i over the sth sub-horizon.
C_i^0	Initial battery energy level of BS i.
C_i^{max}	Battery capacity of BS i.
$P_{b,i}^{\min}$, $P_{b,i}^{\max}$	Minimum and maximum (dis-)charging power at BS i.
ϖ_i	Battery efficiency at BS i.
$P_{c,i}$	Fixed energy consumption at BS i.
ξ	Power amplifier efficiency.
$P_{g,i}^{max}$	Upper limit of the total energy consumption for BS i.
α^t, β^t; ψ^t, ϕ^t	Buying and selling energy prices; and functions thereof.

C. Basic variables

$\mathbf{h}_{ik}^t$	Vector channel from BS i to user k in slot t.
$\mathbf{h}_k^t$	Vector channel from all BSs to user k in slot t.
$\hat{\mathbf{h}}_k^t$	Estimated channel for user k in slot t.
$\mathbf{q}_k^t$	Vector signal transmitted to user k in slot t.
s_k^t	Information-bearing symbol of user k in slot t.
y_k^t	Received signal of user k in slot t.
n_k^t	Channel additive noise of user k in slot t.
$\mathbf{B}_i$	Selection matrix forming the i-th BS's transmit-beamforming vector.
$\boldsymbol{\delta}_k^t$	Channel estimation error of user k in slot t.
$\mathbf{y}_\ell$, ρ_ℓ	Proximal center and weight of the bundle method at iteration ℓ.
η_ℓ	Predicted descent of the objective value for the bundle method at iteration ℓ.
$\mathbf{g}_{i,\ell}$	Subgradient of $\tilde{G}(\mathbf{p}_i)$ at iteration ℓ.
r	Energy buying-to-selling price ratio.
$\hat{E}_i^t$	Expected renewable generation.
$\tilde{E}_i^t$	Realizations of random renewable generation.
κ, U	Scaling variable and uniform random variable relating to $\tilde{E}_i^t$.

D. Decision variables

$\mathbf{w}_k^t$	Transmit beamforming vector from all BSs to user k in slot t.
C_i^t	Stored battery energy at BS i at the beginning of slot t.
$P_{b,i}^t$	Power delivered to or drawn from the batteries of BS i in slot t.
$P_{g,i}^t$	Total energy consumption for BS i in slot t.
$P_{x,i}^t$	Transmit power at BS i in slot t.
E_i^t	Energy harvested at BS i in slot t.
$\mathbf{e}_i$	Vector collecting $\{E_i^t\}_{t=1}^T$.
P_i^t	Auxiliary variable relating $P_{g,i}^t$ and $P_{b,i}^t$.
$\mathbf{p}_i$	Vector collecting $\{P_i^t\}_{t=1}^T$.
$\mathbf{X}_k^t$	Matrix-lifting Beamformer for user k in slot t.
τ_k^t	Auxiliary variable introduced by the S-procedure.
$\mathbf{Z}$	Matrix collecting all primal variables.
λ_i^t	Lagrange multiplier.
$\mathbf{\Lambda}$	Matrix collecting all Lagrange multipliers.

E. Functions

$\mathrm{SINR}_k(\cdot)$	SINR of user k.
$\widetilde{\mathrm{SINR}}_k(\cdot)$	Worst-case SINR of user k.
$G(\cdot,\cdot)$	Worst transaction cost across entire horizon.
$G(\cdot), \tilde{G}(\cdot)$	Modified worst-case transaction cost.
$L(\mathbf{Z},\mathbf{\Lambda})$	Partial Lagrangian function.
$D(\mathbf{\Lambda})$	Dual function.

I. Introduction

To accommodate the explosive demand for wireless services, cellular systems are evolving into what are termed heterogeneous networks (HetNets) consisting of distributed macro/micro/pico base stations (BSs) to cover overlapping areas of different sizes [1]. Close proximity of many HetNet transmitters introduces severe inter-cell interference. For efficient interference management, coordinated multi-point processing (CoMP) has emerged as a promising technique for next-generation cellular networks such as LTE-Advanced [2].

To fully exploit the potential of CoMP at affordable overhead, coordinated beamforming and/or clustered BS cooperation for downlink systems were investigated in [3]–[8]. Multiple BSs cooperate to beamform, and each user's data are only shared among a small number of BSs per cluster, thus greatly reducing the overall backhaul signalling cost.

The rapid development of small cells in HetNets has also driven the need for energy-efficient transmissions. Due to the growing number of BSs, the electricity bill has become a major part of the operational expenditure of cellular operators, and cellular networks contribute a considerable portion of the global "carbon footprint" [9]. These economic and ecological concerns advocate a "green communication" solution, where BSs in cellular networks are powered by the electricity grid [3]–[11]. However, the current grid infrastructure is on the verge of a major paradigm shift, migrating from the traditional electricity grid to the so-termed "smart grid" equipped with a large number of advanced smart meters and state-of-the-art communication and control links. To decrease greenhouse gas emissions, an important feature of future power systems

is integration of renewable energy sources (RES). This leads to high penetration of distributed generators equipped with energy harvesting modules, which can crop energy from the environmental resources (e.g., solar and wind), and possibly trade the harvested energy with the main grid. In addition to distributed generation, distributed storage, and two-way energy trading associated with RES, demand-side management (DSM) including dynamic pricing and demand response, can further improve grid reliability and efficiency. Relying on pertinent tools, optimal energy management and scheduling with RES and/or DSM were proposed in [12]–[15].

To take advantage of the aforementioned smart grid capabilities in next-generation cellular systems, only a few recent works have considered the smart-grid powered CoMP transmissions [16]–[18]. However, [16] only addressed dynamic pricing using a simplified smart grid level game, while [17] and [18] assumed that the energy amounts harvested from RES are precisely available *a priori* (e.g., through forecasting), and the harvested energy cannot be stored at the BSs. In addition, [18] assumed demand (or load) response based on different energy buying/selling prices across BSs without adapting power consumption to *time-varying* energy pricing.

The present paper deals with optimal energy management and transmit-beamforming designs for the smart-grid powered, cluster-based CoMP downlink with the clustering carried by the HetNet's central processor. Each BS has local RES and can perform two-way energy trading with the grid based on *dynamic* buying/selling prices. Different from [16]–[18], we suppose that each BS has a local storage device, which can be charged to store the harvested (and even grid) energy and can be discharged to supply electricity if needed. To account for the stochastic and nondispatchable nature of both RES and wireless channels, we assume that the actual harvested energy amounts and the wireless channel states are *unknown* and *time-varying*, yet lie in some known uncertainty regions. Building on realistic models, we develop robust energy management and transmit-beamforming designs that minimize the worst-case energy cost subject to the worst-case user QoS guarantees for the CoMP downlink. Leveraging a novel formulation accounting for the worst-case transaction cost with two-way energy trading, as well as the S-procedure in robust beamforming designs, we show how to (re-)formulate the said task as a convex problem. Strong duality of the latter is then utilized to develop a Lagrange dual based subgradient solver. It is shown that the resultant algorithm is guaranteed to find the desired optimal energy-management strategy and transmit-beamforming vectors, and could also facilitate distributed implementations among the BSs.

The rest of the paper is organized as follows. The system models are described in Section II. The proposed approach to robust energy management and transmit-beamforming designs for the CoMP downlink is developed in Section III. Numerical results are provided in Section IV to demonstrate the proposed scheme. We conclude the paper in Section V.

Notation: Boldface fonts denote vectors or matrices; $\mathbb{C}^{K \times M}$ is the K-by-M dimensional complex space; $(\cdot)'$ denotes transpose, and $(\cdot)^H$ conjugate transpose; $\mathrm{tr}(\mathbf{A})$ and $\mathrm{rank}(\mathbf{A})$ the trace and rank operators for matrix $\mathbf{A}$, respectively;

$\mathrm{diag}(a_1, \ldots, a_M)$ denotes a diagonal matrix with $a_1, \ldots, a_M$ on the diagonal; $|\cdot|$ represents the magnitude of a complex scalar; $\mathbf{A} \succeq \mathbf{0}$ means that a square matrix $\mathbf{A}$ is positive semi-definite.

II. System Models

Consider a smart-grid powered cluster-based CoMP downlink. A set $\mathcal{I} := \{1, \ldots, I\}$ of distributed BSs (e.g., macro/micro/pico BSs) provides service to a set $\mathcal{K} := \{1, \ldots, K\}$ of mobile users; see Fig. 1. Each BS has $M \geq 1$ transmit antennas and each user has a single receive antenna. Each BS is equipped with one or more energy harvesting devices (solar panels and/or wind turbines), and is also connected to the power grid with a two-way energy trading facility. Different from [16]–[18], each BS has an energy storage device (possibly consisting of several large-capacity batteries) so that it does not have to consume or sell all the harvested energy on the spot, but can save it for later use.

For each CoMP cluster, there is a low-latency backhaul network connecting the set of BSs to a central controller [5], which coordinates energy trading as well as cooperative communication. The central entity can collect both the communication data (transmit messages) from each of the BSs through the cellular backhaul links, as well as the energy information (energy buying/selling prices) from these BSs via smart meters installed at BSs, and the grid-deployed communication/control links connecting them.

Assume slot-based transmissions from the BSs to the users. While the actual harvested energy amounts and wireless channels cannot be accurately predicted, uncertainty regions for the wireless channels and renewable energy arrivals can be obtained, based on historical measurements and/or forecasting techniques. The slot duration is selected equal to the minimum time interval between the changes of the (channel or energy) uncertainty regions. Consider a finite scheduling horizon consisting of T slots, indexed by the set $\mathcal{T} := \{1, \ldots, T\}$. For convenience, the slot duration is normalized to unity unless otherwise specified; thus, the terms "energy" and "power" will hereafter be used interchangeably throughout the paper.

A. Downlink CoMP Transmission Model

Per slot t, let $\mathbf{h}_{ik}^t \in \mathbb{C}^{M \times 1}$ denote the vector channel from BS i to user k, $\forall i \in \mathcal{I}$, $\forall k \in \mathcal{K}$; and let $\mathbf{h}_k^t := [\mathbf{h}_{1k}^t{}', \ldots, \mathbf{h}_{Ik}^t{}']'$ collect the channel vectors from all BSs to user k. With linear transmit beamforming performed across BSs, the vector signal transmitted to user k is

$$\mathbf{q}_k^t = \mathbf{w}_k^t s_k^t, \quad \forall k$$

where s_k^t denotes the information-bearing symbol, and $\mathbf{w}_k^t \in \mathbb{C}^{MI \times 1}$ denotes the beamforming vector across the BSs for user k. For convenience, s_k^t is assumed to be a complex random variable with zero mean and unit variance. The received vector at user k is therefore

$$y_k^t = \mathbf{h}_k^t{}^H \mathbf{q}_k^t + \sum_{l \neq k} \mathbf{h}_k^t{}^H \mathbf{q}_l^t + n_k^t \quad (1)$$

where $\mathbf{h}_k^t{}^H \mathbf{q}_k^t$ is the desired signal for user k, $\sum_{l \neq k} \mathbf{h}_k^t{}^H \mathbf{q}_l^t$ is the inter-user interference from the same cluster, and n_k^t denotes additive noise, which may also include the downlink interference from other BSs outside this cluster. It is assumed that n_k^t is a circularly symmetric complex Gaussian (CSCG) random variable with zero mean and variance σ_k^2.

The signal-to-interference-plus-noise-ratio (SINR) at user k can be expressed as

$$\mathrm{SINR}_k(\{\mathbf{w}_k^t\}) = \frac{|\mathbf{h}_k^t{}^H \mathbf{w}_k^t|^2}{\sum_{l \neq k}(|\mathbf{h}_k^t{}^H \mathbf{w}_l^t|^2) + \sigma_k^2} . \quad (2)$$

The transmit power at each BS i clearly is given by

$$P_{x,i}^t = \sum_{k \in \mathcal{K}} \mathbf{w}_k^t{}^H \mathbf{B}_i \mathbf{w}_k^t \quad (3)$$

where matrix $\mathbf{B}_i := \mathrm{diag}(\underbrace{0, \ldots, 0}_{(i-1)M}, \underbrace{1, \ldots, 1}_{M}, \underbrace{0, \ldots, 0}_{(I-i)M})$ selects the corresponding rows out of $\{\mathbf{w}_k^t\}_{k \in \mathcal{K}}$ to form the i-th BS's transmit-beamforming vector.

The channel state information $\mathbf{h}_k^t$ is seldom precisely available *a priori* in practice. Relying on past channel measurements and/or reliable channel predictions, we adopt the following additive error model: $\mathbf{h}_k^t = \hat{\mathbf{h}}_k^t + \boldsymbol{\delta}_k^t$, where $\hat{\mathbf{h}}_k^t$ is the predicted channel. The uncertainty of this estimate is bounded by a spherical region [19]:

$$\mathcal{H}_k^t := \left\{ \hat{\mathbf{h}}_k^t + \boldsymbol{\delta}_k^t \mid \left\| \boldsymbol{\delta}_k^t \right\| \leq \epsilon_k^t \right\}, \quad \forall k, t \quad (4)$$

where $\epsilon_k^t > 0$ specifies the radius of $\mathcal{H}_k^t$. This leads to the worst-case SINR per user k as [cf. (2)]

$$\widetilde{\mathrm{SINR}}_k(\{\mathbf{w}_k^t\}) := \min_{\mathbf{h}_k^t \in \mathcal{H}_k^t} \frac{|\mathbf{h}_k^t{}^H \mathbf{w}_k^t|^2}{\sum_{l \neq k} |\mathbf{h}_k^t{}^H \mathbf{w}_l^t|^2 + \sigma_k^2} . \quad (5)$$

To guarantee QoS per slot, it is required that

$$\widetilde{\mathrm{SINR}}_k(\{\mathbf{w}_k^t\}) \geq \gamma_k, \quad \forall k \quad (6)$$

where γ_k denotes the target SINR value per user k.

B. Energy Harvesting Model

Let E_i^t denote the energy harvested during the last slot that is available at the beginning of slot t at each BS $i \in \mathcal{I}$; and let $\mathbf{e}_i := [E_i^1, \ldots, E_i^T]'$. Due to the unpredictable and intermittent nature of RES, $\mathbf{e}_i$ is unknown a priori. In general, uncertain quantities can be modeled by postulating either an underlying probability distribution or an uncertainty region. Probability distributions (possibly mixed discrete/continuous) of the RES generation are seldom available in practice. Although (non-)parametric approaches can be used to learn the distributions, the processes can be very complicated due to the spatio-temporal correlations incurred by various meteorological factors. On the other hand, the approach of postulating an uncertainty region provides the decision maker range forecasts instead of point forecasts, which are essentially distribution-free. Suppose that $\mathbf{e}_i$ lies in an uncertainty set $\mathcal{E}_i$, which can be obtained from historical measurements and/or fine forecast techniques. From the perspective of computational tractability, two practical paradigms for $\mathcal{E}_i$ are considered here.
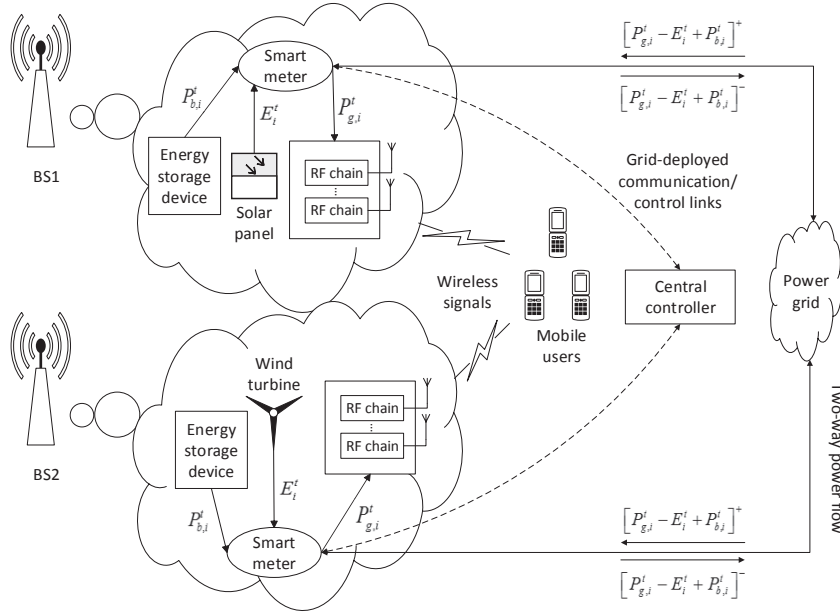
Fig. 1. A two-BS CoMP system powered by smart grids, where BSs with local renewable energy harvesting and storage devices implement two-way energy trading with the main grid.

i) The first model amounts to a polyhedral set [12]:

$$\mathcal{E}_i^{\mathrm{p}} := \left\{ \mathbf{e}_i \mid \underline{E}_i^t \le E_i^t \le \overline{E}_i^t, \right.$$
$$\left. E_{i,s}^{\min} \le \sum_{t \in \mathcal{T}_{i,s}} E_i^t \le E_{i,s}^{\max}, \ \mathcal{T} = \bigcup_{s=1}^S \mathcal{T}_{i,s} \right\} \quad (7)$$

where $\underline{E}_i^t$ ($\overline{E}_i^t$) denotes a lower (upper) bound on E_i^t; time horizon $\mathcal{T}$ is partitioned into consecutive but non-overlapping sub-horizons $\mathcal{T}_{i,s}$, $s = 1, \dots, S$. Each sub-horizon can consist of multiple time slots, and the total energy harvested by BS i over the sth sub-horizon is bounded by $E_{i,s}^{\min}$ and $E_{i,s}^{\max}$.

ii) The second model relies on an ellipsoidal uncertainty set (see e.g. [20])

$$\mathcal{E}_i^{\mathrm{e}} := \left\{ \mathbf{e}_i = \hat{\mathbf{e}}_i + \boldsymbol{\varsigma}_i \mid \boldsymbol{\varsigma}_i' \boldsymbol{\Sigma}^{-1} \boldsymbol{\varsigma}_i \le 1 \right\} \quad (8)$$

where $\hat{\mathbf{e}}_i := [\hat{E}_i^1, \dots, \hat{E}_i^T]'$ denotes the nominal energy harvested at BS i, which can be the forecasted energy, or simply its expected value. Vector $\boldsymbol{\varsigma}_i$ is the corresponding error in forecasting. The known matrix $\boldsymbol{\Sigma} \succ \mathbf{0}$ quantifies the shape of the ellipsoid $\mathcal{E}_i^{\mathrm{e}}$, and hence determines the quality of the forecast.

Remark 1: *(Spatio-temporally correlated energy harvesting models)*. The aforementioned two practical models capture RES uncertainty across the scheduling (sub-)horizons per BS. The parameters required for constructing the sets $\{\mathcal{E}_i^{\mathrm{p}}, \mathcal{E}_i^{\mathrm{e}}\}$ can be obtained offline via statistical learning techniques using historical data. In general, green energy harvested at different BSs can be spatially correlated if some BSs are geographically close. In this case, joint spatio-temporal uncertainty models can be postulated whenever the underlying correlations are

known *a priori*; see details in [12]. In general, a refined uncertainty model quantifying the actual harvested energy in a smaller region with a higher confidence level can be less conservative in the robust optimization formulation. However, the complexity of solving the resulting optimization problems directly depends on the choice of the uncertainty set.

Note that the coherence time of RES arrivals can be much longer than that of wireless channels in practice [17], [21]. Yet, coherence times corresponding to the uncertainty regions of wireless channels can be much larger than those of the channel itself. In Section II-A, we implicitly assume that channel $\mathbf{h}_k^t$ remains unchanged per slot. However, $\widetilde{\mathrm{SINR}}_k$ in (5) can be easily redefined as the worst-case SINR over multiple channel coherence times with the same uncertainty region (and possibly different channel realizations). This way the issue of different time scales becomes less critical. In addition, our models in fact accommodate cases where uncertainty regions for RES arrivals remain unchanged over multiple time slots. The proposed approach presented next readily applies to obtain the robust ahead-of-time schedule in this setup.

C. Energy Storage Model

Let C_i^0 denote the initial energy, and C_i^t the amount of stored energy in the batteries of the i-th BS at the beginning of time slot t. With $C_i^{\max}$ bounding the capacity of batteries, it is clear that $0 \le C_i^t \le C_i^{\max}$, $\forall i \in \mathcal{I}$. Let $P_{b,i}^t$ denote the power delivered to or drawn from the batteries at slot t, which amounts to either charging ($P_{b,i}^t > 0$) or discharging ($P_{b,i}^t < 0$). Hence, the stored energy obeys the dynamic equation

$$C_i^t = C_i^{t-1} + P_{b,i}^t, \quad t \in \mathcal{T}, \ i \in \mathcal{I}. \quad (9)$$

The amount of power (dis-)charged is also bounded by

$$P_{b,i}^{\min} \leq P_{b,i}^t \leq P_{b,i}^{\max}$$
$$-\varpi_i C_i^{t-1} \leq P_{b,i}^t \tag{10}$$

where $P_{b,i}^{\min} < 0$ and $P_{b,i}^{\max} > 0$, while $\varpi_i \in (0,1]$ is the battery efficiency at BS i. The constraint $-\varpi_i C_i^{t-1} \leq P_{b,i}^t$ means that at most a fraction ϖ_i of the stored energy C_i^{t-1} is available for discharge.

D. Energy Cost Model

For the i-th BS per slot t, the total energy consumption $P_{g,i}^t$ includes the transmission-related power $P_{x,i}^t$, and the rest that is due to other components such as air conditioning, data processor, and circuits, which can be generally modeled as a constant power, $P_{c,i} > 0$ [10], [18]; namely,

$$P_{g,i}^t = P_{c,i} + P_{x,i}^t/\xi = P_{c,i} + \sum_{k \in \mathcal{K}} \mathbf{w}_k^{t\,H} \mathbf{B}_i \mathbf{w}_k^t/\xi$$

where $\xi > 0$ denotes the power amplifier efficiency. For notational convenience, we absorb ξ into $\mathbf{B}_i$ by redefining $\mathbf{B}_i := \mathbf{B}_i/\xi$; and further assume that $P_{g,i}^t$ is bounded by $P_{g,i}^{\max}$.

In addition to the harvested RES, the power grid can also supply the needed $P_{g,i}^t$ per BS i. With a two-way energy trading facility, the BS can also sell its surplus energy to the grid at a fair price in order to reduce operational costs. Given the required energy $P_{g,i}^t$, the harvested energy E_i^t, and the battery charging energy $P_{b,i}^t$, the shortage energy that needs to be purchased from the grid for BS i is clearly $[P_{g,i}^t - E_i^t + P_{b,i}^t]^+$; or, the surplus energy (when the harvested energy is abundant) that can be sold to the grid is $[P_{g,i}^t - E_i^t + P_{b,i}^t]^-$, where $[a]^+ := \max\{a,0\}$, and $[a]^- := \max\{-a,0\}$. Note that both the shortage energy and surplus energy are non-negative, and we have at most one of them be positive at any time t for BS i.

Suppose that the energy can be purchased from the grid at price α^t, while the energy is sold to the grid at price β^t per slot t. Notwithstanding, we shall always set $\alpha^t > \beta^t, \forall$, to avoid meaningless buy-and-sell activities of the BSs for profit. Assuming that the prices α^t and β^t are known *a priori*, the *worst-case transaction cost* for BS i for the whole scheduling horizon is therefore given by

$$G(\{P_{g,i}^t\}, \{P_{b,i}^t\}) := \max_{\mathbf{e}_i \in \mathcal{E}_i} \sum_{t=1}^{T} \Big(\alpha^t [P_{g,i}^t - E_i^t + P_{b,i}^t]^+$$
$$-\beta^t [P_{g,i}^t - E_i^t + P_{b,i}^t]^- \Big). \tag{11}$$

III. ENERGY MANAGEMENT FOR CoMP BEAMFORMING

Based on the models of Section II, we consider here robust energy management for transmit beamforming in a CoMP cluster. Knowing only the uncertainty regions of the wireless channels and renewable energy arrivals, the central controller per cluster performs an (e.g. hour-) ahead-of-time schedule to optimize cooperative transmit beamforming vectors $\{\mathbf{w}_k^t\}$ and battery charging energy $\{P_{b,i}^t\}$, in order to minimize the worst-case total cost $\sum_{i \in \mathcal{I}} G(\{P_{g,i}^t\}, \{P_{b,i}^t\})$, while satisfying

the QoS guarantees $\widetilde{\mathrm{SINR}}_k(\{\mathbf{w}_k^t\}) \geq \gamma_k, \forall k$, over the scheduling horizon $\mathcal{T}$. For convenience, we introduce the auxiliary variables $P_i^t := P_{g,i}^t + P_{b,i}^t$, and formulate the problem as

$$\underset{\{\mathbf{w}_k^t, P_{b,i}^t, C_i^t, P_i^t\}}{\text{minimize}} \sum_{i \in \mathcal{I}} G(\{P_i^t\}) \tag{12a}$$

suject to:

$$P_i^t = P_{c,i} + \sum_{k \in \mathcal{K}} \mathbf{w}_k^{t\,H} \mathbf{B}_i \mathbf{w}_k^t + P_{b,i}^t, \quad \forall i \in \mathcal{I}, \ t \in \mathcal{T} \tag{12b}$$

$$0 \leq P_{c,i} + \sum_{k \in \mathcal{K}} \mathbf{w}_k^{t\,H} \mathbf{B}_i \mathbf{w}_k^t \leq P_{g,i}^{\max}, \quad \forall i \in \mathcal{I}, \ t \in \mathcal{T} \tag{12c}$$

$$C_i^t = C_i^{t-1} + P_{b,i}^t, \quad \forall i \in \mathcal{I}, \ t \in \mathcal{T} \tag{12d}$$

$$0 \leq C_i^t \leq C_i^{\max}, \quad \forall i \in \mathcal{I}, \ t \in \mathcal{T} \tag{12e}$$

$$P_{b,i}^{\min} \leq P_{b,i}^t \leq P_{b,i}^{\max}, \quad \forall i \in \mathcal{I}, \ t \in \mathcal{T} \tag{12f}$$

$$-\varpi_i C_i^{t-1} \leq P_{b,i}^t, \quad \forall i \in \mathcal{I}, \ t \in \mathcal{T} \tag{12g}$$

$$\widetilde{\mathrm{SINR}}_k(\{\mathbf{w}_k^t\}) \geq \gamma_k, \quad \forall k \in \mathcal{K}, \ t \in \mathcal{T}. \tag{12h}$$

Consider for simplicity that (12) is feasible. The problem can in fact become infeasible if the SINR thresholds γ_k are high and the wireless channel qualities are not good enough. In this case, the subsequent problems (22a) are always infeasible. Recall that the proposed scheme is proposed to determine the ahead-of-time beamformer and energy schedule (offline). Such an infeasibility, once detected, can naturally lead to an admission control policy, i.e., to a criterion for dropping users or SINR requirements that render the problem infeasible [22].

Solving (12) can provide a robust solution for the smart-grid powered CoMP downlink with worst-case performance guarantees. It is worth mentioning here that thanks to the worst-case cost $G(\{P_i^t\})$, randomness introduced due to the wireless fading propagation and also due to the RES uncertainty has been eliminated; thus (12) contains only deterministic variables. Because of (12b), (12c), and (12h), the problem is nonconvex, which motivates the reformulation pursued next.

A. Convex Reformulation

First, let us consider convexity issues of the objective function $G(\{P_i^t\})$. Define $\psi^t := (\alpha^t - \beta^t)/2$ and $\phi^t := (\alpha^t + \beta^t)/2$, and then rewrite:

$$G(\{P_i^t\}) = \max_{\mathbf{e}_i \in \mathcal{E}_i} \sum_{t=1}^{T} \big(\psi^t |P_i^t - E_i^t| + \phi^t (P_i^t - E_i^t) \big) .$$

Since $\alpha^t > \beta^t \geq 0$, we have $\phi^t > \psi^t > 0$. It is then clear that $\psi^t |P_i^t - E_i^t| + \phi^t (P_i^t - E_i^t)$ is a convex function of P_i^t for any given E_i^t. As a pointwise maximization of these convex functions, $G(\{P_i^t\})$ is also a convex function of $\{P_i^t\}$ (even when the set $\mathcal{E}_i$ is non-convex) [23].

Except for (12b), (12c), and (12h), all other constraints are convex. We next rely on the popular semidefinite program (SDP) relaxation technique to convexify (12h). By the definitions of $\mathcal{H}_k^t$ and $\widetilde{\mathrm{SINR}}_k(\{\mathbf{w}_k^t\})$, the constraint $\widetilde{\mathrm{SINR}}_k(\{\mathbf{w}_k^t\}) \geq \gamma_k$ can be rewritten as:

$$F_k(\boldsymbol{\delta}_k^t) \geq 0 \text{ for all } \boldsymbol{\delta}_k^t \text{ such that } \boldsymbol{\delta}_k^{t\,H} \boldsymbol{\delta}_k^t \leq (\epsilon_k^t)^2 \tag{13}$$

where

$$F_k(\boldsymbol{\delta}_k^t) := (\hat{\mathbf{h}}_k^t + \boldsymbol{\delta}_k^t)^H \Big(\frac{\mathbf{w}_k^t \mathbf{w}_k^{tH}}{\gamma_k^t} - \sum_{l \neq k} \mathbf{w}_l^t \mathbf{w}_l^{tH} \Big)(\hat{\mathbf{h}}_k^t + \boldsymbol{\delta}_k^t) - \sigma_k^2.$$

Using (13) and upon applying the well-known S-procedure in robust optimization [24], the original problem (12) can be reformulated as an SDP with rank constraints.

(S-procedure) Let $\mathbf{A}, \mathbf{B}$ be $n \times n$ Hermitian matrices, $\mathbf{c} \in \mathbb{C}^n$ and $d \in \mathbb{R}$ for which the interior-point condition holds; that is, there exists an $\bar{\mathbf{x}}$ such that $\bar{\mathbf{x}}^H \mathbf{B} \bar{\mathbf{x}} < 1$; and the following are equivalent:

(i) $\mathbf{x}^H \mathbf{A} \mathbf{x} + \mathbf{c}^H \mathbf{x} + \mathbf{x}^H \mathbf{c} + d \geq 0$ for all $\mathbf{x} \in \mathbb{C}^n$ such that $\mathbf{x}^H \mathbf{B} \mathbf{x} \leq 1$;

(ii) There exists a $\nu \geq 0$ such that

$$\begin{pmatrix} \mathbf{A} + \nu \mathbf{B} & \mathbf{c} \\ \mathbf{c}^H & d - \nu \end{pmatrix} \succeq \mathbf{0} \ .$$

For $\mathbf{X}_k^t := \mathbf{w}_k^t \mathbf{w}_k^{tH} \in \mathbb{C}^{MI \times MI}$, it clearly holds that $\mathbf{X}_k^t \succeq \mathbf{0}$ and $\mathrm{rank}(\mathbf{X}_k^t) = 1$. Using the S-procedure, (13) can be transformed to

$$\boldsymbol{\Gamma}_k^t := \begin{pmatrix} \mathbf{Y}_k^t + \tau_k^t \mathbf{I} & \mathbf{Y}_k^t \hat{\mathbf{h}}_k^t \\ \hat{\mathbf{h}}_k^{tH} \mathbf{Y}_k^{tH} & \hat{\mathbf{h}}_k^{tH} \mathbf{Y}_k^t \hat{\mathbf{h}}_k^t - \sigma_k^2 - \tau_k^t (\epsilon_k^t)^2 \end{pmatrix} \succeq \mathbf{0} \quad (14)$$

where $\tau_k^t \geq 0$ and

$$\mathbf{Y}_k^t := \frac{1}{\gamma_k} \mathbf{X}_k^t - \sum_{l \neq k} \mathbf{X}_l^t. \tag{15}$$

Introducing auxiliary variables τ_k^t and dropping the rank constraints $\mathrm{rank}(\mathbf{X}_k^t) = 1$, $\forall k, t$, we can then relax (12) to:

$$\min_{\{\mathbf{X}_k^t, \tau_k^t, P_i^t, P_{b,i}^t, C_i^t\}} \sum_{i \in \mathcal{I}} G(\{P_i^t\}) \tag{16a}$$

$$\text{s. t. } (12d)\text{–}(12g)$$

$$P_i^t = P_{c,i} + \sum_{k \in \mathcal{K}} \mathrm{tr}(\mathbf{B}_i \mathbf{X}_k^t) + P_{b,i}^t, \ \forall i, t \tag{16b}$$

$$0 \leq \sum_{k \in \mathcal{K}} \mathrm{tr}(\mathbf{B}_i \mathbf{X}_k^t) \leq P_{g,i}^{\max} - P_{c,i}, \ \forall i, t \tag{16c}$$

$$\boldsymbol{\Gamma}_k^t \succeq \mathbf{0}, \ \tau_k^t \geq 0, \ \mathbf{X}_k^t \succeq \mathbf{0}, \forall k, t. \tag{16d}$$

In addition to the linear constraints (12d)–(12g), the quadratic power constraints (12b) and (12c) have now become linear, and SINR constraints (12h) become a set of convex SDP constraints in (16). Since the objective function is convex, problem (16) is a convex program that can be tackled by a centralized solver; e.g., using the projected subgradient descent approach. However, the feasible set (16b)–(16d) is the intersection of a semi-definite cone and a polytope, for which the iterative projection is complicated. To reduce computational complexity and enhance resilience to failures, we next develop an efficient algorithm to solve (16) in a distributed fashion coordinated by different agents.

B. Lagrange Dual Approach

Since (16) is convex, a Lagrange dual approach can be developed to efficiently find its solution. Let λ_i^t, $\forall i, t$ denote the Lagrange multipliers associated with the constraints (16b).

With the convenient notation $\mathbf{Z} := \{\mathbf{X}_k^t, \tau_k^t, P_{b,i}^t, C_i^t, P_i^t\}$ and $\boldsymbol{\Lambda} := \{\lambda_i^t\}$, the partial Lagrangian function of (16) is

$$L(\mathbf{Z}, \boldsymbol{\Lambda}) :=$$

$$\sum_{i \in \mathcal{I}} \Big[G(\{P_i^t\}) + \sum_{t=1}^{T} \lambda_i^t \Big(P_{c,i} + \sum_{k \in \mathcal{K}} \mathrm{tr}(\mathbf{B}_i \mathbf{X}_k^t) + P_{b,i}^t - P_i^t \Big) \Big]. \tag{17}$$

The Lagrange dual function is then given by

$$D(\boldsymbol{\Lambda}) := \min_{\mathbf{Z}} \ L(\mathbf{Z}, \boldsymbol{\Lambda})$$
$$\text{s. t. } (12d)-(12g), (16c)-(16d) \tag{18}$$

and the dual problem of (16) is:

$$\max_{\boldsymbol{\Lambda}} \ D(\boldsymbol{\Lambda}). \tag{19}$$

Subgradient iterations: Let j denote the iteration index. To obtain the optimal solution $\boldsymbol{\Lambda}^*$ to the dual problem (19), we resort to the dual subgradient ascent method, which amounts to the following update

$$\lambda_i^t(j+1) = \lambda_i^t(j) + \mu(j) g_{\lambda_i^t}(j), \ \forall i, t \tag{20}$$

where $\mu(j) > 0$ is an appropriate stepsize. The subgradient $\mathbf{g}(j) := [g_{\lambda_i^t}(j) \ \forall i, t]$ can then be expressed as

$$g_{\lambda_i^t}(j) = P_{c,i} + \sum_{k \in \mathcal{K}} \mathrm{tr}(\mathbf{B}_i \mathbf{X}_k^t(j)) + P_{b,i}^t(j) - P_i^t(j), \ \forall i, t \tag{21}$$

where $\mathbf{X}_k^t(j)$, $P_{b,i}^t(j)$, and $P_i^t(j)$ are given by

$$\{\mathbf{X}_k^t(j)\}_{k=1}^K \in \underset{\substack{\{\mathbf{X}_k^t, \boldsymbol{\Gamma}_k^t, \tau_k^t\} \\ \text{s. t. } (16c)-(16d)}}{\arg\min} \sum_{i \in \mathcal{I}, k \in \mathcal{K}} \lambda_i^t(j) \mathrm{tr}(\mathbf{B}_i \mathbf{X}_k^t) \tag{22a}$$

$$\{P_{b,i}^t(j)\}_{t=1}^T \in \underset{\substack{\{P_{b,i}^t, C_i^t\} \\ \text{s. t. } (12d)-(12g)}}{\arg\min} \sum_{t=1}^{T} \lambda_i^t(j) P_{b,i}^t \tag{22b}$$

$$\{P_i^t(j)\}_{t=1}^T \in \underset{\{P_i^t\}}{\arg\min} \Big\{ G(\{P_i^t\}) - \sum_{t=1}^{T} \lambda_i^t(j) P_i^t \Big\}. \tag{22c}$$

Solving the subproblems: Subproblems (22a) are standard SDPs per $t \in \mathcal{T}$; hence, $\{\mathbf{X}_k^t(j)\}_{k=1}^K$ for all t can be efficiently solved by general interior-point methods [23]. The subproblems (22b) are simple linear programs (LPs) over $\{P_{b,i}^t, C_i^t\}_{t=1}^T$, which can be independently obtained for each $i \in \mathcal{I}$ using off-the-shelf LP solvers.

Due to the convexity of $G(\{P_i^t\})$, the subproblems (22c) are convex per $i \in \mathcal{I}$. Yet, because of the non-differentiability of $G(\{P_i^t\})$ due to the absolute value operator and the maximization over $\mathbf{e}_i \in \mathcal{E}_i$, the problem is challenging to be handled by existing general solvers. For this reason, we resort to the proximal bundle method to obtain $\{P_i^t(j)\}_{t=1}^T$. Upon defining $\tilde{G}(\mathbf{p}_i) := G(\mathbf{p}_i) - \sum_{t=1}^{T} \lambda_i^t(j) P_i^t$, where $\mathbf{p}_i := [P_i^1, \ldots, P_i^T]'$, the partial subgradient of $\tilde{G}(\mathbf{p}_i)$ with respect to P_i^t can be obtained as

$$\frac{\partial \tilde{G}(\mathbf{p}_i)}{\partial P_i^t} = \begin{cases} \alpha^t - \lambda_i^t(j), & \text{if } P_i^t \geq E_i^{t*} \\ \beta^t - \lambda_i^t(j), & \text{if } P_i^t < E_i^{t*} \end{cases} \tag{23}$$

where $\mathbf{e}_i^* := [E_i^{1*}, \ldots, E_i^{T*}]'$ for the given $\mathbf{p}_i$ is obtained as

$$\mathbf{e}_i^* \in \arg \max_{\mathbf{e}_i \in \mathcal{E}_i} \sum_{t=1}^{T} \left(\psi^t |P_i^t - E_i^t| + \phi^t (P_i^t - E_i^t) \right). \quad (24)$$

It can be readily checked that the objective function in (24) is convex in $\mathbf{e}_i$ under the condition $\alpha^t > \beta^t$, $\forall t \in \mathcal{T}$. Hence, the globally optimal solution is attainable at the extreme points of $\mathcal{E}_i$. Leveraging the special structure of $\mathcal{E}_i$, we utilize an efficient vertex enumerating algorithm to obtain $\mathbf{e}_i^*$ directly, as detailed next.

C. Proximal Bundle Method

Given the partial subgradient in (23), nonsmooth convex optimization algorithms can be employed to solve the subproblem (22c). A state-of-the-art bundle method [23, Ch. 6], [25] will be developed here with guaranteed convergence to the optimal $\{P_i^t(j)\}_{t=1}^{T}$; see also [12].

Similar to cutting plane methods, the idea of bundle method is to approximate the epigraph of an objective by the intersection of a number of halfspaces. The latter are generated through the supporting hyperplanes, referred to as cuts, by using the subgradients. Specifically, letting ℓ denote the iteration index of the bundle method, the iterate $\mathbf{p}_{i,\ell+1}$ is obtained by minimizing a polyhedral (piecewise linear) approximation of $\tilde{G}(\mathbf{p}_i)$ with a quadratic proximal regularizer

$$\mathbf{p}_{i,\ell+1} := \arg \min_{\mathbf{p}_i \in \mathbb{R}^T} \left\{ \hat{G}_\ell(\mathbf{p}_i) + \frac{\rho_\ell}{2} \|\mathbf{p}_i - \mathbf{y}_\ell\|^2 \right\} \quad (25)$$

where $\hat{G}_\ell(\mathbf{p}_i) := \max\{\tilde{G}(\mathbf{p}_{i,0}) + \mathbf{g}_0'(\mathbf{p}_i - \mathbf{p}_{i,0}), \ldots, \tilde{G}(\mathbf{p}_{i,\ell}) + \mathbf{g}_\ell'(\mathbf{p}_i - \mathbf{p}_{i,\ell})\}$; $\mathbf{g}_{i,\ell}$ is the subgradient of $\tilde{G}(\mathbf{p}_i)$ evaluated at the point $\mathbf{p} = \mathbf{p}_{i,\ell}$ [cf. (23)]; and the proximity weight ρ_ℓ controls the stability of iterates.

Different from the proximal cutting plane method (CPM), the bundle method updates its proximal center $\mathbf{y}_\ell$ according to a descent query

$$\mathbf{y}_{\ell+1} = \begin{cases} \mathbf{p}_{i,\ell+1}, & \text{if } \tilde{G}(\mathbf{y}_\ell) - \tilde{G}(\mathbf{p}_{i,\ell+1}) \geq \theta \eta_\ell \\ \mathbf{y}_\ell, & \text{if } \tilde{G}(\mathbf{y}_\ell) - \tilde{G}(\mathbf{p}_{i,\ell+1}) < \theta \eta_\ell \end{cases} \quad (26)$$

where $\theta \in (0, 1)$ is a pre-selected constant, and $\eta_\ell := \tilde{G}(\mathbf{y}_\ell) - \left(\hat{G}_\ell(\mathbf{p}_{\ell+1}) + \frac{\rho_\ell}{2} \|\mathbf{p}_{\ell+1} - \mathbf{y}_\ell\|^2 \right)$ is the predicted descent of the objective in (25). Essentially, if the actual descent amount $\tilde{G}(\mathbf{y}_\ell) - \tilde{G}(\mathbf{p}_{\ell+1})$ is no less than a θ fraction of the predicted counterpart η_ℓ, then the iterate takes a "serious" step updating its proximal center $\mathbf{y}_{\ell+1}$ to the latest point $\mathbf{p}_{\ell+1}$; otherwise it is just a "null" step with the center unchanged. The intelligent query (26) enables the bundle method to find "good" proximal centers along the iterates, and hence it converges faster than the proximal CPM. In addition, depending on whether a serious or a null step is taken, the proximity weight ρ_ℓ can be updated accordingly to further accelerate convergence [26]; that is,

$$\rho_{\ell+1} = \begin{cases} \max(\rho_\ell/10, \rho_{\min}), & \text{if } \tilde{G}(\mathbf{y}_\ell) - \tilde{G}(\mathbf{p}_{i,\ell+1}) \geq \theta \eta_\ell \\ \min(10\rho_\ell, \rho_{\max}), & \text{if } \tilde{G}(\mathbf{y}_\ell) - \tilde{G}(\mathbf{p}_{i,\ell+1}) < \theta \eta_\ell. \end{cases}$$

The algorithm terminates if $\mathbf{y}_\ell = \mathbf{p}_{i,\ell+1}$, while finite termination is achievable if both the objective and the feasible set are polyhedral [23, Sec. 6.5.3].

Now, to complete the proximal bundle method for solving (22c), we only need solve problem (25). Using an auxiliary variable u, (25) can be re-written as

$$\min_{\mathbf{p}_i, u} \quad u + \frac{\rho_\ell}{2} \|\mathbf{p}_i - \mathbf{y}_\ell\|^2 \quad (27a)$$

$$\text{s.t.} \quad \tilde{G}(\mathbf{p}_{i,n}) + \mathbf{g}_{i,n}'(\mathbf{p}_i - \mathbf{p}_{i,n}) \leq u, \ n = 0, 1, \ldots, \ell. \quad (27b)$$

Introducing multipliers $\boldsymbol{\xi} \in \mathbb{R}_+^{\ell+1}$ and letting $\mathbf{1}$ denote the all-ones vector, the Lagrangian function is given as

$$\mathcal{L}(u, \mathbf{p}_i, \boldsymbol{\xi}) = (1 - \mathbf{1}'\boldsymbol{\xi})u + \frac{\rho_\ell}{2} \|\mathbf{p}_i - \mathbf{y}_\ell\|^2$$

$$+ \sum_{n=0}^{\ell} \xi_n \left(\tilde{G}(\mathbf{p}_{i,n}) + \mathbf{g}_{i,n}'(\mathbf{p}_i - \mathbf{p}_{i,n}) \right). \quad (28)$$

The optimality condition $\nabla_{\mathbf{p}_i} \mathcal{L}(u, \mathbf{p}_i, \boldsymbol{\xi}) = \mathbf{0}$ yields

$$\mathbf{p}_i^* = \mathbf{y}_\ell - \frac{1}{\rho_\ell} \sum_{n=0}^{\ell} \xi_n \mathbf{g}_{i,n}. \quad (29)$$

Substituting (29) into (28), the dual of (27) is

$$\max_{\boldsymbol{\xi}} \ -\frac{1}{2\rho_\ell} \left\| \sum_{n=0}^{\ell} \xi_n \mathbf{g}_{i,n} \right\|^2 + \sum_{n=0}^{\ell} \xi_n \left(\tilde{G}(\mathbf{p}_{i,n}) + \mathbf{g}_{i,n}'(\mathbf{y}_\ell - \mathbf{p}_{i,n}) \right) \quad (30a)$$

$$\text{s.t.} \quad \boldsymbol{\xi} \succeq \mathbf{0}, \ \mathbf{1}'\boldsymbol{\xi} = 1 . \quad (30b)$$

It can be readily seen that (30) is essentially a QP over the probability simplex in the dual space, which can be solved very efficiently as in e.g., [27].

D. Optimality and Distributed Implementation

When a constant stepsize $\mu(j) = \mu$ is adopted, the subgradient iterations (20) are guaranteed to converge to a neighborhood of the optimal $\boldsymbol{\Lambda}^*$ for the dual problem (19) from any initial $\boldsymbol{\Lambda}(0)$. The size of the neighborhood is proportional to the stepsize μ. If we adopt a sequence of nonsummable diminishing stepsizes satisfying $\lim_{j \to \infty} \mu(j) = 0$ and $\sum_{j=0}^{\infty} \mu(j) = \infty$, then the iterations (20) will asymptotically converge to the exact $\boldsymbol{\Lambda}^*$ as $j \to \infty$ [23].

The objective function (16a) is not strictly convex because it does not involve all optimization variables. Hence, when it comes to primal convergence, extra care is necessary [28], [29]. Specifically, the optimal primal can be attained either by adding a strictly convex regularization term, or, by utilizing the augmented Lagrangian. Here, we will simply implement the *Cesáro averaging* method [30] to obtain the optimal power schedules. With $\mu_{\text{sum}}^m := \sum_{j=1}^{m} \mu(j)$, the running average is

$$\bar{\mathbf{Z}}^m := \frac{1}{\mu_{\text{sum}}^m} \sum_{j=1}^{m} \mu(j) \mathbf{Z}(j) \quad (31)$$

which can be efficiently computed in a recursive way

$$\bar{\mathbf{Z}}^m := \frac{\mu(m)}{\mu_{\text{sum}}^m} \mathbf{Z}(m) + \frac{\mu_{\text{sum}}^{m-1}}{\mu_{\text{sum}}^m} \bar{\mathbf{Z}}^{m-1} . \quad (32)$$

Note that if a constant stepsize $\mu(j) \equiv \mu$ is adopted, (31) and (32) boil down to the ordinary running average

$$\bar{\mathbf{Z}}^m = \frac{1}{m} \sum_{j=1}^{m} \mathbf{Z}(j) = \frac{1}{m} \mathbf{Z}(m) + \frac{m-1}{m} \bar{\mathbf{Z}}^{m-1} .$$

TABLE I
GENERATING CAPACITIES, BATTERY INITIAL ENERGY AND CAPACITY,
CHARGING LIMITS AND EFFICIENCY.

Unit	$P_{G_i}^{\min}$	$P_{G_i}^{\max}$	C_i^0	$C_i^{\max}$	$P_{b_i}^{\min}$	$P_{b_i}^{\max}$	ϖ_i
1	0	50	5	30	-10	10	0.95
2	0	45	5	30	-10	10	0.95
3	0	45	5	30	-10	10	0.95
4	0	45	5	30	-10	10	0.95
5	0	50	5	30	-10	10	0.95
6	0	45	5	30	-10	10	0.95

TABLE II
LIMITS OF FORECASTED WIND POWER AND ENERGY PURCHASE PRICES

Slot	1	2	3	4	5	6	7	8
E_1^t	2.47	2.27	2.18	1.97	2.28	2.66	3.10	3.38
E_2^t	2.57	1.88	2.16	1.56	1.95	3.07	3.44	3.11
E_3^t	2.32	2.43	1.27	1.39	2.14	1.98	2.68	4.04
E_4^t	2.04	1.92	2.33	2.07	2.13	2.36	3.13	4.16
E_5^t	2.11	1.19	2.26	2.19	1.55	2.71	3.37	2.45
E_6^t	2.01	2.29	2.20	0.98	2.43	3.22	2.74	3.93
α^t	0.402	0.44	0.724	1.32	1.166	0.798	0.506	0.468

If for the relaxed problem (16), the obtained solution satisfies the condition rank$(\mathbf{X}_k^{t\,*}) = 1$, $\forall k, t$, then it clearly yields the optimal beamforming vectors $\mathbf{w}_k^{t\,*}$ as the (scaled) eigenvector with respect to the only positive eigenvalue of $\mathbf{X}_k^{t\,*}$ for the original problem (12). Fortunately, it was shown in [31, Them. 1] that the S-procedure based SDP (16) always returns a rank-one optimal solution $\mathbf{X}_k^{t\,*}$, $\forall k, t$, when the uncertainty bounds ϵ_k^t are sufficiently small. If ϵ_k is large, the existence of rank-one optimal solutions of (16) cannot be provably guaranteed. In this case, a randomized rounding strategy [32] needs to be adopted to obtain vectors $\mathbf{w}_k^{t\,*}$ from $\mathbf{X}_k^{t\,*}$ to nicely approximate the solution of the original problem (12). Even though no proof is available to ensure a rank-one solution when ϵ_k is large, it has been extensively observed in simulations that the SDP relaxation always returns a rank-one optimal solution [31]. This confirms the view that the optimal beamforming vectors for the original problem (12) will be obtained by our approach with high probability.

The subgradient iterations can be run in a distributed fashion. Specifically, the central controller can maintain the Lagrange multipliers $\mathbf{\Lambda}(j)$ and broadcast them to all BSs via backhaul links. Given the current $\mathbf{\Lambda}(j)$, the central controller solves the subproblems (22a) to obtain the beamforming vectors for all BSs. On the other hand, each BS solves its own subproblems (22b)–(22c), which are decoupled across BSs. The BSs send back to the central controller $\{P_{g,i}^t(j)\}_{t=1}^T$, $\{P_{b,i}^t(j)\}_{t=1}^T$, and $\{P_i^t(j)\}_{t=1}^T$, which are in turn used to update $\mathbf{\Lambda}(j+1)$ through the subgradient iterations (20).
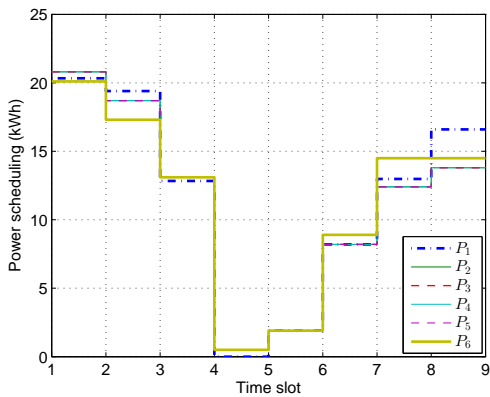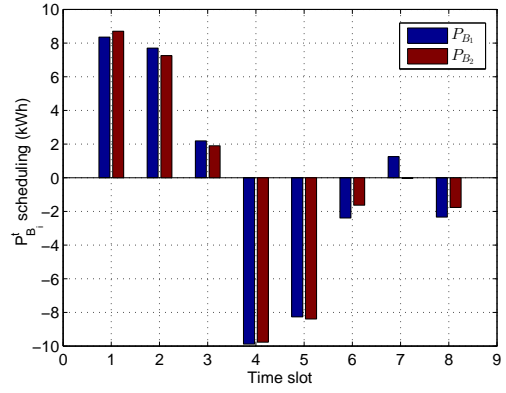
IV. NUMERICAL TESTS

In this section, simulated tests are presented to verify the performance of the proposed approach. The Matlab-based modeling package CVX 2.1 [33] along with the solvers MOSEK 7.0 [34] and Sedumi 1.02 [35] are used to specify and solve the resulting optimization problems. All numerical tests are implemented on a computer workstation with Intel cores 3.40 GHz and 32 GB RAM.
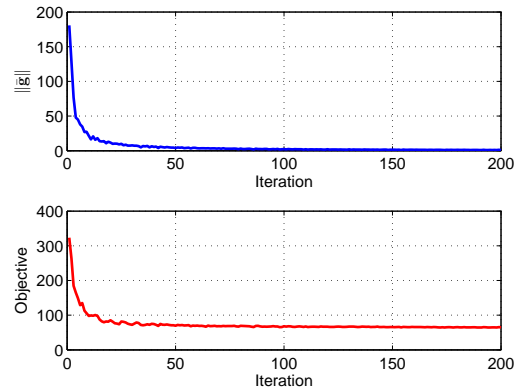
Fig. 2. Convergence of the objective value and the subgradient norm of Lagrange multipliers ($M = 2$, $K = 10$, $r = 1$).
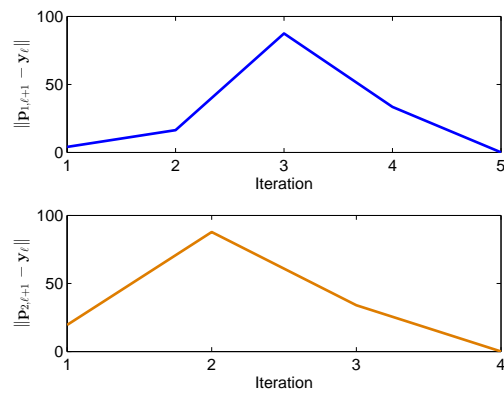
Fig. 3. Convergence of the bundle method solving the subproblem (22c) ($M = 2$, $K = 10$, $r = 1$, $j = 2$).

The scheduling horizons of the considered CoMP network is $T = 8$. Two configurations are tested: (C1) $I = 2$ BSs and $K = 10$ end users (small size); and (C2) $I = 6$ BSs and $K = 30$ end users (large size). All wireless channels are assumed to be flat Rayleigh fading, and normalized to unit power. The noises are modeled as circularly-symmetric Gaussian random vectors. Without loss of generality, the effects of path loss, shadowing, and Doppler fading are ignored. Parameters including the limits of P_{G_i}, C_i, P_{b_i} and the discharging efficiency ϖ_i are listed in Table I. A polyhedral uncertainty set (7) with a single sub-horizon (no time partition) is considered for the RES. In Table II, the energy purchase price α^t is given across the entire time horizon. The selling price is set as $\beta^t = r\alpha^t$ with $r \in [0, 1]$. In addition, the lower limits $\{\underline{E}_i^t\}_{i \in \mathcal{I}}$ are listed therein, which were rescaled from real data we obtained from the Midcontinent Independent System Operator (MISO) [36]. The upper limits were set to $\overline{E}_i^t = 10\underline{E}_i^t$, while the total horizon bounds are $E_i^{\max} = 0.9 \sum_t \overline{E}_i^t$.

First, convergence of the objective value (16a), and the ℓ_2-norm of the subgradient of the running-average Lagrange multiplier (21) is verified in Fig. 2. It can be seen that both metrics converge within a few hundred iterations, which confirms the validity of the dual decomposition approach
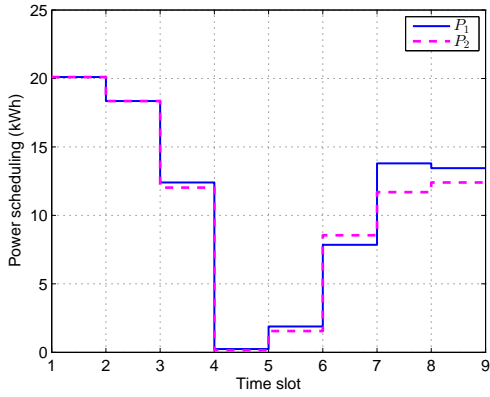
Fig. 4. Optimal power schedules of $\bar{P}_i^t$ ($I = 2$, $M = 2$, $K = 10$, $r = 1$).

Fig. 6. Optimal power schedule of $\bar{P}_{B_i}^t$ ($I = 2$, $M = 2$, $K = 10$, $r = 1$).

Fig. 5. Optimal power schedules of $\bar{P}_i^t$ ($I = 6$, $M = 2$, $K = 20$, $r = 1$).
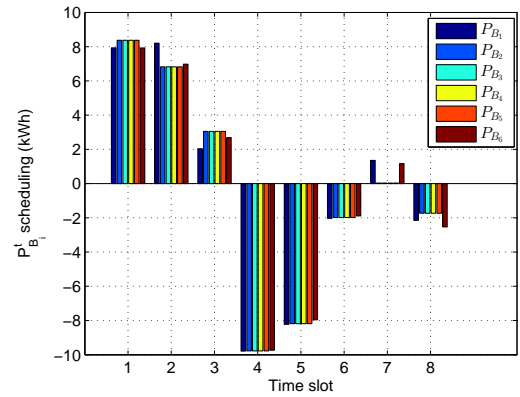
Fig. 7. Optimal power schedule of $\bar{P}_{B_i}^t$ ($I = 6$, $M = 2$, $K = 20$, $r = 1$).

along with the subgradient ascent algorithm. With the Cesàro averages, convergence of the dual and primal variables was also confirmed, but it is omitted due to limited space.

Figure 3 depicts the effectiveness of the proposed bundle method minimizing the convex nonsmooth objective (22c). Clearly, incorporating the scheme of dynamically changing the proximity weight ρ_ℓ, the bundle algorithm converges very fast; typically, within 10 iterations.

The optimal power schedules of $\bar{P}_i^t$ are depicted in Figs. 4 and 5. For both configurations, the stairstep curves show that the lowest levels of $\bar{P}_i^t$ occur from slot 4 to 6. This is because the energy selling and purchase prices are relatively high during these horizons [cf. Sec. II], which drives the BSs's power consumption low in order to minimize the transaction cost.

Figs. 6 and 7 show the optimal battery (dis-)charging amount $\bar{P}_{B_i}^t$, while Figs. 8 and 9 depict the state of charge $\bar{C}_i^t$. As a component part of $\bar{P}_i^t$, $\bar{P}_{B_i}^t$ exhibits a similar trend in response to the price fluctuation; that is, there is a relative large amount of battery discharging ($\bar{P}_{B_i}^t < 0$) when the corresponding price is high. Note that both $\bar{P}_{B_i}^t$ and $\bar{C}_i^t$ never exceed their lower and upper limits [cf. Table I].

Robustness of the worst-case design to the uncertainty of channel estimates [cf. (4)] is confirmed in Figs. 10 and 11. The red solid line indicates the SINR threshold $\gamma_k = 0.1$

that is set common to all users for simplicity. The non-robust scheme simply considers the estimated channel $\hat{\mathbf{h}}_k^t$ as the actual one, and plugs it into the worst-case SINR control design (6). This constraint can be relaxed to a linear matrix inequality: $\hat{\mathbf{h}}_k^{tH} \mathbf{Y}_k^t \hat{\mathbf{h}}_k^t - \sigma_k^2 \geq 0$, which is a relaxed version of our proposed counterpart in (14). For both the robust and non-robust approaches, each transmit beamformer $\mathbf{w}_k^t$ is obtained as the principal eigenvector associated with the largest eigenvalue of the resulting $\bar{\mathbf{X}}_k^t$. The CDF of the actual SINR is obtained by evaluating (2) with $5,000$ independent and identically distributed (i.i.d.) channel realizations. The channel perturbations $\{\boldsymbol{\delta}_k^t\}$ are first generated as complex Gaussian, and then rescaled to the boundary of the spherical region $\mathcal{H}_k^t$ [cf. (4)]. It can be seen that 20% of the realizations of the non-robust scheme does not satisfy the SINR constraint, while only about 2% for the proposed approach. Note that the SDP relaxation is not always exact, which results in violating the SINR threshold for a few channel realizations.

Finally, CDFs of the transaction cost are depicted in Fig. 12. The proposed robust approach is compared with a heuristic scheme that assumes the expected renewable generation $\hat{E}_i^t = \frac{1}{2}(\underline{E}_i^t + \overline{E}_i^t)$ is available and prefixed for problem (16). The CDF curves were plotted by evaluating the transaction cost (16a) with 10^5 realizations of $\{E_i^t\}_{i,t}$ and the obtained optimal solutions $\{\bar{P}_i^t\}_{i,t}$. The RES realizations were gener-
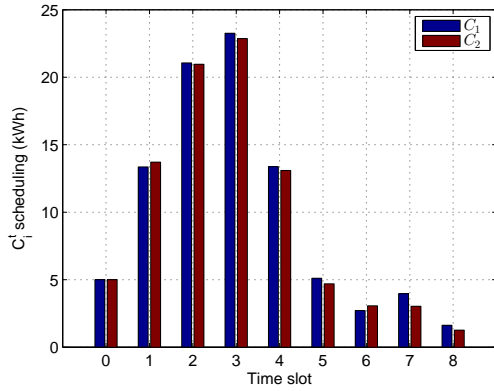
Fig. 8. Optimal power schedule for $\bar{C}_i^t$ ($I = 2$, $M = 2$, $K = 10$, $r = 1$).
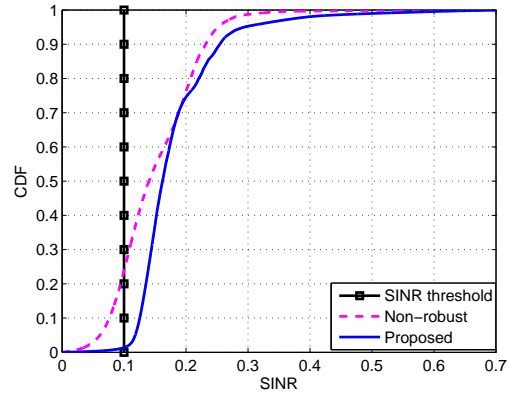
Fig. 10. SINR cumulative distribution function (CDF) ($I = 2$, $M = 2$, $K = 10$, $r = 1$).
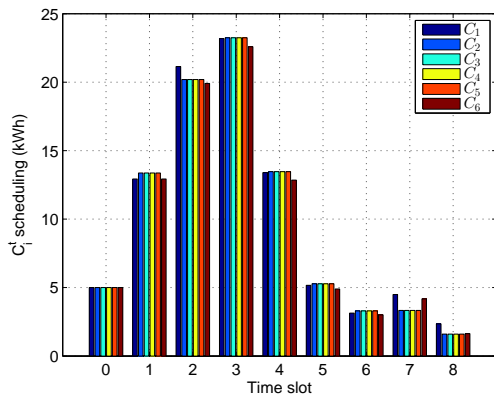
Fig. 9. Optimal power schedule for $\bar{C}_i^t$ ($I = 6$, $M = 2$, $K = 20$, $r = 1$).
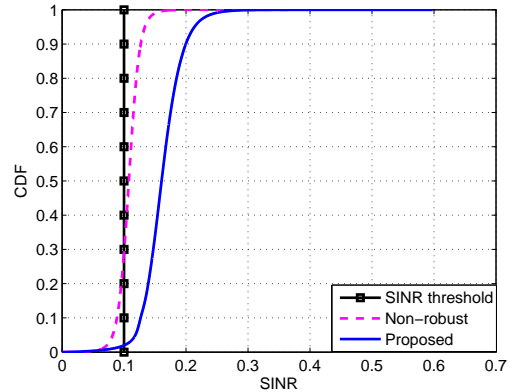
Fig. 11. SINR cumulative distribution function (CDF) ($I = 6$, $M = 2$, $K = 20$, $r = 1$).

ated as $\{\tilde{E}_i^t\}_{i,t} = \underline{E}_i^t + \kappa U(\overline{E}_i^t - \underline{E}_i^t)$, where U is a uniform random variable on $[0, 1]$. Three cases with $\kappa = 0.01, 0.1, 0.5$ were tested. Clearly, transaction costs of both the proposed and the heuristic methods decrease with the increase of κ. Since a larger value of κ implies more harvested renewables energy yielding a reduced transaction cost. Note that negative transaction costs means net profits are obtained by selling surplus renewables back to the smart grid. Interestingly, for all cases, the proposed approach always outperforms the heuristic scheme with less transaction costs. This is because the proposed schedules of $\{P_{g,i}^t, P_{b,i}^t\}_{i,t}$ are robust to the worst-case renewable generation $\{E_i^t\}_{i,t}$. However, in practice more RES is often available than the worst case. Hence, the proposed method has a larger profit-making capability than the heuristic alternative.

V. CONCLUSIONS

Robust ahead-of-time energy management and transmit-beamforming designs were developed to minimize the worst-case energy costs subject to the worst-case user QoS guarantees for the CoMP downlink, which is powered by a grid with smart-meter based dynamic pricing and RES available at the BSs. Building on innovative models, the task was formulated as a convex program. Relying on a Lagrange-dual approach, efficient algorithms were introduced to obtain optimal solutions. The proposed scheme provides the offline ahead-of-time beamformer and energy schedule over a finite time horizon. Development of online management schemes will be an interesting direction to pursue in our future work.

Supported by major programs such as the US Energy Independence and Security Act and European SmartGrids Technology Platform, the smart grid industry has seen fast growth. It is expected that all future communication systems will be powered by a smart grid. As a result, integration of smart-grid technologies into system designs will hold the key to fully leverage energy-efficient communications in next-generation HetNets. The proposed models and approaches can pave a way to further advancing fundamental research on smart-grid powered HetNet transmissions, which will be pursued in future works.

REFERENCES

[1] I. Hwang, B. Song, and S. Soliman, "A holistic view on hyper-dense heterogeneous and small cell networks," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 20–27, Jun. 2013.

[2] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.
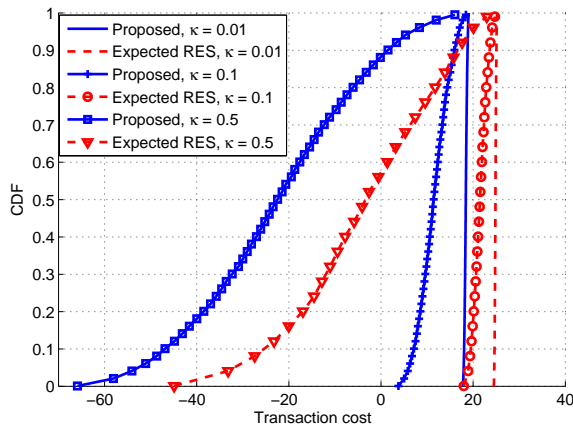
Fig. 12. Cumulative distribution function (CDF) of the transaction cost ($I = 2$, $M = 2$, $K = 10$, $r = 0.3$).

[3] A. Tolli, H. Pennanen, and P. Komulainen, "Decentralized minimum power multi-cell beamforming with limited backhaul signaling," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 570–580, Feb. 2011.

[4] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010.

[5] J. Zhang, R. Chen, J. Andrews, A. Ghosh, and R. W. Heath, "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, Apr. 2009.

[6] S.-J. Kim, S. Jain, and G. B. Giannakis, "Backhaul-constrained multicell cooperation using compressive sensing and spectral clustering," in *Proc. of 13th Wrkshp. on Signal Process. Advances in Wireless Commun.*, Cesme, Turkey, June 2012.

[7] C. Ng and H. Huang, "Linear precoding in cooperative MIMO cellular networks with limited coordination clusters," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1446–1454, Dec. 2010.

[8] M. Hong, R. Sun, H. Baligh, and Z. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 226–240, Feb. 2013.

[9] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 56–61, Jun. 2011.

[10] Y. Shi, J. Zhang, and K. Letaief, "Group sparse beamforming for green cloud-RAN," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809–2823, May 2014.

[11] J. Rao and A. Fapojuwo, "On the tradeoff between spectral efficiency and energy efficiency of homogeneous cellular networks with outage constraint," *IEEE Tans. Veh. Technol.*, vol. 62, no. 4, pp. 1801–1814, May 2013.

[12] Y. Zhang, N. Gatsis, and G. B. Giannakis, "Robust energy management for microgrids with high-penetration renewables," *IEEE Trans. Sustain. Energy*, vol. 4, no. 4, pp. 944–953, Oct. 2013.

[13] X. Liu and W. Xu, "Economic load dispatch constrained by wind power availability: A here-and-now approach," *IEEE Trans. Sustain. Energy*, vol. 1, no. 1, pp. 2–9, Apr. 2010.

[14] X. Guan, Z. Xu, and Q. Jia, "Energy-efficient buildings facilitated by microgrid," *IEEE Trans. Smart Grid*, vol. 1, no. 3, pp. 243–252, Dec. 2010.

[15] G. B. Giannakis, V. Kekatos, N. Gatsis, S. Kim, H. Zhu, and B. Wollenberg, "Monitoring and optimization for power grids: A signal processing perspective," *IEEE Signal Process. Mag.*, vol. 30, no. 5, pp. 107–128, Sept. 2013.

[16] S. Bu, F. Yu, Y. Cai, and X. Liu, "When the smart grid meets energy-efficient communications: Green wireless cellular networks powered by the smart grid," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 3014–3024, Aug. 2012.

[17] J. Xu, Y. Guo, and R. Zhang, "CoMP meets energy harvesting: A new communication and energy cooperation paradigm," in *Proc. of Global Commun. Conf.*, Atlanta, GA, Dec. 2013, pp. 2508–2513.

[18] J. Xu and R. Zhang, "Cooperative energy trading in CoMP systems powered by smart grids," Mar. 2014. [Online]. Available: http://arxiv.org/pdf/1403.5735v1.pdf

[19] N. Vucic and H. Boche, "Robust QoS-constrained optimization of downlink multiuser MISO systems," *IEEE Trans. Signal Process*, vol. 57, no. 2, pp. 714–725, Jun. 2009.

[20] Y. Chen and A. D. Domínguez-García, "A method to study the effect of renewable resource variability on power system dynamics," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 1978–1989, Nov. 2012.

[21] J. Xu, L. Duan, and R. Zhang, "Cost-aware green cellular networks with energy and communication cooperation," Nov. 2014. [Online]. Available: http://arxiv.org/pdf/1411.4139v1.pdf

[22] X. Wang, G. B. Giannakis, and A. G. Marques, "A unified approach to QoS-guaranteed scheduling for channel-adaptive wireless networks," *Proc. IEEE*, vol. 95, no. 12, pp. 2410–2431, Dec. 2007.

[23] D. P. Bertsekas, *Convex Optimization Theory*. Belmont, MA: Athena Scientific, 2009.

[24] I. Polik and T. Terlaky, "A survey of the s-lemma," *SIAM Rev.*, vol. 49, no. 3, pp. 371–418, 2007.

[25] S. Feltenmark and K. C. Kiwiel, "Dual application of proximal bundle methods, including Lagrange relaxation of nonconvex problems," *SIAM J. Optim.*, vol. 10, no. 3, pp. 697–721, Feb./Mar. 2000.

[26] K. C. Kiwiel, "Approximation in proximal bundle methods and decomposition of convex programs," *J. Optim. and Appl.*, vol. 84, no. 3, pp. 529–548, Mar. 1995.

[27] D. P. Bertsekas, "Projected Newton methods for optimization problems with simple constraints," *SIAM J. Contol and Optim.*, vol. 20, no. 2, pp. 221–246, Mar. 1982.

[28] L. Xiao and S. Boyd, "Simultaneous routing and resource allocation via dual decomposition," *IEEE Trans. Commun.*, vol. 52, no. 7, pp. 1136–1144, July 2004.

[29] N. Gatsis and G. B. Giannakis, "Residential load control: Distributed scheduling and convergence with lost AMI message," *IEEE Trans. Smart Grid*, vol. 3, no. 2, pp. 770–786, Jun. 2012.

[30] K. C. Kiwiel, "Convergence of approximate and incremental subgradient methods for convex optimization," *SIAM J. Optim.*, vol. 14, no. 3, pp. 807–840, 2004.

[31] E. Song, Q. Shi, M. Sanjabi, R. Sun, and Z.-Q. Luo, "Robust SINR-constrained MISO downlink beamforming: When is semidefinite programming relaxation tight?" *EURASIP J. Wireless Commun. Netw.*, Feb. 2012.

[32] P. Tseng, "Further results on approximating nonconvex quadratic optimization by semidefinite programming relaxation," *SIAM J. Optim.*, vol. 14, no. 1, pp. 268–283, July 2003.

[33] "CVX: Matlab software for disciplined convex programming, version 2.1," Sep. 2012. [Online]. Available: http://cvxr.com/cvx

[34] "MOSEK ApS," 2014. [Online]. Available: http://www.mosek.com/

[35] J. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Methods and Software*, vol. 11, no. 1–4, pp. 625–633, 1999.

[36] "MISO market data." [Online]. Available: https://www.midwestiso.org/MarketsOperations/RealTimeMarketData/Pages/DayAheadWindForecast.aspx

PLACE PHOTO HERE

Xin Wang (SM'09) received the B.Sc. degree and the M.Sc. degree from Fudan University, Shanghai, China, in 1997 and 2000, respectively, and the Ph.D. degree from Auburn University, Auburn, AL, in 2004, all in electrical engineering.

From September 2004 to August 2006, he was a Postdoctoral Research Associate with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis. In August 2006, he joined the Department of Computer & Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, as an Assistant Professor, and then an Associate Professor from August 2010. He is now a Professor with the Department of Communication Science and Engineering, Fudan University, China. His research interests include stochastic network optimization, energy-efficient communications, cross-layer design, and signal processing for communications. He is an Associate Editor for the IEEE Transactions on Signal Processing, and an Editor for the IEEE Transactions on Vehicular Technology. He was an Associate Editor for the IEEE Signal Processing Letters.

Yu Zhang (S'11) received his B.Eng. and M.Sc. degrees (both with highest honors) from Wuhan University of Technology, Wuhan, China, and from Shanghai Jiao Tong University, Shanghai, China, in 2006 and 2010, respectively, and the Ph.D. degree from the University of Minnesota, Minneapolis, MN, USA, in 2015, all in electrical engineering. His research interests span the areas of smart power grids, optimization theory, machine learning, and wireless networks. Current research focuses on distributed energy management with renewables and smart grid data analytics. He received the Huawei Scholarship and the Infineon Scholarship from the Shanghai Jiao Tong University (2009), the ECE Dept. Fellowship from the University of Minnesota (2010), and the Student Travel Awards from the SIAM and the IEEE Signal Processing Society (2014).