# Optimal Placement of Energy Storage in Distribution Networks

Yujie Tang, *Student Member, IEEE,* and Steven H. Low, *Fellow, IEEE*

*Abstract*—We study the problem of optimal placement and capacity of energy storage devices in a distribution network to minimize total energy loss. A continuous tree with linearized DistFlow model is developed to model the distribution network. We analyze structural properties of the optimal solution when all loads have the same shape. We prove that it is optimal to place storage devices near the leaves of the network away from the substation, and that the scaled storage capacity monotonically increases towards the leaves. Moreover, under optimal storage placement, the locational marginal value of storage is equalized wherever nonzero storage is deployed and increases from the substation towards any leaf node over places where there is zero storage deployment. We illustrate through simulations that these structural properties are robust in that they hold approximately when some of our modeling assumptions are relaxed.

*Index Terms*—Energy storage, distribution networks, optimal power flow

## I. INTRODUCTION

Energy storage devices shift energy generation and consumption across time, which helps integrate intermittent generations and loads. There is a large amount of literature on energy storage integration, and one important direction is to study the use of storage to minimize operational cost of the power system as well as to maintain power grid reliability, assuming that the storage devices are operated by system operators. For example, [1] treats the stochastic control of distributed energy storage devices using a surrogate LQ problem; [2], [3] propose methods to incorporate distributed energy storage systems with wind generation, and conduct case studies to evaluate the benefit of storage; [4]–[6] propose algorithms to find the optimal placement of energy storage using different power flow models; [7] proves a particular structural property of optimal storage placement using DC power flow model; [8] proposes a discrete optimization formulation for energy storage placement and characterizes conditions under which the placement value function is submodular; [9] relates the locational marginal value of storage with the upward variation of the locational marginal price at each bus when the storage capacity is sufficiently small.

In this paper we study the problem of optimally placing and sizing energy storage devices in distribution networks, where the system operator has a certain budget for energy

storage and needs to decide where to deploy storage devices and their capacities so as to minimize the total energy loss of the network. We focus on analytically deriving structural properties of the optimal solution that can provide insight for deployment. This problem is difficult in that the cost function depends on the detailed charging and discharging schedule of storage, and so the placement of storage devices and their charging and discharging schedules over time should be determined jointly, resulting in an optimization problem whose time and space dimensions are strongly coupled.

Our main contribution is two-fold. First, a continuous tree model equipped with DistFlow equations [10], [11] is proposed in Section II for modeling single-phase distribution networks. A similar model is used in [12] for analyzing a single feeder; our model here extends it to a radial network. It turns out that the proposed continuous model is very powerful for theoretical analysis, and can be used to prove results which seem hard to derive by the commonly-used discrete models.

Second, we derive some structural results for the optimal storage placement and capacity where the total loss of the network is minimized. They are presented in Section III for a line network and in Section IV for a radial network. Specifically, we show that, when all loads have the same shape, on each path connecting the substation to a leaf, there exists a threshold location that splits the path into two parts: the optimal strategy deploys zero storage in the part near the substation and strictly positive amount of storage at every point in the other part. Moreover, the optimal capacity scaled by the variation of the load increases as we move from the substation towards the leaf. We also show that, under the optimal storage placement, the locational marginal value of storage strictly increases from the substation towards the threshold location over the part where zero storage is deployed, and is equalized over the entire network everywhere nonzero storage is deployed.

We employ models and assumptions that simplify our analysis but still capture the core features of the storage placement problem. We present in Section V simulation results to demonstrate that these structural properties continue to hold approximately when some of our modeling assumptions are relaxed. We conclude in Section VI. Due to space limit, all proofs are relegated to [13].

## II. PROBLEM FORMULATION

### A. Continuous Trees

A continuous tree is a continuum of nodes that embeds a discrete tree structure. To be precise, let $\mathcal{N}$ be a set with partial

order $\leq$. Let $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ be a discrete tree graph, with $\mathcal{V}$ being the set of vertices and $\mathcal{E}$ being the set of edges. We say that $\mathcal{N}$ is a continuous tree with the underlying tree structure $\mathcal{T}$ if for each $k \in \mathcal{V}$, there is an associated subset $\mathcal{N}_k \subseteq \mathcal{N}$ such that

1) $\bigcup_{k \in \mathcal{V}} \mathcal{N}_k = \mathcal{N}$.
2) $\leq$ is a total order on each $\mathcal{N}_k$ so that $\mathcal{N}_k$ admits the order topology generated by $\leq$.
3) Each $\mathcal{N}_k$ is equipped with a (positive) Borel measure $\mu_k$, and there is a bijection $\psi_k : \mathcal{N}_k \to [0, \mu_k(\mathcal{N}_k)]$ such that for any $x_1, x_2 \in \mathcal{N}_k$ with $x_1 < x_2$, we have

$$\mu_k(\{y \in \mathcal{N}_k : x_1 < y < x_2\})$$
$$= \psi_k(x_2) - \psi_k(x_1) > 0.$$

We denote $s_k = \psi_k^{-1}(0)$ and $e_k = \psi_k^{-1}(\mu_k(\mathcal{N}_k))$.
4) Let $k, k' \in \mathcal{V}$ be two different vertices. Then

$$\mathcal{N}_k \cap \mathcal{N}_{k'} = \begin{cases} \{e_k\} = \{s_{k'}\}, & k \text{ is the parent of } k', \\ \{s_k\} = \{s_{k'}\}, & k \text{ and } k' \text{ are siblings}, \\ \varnothing, & \text{otherwise.} \end{cases}$$

5) For any $x \in \mathcal{N}_k$ and $y \in \mathcal{N}_{k'}$ with $k \neq k'$, if $x \leq y$, then either $k'$ is a descendant of $k$ in $\mathcal{T}$, or $k'$ is not a descendant of $k$ but $x = s_k$ and $k'$ is a descendant of the parent of $k$.

Roughly speaking, the first condition says that the continuous tree consists of a collection of *segments*, denoted by $\mathcal{N}_k$. The second and third conditions state that each individual segment can be treated just like a closed interval in $\mathbb{R}$ with a total order and Lebesgue measure. Then the forth condition states that the individual segments are connected according to the structure of an ordinary discrete tree. The fifth condition indicates that the partial order relation is compatible with the discrete tree structure $\mathcal{T}$, and so $x \leq y$ can be interpreted as $y$ being in the continuous subtree rooted at $x$.

The point $s_k$ for a non-root vertex $k$ will be called a *branch point*. The unique path connecting $x$ and $y$ will be denoted by $[x, y]$, i.e.,

$$[x, y] = \{w \in \mathcal{N} : x \leq w \leq y\}.$$

We will use 0 to denote the root node, i.e., $0 \leq x$ for all $x \in \mathcal{N}$.

As we have pointed out, each segment $\mathcal{N}_k$ can be treated like a closed interval. For this reason we will simply write $\int_{\mathcal{N}_k} f(x) \, \mu_k(dx)$ as $\int_{\mathcal{N}_k} f(x) \, dx$. For an arbitrary subset $B \subseteq \mathcal{N}$, we define

$$\int_B f(x) \, dx = \sum_{k \in \mathcal{V}_k} \int_{B \cap \mathcal{N}_k} f(x) \, dx.$$

Figure 1 gives an example of a continuous tree and its underlying tree structure.

### B. Power Flows on Continuous Trees

We use continuous trees to model single-phase distribution networks.

Let $z(x)$ be the impedance per unit length in the sense that $\int_{[x,y]} z(x) \, dx$ gives the impedance of the line between $x$ and $y$.
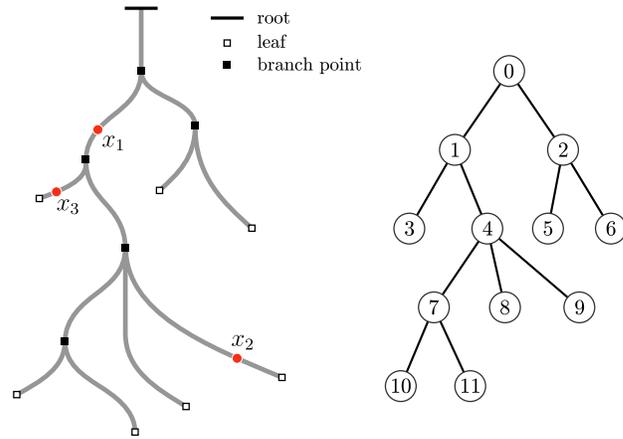


Fig. 1. Illustration of a continuous tree and its underlying tree structure. Here the set of vertices is $\mathcal{V} = \{0, 1, 2, \ldots, 11\}$. It can be seen that $x_1 \in \mathcal{N}_1$, $x_2 \in \mathcal{N}_9$ and $x_3 \in \mathcal{N}_3$. Moreover, $x_1 < e_1 = s_4 < e_4 = s_9 < x_2$ and $x_1 < e_1 = s_3 < x_3$, but $x_2$ and $x_3$ are not comparable.

We assume that $z$ is continuously differentiable when restricted to any $\mathcal{N}_k$. We denote the real part of $z(x)$ (the resistance per unit length) by $r(x)$, and assume that $r(x) > 0$ for all $x \in \mathcal{N}$.

Let $s(x) = p(x) + jq(x)$ be the complex power injection, and $v(x)$ be the squared voltage magnitude at node $x$. For simplicity we assume $v(0) = 1$ pu. Let $S(x) = P(x) + jQ(x)$ be the complex power flow towards the substation (the root node) on the distribution line at node $x$. Then the continuous power flow equations are given by

$$S(x) = \int_{y \geq x} \left( s(y) - z(y) \frac{|S(y)|^2}{v(y)} \right) \, dy,$$
$$v(x) = 1 + \int_{[0,x]} 2 \operatorname{Re}(z^*(y) S(y)) \, dy, \tag{1}$$

from which we can see that the total power loss is given by

$$P_{\text{loss}} = \int_{\mathcal{N}} p(x) \, dx - P(0) = \int_{\mathcal{N}} r(x) \frac{|S(x)|^2}{v(x)} \, dx.$$

These equations are the continuous version of the DistFlow model for radial networks [10], [11], and can be obtained by taking limits of the original DistFlow equations; see [12] for a version modeling a continuous feeder line. Furthermore, the continuous model reduces to the discrete model if we set $s(x) = \sum_{j=1}^{N} s_j h_j(x; n)$ and take the limit $n \to \infty$, where $s_j$ is the power injection at bus $j$, $h_j(x; n)$ tends to a Dirac delta function located at $x_j$ as $n \to \infty$ and $x_j$ is the location of bus $j$. This means that we don't lose any modeling capability when using the continuous model.

In most cases the continuous DistFlow equations can be rewritten as a set of nonlinear ordinary differential equations with boundary conditions at branch points, leaf nodes and the root node.

### C. Storage Model and Load Profiles

Load profiles often exhibit a cyclic behavior, and so we use a finite time horizon in this paper and equip it with a cyclic structure. The time horizon $\mathbb{T}$ will be defined as

$$\mathbb{T} = [0, T),$$

where $T$ is the length of the whole period. To incorporate cyclic structure of the time domain, for $0 \leq t_2 < t_1 < T$, we define the intervals

$$(t_1, t_2) := [0, t_2) \cup (t_1, T),$$
$$[t_1, t_2] := [0, t_2] \cup [t_1, T).$$

The topology of $\mathbb{T}$ is then generated by the family $\{(t_1, t_2) : t_1, t_2 \in \mathbb{T}\}$, and we denote $\int_{t_1}^{t_2} = \int_{[t_1, t_2]}$. It can be shown that a continuous function $f$ on domain $\mathbb{T}$ will automatically be cyclic in the sense that $f(0) = \lim_{t \uparrow T} f(t)$.

We assume that at each node $x$ at time $t$, the real power injection consists of two parts, one being the power consumption by the storage device which we denote by $u(x, t)$, and the other being the sum of all other types of background injections which we denote by $\tilde{p}(x, t)$.

For storage devices, we employ the simple model

$$u(x, t) = \frac{\partial b(x, t)}{\partial t}, \quad 0 \leq b(x, t) \leq B(x),$$

where $b(x, t)$ is the energy level (SoC) at time $t$, and $B(x)$ is the energy capacity of the storage device at node $x$. We ignore reactive power injections from storage. Although some fine features of practical storage devices, such as power limit and round-trip efficiency, are not considered in this simple linear model, this model will be sufficient for demonstrating broad structures of the optimal placement and sizing strategy.

For the background injections, we assume they are deterministic and fixed. In practice $\tilde{p}(x, t)$ can be obtained from historical data. We further assume that the background injection $\tilde{p}(x, t)$ takes the form

$$\tilde{p}(x, t) = \alpha(x)p(t) + \beta(x),$$

where $\alpha(x) > 0$. This means that load profiles at different nodes have a common shape $p(t)$, while $\alpha(x)$ represents the relative variation and $\beta(x)$ is an offset. This assumption is based on the observation that the electricity usage pattern of a large group of residential consumers can be categorized into a relatively small set of signature load shapes [14]–[16]. In practice, the assumption that all load profiles have the same shape rarely holds with accuracy, but we will show by simulation that it can be relaxed to some extent and our main results will hold approximately.

Without loss of generality we assume that $\alpha(x)$ and $\beta(x)$ are continuously differentiable when restricted to an arbitrary segment. For the load shape $p(t)$, we assume that

1) It is twice continuously differentiable on $\mathbb{T}$.
2) There are finitely many solutions to $\dot{p}(t) = 0$, each of which is either a local maximum or a local minimum with a non-vanishing second-order derivative. The set of solutions to $\dot{p}(t) = 0$ will be denoted by

$$t_0^{\max}, t_0^{\min}, t_1^{\max}, t_1^{\min}, \ldots, t_{M-1}^{\max}, t_{M-1}^{\min}$$

such that $\dot{p}(t) < 0$ for $t \in (t_m^{\max}, t_m^{\min})$ and $\dot{p}(t) > 0$ for $t \in (t_{m-1}^{\min}, t_m^{\max})$ for each $m$ (addition and subtraction on the subscript $m$ are modulo $M$).
3) It has a vanishing time average, i.e., $\int_{\mathbb{T}} p(t)\, dt = 0$.

As any continuous function on $\mathbb{T}$ with a vanishing average can be approximated by a function satisfying the above properties

to any given precision, these assumptions impose no essential but only technical restrictions on $p(t)$.

The net power injection is then given by

$$p(x, t) = \tilde{p}(x, t) - u(x, t)$$
$$= \alpha(x)p(t) + \beta(x) - \frac{\partial b(x, t)}{\partial t}. \tag{2}$$

### D. Optimal Placement and Sizing

The cost we wish to minimize is the total energy loss of the network during $\mathbb{T}$, which is given by

$$E_{\text{loss}} = \int_{\mathbb{T}} P_{\text{loss}}(t)\, dt = \int_{\mathbb{T}} \int_{\mathcal{N}} r(x) \frac{|S(x, t)|^2}{v(x, t)}\, dx\, dt.$$

This cost function has a rather complicated form, making its minimization hard to analyze. We adopt the following approximation: The voltage $v(x, t)$ is not far from some fixed nominal value $v_{\text{nom}}(x)$ as time proceeds. As a consequence, the energy loss can be approximately given by

$$E_{\text{loss}} = \int_{\mathbb{T}} \int_{\mathcal{N}} r(x) \frac{|S(x, t)|^2}{v(x, t)}\, dx\, dt$$
$$\approx \int_{\mathbb{T}} \int_{\mathcal{N}} \frac{r(x)}{v_{\text{nom}}(x)} P^2(x, t)\, dx\, dt + \text{const}, \tag{3}$$

where we have used the assumption that storage does not influence reactive power. We assume that $v_{\text{nom}}(x)$ is positive and continuously differentiable when restricted to any $\mathcal{N}_k$. Similar approximation has been employed in, for instance, [11], [17], [18].

Loss minimization is commonly employed for operating distribution networks. Even if the loss is small compared to the loads, loss minimization can lead to other advantages such as increased voltage stability [19] and reduced branch power flows (as can be seen from (3)). This is partly why voltage and line limits are sometimes relaxed in loss minimization.

The nonlinearity of the power flow equations (1) also adds significant difficulty to analysis. We adopt the approximation that the power loss $\int_{y \geq x} z(y)|S(y, t)|^2/v(y, t)\, dy$ is small compared to the line flow $S(x, t)$ for each $t \in \mathbb{T}$. Then

$$S(x, t) \approx \int_{y \geq x} s(y, t)\, dy,$$

and so

$$P(x, t) \approx \int_{y \geq x} \left( \alpha(y)p(t) + \beta(y) - \frac{\partial b(y, t)}{\partial t} \right) dy \tag{4}$$

by (2). It can be seen that the power flow along the lines is now linear in power injections. Similar linearization has been employed in [10], [17], [20].

Now let $B_{\text{tot}}$ be the total storage budget. The optimal storage placement problem is then formulated as

$$\min_{b(\cdot, \cdot), B(\cdot)} \int_{\mathbb{T}} \int_{\mathcal{N}} w(x) P^2(x, t)\, dx\, dt \tag{5a}$$

$$\text{s.t.} \quad P(x, t) = \int_{y \geq x} \left( \alpha(y)p(t) + \beta(y) - \frac{\partial b(y, t)}{\partial t} \right) dy \tag{5b}$$

$$0 \leq b(x, t) \leq B(x), \quad \forall x \in \mathcal{N}, t \in \mathbb{T}, \tag{5c}$$

$$\int_{\mathcal{N}} B(x)\, dx \leq B_{\text{tot}}. \tag{5d}$$

Here we denote $w(x) := r(x)/v_{\mathrm{nom}}(x)$. We call $b^*(x,t)$ the optimal schedule and $B^*(x)$ the optimal storage capacity if they solve the problem (5). The optimal value, which is a function of $B_{\mathrm{tot}}$, will be denoted by $F^*(B_{\mathrm{tot}})$.

### E. Locational Marginal Value of Storage

The locational marginal value of storage characterizes the marginal decrease in the total cost when a small amount of storage capacity is added at a specific location [9]. To be precise, we define the locational marginal value of storage as follows: Suppose the capacity function $B(x)$ is fixed and we consider the following optimization problem:

$$\min_{b(\cdot,\cdot)} \quad \int_{\mathbb{T}} \int_{\mathcal{N}} w(x) P^2(x,t)\, dx\, dt \tag{6a}$$

$$\text{s.t.} \quad P(x,t) = \int_{y \geq x} \left( \alpha(y)p(t) + \beta(y) - \frac{\partial b(y,t)}{\partial t} \right) dy \tag{6b}$$

$$0 \leq b(x,t) \leq B(x), \quad \forall x \in \mathcal{N}, t \in \mathbb{T}, \tag{6c}$$

and denote its optimal value by $J^*(B)$. We say that $\mathrm{lmv}(\cdot; B) : \mathcal{N} \to \mathbb{R}$ is the locational marginal value function of storage if for every nonnegative continuous function $\delta$ defined on $\mathcal{N}$, we have

$$J^*(B + \epsilon\delta) = J^*(B) - \epsilon \int_{\mathcal{N}} \delta(x)\, \mathrm{lmv}(x; B)\, dx + o(\epsilon) \tag{7}$$

as $\epsilon \downarrow 0$, where $\epsilon$ is a positive real number.

This definition is the natural generalization of locational marginal value of storage in discrete networks where it is defined by the negative gradient of $J^*$ with respect to storage capacities as in [9].

## III. MAIN RESULTS FOR LINE NETWORKS

A continuous line network is essentially a closed interval $[0, L]$ where $L$ is the total length of the network. The structural results for the optimal storage placement problem on line networks have been presented in [18] and the detailed theory can be found in [13]. Here we give a summary of these results.

**Theorem** (Line Networks). *Suppose the shape $p(t)$ has only one minimum and one maximum, and the storage budget satisfies*

$$B_{\mathrm{tot}} < \frac{1}{2} \int_0^L \alpha(x)\, dx \int_{\mathbb{T}} |p(t)|\, dt.$$

*Then the optimal storage capacity $B^*(x)$ satisfies the following properties.*

1) *There exists a location $\xi \in (0, L]$ such that $B^*(x) = 0$ for every location $x \leq \xi$, while $B^*(x) > 0$ for every $x > \xi$.*
2) *For $x \geq \xi$, the scaled optimal storage capacity $B^*(x)/\alpha(x)$ increases as one moves away from the substation, i.e., $B^*(x)/\alpha(x)$ is an increasing function of $x$ on $[\xi, L]$.*
3) *The larger the storage budget $B_{\mathrm{tot}}$, the closer $\xi$ is to the substation, i.e., there is a continuous bijection between $B_{\mathrm{tot}}$ and the corresponding $\xi \in (0, L]$, and $\xi$ decreases as $B_{\mathrm{tot}}$ increases.*

4) *The larger the storage budget $B_{\mathrm{tot}}$, the higher the storage capacity $B^*(x)$ for $x > \xi$.*

Although line networks are a special case of continuous trees, the approach of analyzing line networks is quite different from that for general radial networks, and the theorems will rely on different conditions and give slightly different conclusions.

## IV. MAIN RESULTS FOR RADIAL NETWORKS

The main results for radial networks are derived by analyzing the generalized KKT conditions[1] of (5). In order to apply the generalized KKT conditions to obtain a set of Lagrange multipliers, we need certain technical assumptions on the existence and regularity of the optimal solution. Those technical assumptions are summarized in [13], and we'll assume throughout this paper that those assumptions hold.

We present three structural results in the following subsections. The first result generalizes the theorem in Section III on the monotonicity of optimal placement and sizing from a line network to a radial network. The second result establishes the monotonicity of locational marginal value of storage. The last result describes the optimal charging and discharging schedule. The proofs of these results are presented in [13].

Define

$$B_{\mathsf{m}} := \int_{\mathcal{N}} \alpha(x)\, dx \cdot \max_{t_1, t_2 \in \mathbb{T}} \int_{t_1}^{t_2} p(t)\, dt.$$

### A. Optimal Placement and Sizing

**Theorem 1.**  1) *If $B_{\mathrm{tot}} < B_{\mathsf{m}}$, then the optimal storage capacity $B^*(x)$ is unique. Furthermore, for any leaf node $\ell \in \mathcal{N}$, there exists some $\xi_\ell \in (0, \ell]$ such that*

$$B^*(x) = 0, \qquad \text{for all } x \leq \xi_\ell,$$
$$B^*(x) > 0, \qquad \text{for almost all } x > \xi_\ell.$$

2) *If $B_{\mathrm{tot}} < B_{\mathsf{m}}$ and $p(t)$ has only one minimum and one maximum, then for any leaf node $\ell \in \mathcal{N}$, the scaled optimal capacity $B^*(x)/\alpha(x)$ is monotonically increasing along $[\xi_\ell, \ell]$; in other words, for each $x_1, x_2 \in [\xi_\ell, \ell]$,*

$$x_1 < x_2 \quad \Longrightarrow \quad \frac{B^*(x_1)}{\alpha(x_1)} \leq \frac{B^*(x_2)}{\alpha(x_2)}.$$

3) *If $B_{\mathrm{tot}} \geq B_{\mathsf{m}}$, then we can choose*

$$B^*(x) \geq \alpha(x) \max_{t_1, t_2 \in \mathbb{T}} \int_{t_1}^{t_2} p(t)\, dt,$$
$$\frac{\partial b^*(x,t)}{\partial t} = \alpha(x)p(t)$$

*for all $x \in \mathcal{N}$.*

We now discuss the implications of Theorems 1 on optimal storage placement and capacity.

The first part of Theorem 1 establishes the existence of the threshold point $\xi_\ell$ that divides the path $[0, \ell]$ into two sections, one with no storage deployed and the other with

---

[1]See [21] for the generalized KKT conditions of convex optimization problems over general linear spaces.

storage allocated almost everywhere, when the total budget does not exceed $B_{\mathrm{m}}$. The third part of Theorem 1 then shows that when the total budget exceeds $B_{\mathrm{m}}$, the optimal capacities and schedules can be explicitly given, and the net power injection is then $p(x, t) = \beta(x)$, which is a constant with respect to time. In other words, the storage devices can flatten the power injections completely at every bus.

The second part of Theorem 1 generalizes the monotonicity of $B^*(x)/\alpha(x)$ from line networks [18] to radial networks. It indicates that priorities should be given to nodes near the leaves or far from the substation when allocating the storage budget; if $\alpha(x)$ is roughly uniform, then more storage capacity should be placed as we move from the substation towards the leaf nodes of the network. But even when $\alpha(x)$ is not uniform at all, the scaled optimal capacity $B^*(x)/\alpha(x)$ is still monotonically increasing. Although the second part of Theorem 1 further relies on the assumption that $p(t)$ has only one minimum and one maximum, simulation suggests that the monotonicity of $B^*(x)/\alpha(x)$ still holds when $p(t)$ has more complicated patterns. Figure 2 gives an example that demonstrates the monotonicity of $B^*(x)/\alpha(x)$ for radial networks.

Some of the results in Theorem 1 are in accordance with the intuition that, in order to reduce losses optimally, storage should be put near the leaf nodes because higher losses and voltage drops are observed as power travels further along the network towards the leaf nodes. This simple intuition, however, does not always hold, and the structure of the optimal storage placement can be complicated if we allow arbitrary load shapes at different nodes (see Section V-B and Figure 7). Theorem 1 proves rigorously sufficient conditions under which such intuitions hold. As far as we know, this is the first time monotonicity of scaled optimal capacity in radial networks is formalized and proved rigorously.

In contrast to the theorem for line networks, in Theorem 1 the existence of the threshold location $\xi_\ell$ can be proved without assuming that $p(t)$ has only one minimum and one maximum.

As we have mentioned before, discrete radial networks can be viewed as a limiting case of continuous trees. Consider the case where we have a discrete network, and the background power injection at bus $i$ is given by $\tilde{p}_i(t) = \alpha_i p(t) + \beta_i$. We let

$$\alpha(x) = \sum_j \alpha_j h_j(x; n), \qquad \beta(x) = \sum_j \beta_j h_j(x; n), \quad (8)$$

where $h_j(x; n)$ tends to a Dirac delta function located at $x_j$ as $n \to \infty$ and the sequence of sets

$$\{x \in \mathcal{N} : h_j(x; n) \ge n^{-1}\}, \quad n = 1, 2, \ldots$$

converges to $\{x_j\}$. Then we can obtain the optimal storage capacity at bus $i$ by

$$B_i^* = \lim_{n \to \infty} \int_{h_i(x; n) \ge n^{-1}} B^*(x)\, dx,$$

and it can be shown that the discrete version of Theorem 1 holds for $B_i^*$.
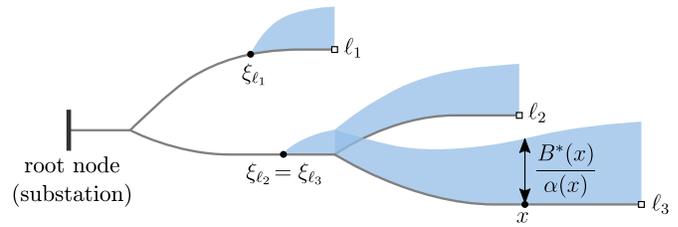


Fig. 2. An example of $B^*(x)/\alpha(x)$ on a radial network.



Fig. 3. An example of the location marginal value of storage $\mathrm{lmv}(x; B^*)$.

### B. Locational Marginal Value

**Theorem 2.** *Suppose $B_{\mathrm{tot}} < B_{\mathrm{m}}$. Then the locational marginal value function $\mathrm{lmv}(x; B^*)$ for the optimal capacity $B^*(x)$ exists and is unique. Furthermore, for any leaf node $\ell \in \mathcal{N}$,*

1) *$\mathrm{lmv}(x; B^*)$ is continuous for $x \in [0, \ell]$.*
2) *For any $x \in [\xi_\ell, \ell]$,*

$$\mathrm{lmv}(x; B^*) = -F^{*\prime}(B_{\mathrm{tot}}).$$

3) *$\mathrm{lmv}(x; B^*)$ is strictly increasing on $[0, \xi_\ell]$.*
4) *$\mathrm{lmv}(0; B^*) = 0$.*

Theorem 2 gives structural properties of the locational marginal value of storage with optimal storage placement. It shows that the locational marginal values are equalized over the entire network where $B^*(x) > 0$; otherwise we can further decrease the total cost by moving storage in regions with lower values to regions with the highest value. Then on $[0, \xi_\ell]$ where there is no storage deployed, the locational marginal value is strictly increasing from the substation towards $\xi_\ell$; this suggests that if we want to further add a small amount of storage to new places, it should be placed far from the substation to achieve higher reduction in the objective. Moreover, with optimal storage placement, the locational marginal values at places with $B^*(x) > 0$ will always be higher than those at places with $B^*(x) = 0$.

Figure 3 shows a typical example of the locational marginal value function $\mathrm{lmv}(x; B^*)$ on the path $[0, \ell]$, where $\ell$ is an arbitrary leaf node.

On the other hand, since locational marginal values only give first order approximation of the functional $J^*(B)$,

$$J^*(B^* + \delta) \approx J^*(B^*) - \int_{\mathcal{N}} \delta(x)\, \mathrm{lmv}(x; B^*)\, dx,$$

decisions based purely on locational marginal values will in general be approximations of the true optimal decisions. For example, suppose we want to add a small amount of storage to the network. If we only base our decision on the current locational marginal values $\mathrm{lmv}(x; B^*)$, then it seems to suggest that the extra storage can be arbitrarily placed in

regions with $B^*(x) > 0$ because those regions have the *same* and highest locational marginal values. Although the resulting cost $J^*$ of such placement will be close to the optimal cost $F^*$, the placement itself can be very different from the optimal placement. In order to determine the optimal placement, we may need second order or even higher order information of $J^*$.

There is also a discrete counterpart of Theorem 2 that holds for discrete networks.

### C. Optimal Schedule

**Theorem 3.** *Suppose $B_{\text{tot}} < B_{\text{m}}$. Let $x \in \mathcal{N}$ be arbitrary with $B^*(x) > 0$. Then there exist $D(x)$ disjoint open intervals*

$$(\tau_d^l(x), \tau_d^r(x)) \subset \mathbb{T}, \quad d = 1, \ldots, D(x)$$

*for some $D(x) \in \mathbb{N}$, satisfying*

$$\{t_m^{\max}, t_m^{\min}\}_{m=0}^{M-1} \subset \bigcup_{d=1}^{D(x)} (\tau_d^l(x), \tau_d^r(x)),$$

*and $D(x)$ offsets $\delta_d(x) \in \mathbb{R}, d = 1, \ldots, D(x)$ such that*

1) *The optimal schedule satisfies*

$$\frac{\partial b^*(x,t)}{\partial t} = \begin{cases} \alpha(x)(p(t) - \delta_d(x)), & t \in (\tau_d^l(x), \tau_d^r(x)), \\ 0, & \text{otherwise,} \end{cases}$$

*and*

$$\sum_{m=0}^{M-1} \left( \frac{\partial b^*}{\partial t}(x, t_m^{\max}) - \frac{\partial b^*}{\partial t}(x, t_m^{\min}) \right)$$

$$= \alpha(x) \sum_{m=0}^{M-1} (p(t_m^{\max}) - p(t_m^{\min})).$$

2) *For any $t \notin \bigcup_{d=1}^{D(x)} (\tau_d^l(x), \tau_d^r(x))$,*

$$b^*(x,t) = \begin{cases} B^*(x), & \dot{p}(t) < 0, \\ 0, & \dot{p}(t) > 0. \end{cases}$$

This paper focuses on the planning problem rather than the real-time operation of storage. In addition, the optimal schedule derived here is non-causal and cannot be directly used for real-time operation of storage. Nevertheless, some of its structures may be useful for designing real-time control policies of storage.

Theorem 3 shows that, the time horizon $\mathbb{T}$ can be divided into several intervals, and on each interval $\partial b^*(x,t)/\partial t$ is either zero or follows the background injection $\tilde{p}(x,t)$ except for an offset. When $\partial b^*(x,t)/\partial t$ does not follow the back ground injection, the SoC $b^*(x,t)$ either is zero (when $p(t)$ is increasing) or reaches full capacity (when $p(t)$ is decreasing). We see that the optimal schedule is a load shifting policy that shaves the peaks and fills the valleys.

Figure 4 shows a typical example of the optimal schedule which illustrates Theorem 3, where we have $D(x) = 5$.

It should be noted that, for discrete networks, the structure of the optimal schedule is more complicated than in the
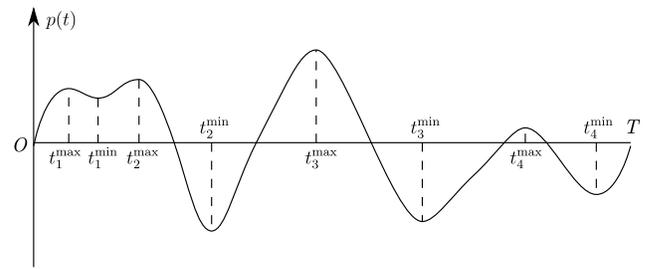


(a) $p(t)$ and its maximum and minimum points.
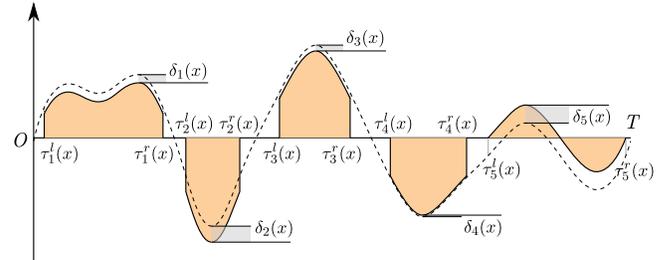


(b) Illustration of $\tau_d^l(x)$, $\tau_d^r(x)$ and $\delta_d(x)$. The dashed curve is $p(t)$ and the solid curve is $(\alpha(x))^{-1} \partial b^*(x,t)/\partial t$.
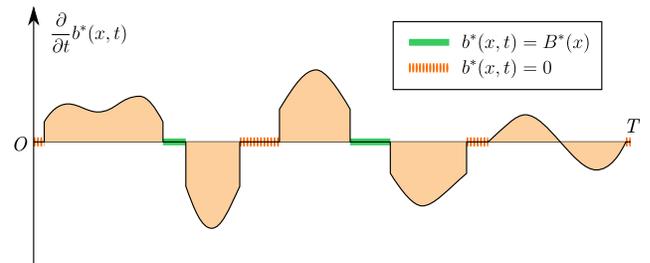


(c) The optimal schedule $\partial b^*(x,t)/\partial t$. We have $b^*(x,t) = B^*(x)$ on the intervals marked by solid green bars and $b^*(x,t) = 0$ on the intervals marked by striped orange bars.

Fig. 4. An example of a typical charging and discharging curve $\partial b^*(x,t)/\partial t$.

continuous case. By employing (8), the optimal schedule for bus $i$ will be given by

$$\frac{\partial b_i^*(t)}{\partial t} = \lim_{n \to \infty} \int_{h_i(x;n) \geq n^{-1}} \frac{\partial b^*(x,t)}{\partial t} \, dx.$$

It can be shown by theory and simulation that for each $i$, the sequence of sets

$$\{\tau_d^l(x) : h_i(x;n) \geq n^{-1}\}, \quad n = 1, 2, \ldots,$$

will in general converge to a non-singleton set; in other words, the values of $\tau_d^l(x)$ over $\{x : h_i(x;n) \geq n^{-1}\}$ do not shrink to a single element as $n \to \infty$ (and similarly for $\tau_d^r$). Therefore the resulting $\partial b_i^*(t)/\partial t$ will be a mix of schedules with different $\tau_d^l$ and $\tau_d^r$, meaning that there will be transition periods between the intervals on which $\partial b_i^*(t)/\partial t = 0$ and the intervals on which $\partial b_i^*(t)/\partial t$ follows the background injection.

## V. SIMULATIONS

In this section, we present simulation results and check whether Theorem 1 holds or approximately holds when some of the restrictive assumptions are relaxed.
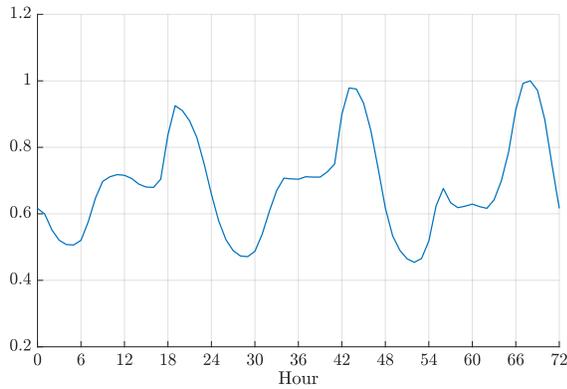
Fig. 5. Load profile with multiple peaks and valleys.

The network for simulation is a modified version of the IEEE 123 node test feeder [22]. We employ its network topology, line impedances, shunt capacitors and default switch settings, and simplify it as a single-phase network with all constant power loads. We also adopt two of the voltage regulators and fix their taps, so that the voltage regulator near the substation has tap ratio 1 : 1.04375 and the one near bus 60 has tap ratio 1 : 1.03125.

The network will be treated as a discrete network. The objective is still minimizing the total energy loss with a limited storage budget given by $B_{\text{tot}}$. We denote the optimal capacity at bus $i$ by $B_i^*$.

### A. Load Shapes with Multiple Extrema

We first show that the monotonicity of the scaled optimal capacity holds when $p(t)$ is allowed to have multiple maxima and minima. The load shape shown in Figure 5 is taken from Southern California Edison [23], which consists of (normalized) load consumptions spanning a 72-hour period, and has multiple peaks and valleys. We employ this load shape as $-p(t)$, and assume that at bus $i$, the background injection is given by $\alpha_i p(t)$, where the scaling factor $\alpha_i$ is the original load consumption of bus $i$ from the IEEE 123 node test case. To make results clearer, we add a small amount of load to each bus that originally has no load consumption, so that $\alpha_i > 0$ and the scaled optimal capacities are well defined for all $i$. The amount of load added is $1/4$ of the smallest original load consumption of a single bus. Approximation and linearization analogous to (3) and (4) are used in the simulation.

The scaled optimal capacities $B_i^*/\alpha_i$ are shown in Figure 6. It can be verified that $B_i^*/\alpha_i \geq B_j^*/\alpha_j$ whenever $B_j^* > 0$ and node $i$ is in the subtree rooted at node $j$. This suggests that the monotonicity of the scaled optimal capacity will hold without assuming that $p(t)$ has only one maximum and one minimum.

### B. Nonlinear Model with More General Load Profiles

In this part we run simulations to check whether the structural properties in Theorem 1 will hold when we use the more accurate nonlinear power flow model, and also allow different shapes of background injections.

Through our simulations, we find that how the background injection profiles at different nodes are correlated is crucial to the structure of the optimal storage placement, and if we allow arbitrary shapes of background injections for different buses, the structural properties in Theorem 1 may not hold. Figure 7 shows such a "counterexample", where the network has 31 buses connected as a line topology $0 - 1 - 2 - \cdots - 29 - 30$, with bus 0 being the substation and bus 30 being the leaf node. Injections at different buses have different shapes of background injections but the same variation $\max_{t_1,t_2 \in \mathbb{T}} \int_{t_1}^{t_2} (\tilde{p}_i(t) - \overline{p}_i)\, dt$ where $\overline{p}_i = T^{-1} \int_{\mathbb{T}} \tilde{p}_i(t)\, dt$. It can be seen that most of the storage is placed in the middle of the line rather than near the leaf.

However, if we consider the case where the background injections at each bus do not precisely have a common shape, but are given by

$$\tilde{p}_i(t) = \alpha_i(p(t) + \delta p_i(t)),$$

where $\delta p_i(t)$ represents some small deviation from the assumption of uniform shape, then simulation shows that, the structural properties in Theorem 1 approximately hold.

In this simulation, for each bus $i$, we generate $\delta p_i(t)$ from an interpolated Gaussian random noise originally sampled every 2 hours, and for each bus $i$, the stochastic processes $(\delta p_i(t))_{t \in \mathbb{T}}$ are independent of each other and identically distributed. We also scale the deviations by a common factor so that

$$\frac{\max_t |\delta p_i(t)|}{\max_t p(t) - \min_t p(t)} \leq 1/3$$

for all $i$. Figure 8 gives a typical deviation pattern for all the buses.

We employ the full nonlinear discrete DistFlow model to do the simulation in this part. The reactive background injection follows the same shape as the real background injection at each bus, and we adopt the ratio of reactive load to real load from the original IEEE 123 bus test case. We use the second-order cone relaxation developed in [24]–[26] to find the optimal solutions. All relaxations have been checked to be exact.

Figure 9 shows the scaled optimal capacities for a particular instance of $\delta p_i(t)$. It can be seen that the storage should still be distributed near the leaf nodes and far from the substation, and a threshold point can still be found for each path connecting the substation and a leaf node that divides the path into a section with no storage and a section with storage everywhere. This suggests that the first part of Theorem 1 is still a good characterization of the optimal placement in the more realistic situations. On the other hand, the monotonicity of $B_i^*/\alpha_i$ does not hold strictly, but is not violated severely either (especially when $B_{\text{tot}}$ is small); we can still find some monotonic pattern which roughly characterizes the optimal placement. This suggests that the second part of Theorem 1 does not strictly hold in the situations where the background injections deviate from the common load shape and the nonlinear power flow model is used, but is an approximate characterization of the true optimal placement strategy. Further simulations suggest that it will be a satisfactory approximation as long as $\max_t |\delta p_i(t)|$ are small and $B_{\text{tot}}$ is small.
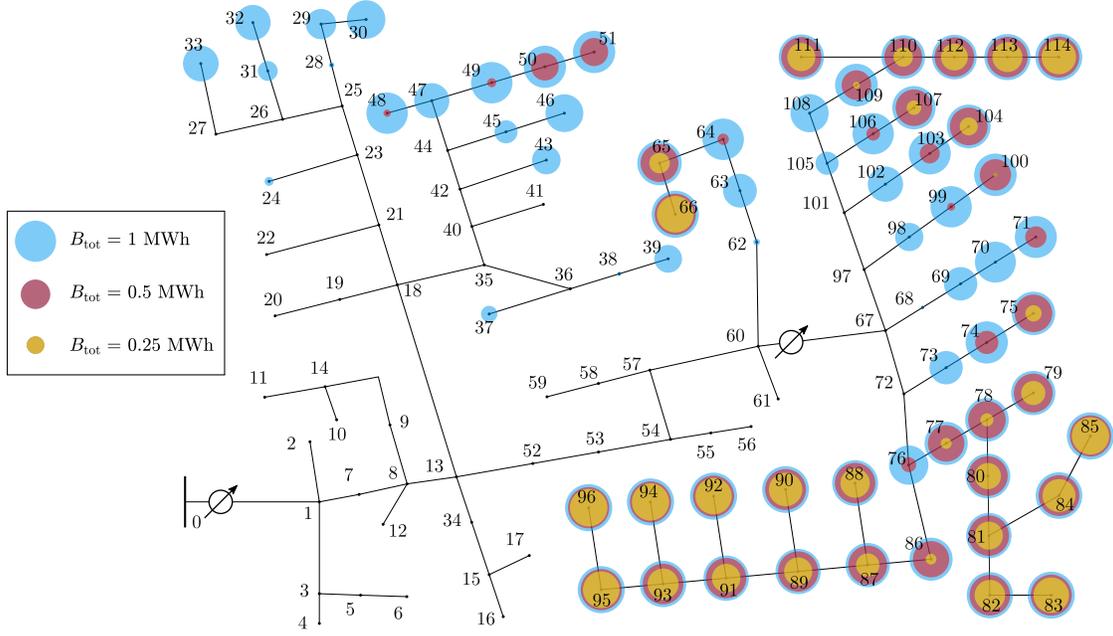
Fig. 6. Optimal storage placement for the modified IEEE 123 node test feeder with linearized power flow when $p(t)$ has multiple peaks and valleys. The radius of each colored solid circle is proportional to $B_i^*/\alpha_i$, and different colors correspond to different $B_{\text{tot}}$.
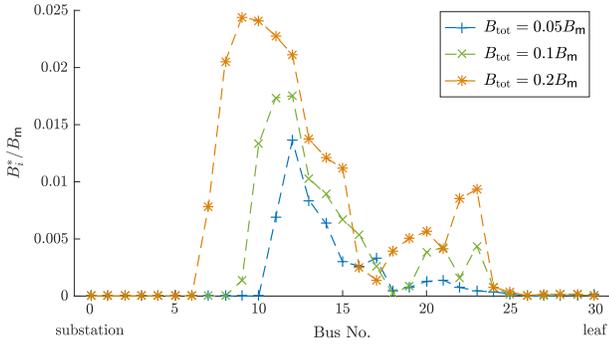


Fig. 7. A "counterexample" of Theorem 1 when different nodes have different shapes of background injections.
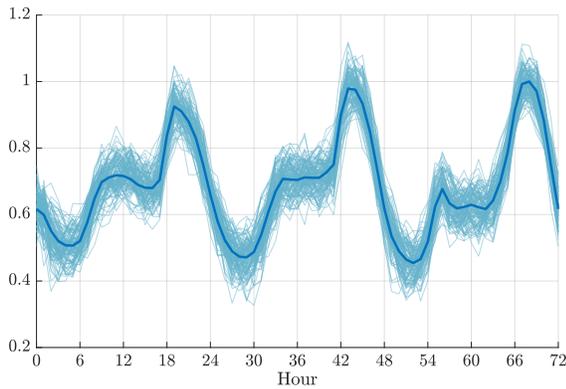


Fig. 8. Load profiles that deviate from the common load shape. The light-colored curves represent $-(p(t) + \delta p_i(t))$.

## C. Applying Monotone Placement to More Realistic Situations

In this part, we consider the situation where we apply the optimal capacity obtained from the simplified formulation (5) to more realistic situations and evaluate its performance.

The simulation procedure involves two stages: the planning stage and the operating stage. In the planning stage, we assume linearized DistFlow and the same shape of background injections, and solve (5) to get the optimal capacity, just as what we did in Section V-A. The resulting optimal storage capacity will be denoted by $\hat{B}_i^*$, and we apply this monotone optimal capacity to the network. Then in the operating stage, we assume nonlinear power flow model and that the background injections deviate from the common shape in the form $\tilde{p}_i(t) = \alpha_i(p(t) + \delta p_i(t))$ as in Section V-B. We fix the storage placement $\hat{B}_i^*$ and optimize only over the charging schedule $b_i(t)$. After we obtain the optimal charging schedule, we calculate the corresponding reduction in total energy loss.

The above procedure is then repeated, but we instead assume nonlinear power flow model and background injections of the form $\tilde{p}_i(t) = \alpha_i(p(t) + \delta p_i(t))$ in the planning stage, just as what we did in Section V-B. The operating stage will remain the same, and we calculate the resulting reduction in total energy loss, which is the optimal loss reduction we can achieve.

Finally, we compare these two loss reductions to see if they are close to each other.

We carry out this experiment on three instances of $\delta p_i(t)$, and the results are shown in Table I. It can be seen that, the differences between the loss reduction with $\hat{B}_i^*$ and the optimal loss reduction are very small, showing that $\hat{B}_i^*$ is a very good suboptimal solution to the more realistic situations with load profile deviations and nonlinear DistFlow model.
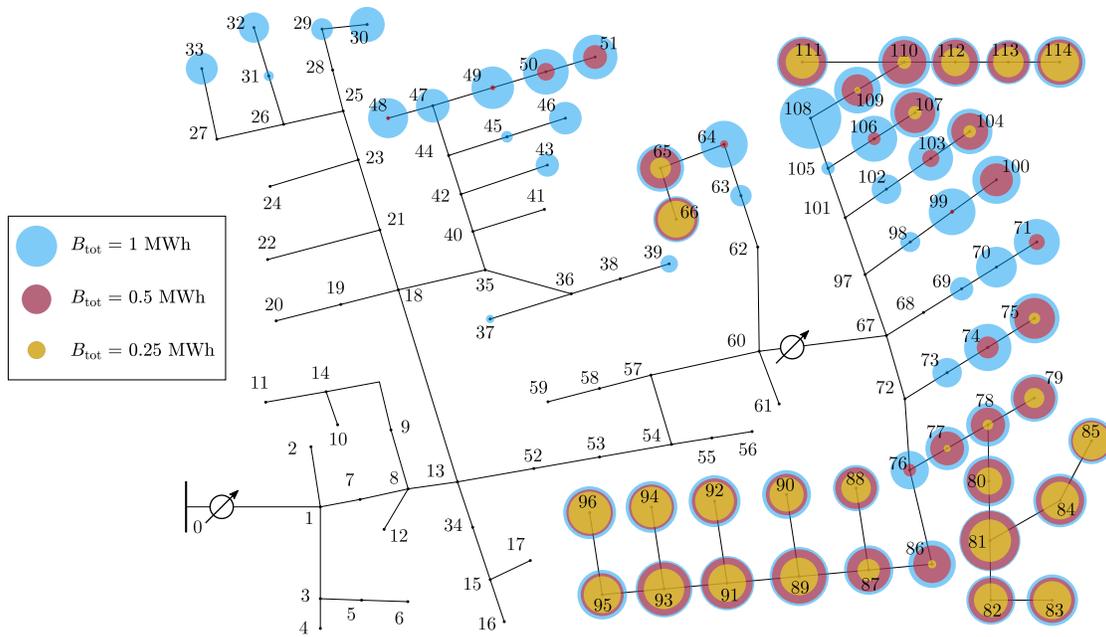
Fig. 9. Optimal storage placement for the modified IEEE 123 node test feeder with nonlinear DistFlow model when $\tilde{p}_i(t)$ deviates from the common load shape. The radius of each colored solid circle is proportional to $B_i^*/\alpha_i$, and different colors correspond to different $B_{\text{tot}}$.

TABLE I
COMPARISON OF LOSS REDUCTION

| $B_{\text{tot}} = 1$ MWh | | |
|---|---|---|
| | Loss reduction with $\hat{B}_i^*$ | Optimal loss reduction |
| Instance 1 | 45.457 kWh | 45.484 kWh |
| Instance 2 | 45.037 kWh | 45.054 kWh |
| Instance 3 | 46.303 kWh | 46.323 kWh |
| $B_{\text{tot}} = 0.5$ MWh | | |
| | Loss reduction with $\hat{B}_i^*$ | Optimal loss reduction |
| Instance 1 | 32.123 kWh | 32.148 kWh |
| Instance 2 | 31.828 kWh | 31.846 kWh |
| Instance 3 | 32.546 kWh | 32.556 kWh |
| $B_{\text{tot}} = 0.25$ MWh | | |
| | Loss reduction with $\hat{B}_i^*$ | Optimal loss reduction |
| Instance 1 | 19.639 kWh | 19.640 kWh |
| Instance 2 | 19.485 kWh | 19.489 kWh |
| Instance 3 | 19.880 kWh | 19.882 kWh |

These simulation results suggest that, although in practice the background injections at each bus do not exactly follow the same shape and the nonlinear DistFlow model is more accurate in characterizing the power flow, Theorem 1, which is derived by assuming all loads have the same shape and using the linearized DistFlow model, is still very useful for practical applications.

## VI. CONCLUSION

We study the problem of optimal placement and sizing of energy storage in distribution networks. We model a distribution network as a continuous tree with the linearized DistFlow model, and formulate the problem as an optimal

power flow problem. The structure of the optimal solution has been analyzed when all loads have the same shape, which demonstrates that storage devices should be placed near the leaves of the network and far from the substation, and that the scaled capacity increases towards the leaves. Then, under optimal storage placement, the locational marginal value of storage is increasing as we move from the substation towards a leaf node and reaches the highest value in places with storage deployed. Moreover the locational marginal value is equalized over the entire network where storage is deployed. Simulations show that these structural properties still hold (or approximately hold) when some of the modeling assumptions are relaxed.

Some interesting future directions are as follows.

1) We have proved that $B^*(x)$ increases and $\xi_\ell$ moves towards the substation as $B_{\text{tot}}$ increases for line networks, and simulation suggests that this also holds for general radial networks. It will be interesting to formally prove this property for radial networks.

2) We have made some simplifying assumptions in this paper. It will be interesting to generalize the theory here to more realistic models. We are especially interested in a theory with line flow constraints like $|S(x)| \le S_{\max}$, and capacity constraints like $B(x) \le B_{\max}(x)$ where $B_{\max}(x)$ is a predetermined function.

3) As mentioned in previous sections, through simulations, we have found that the pattern of the optimal storage placement is closely related to how the background injection profiles at different nodes are correlated spatially. Here we use the simplest assumption that they have the same shape. It will be interesting to study if more complicated spatial structure of background injections

can lead to other or more general patterns for the problem of optimal storage placement and sizing.

## REFERENCES

[1] Y. Kanoria, A. Montanari, D. Tse, and B. Zhang, "Distributed storage for intermittent energy sources: Control design and performance limits," in *Proc. 49th Allerton Conf. Communication, Control, and Computing*, 2011, pp. 1310–1317.

[2] Y. M. Atwa and E. F. El-Saadany, "Optimal allocation of ESS in distribution systems with a high penetration of wind energy," *IEEE Trans. Power Systems*, vol. 25, no. 4, pp. 1815–1822, Nov. 2010.

[3] H. Pandzic, Y. Wang, T. Qiu, Y. Dvorkin, and D. S. Kirschen, "Near-optimal method for siting and sizing of distributed storage in a transmission network," *IEEE Trans. Power Systems*, vol. 30, no. 5, pp. 2288–2300, Sept. 2015.

[4] S. Bose, D. F. Gayme, U. Topcu, and K. M. Chandy, "Optimal placement of energy storage in the grid," in *Proc. 51st IEEE Conf. Decision and Control (CDC)*, 2012, pp. 5605–5612.

[5] A. Castillo, X. Jiang, and D. F. Gayme, "Lossy DCOPF for optimizing congested grids with renewable energy and storage," in *Proc. American Control Conf.*, 2014, pp. 4342–4347.

[6] D. Gayme and U. Topcu, "Optimal power flow with large-scale storage integration," *IEEE Trans. Power Systems*, vol. 28, no. 2, pp. 709–717, May 2013.

[7] C. Thrampoulidis, S. Bose, and B. Hassibi, "Optimal placement of distributed energy storage in power networks," *IEEE Trans. Automatic Control*, vol. 61, no. 2, pp. 416–429, Feb. 2016.

[8] J. Qin, I. Yang, and R. Rajagopal, "Submodularity of energy storage placement in power networks," in *Proc. 55th IEEE Conf. Decision and Control (CDC)*, 2016, pp. 686–693.

[9] S. Bose and E. Bitar, "Variability and the locational marginal value of energy storage," in *Proc. 53rd IEEE Conf. Decision and Control (CDC)*, 2014, pp. 3259–3265.

[10] M. E. Baran and F. F. Wu, "Optimal sizing of capacitors placed on a radial distribution system," *IEEE Trans. Power Delivery*, vol. 4, no. 1, pp. 735–743, Jan. 1989.

[11] ——, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Trans. Power Delivery*, vol. 4, no. 2, pp. 1401–1407, Apr. 1989.

[12] D. Wang, K. Turitsyn, and M. Chertkov, "Distflow ODE: Modeling, analyzing and controlling long distribution feeder," in *Proc. 51st IEEE Conf. Decision and Control (CDC)*, 2012, pp. 5613–5618.

[13] Y. Tang and S. H. Low, "Optimal placement of energy storage in distribution networks using continuous models," in preparation.

[14] B. A. Smith, J. Wong, and R. Rajagopal, "A simple way to use interval data to segment residential customers for energy efficiency and demand response program targeting," in *ACEEE Summer Study on Energy Efficiency in Buildings*, vol. 5, 2012, pp. 374–386.

[15] J. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 420–430, Jan. 2014.

[16] C. Flath, D. Nicolay, T. Conte, C. van Dinther, and L. Filipova-Neumann, "Cluster analysis of smart metering data," *Business & Information Systems Engineering*, vol. 4, no. 1, pp. 31–39, 2012.

[17] J. A. Taylor and F. S. Hover, "Convex models of distribution system reconfiguration," *IEEE Trans. Power Systems*, vol. 27, no. 3, pp. 1407–1413, Aug. 2012.

[18] Y. Tang and S. H. Low, "Optimal placement of energy storage in distribution networks," in *Proc. 55th IEEE Conf. Decision and Control (CDC)*, Dec. 2016.

[19] M. A. Kashem and M. Moghavvemi, "Maximizing radial voltage stability and load balancing via loss minimization in distribution networks," in *Proc. Int. Conf. Energy Management and Power Delivery*, vol. 1, 1998, pp. 91–96.

[20] S. H. Low, "Convex relaxation of optimal power flow, I: Formulations and equivalence," *IEEE Trans. Control of Network Systems*, vol. 1, no. 1, pp. 15–27, Mar. 2014.

[21] D. G. Luenberger, *Optimization by Vector Space Methods*. John Wiley & Sons, 1969.

[22] W. H. Kersting, "Radial distribution test feeders," in *Proc. IEEE Power Engineering Soc. Winter Meeting*, vol. 2, 2001, pp. 908–912.

[23] [Online]. Available: https://www.sce.com/wps/portal/home/regulatory/load-profiles/dynamic-load-profiles/

[24] M. Farivar, C. R. Clarke, S. H. Low, and K. M. Chandy, "Inverter var control for distribution systems with renewables," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, 2011, pp. 457–462.

[25] M. Farivar and S. H. Low, "Branch flow model: Relaxations and convexification (parts I, II)," *IEEE Trans. Power Systems*, vol. 28, no. 3, pp. 2554–2572, Aug. 2013.

[26] L. Gan, N. Li, U. Topcu, and S. H. Low, "Exact convex relaxation of optimal power flow in radial networks," *IEEE Trans. Automatic Control*, vol. 60, no. 1, pp. 72–87, Jan. 2015.
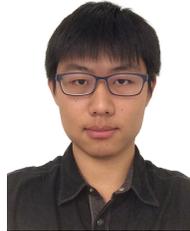
**Yujie Tang** (S'17) received the Bachelor's degree in electronic engineering from Tsinghua University, Beijing, China, in 2013, and is currently pursuing the Ph.D. degree in electrical engineering at California Institute of Technology, Pasadena, CA, USA.

His research interests include optimization and control of power systems, with a particular focus on the planning and control of distributed energy resources in smart grids.
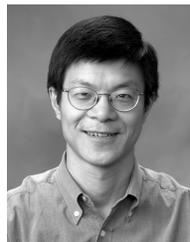
**Steven H. Low** (F'08) received the B.S. degree from Cornell University, Ithaca, NY, USA, in 1987, and the Ph.D. degree from the University of California, Berkeley, CA, USA, in 1992, both in electrical engineering.

He is a Professor of the Department of Computing & Mathematical Sciences and the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA. Before that, he was with AT&T Bell Laboratories, Murray Hill, NJ, and the University of Melbourne, Australia. His research on communication networks was accelerating more than 1TB of Internet traffic every second in 2014.

Prof. Low was a co-recipient of IEEE best paper awards. He was a member of the Networking and Information Technology Technical Advisory Group for the US President's Council of Advisors on Science and Technology (PCAST) in 2006.