Optimal Energy Trading with Battery Energy Storage under Dynamic Pricing

Xiaoqi Tan§, Yuan Wu†‡, Danny H.K. Tsang§
§Department of Electronic and Computer Engineering
Hong Kong University of Science and Technology, Hong Kong. E-mail:{xtanaa@ust.hk, eetsang@ece.ust.hk}
†College of Information Engineering, Zhejiang University of Technology, Hangzhou, China. E-mail: iewuy@zjut.edu.cn
‡State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China

***Abstract*—This paper proposes a mathematical framework for finding the optimal energy trading policy with battery energy storage (BES) under a dynamic pricing environment. We have previously shown that finding the arbitrage value of BES with known historical price data can be solved by iterative linear programming. The objective of the present paper is to show that, when the price information remains unknown, finding the optimal economic value of lifetime-constrained BES falls within the purview of *stochastic shortest path* problems, and the optimal policy presents the property of a *threshold structure*. To overcome the dimensionality difficulty, we propose a structure-based aggregation method, i.e., *Layer and Group*, to construct optimal trading policies. The elegance of this approach lies in its circumventing of the need for exhausted value iteration over the entire state space. Instead, the approach works in a hierarchical and parallel fashion, thus significantly speeding up the convergence to the optimality. Extensive experimental results show that this approach can dramatically reduce the computational complexity, thus contributing to the computationally tractable optimality without requiring any approximation. Numerical simulation also demonstrates the validity of the proposed framework, and various trading insights for practical BES systems have been formed.**

I. Introduction

Battery Energy Storage (BES) technologies have been receiving increasing attention in recent years due to their widespread functionalities. Energy trading with BES is a functionality of BES linked with commercial trading of energy, and it is playing an increasingly important role in the modern deregulated electricity market. For instance, in cities such as San Francisco and New York, where power is costly, large commercial buildings are installed with batteries in their basements, for the purpose of buying and storing electricity at night when prices are low, and tapping into the batteries during peak afternoons when prices are high [1]. However, without an appropriate configuration before deployment and a beneficial operation after deployment, the applications of BES will be hindered by both the high capital cost and the uncertain economic value. This situation will become even worse when the limited lifetime of BES (which is typically unknown or not exactly known by the BES operator) is taken into account in addition to its ramp and capacity constraints [2].

The existing extensive studies have covered a large dimension of the energy storage problems: renewable energy integration into the smart grid supported by BES (e.g., smoothing out the intermittent power from wind farms [4]), optimal power cost management when using BES as an energy backup (e.g., in the form of UPS) in data centers [6], and *arbitrage* analysis (i.e., buying low and selling high) of BES [3], [5], [7]. In particular, the authors of [5] note that an increasing number of retail energy markets show price fluctuations, which provides end users with the opportunity to buy energy at lower than average prices. Based on this observation, the authors study a battery control policy which minimizes the total discounted costs. Furthermore, the optimal policy is shown to be threshold-structured, and they derive these thresholds in a few special cases. However, when the battery lifetime is taken into account, as what we have done in [3], the policy structure can be inevitably changed, but this, however, has been ignored by most of the existing studies.

Based on the state-of-the-art *Ah-throughput* model [2], we have previously shown that the optimization problem of maximizing the economic value when directly linking the battery lifetime to detailed operational behaviors can be solved by iterative linear programming based on given historical price data [3]. In this paper, however, we study a more general case where the price is stochastic. We show that the problem falls within the purview of Stochastic Shortest Path (SSP) problems. Different from most of the existing battery models, such in [5]-[7], we explicitly take the lifetime issue into account, and propose to *leverage the SSP model to capture the random expiration time as well as the stochastic nature of the market price*. As one of the main results, we establish a threshold-structured policy which extends the similar policy structure in [5]. The policy structure in this paper, however, shows some more interesting insights due to the consideration of battery lifetime. Another main contribution of this paper is that we discover the underlying layered and grouped structure of the state space of the proposed SSP problem. Based on this observation, we propose an elegant dynamic programming based algorithm which significantly outperforms the traditional value iteration approach. *Note that our algorithm is still optimal since no approximation is introduced at all*. Meanwhile, it is worth pointing out here that the proposed dynamic programming based algorithm is general to some extent, which could make it applicable to many other SSP related problems.

The rest of this paper is organized as follows. In Sec. II, we introduce the BES model and the operational objective. In Sec. III, we formulate the BES trading problem as an SSP problem. Subsequently, in Sec. IV, a computationally tractable algorithm based on *Layer and Group* is proposed, and in Sec. V, we analyze the economic value and introduce some system

This work was supported, in part, by the Hong Kong Research Grants Council's General Research Fund (Project No. 619312 and 16209814) and Specialized Research Fund for the Doctoral Program of Higher Education (Project No. 20133317120002), MOE. P.R. China.

978-1-4799-4934-2/14/$31.00 ©2014 IEEE

implementation insights based on our extensive experimental results. The paper is concluded in Sec. VI.

II. The Model

We consider the general system model in Fig. 1. BES operators obtain system parameters including the energy level b_t, the real-time price p_t and the remaining throughput θ_t (definition comes later) through a Control and Measurement Center (CMC). They then make decisions on buying electricity c_t from the real-time market to charge the BES with $\eta_c c_t$, or discharge the BES with $\frac{1}{\eta_d} d_t$ to sell electricity d_t back to the market, where $\eta_c, \eta_d \in (0, 1]$ are charging and discharging efficiencies, respectively. We consider that sequential decisions are made periodically over the time horizon denoted by $\mathcal{T} = \{0, 1, ...\}$. Let $t \in \mathcal{T}$ denotes the discrete time index corresponding to the decision epoch for time interval $(t, t+1]$ (e.g., one hour in this paper).

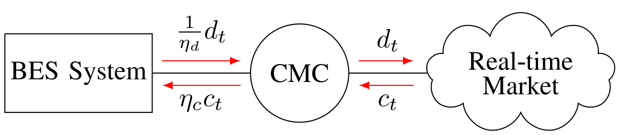


Fig. 1. Illustration of the system model. We assume that the BES operator is a price-taker, thus rigorously eliminating any room for market manipulation.

Based on the system model in Fig. 1, we consider the parameters for a general BES and the objective for BES operators with the notations specified in the following two subsections.

A. Charging/Discharging Model of BES

We adopt the same BES model as [3]. We use s_t to denote the current capacity of the BES at time t. The energy level b_t evolves according to $b_{t+1} = b_t + \eta_c c_t - \frac{1}{\eta_d} d_t$, which is always bounded within the region $[\underline{b}_t, \overline{b}_t]$, where $\underline{b}_t = \gamma_1 s_t, \overline{b}_t = \gamma_2 s_t$ with γ_1, γ_2 determined by the *depth of discharge*. We define $\mu_t = \eta_c c_t - \frac{1}{\eta_d} d_t$ as the net energy flow through the BES (*i.e.*, $c_t = \frac{\max\{0, \mu_t\}}{\eta_c}, d_t = \eta_d \max\{0, -\mu_t\}$). Note that both c_t and d_t are bounded by the power rating, i.e., $c_t \in [0, c^{\max}]$ and $d_t \in [0, d^{\max}]$, where $c^{\max}$ and $d^{\max}$ are the charging and discharging power rating, respectively.

As one of the methods for battery lifetime modeling, the *Ah-throughput* model assumes that there is a *fixed amount of energy* that can be cycled through a battery before it requires replacement (details can be found in [2], [3]). We denote this fixed amount of energy by the initial throughput θ_m and use θ_t to denote the remaining throughput at time t during usage $(t \geq 1)$. Therefore, we have

$$\theta_t = \theta_m - \sum_{\tau=0}^{t-1} (\eta_c c_\tau + d_\tau / \eta_d). \quad (1)$$

We illustrate the feasible charging/discharging region in Fig. 2. It can be noted that for each point (or action) in the shaded area, we always have $c_t \cdot d_t = 0$, which is consistent with the reality that being charged and discharged at the same time is suboptimal. Therefore, the battery can only be in one of the working modes (i.e., charge, discharge or idle mode).The two important points $K_{c^{\max}}$ and $K_{d^{\max}}$ are determined by the charge power rating $c^{\max}$ and discharge power rating $d^{\max}$, respectively. The shaded area is the feasible action region (denoted by $\mathcal{A}(b_t, \theta_t, p_t)$, which will be specified in the next section). Note that the feasible action region $\mathcal{A}(b_t, \theta_t, p_t)$ is a convex polyhedron, which actually serves as the action space in our following SSP formulation.

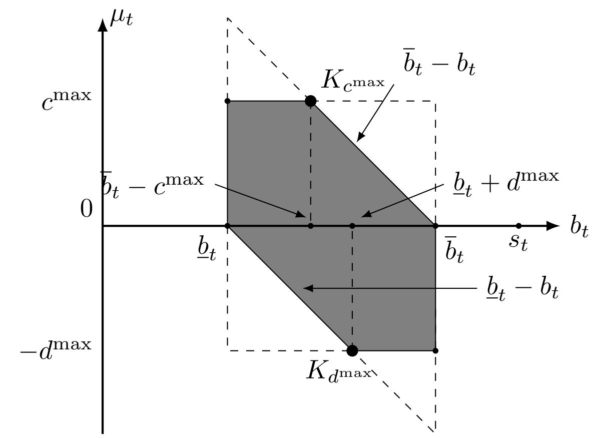


Fig. 2. Illustration of the BES model and the feasible action region. In order to avoid unnecessary clutter, we assume that the charging and discharging efficiency $\eta_c = \eta_d = 1$ and the remaining throughput θ_t is large.

B. Objective for the BES Operator

The objective function of the BES operator is modeled as the total economic reward made by selling electricity minus the total expense of buying electricity from the market and the holding cost. Thus, the reward $\mathcal{R}(b_t, \theta_t, p_t, \mu_t)$ in period t is a function of the four tuple $(b_t, \theta_t, p_t, \mu_t)$, and it can be formulated as

$$\begin{aligned}\mathcal{R}(b_t, \theta_t, & p_t, \mu_t) \\ &= \Big\{ (p_t - \frac{\alpha}{\eta_d}) d_t - (p_t + \alpha \eta_c) c_t - h_t b_t \Big\} \cdot \mathbb{I}_{\{\theta_t > 0\}},\end{aligned}$$

where the cost of *per unit electricity use* of the BES is denoted by $\alpha = \frac{\text{Capital Cost}}{\theta_m}$, which is a proportional coefficient mapping the charge and discharge profile into the monetary cost. The parameter h_t denotes the holding cost of BES at slot t [7], which depends on working conditions like weather, temperature, etc. Both α and h_t are assumed to be known by the operators[1]. Note that $\mathbb{I}_{\{\cdot\}}$ is an indicator function. The case of $\theta_t = 0$, having zero reward, means no profit for the dead BES, and the operator is assumed to be obligated to replace it.

III. SSP Formulation and Policy Structure

An SSP problem is a special case of the expected total-reward (or expected total-cost) infinite-horizon Markov decision process (MDP) problems [12]. For SSP problems, there is a (or a set of) special cost-free termination state. Once the

[1]One of the extensions of the current model in our ongoing work is to further consider h_t as a predictable random variable, e.g., by means of temperature and humidity forecast. Note that the fundamental SSP model will not be changed.

system reaches the termination state, it stops evolving, and thus no further system cost will be incurred. We next present the formal SSP formulation in subsection III-A.

A. Formal SSP Formulation

System state: the system at each slot t can be characterized by a three tuple $\boldsymbol{\omega}_t = (b_t, \theta_t, p_t) \in \Omega = \mathcal{B} \times \Theta \times \mathcal{P}$. $\mathcal{B} = \{\underline{b}, \underline{b}+\Delta, \underline{b}+2\Delta, \cdots, \overline{b}\}$ is the state space for the battery energy level and $\Theta = \{0, \Delta, 2\Delta, 3\Delta, \cdots, \theta_m\}$ is the state space for the remaining throughput. $\mathcal{P}$ will be specified later in this subsection. We use Δ to denote the discretization step-size, and without loss of generality, we assume θ_m is completely divisible by Δ. We drop the time index t and use $\boldsymbol{\omega} = (b, \theta, p)$ to denote the current system state, for simplicity.

Action space: given the system state $\boldsymbol{\omega} = (b, \theta, p)$, the operator determines its *charge*, *discharge* or *idle* action $\mu \in \mathcal{A}(\boldsymbol{\omega})$. Recall that given a specific system state $\boldsymbol{\omega}$, the corresponding action space $\mathcal{A}(\boldsymbol{\omega})$ is the shaded area, as shown in Fig. 2. Mathematically, it is given by

$$\mathcal{A}(\boldsymbol{\omega}) = \left\{ \mu \Big| -\frac{\min\left\{b - \gamma_1 s, d^{\max}, \theta\right\}}{\eta_d} \leq \mu \leq \eta_c \min\left\{\gamma_2 s - b, c^{\max}, \theta\right\} \right\}, \tag{2}$$

where s is the current battery capacity. (Note that s_t is decaying during the usage, and in this paper we use the same decaying model as [3].) Note that when $\theta = 0$, we have $\mathcal{A}(b, 0, p) = \{0\}$, which is consistent with the *Ah-throughput* model.

Exogenous process: we denote p_t as a real-time price during period $(t, t+1]$, and assume that for all t, $p_t \in [p^{\min}, p^{\max}]$. The price space is further discretized into $\mathcal{P} = \{p^{\min}, p^{\min} + \Delta_p, p^{\min} + 2\Delta_p, \cdots, p^{\max}\}$ with step size Δ_p. The price is declared at the beginning of each interval $(t, t+1]$ and remains constant during that interval. We further assume that at each slot, the electricity price evolves according to a distribution which may only depend on the price in the current time slot (i.e., Markovian). Thus, we can define a Markov chain with state space $\mathcal{P}$ and transition probability given by $\vartheta_{pp'} = \Pr\{p_{t+1} = p' | p_t = p\}, p, p' \in \mathcal{P}$. We adopt the same assumption for modeling the exogenous electricity price process as [8] by introducing two properties for the Markov chain we define above: 1) *Markov-Contractivity* in the mean and 2) *Stochastic Monotonicity*. Basically, the *Markov-Contractivity* captures the fact that prices tend to be mean reverting and the *Stochastic Monotonicity* describes the "stickiness" of the price: a low price at time t is more likely to lead to a low price at time $t+1$, and, likewise, a high price at time t is more likely to lead to a high price at time $t+1$. Both of these definitions are relatively mild conditions for representing simple forms of price dynamics [8]. Similar approaches also appear in [9] and [10].

Transition kernel: the transition kernel reflects the uncertainty in the SSP problem, and we represent the state transitons as the following conditional probability:

$$\begin{aligned}\Pr[\boldsymbol{\omega}' | \boldsymbol{\omega}, \mu] &= \Pr\Big[(b', \theta', p') | (b, \theta, p), \mu\Big] \\ &= \vartheta_{pp'} \cdot \mathbb{I}_{\{b' = b + \mu\}} \cdot \mathbb{I}_{\{\theta' = \theta - |\mu|\}}.\end{aligned} \tag{3}$$

Based on the parameters specified above, the formal definition of maximizing the expected total economic reward for the BES operator is given as follows:

$$\underset{\{\pi_0, \pi_1, \cdots, \pi_t, \cdots\}}{\text{maximize}} \quad \mathbb{E}\Big\{ \sum_{t=0}^{\mathcal{T}(\Pi)-1} \mathcal{R}(\boldsymbol{\omega}_t, \mu_t) | (b_0, \theta_0, p_0) \Big\}, \tag{4}$$

where $\Pi = \{\pi_0, \pi_1, \cdots, \pi_t, \cdots\}$ is the policy that mapps the state space into its action space. Note that *we maximize the reward over the time horizon which stops at epoch* $\mathcal{T}(\Pi) - 1$, *where* $\mathcal{T}(\Pi) = \arg\min_t \{\theta_t = 0\}$ *is a function of policy* Π *and also a random variable that characterizes the terminal time of the BES*. It is worth mentioning here that the intrinsic coupling of the operational policy and lifetime (or the termination time of this SSP problem, which is denoted by the random variable $\mathcal{T}(\Pi)$), makes the problem different from the traditional finite-horizon MDP problems, and it is also the key difference between our work and most of the existing literature.

B. Value Iteration and Optimal Trading Policy

We are interested in the optimal policy that can achieve the maximum expected reward when reaching the termination state. We have the following result regarding the optimal policy, and the details are illustrated in the following Theorem 1:

Theorem 1. *Problem (4) admits an optimal policy that is deterministic and stationary. For all* $\boldsymbol{\omega} \in \Omega$, *the optimal action* $\mu^* = \pi^*(\boldsymbol{\omega})$, *where* π^* *denotes the deterministic and stationary optimal policy. The optimal operational policies* Π^* *are given by* $\Pi^* = [\underbrace{\pi^*, \cdots, \pi^*, \cdots, \pi^*}_{\mathcal{T}(\Pi^*)}]$.

Proof. The proof of the stationary policy that is optimal and deterministic is the direct result following from Proposition 7.2.1 [12], and thus it is omitted here for brevity. □

We use $J(\boldsymbol{\omega})$ to denote the optimal reward starting from the system state $\boldsymbol{\omega}$ before termination. $J(\boldsymbol{\omega})$ is known to be the optimal cost-to-go function from the dynamic programming perspective. The optimal policy of Problem (4) satisfies the Bellman's equation as

$$J(\boldsymbol{\omega}) = \mathcal{R}\Big(\boldsymbol{\omega}, \mu^*\Big) + \mathbb{E}\Big[J\Big(b + \mu^*, \theta - |\mu^*|, p'\Big)\Big]. \tag{5}$$

According to the definition of SSP, the optimal value function $J(\boldsymbol{\omega}) = 0$ for all termination states. The optimal policy for Problem (4) can be derived through the standard value iteration algorithm, in which the optimal cost-to-go function $J(\boldsymbol{\omega})$ is iteratively determined by

$$J_n(\boldsymbol{\omega}) = \max_{\mu \in \mathcal{A}(\boldsymbol{\omega})} \Big\{ \mathcal{R}(\boldsymbol{\omega}, \mu) + \sum_{\boldsymbol{\omega}' \in \Omega} \Pr[\boldsymbol{\omega}' | \boldsymbol{\omega}, \mu] J_{n-1}(\boldsymbol{\omega}') \Big\}, \tag{6}$$

where n denotes the iteration index in the standard value iteration algorithm.

C. Threshold Structure Property of the Optimal Policy

Based on the value iteration in (6), by further leveraging the fact that the SSP problem is a finite horizon MDP problem in nature[2], the optimal policy we obtained from the SSP problem can be proved to present some threshold-structured properties.

Theorem 2. *In every time slot $t \in \mathcal{T}$, two energy levels, i.e., an upper threshold boundary $b_t^{(up)}(\theta, p)$ and a lower threshold boundary $b_t^{(low)}(\theta, p)$ exist. They serve as two thresholds to determine charge, discharge and idle with $b_t^{(low)}(\theta, p) \leq b_t^{(up)}(\theta, p)$. Specifically, the optimal policy π_t^* is given by*

$$\begin{cases} \frac{\min\left\{b_t - b_t^{(up)}(\theta,p), d^{\max}, \theta\right\}}{-\eta_d} & \text{if } b_t \in \left(b_t^{(up)}(\theta, p), \overline{b}_t\right], \\ 0 & \text{if } b_t \in \left[b_t^{(low)}(\theta, p), b_t^{(up)}(\theta, p)\right], \\ \frac{\min\left\{b_t^{(low)}(\theta,p) - b_t, c^{\max}, \theta\right\}}{1/\eta_c} & \text{if } b_t \in \left[\underline{b}_t, b_t^{(low)}(\theta, p)\right). \end{cases}$$

Proof. The establishment of this threshold-structured policy relies on transforming the original problem of deciding the optimal action μ^* into deciding the optimal energy level b^*, and further leveraging the piecewise linearity and concavity of the value function. Please refer to [13] for the detailed proof. □

Recall that in Fig. 2, point $K_{c^{\max}}$ will shift to the left hand side if the charge power rating $c^{\max}$ becomes larger, and point $K_{d^{\max}}$ will shift to the right hand side when the discharge power rating $d^{\max}$ becomes larger. Based on Theorem 2, we can see that these two shifts play very important roles in determining the threshold structure of the optimal policy. *Meanwhile, the remaining throughput θ also determines the policy by either deciding the action directly (e.g., consider when θ is smaller than $d^{\max}$ and $b_t - b_t^{(\text{up})}(\theta, p)$) or indirectly by influencing the threshold $b_t^{(\text{up})}(\theta, p)$ and $b_t^{(\text{low})}(\theta, p)$.* We will demonstrate all these observations by simulation in Sec. V, which turns out to present many important insights for real BES systems. Note that by rigorous mathematics, Theorem 2 extends the similar threshold property in [5] and other literature on the warehouse management area [11] to be more general. We believe the existence of the threshold-structured policy, even when taking the lifetime factor into account, brings more instructive insight into a real system's operation.

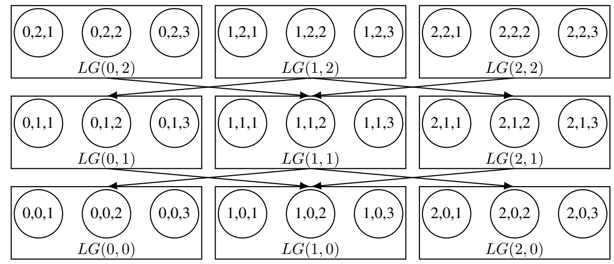
Although it is shown in [12] that the value iteration for our proposed SSP problem will converge to the optimal value within a finite number of iterations, this approach is unsuitable for systems with a large state space due to the famous "curse of dimensionality". Meanwhile, analytically calculating the two thresholds in Theorem 2 is very difficult in general (if not impossible). Therefore, we propose to further exploit a computationally tractable algorithm to solve Problem (4) in the next section.

IV. A Fast Algorithm based on Layer and Group

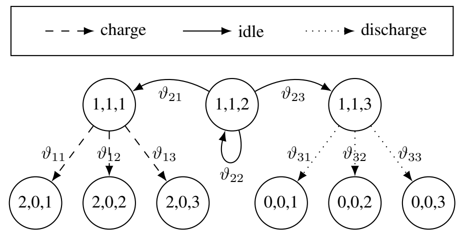
A. Overview of Layer and Group

We first briefly introduce how we aggregate the state space. For any states $\boldsymbol{\omega} = (b, \theta, p)$, according to the non-increasing property of the throughput during the usage of the BES, we can

[2]According to [12], SSP is a special case of the total expected cost infinite horizon MDP problem; however, due to the existence of absorbing states, SSP is finite in nature. Note that there is no contradiction since absorbing states are assumed to have zero cost.



(a) The Layer-and-Group state space subdivision.



(b) A feasible incomplete transition digram.

Fig. 3. An illustration of the layered and grouped structure of the proposed CSSP. Each state is denoted by (b, θ, p), where the state space for this simple example is $\{0,1,2\}\times\{0,1,2\}\times\{1,2,3\}$. The action space is $\mu \in \{-1, 0, 1\}$. We also illustrate an incomplete transition diagram for $LG(1, 1)$ when following the given policy shown in the subfigure (b). The Markov chain for the price transition is given by $\Pr(j|i) = \vartheta_{ij}, \forall i, j \in \mathcal{P} = \{1, 2, 3\}$.

layer the whole system state space into a hierarchical structure by their remaining throughput θ. Furthermore, for the states in the same layer, we can *group* them into a block structure by their current energy level b. If we regard our system state as (*physical state*, *information state*), then all the states from the same group have the same physical condition.

We exemplify this *Layer and Group* approach in Fig. 3. For the system state space (b, θ, p) denoted by $\{0, 1, 2\} \times \{0, 1, 2\} \times \{1, 2, 3\}$, we first layer the whole space into three layers according to their θ value. For example, all the states in *Layer 0* are termination states with zero throughput (i.e., $\theta = 0$), the states in *Layer 1* have remaining throughput 1, and it is 2 for the states in *Layer 2*. Furthermore, for all the states in the same layer, we group them according to their energy level b. For example, the leftmost group in *Layer 1* consists of three states with energy level $b = 0$, and likewise for the other two groups. We label each group with $LG(b, \theta)$ (which are shown at the leftmost part of Fig. 3 for each group correspondingly) and give the following properties of this construction that contribute to our computationally efficient algorithm:

Property 1: In each layer, the states from group $LG(b, \theta)$ cannot transit to any other groups $LG(b', \theta), b' \in \mathcal{B}\backslash b$ since the BES energy level can never increase or decrease without decreasing the remaining throughput. However, they can move to the states within the same group by keeping the BES idle.

Property 2: Between different layers, the states from group $LG(b, \theta)$ cannot transit to any other groups, $LG(b, \theta'), \theta' \in \Theta\backslash\theta$, since the remaining throughput can never increase or decrease without changing the energy level. However, they can move to the states from groups $LG(b', \theta'), b' \in \mathcal{B}\backslash b$ or $\theta' \in \Theta\backslash\theta$ within the allowance of power constraints.

We will show in the next subsection that our proposed efficient dynamic programming algorithm exactly leverages the above two properties, and this is also the main reason for performing layering and grouping aggregations.

B. Dynamic Programming Based on Layer and Group

Recall that the optimal value function $J(\boldsymbol{\omega}) = 0, \forall \boldsymbol{\omega} \in LG(b, 0)$. For any system state $\boldsymbol{\omega}$, we denote the states from the same group as $\boldsymbol{\omega}$ by $\mathcal{G}(\boldsymbol{\omega})$. We also denote those states by $\mathcal{M}(\boldsymbol{\omega})$, to which the system can move from state $\boldsymbol{\omega}$. According to the above two properties, the optimal value function $J(\boldsymbol{\omega})$ only depends on all the optimal value functions $J(\boldsymbol{\omega'})$ for $\boldsymbol{\omega'}$ from the lower layers and the optimal value function for the states in the same group $\mathcal{G}(\boldsymbol{\omega})$. Therefore, the value iteration (6) can be reformulated as

$$\overbrace{J_n(\boldsymbol{\omega}) = \max_{\mu \in \mathcal{A}(\boldsymbol{\omega})} \Big\{ \mathcal{R}(\boldsymbol{\omega}, \mu) + \sum_{\boldsymbol{\omega}' \in \mathcal{G}(\boldsymbol{\omega})} \vartheta_{pp'} J_{n-1}\Big(\boldsymbol{\omega}'\Big) \Big\}}^{\text{modified value iteration}} + \underbrace{\sum_{\boldsymbol{\omega}' \in \mathcal{M}(\boldsymbol{\omega}) \setminus \mathcal{G}(\boldsymbol{\omega})} \vartheta_{pp'} J\Big(\boldsymbol{\omega}'\Big)}_{\text{pre-computed information}}. \tag{7}$$

The idea of (7) is that *starting from Layer 0, we sequentially apply the value iteration for each group $LG(b, \theta)$ until arriving at the uppermost Layer θ_m*. Note that for the value iteration (which we refer to as the modified value iteration) within each group, only those states from the same group are involved, while all the value functions of those states from the lower layers will have already been computed (which we refer to as the pre-computed information). Following this approach, it should be noted that the value iteration converges to its optimality as (6), since we do not violate the *principle of optimality* in (5).

Remark 1. *The original state space dimension is $M = |\mathcal{B}| \cdot |\Theta| \cdot |\mathcal{P}|$. Assume that the dimension of the action space is fixed to be $|\mathcal{A}(\boldsymbol{\omega})| = A$, then the total complexity of solving (5) will be in the order of $O(M^2 A)$. However, the computational complexity of solving (7) will be significantly reduced to be linear in $O(|\mathcal{P}|^2 A)$. In practice, we may have very large dimension spaces for $\mathcal{B}$ and Θ due to large BES configurations but often a small dimension for the price space $\mathcal{P}$*[3].

V. Computational Experiment

In this section, we will illustrate our theoretical results through extensive simulation experiments[4]. The main simulation results are summarized as follows.

A. Optimal Policy with Threshold Structure

Figure 4 shows the optimal policy for different holding cost scenarios and different remaining throughputs. Note that for $h = 0, \theta = 100$, Fig. 4(a) validates our theoretical

TABLE I
CPU Time Comparison and Expected Lifetime[5]

	Time by GS (sec.)	Time by LG (sec.)	Expected Lifetime (hr.)
$\theta_m = 50$	153.328	12.231	221.367
$\theta_m = 75$	368.097	22.553	321.949
$\theta_m = 100$	622.522	38.138	427.177
$\theta_m = 125$	1299.557	57.488	531.773
$\theta_m = 150$	2127.213	86.139	644.260
$\theta_m = 175$	3081.675	111.311	748.174
$\theta_m = 200$	4532.047	148.927	860.431

results in Theorem 2 by demonstrating a very clear threshold structure property. When the holding cost increases, both the *upper threshold boundary* $b_t^{(up)}(\theta, p)$ and the *lower threshold boundary* $b_t^{(low)}(\theta, p)$ shift to the left, which is depicted by Fig. 4(b). Note that the degree of movement for each part of the boundaries is different. An interesting analogy is as follows: Imagine that there exists some strength to drag the threshold boundaries to move. Then, when h increases, there is more strength on the lower (upper) part of the dotted-blue (dotted-red) line. We show this property by different sizes of arrows, meaning different levels of strength to drag the boundaries to move. The insight for real systems behind this phenomenon is that, in order to achieve the optimal policy with maximal expected economic value, the BES should be charged more conservatively when the energy level is low, but it should be discharged more aggressively when the energy level is high; both are trying to keep the energy level low enough so as to reduce the high holding cost. Another interesting insight for real systems is that, when the BES is close to its lifetime, by comparing the optimal actions in the circles between Fig. 4(b) and Fig. 4(c), we observe that trying to charge (discharge) more aggressively when the energy level is low (high) is more beneficial. Meanwhile, we should also avoid too much charging and discharging for those states with average energy levels.

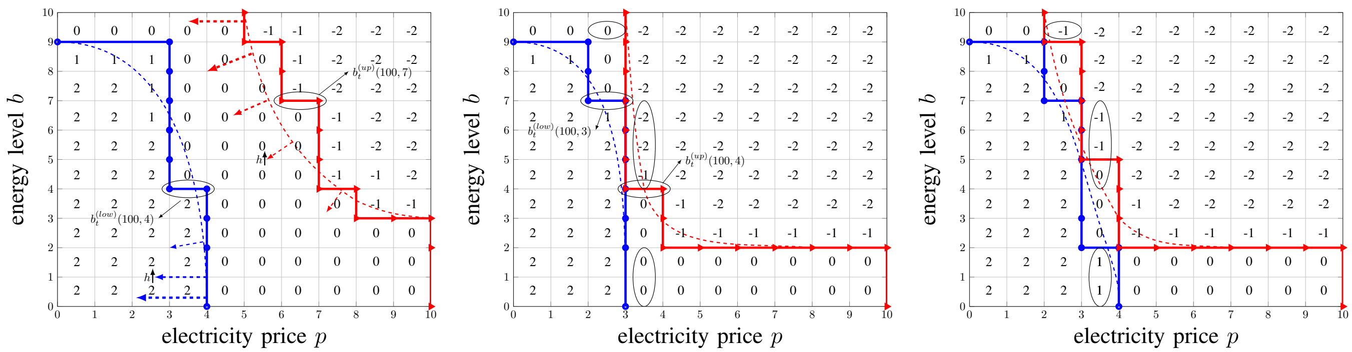
B. Economic Value under Optimal Trading Policy

Figure 5 shows the economic value for the different power ratings $c^{\max}$ and $d^{\max}$. Fig. 5(a) illustrates that in the low holding cost scenario ($h = 0$), it is always beneficial to increase the charge power rating $c^{\max}$. However, a larger discharge power rating d^{max} does not significantly yield a better result due to the quick saturation of the BES's economic value. Therefore, BES operators should carefully consider their requirements for $c^{\max}$ and $d^{\max}$ according to their different holding cost environments, thus avoiding unnecessary extra capital costs. However, when the holding cost is high ($h = 0.4$), unilaterally increasing the discharge power rating will induce some undesirable decreases in economic value, especially when the charge power rating $c^{\max}$ is small. This motivates some potential BES operators to avoid a mis-balance between $c^{\max}$ and $d^{\max}$ when their holding cost is high. Meanwhile, it is always beneficial to have a larger $d^{\max}$ for BES with a large $c^{\max}$ since it helps to discharge previously accumulated electricity out of the BES.

[3]Of course, more fine-grained discretization of price p will introduce larger $|\mathcal{P}|$; however, in a practical electricity pricing scenario, it is rational to assume that p is not changing in a very fine-grained manner.

[4]All the algorithms in this paper are implemented in MATLAB R2013A on an INTEL CORE i7-4770K Haswell 3.5GHz CPU, 16GB RAM PC.

[5]In the table, GS denotes the Gauss-Seidel flavor of the value iteration and LG denotes our proposed layer and group approach.



(a) $h = 0, \theta_t = 100$ (b) $h = 0.4, \theta_t = 100$ (c) $h = 0.4, \theta_t = 10$

Fig. 4. Optimal policy with different holding cost factors h and remaining throughputs θ_t. We assume the energy level b is discretized into 10 levels, and the whole action space is $\mu = \{-2, -1, 0, 1, 2\}$. The blue (red) polygonal lines are the *lower threshold boundary* (*upper threshold boundary*) below (above) which the BES should work in charge (discharge) mode with the amount of power specified in each grid. We use the dotted-blue and dotted-red lines to represent the possible thresholds for a real system with a continuous system space.

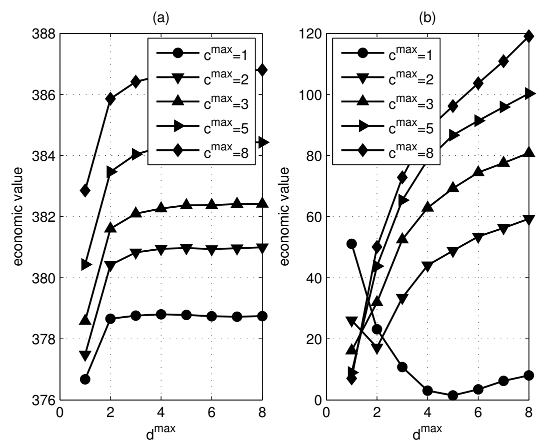


Fig. 5. Economic value with different power ratings $c^{\max}$ and $d^{\max}$ under (a) low holding cost environment and (b) high holding cost environment.

C. Computational Complexity Comparison

We compare our *Layer and Group* approach and the traditional Gauss-Seidel value iteration approach in Table I under different initial throughput. The system state space is increased by varying θ_m, while the other two entries b and p are kept fixed. Note that we use the same parameters as in Fig. 4(a) except θ. The last row of the table shows the expected lifetime obtained based on the underlying absorbing Markov chain of the SSP. As shown in Table I, the CPU time of our *Layer and Group* approach completely outperforms that of the traditional Gauss-Seidel algorithm, with a minimum speedup of 12.5 times (when $\theta_m = 50$). Note that this speedup becomes more substantial when the state space is larger. For example, when $\theta_m = 200$, the *Layer and Group* approach achieves a speedup of more than 30 times.

VI. Conclusion

In this paper, we have proposed a novel model to exploit the arbitrage value of BES. In addition to the ramp and capacity constraints, we explicitly took the lifetime constraint into consideration and formulated the BES profit maximization problem as an SSP problem. By exploiting the structure of the problem, we proposed a *Layer and Group* approach to make the SSP computationally tractable. We also proved the threshold structure property of the optimal policy, and we demonstrated various implementation insights based on our extensive simulation experiments. Note that simply choosing BES with a larger discharge power rating will not always be beneficial, and that maximizing the total expected economic value of BES along its entire lifetime requires an optimal tradeoff between exogenous price, capital cost and holding cost.

References

[1] http://www.scientificamerican.com/article.cfm?id=cities-show-the-way-with-energy-storage.

[2] H. Bindner, T. Cronin, P. Lundsager, J. F. Manwell, U. Abdulwahid, and I. Baring-Gould, "Lifetime modelling of lead acid batteries," Risø Nat.Lab., Roskilde, Denmark, Apr. 2005. Risø Rep.

[3] X. Tan, Y. Wu, and D.H.K. Tsang, "Economic analysis of lifetime-constrained battery storage under dynamic pricing," *in Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Oct. 2013.

[4] S. Teleke, M. E. Baran, A. Huang, S. Bhattacharya and L. Anderson, "Control strategies for battery energy storage for wind farm dispatching," *IEEE Trans. Energy Convers.*, vol. 24, no. 3, pp.725-732, 2009.

[5] P.M. van de Ven, N. Hegde, L. Massoulie and T. Salonidis, "Optimal control of end-user energy storage", [Available]: http://arxiv.org/pdf/1203.1891v3.pdf.

[6] R. Urgaonkar, B. Urgaonkar, M. Neely, and A. Sivasubramaniam, "optimal power cost management using stored energy in data centers," *in Proc. of ACM Sigmetrics*, 2011.

[7] P. Mokrian and M. Stephen, "A stochastic programming framework for the valuation of electricity storage," in *Proc. 26th USAEE/IAEE North Amer. Conf.*, pp.24-27, 2006.

[8] M. Roozbehani, M. Ohannessian, D. Materassi, and M. A. Dahleh, "Load-shifting under perfect and partial information: models, robust policies, and economic value," *Operations Research*, 2012 (Submitted).

[9] M. Olsson and L. Söder, "Modeling real-time balancing power market prices using combined SARIMA and Markov processes," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 443-450, May 2008.

[10] G. Gutiérrez-Alcarax and G. B. Sheblé, "Electricity market price dynamics: Markov process analysis," in *Proc. Int. Conf. Probabilistic Methods Applied to Power Systems (PMAPS)*, 2004.

[11] N. Secomandi, "Optimal commodity trading with a capacitated storage asset," *Management Science*, 56(3):449-467, 2010.

[12] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Belmont, MA: Athena Scientific, 2005.

[13] X. Tan, Y. Wu, and D.H.K. Tsang, "Optimal energy trading with battery energy storage under dynamic pricing: A fast stochastic shortest path based algorithm," [Online]. http://ihome.ust.hk/~xtanaa/publications/TR_energy_trading.pdf