A Stochastic Sizing Approach for Sharing-Based Energy Storage Applications

Islam Safak Bayram, Member, IEEE, Mohamed Abdallah, Senior Member, IEEE, Ali Tajer, Senior Member, IEEE, and Khalid A. Qaraqe, Senior Member, IEEE

Abstract—In order to foster renewable energy integration, improve power quality and reliability, and reduce hydrocarbon emissions, there is a strong need to deploy energy storage systems (ESSs), which can provide a control medium for peak hour utility operations. ESSs are especially desirable at the residential level, as this sector has the most untapped demand response potential. However, considering their high acquisition, operation, and maintenance costs, isolated deployment of ESSs is not economically viable. Hence, this paper proposes a sharing-based ESS architecture, in which the demand of each customer is modeled stochastically and the aggregate demand is accommodated by a combination of power drawn from the grid and the storage unit when the demand exceeds grid capacity. The optimal size of ESSs is analyzed and an analytical method is developed for a group of customers with a single type of appliances. This framework is also extended to any network size with an arbitrary number of customers and appliance types, where the analytical method provides a tractable solution to the ESS sizing problem. Finally, a detailed cost-benefit analysis is provided, where the results indicate that sharing-based ESSs are practical and yield significant savings in terms of ESS size.

Index Terms—Demand response, energy storage systems, stochastic sizing.

I. INTRODUCTION

OVER THE past few years, the power grids have become more stressed due to the steady increase in peak demand and more fragile due to the integration of intermittent renewable energy resources. Such a negative trajectory poses a profound threat to power grids, leads to a rise in hydrocarbon emissions, and exposes new economic challenges. One effective solution to mitigate the aforementioned issues, is to deploy energy storage systems (ESSs), which can act as an energy buffer and decouple the time of generation and demand by storing cheaper and cleaner off-peak hour electricity and delivering it during the peak load periods [1]. Integration of such storage units has multifaceted monetary benefits for different players. For instance, end users can enjoy the reduced electricity cost and lessened interruptions, while utility operators can enjoy improved grid reliability and asset utilization, and possibly deferral of system upgrades. Moreover, policymakers benefit from reduced hydrocarbon emissions and the mainstream integration of renewables, which are closely linked to energy security [1], [2].

While employing storage devices at microgrids has certain benefits, their deployment based on the current technologies may still not be economically viable [2]. Therefore, optimal sizing of the storage units based on the realistic needs of the grids is a critical step toward the efficient operation of the grid. Specifically, over-provisioning ESS size entails costly and underutilized assets, whereas under-provisioning reduces its operating lifetime (e.g., frequently exceeding the allowable depth of charge level degrades its health). Hence, there is a strong need to develop analytical models to solve the sizing problem. Furthermore, the sizing requirements for storage units are usually application-dependent. For instance, energy arbitrage applications for independent system operators require 100s MW in size, while storage units employed for energy management for communities or microgrids require 10s kW. In this paper, our main focus is on the latter one, as the residential sector has a great potential for peak demand reduction in the U.S. [3].

ESS sizing has received some attention in the literature. Oudalov et al. [4] presented a sizing approach for single industrial customers for peak saving applications. The sizing problem is solved by maximizing the net benefit, which is the sum of reductions in the electricity bills minus the operation costs and the one-time acquisition cost. Similarly, Chen et al. [5] proposed a sizing framework using similar cost models for a microgrid, but it also considers savings due to the storage of energy generated from renewable resources. Ghiassi-Farrokhfal et al. [6] developed a sizing approach based on stochastic network calculus to size the storage units coupled with solar power generation. It employs loss of load probability as the main performance metric to provision the resources. Another probabilistic approach
is presented in [7], which couples forecasting errors in wind generation with the storage unit and sizes the storage unit according to a desired level of accuracy. Moreover, Harsha and Dahleh [8] solved the sizing problem by using stochastic dynamic programming to minimize the operation cost and employs the storage device for balancing the load in a wind farm application. From the power engineering point of view, the sizing problem is usually solved via simulations [9]–[11]. However, simulations techniques are usually computationally expensive, and their accuracy depends upon the availability of data traces.

In this paper, we develop an analytical framework for optimal energy storage sizing for sharing-based end-user applications. The proposed framework contributes to the existing literature in the following ways.

1) The existing analytical methods for storage sizing focus on settings with one customer, or are developed for renewable integration and do not explore the underlying user dynamics. The proposed framework can cope with any network with arbitrary number of consumers with different levels of demand. Our analytical results show that a sharing-based design of storage units exhibits substantial gains over user-level designs.

2) The existing methods for multiconsumer settings are simulation-based. Lack of closed form expressions for the storage size-cost model limits the applicability of the model in optimizing the operations. The advantage of the proposed analytical method is that it establishes the exact optimal sizing, and subsequently, is computationally less expensive.

3) Majority of the existing literature assumes that residential users employ storage units for energy management applications without considering the associated costs. However, the studies conducted by Electric Power Research Institute [1] and Sandia Laboratories [2] reveal that the cost of employing stand-alone ESS is very high. Hence, we provide a detailed economic analysis for ESS applications in residential usage and show that storage technologies can be economically viable, only if operated in a shared manner.

Proper sizing of storage units is ultimately linked to the customer electricity consumption model. Several measurement-based studies [12]–[15] show that electricity consumption of households can be represented by On–Off models. For instance, Ardakanian et al. [12] measured the electricity consumption of 20 households at fine-grained intervals (every 6 s) and models the loads with continuous-time Markov chains. Driven by this observation, we adopt a Markovian fluid model to represent consumer’s demands, and the analyzes rely on the stochastic theory of fluid dynamics. We establish the interplay among the minimum amount of storage size, the grid capacity, number of consumers, and the stochastic guarantees on outage events.

We note that studying storage units at a network level is of paramount significance as they are expected to become integral to smart energy grids. More specifically, ESSs will be employed at smart residential, business complexes, and university campuses, to name a few, to reduce peak hour consumption. Clearly, in such sharing-based applications, the size of the energy storage is linked to the customer population, the power drawn from the grid, and the load profile. This relation is highly nonlinear due to multiplexing gains which are computed by the percentage of reduction in the required amount of resources with respect to the baseline case of assigning peak demand to each user.

II. SYSTEM DESCRIPTION

We consider a community of consumers in which the demands of N users are accommodated by the power grid capacity along with a shared ESS unit, in which capacity fluctuates over time. The grid capacity at time \( t \) is denoted by \( C_t \) and ESS depletion level is \( \mathbf{S}(t) \), for \( t \in \mathbb{R}^+ \). Furthermore, the energy storage unit has the following parameters.

1) Energy rating or the size of the storage is denoted by \( B \) (kWh).

2) Power rating is the rate at which storage can be charged or discharged. The charging rate is assumed to be \( P_c \leq C_t \), \( \forall t \), and discharge rating \( P_d \) is related to the energy rating \( B \) since \( B = P_d \times (\text{desired support duration}) \). Desired support duration is typically equivalent to the length of the peak hour.

3) The efficiency of a charge-discharge cycle is modeled by \( \eta \in [0, 1] \) to capture the percentage of stored energy or the fraction of energy that is transferred after losses are extracted.

4) Dissipation losses represent a small percentage of losses that occur due to leakage. For simplicity in notations, dissipation losses are ignored.

5) For economic analysis, we follow the standard assumptions made in [1] and assume that the project will be sited for 15 years, with 10% discount rate and the economic analysis is carried out by computing the net present value of the storage in the first year by using the cost per kW or kWh.

Furthermore, since the consumers do not necessarily have identical demands, we consider \( K \) customer classes, which are distinguished by the amount of electricity demands, represented by \( \{R_k\} \), where \( R_k \) denotes the energy demand per time unit for customer type \( k \). Let \( N_k \) denote the number of consumers of type \( k \) and the vector \( N \) represents the number of consumers of each type, that is \( N = (N_1, \ldots, N_K) \). As established in [12], [13], and [15]–[17] the consumption pattern of each consumer can be well-represented by a two-state “On/Off” process. We define the binary variable \( s^k_{it} \) to represent the state of consumer \( i \) of type \( k \) at time \( t \) such that

\[
s^k_{it} = \begin{cases} 
1 & \text{if consumer } i \text{ is On at time } t \\
0 & \text{if consumer } i \text{ is Off at time } t 
\end{cases}
\]

When a customer is in the On state, it initiates an energy demand. The duration of demand is modeled statistically, which is adopted to capture the variety types of consumers’

\footnotesize{\textsuperscript{1}The cost/benefit regime depends on different assumptions such as availability of solar rooftops, energy consumption statistics, project duration, and utility tariffs, to name a few. See references for details.}
demands. Specifically, the duration of type \(k\) customer’s demand is assumed to be exponentially distributed with parameter \(\mu_k\). Furthermore, we assume that the requests, that are transitions from “Off” to “On,” are generated randomly and according to a Poisson process with parameter \(\lambda_k\). Hence, for each consumer \(i\) of type \(k\) at any time \(t\) we have
\[
P(s^k_i(t) = 1) = \frac{\lambda_k}{\lambda_k + \mu_k}. \tag{2}
\]

In order to formalize the dynamics of ESS, we define \(L_k(t)\) as the charge request of consumer \(i\) of type \(k\). Then, there are exactly three cases that define the rate of change in the storage unit.

1) The storage can be in the fully charged state and the aggregate demand is less than grid power \(C_t\).

2) The storage can be completely discharged and the aggregate demand is more than the grid power \(C_t\).

3) Any partially charged state with any customer demand.

Therefore, for the rate of change in the storage level of the ESS, we have
\[
\frac{dS(t)}{dt} = \begin{cases} 
0, & \text{if } S(t) = B \& \sum_{k=1}^{K} \sum_{i=1}^{N_k} \eta L_k(t) < C_t \\
0, & \text{if } S(t) = 0 \& \sum_{k=1}^{K} \sum_{i=1}^{N_k} \eta L_k(t) > C_t \\
\eta(C_t - \sum_k \sum_i L_k(t)), & \text{otherwise}.
\end{cases} \tag{3}
\]

Due to stochasticities involved (in consumption and generation), by choosing any storage capacity \(B\), only stochastic guarantees can be provided for the reliability of the system and always there exists a chance of outage, which occurs when available resources fall below the aggregate demands by the consumers. By noting that \(S(t)\) denotes the energy level that the storage unit needs to feed into the grid to avoid outage, an outage event occurs when the necessary load from the storage unit exceeds the maximum available \(B\). Hence, we define \(\varepsilon\)-outage storage capacity, denoted by \(B(\varepsilon)\), as the smallest choice of \(B\) corresponding to which the probability of outage does not exceed \(\varepsilon \in (0, 1)\), that is
\[
B(\varepsilon) = \min \left\{ B \mid P(S_t \geq B) \leq \varepsilon \right\}. \tag{4}
\]

Our goal is to determine the \(\varepsilon\)-outage storage capacity \(B(\varepsilon)\) based on grid capacity \(C_t\), the number of users \(N\), and their associated consumption dynamics. For the simplicity of mathematical expressions, we scale the storage parameters and instead of \(B/\eta\) we redefine \(B\) as the maximum amount of energy that can be stored, and similarly \(P_d\) and \(P_p\) represent the actual power ratings. The notations are summarized in Table I.

We start our analysis for deriving \(B(\varepsilon)\) for a single class case.

III. STORAGE CAPACITY ANALYSIS FOR SINGLE CLASS CUSTOMERS (\(K = 1\))

A. Storage Access Dynamics

When the grid can serve all the consumers’ demands, there will be no consumer served by the storage unit. On the other hand, when the grid capacity falls below the aggregate demand, the consumers access the storage unit. Since the requests of the consumers arrive randomly, the number of consumers accessing the unit also varies randomly.

Since we have \(N\) independent consumers each with a two-state model, by taking into account their underlying arrival and consumption processes, the composite model counting the number of users accessing the storage unit at a given time can be modeled as a continuous-time birth-death process. Specifically, this process consists of \((N+1)\) states, in which state \(n \in \{0, \ldots, N\}\) models \(n\) consumers being active and accessing the storage unit, that is
\[
\text{state at time } t \text{ is } n \text{ if } \sum_{i=1}^{N} s^i_t = n \tag{5}
\]
and drawing \(nR_p\) units of power from the storage unit. As depicted in Fig. 1, the transition rate from state \(n\) to state \(n+1\) is \((N-n)\lambda\) and, conversely, the transition from state \(n+1\) to state \(n\) is \((n+1)\mu\). Hence, for the associated infinitesimal generator matrix \(M\), in which the row elements sum to zero, for \(i, n \in \{0, \ldots, N\}\) we have
\[
M[i, n] = \begin{cases} 
-(N-i)\lambda & n = i \\
i\mu & n = i + 1, i > 0 \\
(N-i)\lambda & n = i - 1, i > 0 \\
0 & \text{otherwise}.
\end{cases} \tag{6}
\]

By denoting the stationary probabilities of state \(n \in \{0, \ldots, N\}\) by \(\pi_n\) and according defining \(\pi = [\pi_0, \pi_1, \ldots, \pi_N]\), these stationary probability values satisfy \(\pi M = 0\).
B. Analyzing Distributions

Given the dynamics of accessing the storage unit, in the next step we analyze the statistical behavior of the ESS charge level. Specifically, we define \( F_i(t, x) \) as the cumulative distribution function (cdf) of the ESS charge level when \( i \in \{0, \ldots, N\} \) consumers are depleting the storage unit at time \( t \), that is

\[
F_i(t, x) = \mathbb{P}\left(S(t) \leq x \text{ and } \sum_{j=1}^{N} s_j = i\right).
\]

(7)

Accordingly, we define the vector of cdfs as

\[
F(t, x) \triangleq \left[F_0(t, x), F_1(t, x), \ldots, F_N(t, x)\right].
\]

(8)

Based on this definition, the next lemma delineates a differential equation which admits the cdf vector as its solution and is instrumental for analyzing the probability of outage events, that is \( \mathbb{P}(\sum_{i=1}^{N} L_i(t) > C_t + B) \).

**Lemma 1:** The cdf vector \( F(t, x) \) satisfies

\[
\frac{dF(t, x)}{dx} \cdot D = F(t, x) \cdot M
\]

(9)

where \( D \) is a diagonal matrix defined as

\[
D \triangleq \text{diag}[-C_t \mu, (1 - C_t) \mu, \ldots, (N - C_t) \mu]
\]

(10)

and matrix \( M \) is defined in (6).

**Proof:** In order to compute the probability density functions, we find the expansion of \( F_i(t, x) \) for an incremental change \( \Delta t \) in \( t \), i.e., \( F_i(t + \Delta t, x) \). Note that during incremental time \( \Delta t \), three elementary events can occur:

1. One inactive consumer might become active, i.e., \( i \) increases to \( i + 1 \).
2. One active consumer becomes inactive, i.e., \( i \) reduces to \( i - 1 \).
3. The number of active consumers remains unchanged.

Since the durations of arrival and departure of consumers are exponentially distributed as cdf \( F_i(t, x) \) can be expanded

\[
F_i(t + \Delta t, x) = \frac{[N - (i - 1)] \cdot (\lambda \Delta t) \cdot F_{i-1}(t, x) + [i + 1] \cdot (\mu \Delta t) \cdot F_{i+1}(t, x) + [1 - ((N - i) \lambda + i \mu) \Delta t] \cdot F_i(t, x - (i - C_t) \cdot \mu \Delta t)}{\text{no change}} + o(\Delta t^2)
\]

(11)

where \( o(\Delta t^2) \) represents the probabilities of the compound events and tends to zero more rapidly than \( \Delta t^2 \) (and \( \Delta t \)) as \( \Delta t \to 0 \). Next, by passing the limit

\[
\lim_{\Delta t \to 0} \frac{F_i(t + \Delta t, x)}{\Delta t} = \frac{[N - (i - 1)] \cdot (\lambda) \cdot F_{i-1}(t, x) + [i + 1] \cdot (\mu) \cdot F_{i+1}(t, x) - [(N - i) \lambda + i \mu] \cdot F_i(t, x) - (i - C_t) \cdot (\mu) \cdot \frac{\partial F_i(t, x)}{\partial x}}{\text{no change}}
\]

(12)

where we have defined \( F_{i-1}(t, x) = F_{N+1}(t, x) = 0 \). By recalling that the design is intended for a long-term steady-state operation we have \( \partial F_i(x, t)/\partial t = 0 \). Hence, (12) can be rewritten as

\[
(i - C_i) \cdot (\mu) \cdot \frac{\partial F_i(t, x)}{\partial x} = [N - (i - 1)] \cdot (\lambda) \cdot F_{i-1}(t, x) + [i + 1] \cdot (\mu) \cdot F_{i+1}(t, x) - [(N - i) \lambda + i \mu] \cdot F_i(t, x).
\]

(13)

By concatenating (13) for all \( i \in \{0, \ldots, N\} \), we obtain the compact form

\[
\frac{dF(t, x)}{dx} \cdot D = F(t, x) \cdot M.
\]

(14)

The solution of the first order differential equation given in (14) can be expressed as a sum of exponential terms. The general solution requires computing \( (N + 1) \) eigenvalues of the matrix \( MD^{-1} \) and the general solution can be expressed as [18]

\[
F(t, x) = \sum_{i=0}^{N} \alpha_i \cdot \phi_i \cdot \exp(z_i x)
\]

(15)

where \( z_i \) is the \( i \)th eigenvalue of \( MD^{-1} \) with the associated eigenvector \( \phi_i \), which satisfy \( z_i \phi_i D = \phi_i M \). The coefficients \( \{\alpha_0, \ldots, \alpha_N\} \) are determined by the boundary conditions, e.g., \( F_i(t, 0) = 0 \) and \( F_i(t, \infty) = 1 \).

In order to compute the probability distribution in (15), we need to determine the eigenvalues of \( MD^{-1} \), the eigenvectors \( \{\phi_i\} \), and coefficients \( \{\alpha_i\} \). Note that, since \( x \geq 0 \) and \( F_i(t, x) \) is upper bounded by 1, all of the positive eigenvalues and the corresponding \( \alpha_i \) must be set to zero, hence this reduces the computational complexity and (15) simplifies to

\[
F(t, x) = \sum_{i: \text{Re}(z_i) \leq 0} \alpha_i \cdot \phi_i \cdot \exp(z_i x).
\]

(16)

It can be further observed that since \( z_i \phi_i D = \phi_i M \), one of the eigenvalues must be zero. Then by setting \( z_0 = 0 \), the corresponding eigenvector can be computed from \( \phi_0 M = 0 \). We also showed earlier that the steady state probability distribution \( \pi \) of the \( N + 1 \) state Markov chain can be computed from the same equation, that is \( \pi M = 0 \). Since, the eigenvector \( \phi_0 \) is known and one of the eigenvalues is \( z_0 = 0 \), we can write \( \phi_0 = \pi \). Therefore, (16) further simplifies to [19]

\[
F(t, x) = \pi + \sum_{i: \text{Re}(z_i) < 0} \alpha_i \cdot \phi_i \cdot \exp(z_i x).
\]

(17)

C. Single User Storage Capacity (\( K = 1 \) and \( N = 1 \))

For computing the desired \( \varepsilon \)-outage storage capacity \( B(\varepsilon) \) by leveraging the cdf vector found in (17) we start by a simple network with one user \( (N = 1) \). The insights gained can be leveraged to generalize the approach for networks with any arbitrary size \( N \). When \( N = 1 \) the infinitesimal generator matrix \( M \) defined in (6) is

\[
M = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}.
\]

(18)
For finding the expansion of $F(t, x)$ as given in (15) we need to find the eigenvalues of $MD^{-1}$, i.e., $z_0$ and $z_1$, where $D$ is defined in (10). Based on (10), we find that

$$MD^{-1} = \begin{bmatrix} \frac{1}{C_1} & \frac{1}{\mu} \\
\frac{1}{\mu} & -\frac{1}{C_1} \end{bmatrix}. \tag{19}$$

Hence, the eigenvalues are

$$z_0 = 0 \quad \text{and} \quad z_1 = \frac{\chi}{C_1} - \frac{1}{1 - C_1} \tag{20}$$

where we have defined $\chi \triangleq \lambda / \mu$. It can be readily verified that the eigenvector associated with $z_1$ is $\phi_1 = [1 - C_1, C_1]$. Therefore, according to (17) we have

$$F(t, x) = \pi + \alpha_1 \phi_1 \exp(z_1 x). \tag{21}$$

Finally, by finding the coefficient $\alpha_1$ we can fully characterize $F(t, x)$. This can be facilitated by leveraging the boundary condition $F(t, 0) = 0$, which yields

$$F(t, 0) = \pi_1 + \alpha_1 C_1 = 0 \tag{22}$$

where we have that $\pi_1 = \lambda / (\lambda + \mu)$. Therefore

$$\alpha_1 = -\frac{\chi}{C_1(1 + \chi)} \tag{23}$$

which, subsequently, fully characterizes both cdfs $F_0(t, x)$ and $F_1(t, x)$ according to

$$F_0(t, x) = \pi_0 + \alpha_1 (1 - C_1) \exp(z_1 x)$$

and

$$F_1(t, x) = \pi_1 + \alpha_1 C_1 \exp(z_1 x).$$

As a result, by recalling the definition of $F_i(t, x)$ in (7), the probability that the storage level $S_i$ falls below a target level $x$ is given by

$$P(S_i \leq x) = F_0(x) + F_1(x) = 1 + \alpha_1 \exp(z_1 x). \tag{24}$$

Given this closed-form characterization for the distribution of $S_i$, we can now evaluate the probability term

$$P(S_i > B) \tag{25}$$

which is the core constraint in the storage sizing problem formalized in (4). Specifically, for any instantaneous realization of $C_i$ denoted by $c$ we have

$$P(S_i > B) = \int_{C_i} \int \frac{\alpha_1 \exp(z_1 x) f_C(c)}{f_C(c) \exp(z_1 x)} \, dc \, dx$$

Therefore, by noting that $z_1 = (\chi / c) - 1/(1 - c)$ is negative, the probability term $P(S_i > B)$ becomes strictly decreasing in $B$. Hence, the smallest storage capacity $B$ that satisfies the stochastic guarantee $P(S_i > B) \leq \varepsilon$ has a unique solution corresponding to which this constraint holds with equality. In the simplest settings in which grid capacity $C_i$ is constant $c$, we find

$$B(\varepsilon) = \frac{c(1 - c)}{\chi - \chi c - c} \cdot \log \frac{c(1 + \chi)}{\chi}. \tag{26}$$

D. Multiuser Storage Capacity ($K = 1$ and $N > 1$)

In this section, we provide a closed-form for the probability term $P(S_i \leq x)$ for arbitrary values of $N$, which we denote by $F_N(x)$. Computing all $F_N(x)$ terms through computing their constituent terms $F_i(t, x)$, especially as $N$ grows, becomes computationally expensive, and possibly prohibitive as it involves computing the eigenvalues and eigenvectors of $MD^{-1}$. By capitalizing on the observation that for large number of users $N \gg 1$, the largest eigenvalues are the main contributors to the probability distribution, Morrison [20] showed that, the asymptotic expression for $F_N(x)$ is given by

$$F_N(x) = \frac{1}{2\sqrt{\pi}} \frac{u}{(\xi + \lambda(1 - \xi))N} \exp(-N\Psi(\xi) - g(\xi)/x) \times \exp(-2\sqrt{|f(\xi)(\xi + \lambda(1 - \xi))N(x)|}) \tag{27}$$

where

$$f(\xi) \triangleq \log \left( \frac{\xi}{\lambda(1 - \xi)} \right) - 2 \frac{\xi(1 + \lambda) - \lambda}{\xi + \lambda(1 - \xi)}$$

$$u \triangleq \frac{\xi(1 + \lambda) - \lambda}{\xi(1 - \lambda)}$$

$$\psi(\xi) \triangleq \xi \log(\xi) + (1 - \xi) \log(1 - \xi) - \xi \log(\xi) + \log(1 + \lambda)$$

$$g(\xi) \triangleq z + 0.5(\xi + \lambda(1 - \xi) - \psi(1 - \xi)$$

$$z \triangleq (1 - \lambda) + \lambda(1 - 2\xi)$$

and $\Psi(\xi) = (2\xi - 1)(\xi(1 + \lambda) - \lambda)^2$.

In this set of equations, time is measured in units of a single average On time $(1/\mu)$. Furthermore, $\kappa$ and $\xi$ are defined as the ESS per user $(B/N)$ and the grid power allocated per one source, respectively. Furthermore, we denote the variable $\nu$ as the power above the mean demand allocated per user as $\nu = \kappa - \lambda / (1 + \lambda)$.

IV. Storage Capacity Analysis for Multiclass Customers ($K > 1$ and $N > 1$)

In this section, we consider the case where the storage serves more than one customer type. We start our analysis by noting that the continuous-time Markov chain model, presented in Fig. 1, becomes a $K$-dimensional Markov process. Recall that total number of users of each class is represented by $N = (N_1, \ldots, N_K)$ and we let $n(t) = [n_1(t) \ldots n_k(t)]$ represent the number of sources of type $k$ that are On at time $t$. Furthermore, $n(t)$ represents the state space of the Markov process and similar to single class case given in (6), the transitions can occur between neighboring states, and the transition matrix $M = \{M(i, n)\}$ is given by

$$M(n, \Delta^{+}_k(n)) = (N_k - n_k) \lambda_k \tag{28}$$

$$M(n, \Delta^{-}_k(n)) = n_k \mu_k \tag{29}$$

$$\bar{M}(n, n') = 0, \text{ otherwise} \tag{30}$$
where
\[ \Delta_{k}^1(n_1, \ldots, n_k, \ldots, n_K) = (n_1, \ldots, n_k + 1, \ldots, n_K) \]
\[ \Delta_{k}^0(n_1, \ldots, n_k, \ldots, n_K) = (n_1, \ldots, n_k - 1, \ldots, n_K). \]

In the single class case, the transitions can only occur between two neighboring states. In the multiclass case, however, the transitions can occur in \(2 \times k\) different neighbors. For instance, assume that there are two classes, \(k = 2\), and the Markov chain is in state \((1, 1)\) (i.e., one active user from each class). Therefore, there exists four possible transitions \((1, 0), (0, 1), (2, 1),\) and \((2, 2)\), and the transition rates depend on the \((\lambda_k, \mu_k)\) pairs. To that end, the transition rates given in (28) represent customer arrivals, whereas the transitions rates in (29) account for new customer arrivals, and similar to single class case there can not be any transitions beyond neighboring states. This is given in (30).

### A. Effective Demand

The analysis in Section III-B provided for the single-class settings is also valid for the multiclass case. However, computing the eigenvalues becomes infeasible when \(K > 1\). Hence, we follow the decomposition method proposed in [21] to efficiently evaluate the eigenvalues \(z_k\) and the coefficients \(\alpha_k\). Our goal in this section is to compute \(\omega_k\) which is a deterministic quantity acting as a surrogate for the actual aggregate stochastic demand [22]. In other words, assigning \(\omega_k\) amount of resources to each class will satisfy (4).

We define parameter \(\xi = 1 - e^{1/B} \in [0, 1]\) to account for the susceptibility of the storage to customer demand pattern. This parameter will be an essential part of \(\omega_k\). First we will consider large storage units, then we will show that the proposed scheme is also valid for any storage size. Assume that we are able to choose \(N = (N_1, \ldots, N_K)\) in a way that ensures \(P(S_t > B) \leq \xi\) for small \(\varepsilon\) and the eigenvalues are relabeled such that \(z_0 \geq z_1 \geq \cdots \geq z_N \geq 0\) according to the following theorem holds.

**Theorem 1:** Let \(B(B, \varepsilon) = \{N : P(S_t > B) \leq \varepsilon\}\). For large storage size \(B\) and for small \(\varepsilon\), we have
\[
\lim_{B \to \infty, \varepsilon \to 0} \frac{\log \varepsilon}{B} = \xi \in [-\infty, 0].
\]

Furthermore, let
\[
\tilde{B} = \left\{ N : \sum_k \omega_k(\xi)N_k \leq C \right\}
\]
and
\[
\tilde{B} = \left\{ N : \sum_k \omega_k(\xi)N_k \leq C \right\}
\]
where effective demand can be computed by [21]
\[
\omega_k(\xi) = \frac{\xi R_k + \mu_k + \lambda_k - \sqrt{(\xi R_k + \mu_k - \lambda_k)^2 + 4\lambda_k \mu_k}}{2\xi}.
\]

Then, \(\tilde{B} \subseteq B(B, \varepsilon) \subseteq \tilde{B}\).

**Proof:** For a given set of users, \(N\), the following holds:
\[
P(S_t \geq B) = \frac{\sum_k \omega_k(1^T \phi_k) \exp(z_k B)}{\varepsilon}
\]
where \(z_k\)’s are the nonpositive eigenvalues of \(\tilde{MD}^{-1}\), \(\phi_k\) are the corresponding eigenvectors, and \(1^T = [1, 1, \ldots, 1]\). Recall that, the coefficients \(\alpha_k\) can be obtained by solving the following set of equations:
\[
\begin{cases}
\sum_k \alpha_k \phi_k = 0, \text{ if } D_k > 0 \\
\alpha_k = 0, \text{ if } z_k > 0 \\
\alpha_0 1^T \phi_0 = 1, \text{ otherwise.}
\end{cases}
\]

In [23], it is shown that \(z_0 > z_k\), for \(\forall k \geq 1\), and \(z_0\) satisfies
\[
C = \sum_k \alpha_k(z_0)N_k.
\]

### V. ECONOMIES OF ESS DEPLOYMENT

The proposed framework targets enhancing energy management via deploying for end users and the monetary associated benefits of ESS deployment [1].

1) Improved power quality refers to voltage sags and outages experienced by customers, which in many cases remains unnoticed. However, if occur for sufficiently large durations, they damage customers’ appliances. The costs of momentary outages, according to a survey study conducted in [24], are summarized in Table II.
be additional benefits due to demand charges. The amount of savings depends on the specific policies and regulations, and the exact benefits will include profits by energy trading and reduced age is coupled with distributed generation options. Additional application-dependent and there can be more benefits if storage (due to the elimination of peaking generators). Using ESSs is mostly economically viable. Using ESSs is mostly economically viable.

2) Improved power reliability refers to the usage of storage units during outages and blackouts. In order to quantify the cost of power reliability, we adopt the measure used in [25], i.e., the cost per kW shown in Table III. Obviously, the impact of the outages depends on the duration and the frequency of the events. We adopt the statistics conducted by [1] and assume that the consumer average interruption duration in one year is 88 min.

3) Reduced time of use (TOU) charges include the savings occurred by eliminating the use of peak hour electricity, and using storage units. “TOU” tariffs may vary in different territories and for different seasons. In Table IV, we present the average TOU tariffs in the U.S.

4) Demand charges are a significant portion of the commercial and industrial customers’ bill. Some utilities also apply these charges to residential customers. It is usually computed by the amount and the duration of the peak usage [4]. Since it is not easy to compute these charges without knowing the load profile, we did not consider this in our calculations. An overview of demand charges is presented in Table V.

It is noteworthy that our goal in discussing this example scenario is to show how our framework can make the ESS deployment economically viable. Using ESSs is mostly application-dependent and there can be more benefits if storage is coupled with distributed generation options. Additional benefits will include profits by energy trading and reduced electricity bills. Also, the cost of carbon emission decreases (due to the elimination of peaking generators). The exact amount of savings depends on the specific policies and regulations. For commercial and industrial customers, there will be additional benefits due to demand charges.

VI. NUMERICAL EXAMPLES

A. Single Class Customers

In this section, we provide numerical examples to explain the system dynamics and show how the proposed framework can be used in typical peak shaving applications. In the first group of setting, uniform customer demand is considered (K = 1). We use the aforementioned normalized values (unit time is measured in μ⁻¹ and unit demand is measured in peak demand—R_p). We start by exploring the relationships between the number of users (N), ESS size (in R_pμ⁻¹ units) and the corresponding underflow probability for a given system capacity, which remains constant over time. C. Charge request rate per single user λ is set to 0.3, and the mean capacity above the mean demand per user is set to ω = 0.035 (assumptions are valid for all single class case studies). Therefore, the total system capacity becomes C = 0.2658N units. In Fig. 2(a), sizing problem is evaluated for user population N ranging from 100 to 350. Considering the fact that, one household can employ 10–12 appliances, this interval is chosen to represent a multidwelling building or a typical building on a university campus. These findings can be used in the following ways. First, system operators can provision the storage units and provide statistical guarantees (underflow probability) to their customers for a given user population N. For example, for a large-scale electric vehicle charging lot (e.g., located in shopping mall or airports [26]) with N = 350 charging slots in order to accommodate 99.95% of the customer demand the ESS size should be selected as B = 7 × R_p × μ⁻¹ = 7 × 6.6 × 2 = 92.4 kWh (considering level-II chargers and 30 min as the unit time). It is noteworthy that the required storage size decreases as the user population increases.

Next, we consider the case where the system operator employs an already acquired ESS of sizes B = 10, 15, or 20 kWh. In this case, the critical step is to calculate the minimum amount of power to draw from the grid so that performance guarantees can be achieved. This case is evaluated in Fig. 2(b) for N = 200 appliances. For instance, suppose ESS size B = 10 kWh is already acquired and the goal is to meet 95% of the demand, and then system operator should draw 0.26 × 200 = 52 kW from the grid. The primary motivation for the employment of the ESS is to reduce the stress on the grid and improve the utilization of power system components in a cost effective manner. Hence, our last evaluation is on computing the amount of savings in ESS size with respect to current common sizing practice, allocating peak demand. The results depicted in Fig. 2(c) show that instead of sizing the ESS to meet the entire customer demand, just by rejecting a few percentage of customers, considerable savings in the storage size can be achieved.

B. Multiclass Customers

Next, we consider the case where customers can request different demand levels, hence different customer classes. In this section, our main objective is to compute the ω_k (effective demand) parameter for each class such that ω_k replaces R_k and the overflow probability targets are met. We start by discussing a toy example. Suppose that users can generate

<table>
<thead>
<tr>
<th>TABLE III</th>
<th>AVERAGE OUTAGE COST [25]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$/kW</td>
<td>15 Min.</td>
</tr>
<tr>
<td>Residential</td>
<td>$0.05</td>
</tr>
<tr>
<td>Small C&amp;I</td>
<td>$8.65</td>
</tr>
<tr>
<td>Large C&amp;I</td>
<td>$4.79</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE IV</th>
<th>AVERAGE TOU RATES IN THE U.S. [1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$/kWh</td>
<td>Summer</td>
</tr>
<tr>
<td>Residential</td>
<td>$0.25</td>
</tr>
<tr>
<td>Small C&amp;I</td>
<td>$0.18</td>
</tr>
<tr>
<td>Large C&amp;I</td>
<td>$0.06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE V</th>
<th>DEMAND CHARGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>$/kWh-Month</td>
<td>Summer</td>
</tr>
<tr>
<td>Residential</td>
<td>$0.00</td>
</tr>
<tr>
<td>Small C&amp;I</td>
<td>$15.00</td>
</tr>
<tr>
<td>Large C&amp;I</td>
<td>$12.00</td>
</tr>
</tbody>
</table>
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

IEEE TRANSACTIONS ON SMART GRID

Fig. 2. Evaluation of single-class case. Community storage sizing for (a) different user population and (b) varying grid power, $N = 200$ appliances (users). (c) Savings in ESS size for varying user population.

Fig. 3. Evaluation of multiclass case. (a) Admission region for $K = 2$. (b) Effective demand calculation for $K = 2$. (c) ESS sizing for $K = 2$.

four different levels of demands, with the following parameters $\lambda = \{0.3, 0.5, 0.7, 0.9\}$, $\mu = \{1, 1, 1, 1\}$, and $R^k = 0.2, 0.4, 0.6, 0.8$ (in kW). Then for the target outage $\varepsilon = 10^{-4}$, and ESS size $B = 10$ kWh, the corresponding effective demand becomes $\omega = \{0.0515, 0.1567, 0.2958, 0.4550\}$. Alternatively, the system operator can find the minimum storage size with respect to the available grid resources and target performance metric. For the simplicity of presentation, the remaining numerical results consider two classes of customers with the following parameters $\lambda_1 = 0.5$, $\mu_1 = 1$, $R_1 = 0.5$ kW and $\lambda_2 = 0.7$, $\mu_2 = 1$, and $R_2 = 1$ kW. Our first evaluation shows the admission set for $C = 50$ kW and $B = 10$ kWh, for a target overflow probability of $\varepsilon = 0.001$. Results presented in Fig. 3(a) show that station operator can choose from any set of customer numbers from the shaded region. Obviously, since the demand for class 1 customers is less than the other class, the system operator can accept more customers from class 1. As a second evaluation, we calculate the effective demand for varying storage size. This time the system underflow probability is set as $\varepsilon = 0.0005$ and the customer population is chosen as $N_1 = 100$ and $N_2 = 45$. The results depicted in Fig. 3(b) show that instead of provisioning the system according to peak demand ($R_1 = 0.6$ kW and $R_2 = 1$ kW), the use of effective demands (e.g., for $B = 10$ kWh, $\omega_1 = 0.25$, and $\omega_2 = 0.5$) reduces the provisioning of resources tremendously. Our final results are on the ESS sizing for multiclasses. For the same set of parameters, we evaluate the storage size with respect to power drawn from the grid. Similar to the single class case, this result can be used to size the storage unit for a given grid power, or it can be used to compute the required grid resources for a given storage size.

C. ESS Economic Analysis

In this section, we provide several numerical examples with actual real-world scenarios to show how the proposed framework reduces the cost of the customers. In our evaluations, we choose the average cost per residential customer. We assume that peak appliance demand is set to $R_1 = 1.2$ kW, each user employs 15 appliances, and there are 250 days of peak usage in one year. For simplicity, the peak hour duration is assumed to 1 h. All the cost/benefit calculations are normalized to peak hour usage. We consider three cases and compare the cost of system operation for different customer population and the frequency of appliance usage.

1) Case 1—Grid Power Only: In this setting, all of the peak hour demand is met by the power demand. This case reflects the current state of affairs. The cost includes the cost of peak charging ($0.25$ kWh) and all of the listed benefits except the demand charges. Obviously, since there is no storage unit, users suffer from power quality and reliability, and high peak-hour charges.

2) Case 2—ESS Only: In this case, we consider the case where all of the peak hour demand is met by the storage unit. The cost includes the cost of off-peak hour charging and the storage cost. In this case, all of the listed benefits except the demand charges contribute to reducing the total cost of the system.
Case 3—Proposed Framework: This is the case where we compute the storage size based on the mathematical proposed framework. We set the probability target as $\varepsilon = 0.001$. By noting that, there can be more than one $(C, B)$ combinations that satisfy the overflow targets for the evaluations, we assume that grid power is 20% more than the average customer demand.

According to Electric Power Research Institute [1], the greatest cost of such of energy management and TOU pricing applications stems from the acquisition and the operation and management costs of the storage units. Generally, the lifetime of a storage unit is assumed to be 15 years with 10% discount rate. In this paper, we assume that the unit cost of the storage unit per year is $1500$, which is the typical cost for a Li-ion battery.

We compute the aggregate cost of each case per user per month for three different demand frequencies $\lambda = 0.25$, $\lambda = 0.45$, and $\lambda = 0.65$. The results depicted in Fig. 4 have the following interpretations. Since the cost of ESS technologies is high, the stand-alone ESS setting (case 2) is extremely costly and not useful practically, which is in line with the conclusion of EPRI [1] and SANDIA [2] studies. On the other hand, users can enjoy the aforementioned benefits by sharing the storage (e.g., in a neighborhood or on a campus etc.), as more users share the same resource pool, the cost per user decreases. Obviously, the population of users who share the same ESS depends on the physical deployments. For instance, 100 apartments in highly dense regions can be fed by the same storage, while this may not be possible in a wide area due to losses. Further, as the storage usage increases (higher $\lambda$) the economic benefits improve faster. These evaluations show that, it may be practical to share the same storage unit so that the total cost of ownership reduces, and customers enjoy a variety of benefits. One final note is that the same analysis can be applied to the multiclass case by computing the effective demands and the corresponding storage size. However, our main goal is to conduct the economic analysis per user based.

3 Another highly relevant report is presented in [2], however, we consider [1] as it is more recent.

References


Islam Safak Bayram (S’10–M’14) received the B.S. degree in electrical and electronic engineering from Dokuz Eylul University, Izmir, Turkey, in 2007; the M.S. degree in telecommunications from the University of Pittsburgh, Pittsburgh, PA, USA, in 2010; and the Ph.D. degree in computer engineering from North Carolina State University, Raleigh, NC, USA, in 2013.

In 2014, he was a Postdoctoral Research Scientist with Texas A&M University at Qatar, Doha, Qatar. He is currently an Assistant Professor with the College of Science and Engineering, Hamad Bin Khalifa University, Doha, and a Scientist with the Qatar Environment and Energy Research Institute, Doha. His current research interests include stochastic modeling and control of communications, and power networks.

Dr. Bayram was a recipient of the Best Paper Award at the Third IEEE International Conference on Smart Grid Communications.

Mohamed Abdallah (S’95–M’08–SM’12) was born in Giza, Egypt. He received the B.Sc. (Hons.) degree in electrical engineering from Cairo University, Giza, in 1996, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Maryland at College Park, College Park, MD, USA, in 2001 and 2006, respectively.

He joined Cairo University, in 2006, where he is an Associate Professor with the Department of Electronics and Electrical Communication. He is currently an Associate Research Scientist with Texas A&M University at Qatar, Doha, Qatar. His current research interests include design and performance of physical layer algorithms for cognitive networks, cellular heterogeneous networks, sensor networks, smart grids, visible light and free-space optical communication systems, and reconfigurable smart antenna systems.

Dr. Tajer serves as an Editor for the IEEE TRANSACTIONS ON SMART GRID and the IEEE TRANSACTIONS ON COMMUNICATIONS, and the Guest Editor-in-Chief for the IEEE TRANSACTIONS ON SMART GRID Special Issue on Theory of Complex Systems With Applications to Smart Grid Operations.


From 2007 to 2010, he was with Columbia University. He was a Postdoctoral Research Associate with Princeton University, Princeton, NJ, USA, from 2010 to 2012. He is currently an Assistant Professor of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY. His current research interests include estimation and detection theory, network information theory, wireless communications, and smart grids.

Dr. Tajer serves as an Editor for the IEEE TRANSACTIONS ON SMART GRID and the IEEE TRANSACTIONS ON COMMUNICATIONS, and the Guest Editor-in-Chief for the IEEE TRANSACTIONS ON SMART GRID Special Issue on Theory of Complex Systems With Applications to Smart Grid Operations.

Khalid A. Qaraqe (M’97–SM’00) received the B.S. (Hons.) degree from the University of Technology, Baghdad, Iraq, in 1986; the M.S. degree from the University of Jordan, Amman, Jordan, in 1989; and the Ph.D. degree from Texas A&M University, College Station, TX, USA, in 1997, all in electrical engineering.

From 1989 to 2004, he held a variety of positions in several companies and he has over 12 years of experience in the telecommunication industry. He has researched on numerous groupe spécial mobile, code division multiple access, and wideband code division multiple access projects, and has experience in product development, design, deployments, testing, and integration. He joined the Department of Electrical and Computer Engineering, Texas A&M University at Qatar, Doha, Qatar, in 2004, where he is currently a Professor. His current research interests include communication theory and its application to design and performance, analysis of cellular systems and indoor communication systems, mobile networks, broadband wireless access, cooperative networks, cognitive radio, diversity techniques, and beyond 4G systems.