Scheduling of Battery Charge, Discharge, and Rest

Hahnsang Kim and Kang G. Shin

Real-Time Computing Laboratory
Department of Electrical Engineering and Computer Science
The University of Michigan
Ann Arbor, MI 48109-2121, U.S.A.
Email: {hahnsang, kgshin}@eecs.umich.edu

Abstract—Electric vehicles operate inefficiently with a naive battery management system that charges or discharges battery cells in a pack based solely on application load demands. The battery pack’s operation-time and lifetime can be extended significantly by effectively scheduling (the cyber part) battery charge, discharge, and rest activities, based on the battery characteristics (the physical part). We propose a set of policies for scheduling battery-cell activities, called the weighted-k round-robin (kRR) scheduling framework. This framework dynamically adapts battery-cell activities to load demands and the condition of individual cells, thereby extending the battery pack’s operation-time and making them robust to anomalous voltage-imbalance.

The framework comprises two key components. First, an adaptive filter estimates the upcoming load demand. Then, based on the estimated load demand, the kRR scheduler determines the number of parallel-connected cells to be discharged simultaneously. The scheduler also effectively partitions the cells in the pack, allowing the cells to be simultaneously charged and discharged in coordination with the battery reconfiguration system we developed earlier [17]. Besides the kRR scheduling framework, we characterize the discharge and recovery efficiency of a Lithium-ion battery cell. The kRR scheduling framework is shown to outperform three alternative scheduling mechanisms with respect to the operation-time by 7–56%, and improve the tolerance of voltage-imbalance by up to 50%.

I. INTRODUCTION

On April 2, 2007, the US Supreme Court identified CO₂ as an air pollutant, authorizing the US Environmental Protection Agency (EPA) to regulate this greenhouse gas [4]. Greenhouse gas emissions could be reduced by half by increasing use of hybrid, full, and/or fuel-cell electric vehicles (EVs). For instance, we can achieve an estimated 35% efficiency for a 5-passenger EV with electricity generated from fossil-based energy, while achieving a 16% efficiency in converting gasoline-based energy to vehicle motion. According to a recent survey [8], due mainly to a sharp increase in fuel cost in 2008, 36% of motorists worldwide wish to buy a car with hybrid drive, while 46% of them are interested in buying full-electric cars. To meet this demand, an estimated 10× improvement of the battery capacity and power is required, making the performance of rechargeable batteries competitive with, and attractive alternatives to, conventional gasoline engines.

An important requirement is to prevent the overcharge and deep-discharge of battery cells. For instance, when a Lithium-ion cell (that has high electrical energy concentrated in a small volume in the cell) is overcharged, active materials therein will most likely react with other materials and electrolytes, potentially causing an explosion, let alone damaging the cell itself. When the cell is deep-discharged, or it continues to be discharged, despite its terminal voltage below a certain threshold called the cutoff voltage, it may become short-circuited, transitioning the cell into an irreversible condition. When the cells are connected in parallel, on the other hand, it is important to balance their voltages, since their interactions and dependencies make their voltages drift apart. Higher-voltage cells may then inversely charge the lower-voltage cells, causing the entire terminal voltage to drop from the desired value of the parallel-connected cells. Moreover, a Lithium-ion cell has unique characteristics [18], such as discharge efficiency (the higher the discharge rate, the lower the deliverable capacity), and recovery efficiency (the interface-concentrated gradient inside the cell is diffused during a ‘rest,’ after which the cell can be charged with large electric current over a short time).

Given parallel-connected cells, we can schedule their charge, discharge, and rest activities. For instance, each cell can be discharged in a round-robin fashion. Furthermore, the amount of discharge time can be scheduled in proportion to the remaining charge current in the corresponding cell, indicated by the State-of-Charge (SoC) level. This scheduling leads to weighted (with SoC levels) round-robin scheduling, called the weighted-IRR scheduling. Similar concepts are found in areas of scheduling tasks in distributed real-time systems, as well as scheduling packets in the differentiated service architecture, such as those in [15, 16, 21, 22, 25]. Alternatively, one or all cells can be discharged sequentially or simultaneously in parallel. This type of cell discharge has been used widely, especially for mobile devices such as PDAs and laptops. Also, using an analytical method, these two mechanisms have been compared in [2, 5]. In general, however, no single mechanism outperforms the others in all circumstances, thus calling for a thorough study of this issue.

Two main challenges exist in scheduling charge, discharge, and rest activities for large-scale battery systems. First, a scheduling framework should operate reasonably well in all circumstances. That is, using the framework, one should be able to extend a battery cell’s operation-time as much as any other scheduling mechanism can. By ‘operation-time,’ we mean the cumulative time of the charge current drawn from a battery cell until the cell no longer delivers the required charge current to applications. That is, the operation-time ends when the terminal voltage of the cell falls below the cutoff voltage.

To extend the cell’s operation-time, we need to understand the battery characteristics, such as the discharge and recovery efficiency. Second, a scheduling framework should be robust to
discharge and recovery efficiency. Section IV describes the battery reconfiguration system, the background of battery characteristics, and the motivation of this work. Section III models the battery characteristics, including the discharge and recovery efficiency. Section IV describes the design of the \textit{kRR} scheduling framework that consists of the adaptive filter and \textit{kRR} scheduler, based on the battery reconfiguration system. Section V presents the analysis of cell arrangements and discharge profiling. Section VI evaluates the performance of the \textit{kRR} scheduling framework. We discuss the related work in Section VII and conclude the paper in Section VIII.

II. RECONFIGURABLE BATTERY SYSTEM, BATTERY CHARACTERISTICS, AND MOTIVATION

This section introduces a reconfigurable battery system, provides a physical insight into rechargeable battery cells, and then describes the motivation behind our work.

A. Reconfigurable Battery System

A reconfigurable battery system that we developed earlier [17] is composed of a controller that manages a set of control units, and an array of battery cells. Each cell is equipped with a set of switches that each control unit is responsible for turning on or off, so that the cells can be connected online in series, in parallel, or both. For instance, as shown in Fig. 1, when we want the cells to be connected in parallel, each control unit turns on switches (1) and (4). When switch (3) instead of (4) is turned on, the cells become a series chain. We can also make multiple parallel groups, e.g., the simplest parallel group is made by turning on a first cell’s switches (1) and (3), and the following cell’s switch (4). More cells can be added in between. This group can then be separated by terminal switches (5) and (6). Terminals are connected to the load, or disconnected if needed.

![Fig. 1. The schematic diagram of the battery reconfiguration system that consists of a set of battery cells and a controller that encompasses multiple control units (CUs). Each CU is responsible for controlling 6 switches.](image)

Here we rely on the system’s three types of capability. First, the coordination of terminals’ connectivity and the terminal switches allows the cells to be charged and discharged simultaneously. That is, we can virtually partition the cells into two groups for charge and discharge activities, respectively. Second, an appropriate combination of on-switches allows for parallel-connected groups of the cells. These groups can then selectively be discharged at a time. Third, a single battery pack can be treated as one module, like a single cell, by connecting all the cells in the battery pack in series. These battery packs can then be connected in series, in parallel, or both. For simplicity, a cell is regarded as a module on which charge and discharge activities are scheduled. Based on these three types of capability, we build a framework of scheduling the charge, discharge, and rest activities for battery cells.

B. Battery Characteristics and Motivation

A rechargeable battery cell (e.g., NiCd, NiMH, and Lithium-ion) is capable of converting chemical energy to electrical energy, and vice versa, via electrochemical oxidation and reduction reactions [3, 18]. These reactions involve the exchange of electrons through the load between electro-active species in two electrodes inside the cell, generating a flow of electric current. Ideally, the total charge, counted in coulomb, from the cell will always be the same throughout its entire life cycle. In reality, however, the characteristics of a cell are nowhere close to being ideal due to the uncertainty of reaction kinetics, diffusion process, aging, and side effects (e.g., active material dissolution, electrolyte decomposition, and passive film formation [18]) inside the cell over time. The battery
type considered in this paper is assumed to be Lithium-ion or Lithium-polymer, for it has been widely used for power devices ranging from mobile handsets to electric vehicles.

We have observed two important aspects of the real-life battery cell characteristics. First, a high discharge rate affects greatly the cell’s operation-time. That is, the higher the discharge rate, the less efficient the conversion of the cell’s chemically-stored energy to electrical energy, thus degrading more in the deliverable capacity. According to our preliminary study (see Fig. 2), the cell of 3600 mAh capacity takes 1 hour to be fully discharged at rate C—1 coulomb is drawn every second. If the discharge rate doubles, its deliverable capacity decreases by 4.7%.

![Voltage degradation at various discharge rates](image)

Fig. 2. Voltage degradation at various discharge rates

Second, the cell has a limited charge recovery effect at a high discharge rate; a high load for a short period of time causes a high interface-concentrated gradient among electro-active species inside the cell, making the usable charge temporarily unavailable due to the lag between reaction and diffusion rates [18]. Thus, when the cell is allowed to ‘rest’ for some time at a low (or zero) discharge rate, the voltage that dropped temporarily goes back up, referred to as the recovery efficiency.

This recovery efficiency is instrumental in extending the operation-time of a large-scale battery pack for powering heavy workloads, such as electric vehicles. For instance, as shown in Fig. 3, a high load causes a temporary voltage drop, and then part of the cell’s voltage is recovered after a certain period of rest. When the cell is discharged for 15 minutes at 4C, its terminal voltage drops 7%. Then, after the cell is rested for 15 minutes, the dropped voltage is restored up to 5.5%. Moreover, when the cell capacity is low, at a high discharge rate, the concentration gradient of active species inside the cell gets steeper, reaching the cutoff voltage (see the large voltage drop at the 7-th discharge activity in Fig. 3). This drop below the cutoff voltage causes the disconnection of the cell from the load, assuming that the cell has been fully discharged. Instead of immediate disconnection, by diverting the load demand to other cells and resting the exhausted cell, it is recovered quickly, lasting for 40 minutes at C/4.

The first question is then “how can the discharge rate be adjusted for individual cells?” Although the battery management cannot directly control the loads, their connection can be diverted to selected cells. By dynamically selecting the number of cells, the discharge rates can be adjusted independently of the loads, and the rest time for an exhausted cell can also be scheduled. The second question is then “how long does an exhausted cell have to rest for its recovery?” To answer this question, we must understand the recovery efficiency in detail. The remainder of this paper will focus on answering these questions.

III. MODELING OF BATTERY CHARACTERISTICS

An accurate estimation of battery-cell characteristics is essential for scheduling charge, discharge, and rest activities. In this section we model the battery characteristics including the discharge and recovery efficiency, with the goal of generating a reference model for the cell characteristics.

A. Discharge Efficiency

As shown in Fig. 2, the voltage-drop curves for different discharge rates are very similar. That is, increasing the discharge rate by 1C decreases the operation-time by 50%. Given the curve-shape similarity, we want to derive the relationship between the voltage-drop curves with respect to the operation-time. First, a nonlinear voltage curve at a constant discharge rate iC is defined as an invertible function, \( F_i : t \rightarrow V \). We specify the curve \( F_{\text{ref}} \) with the constant discharge rate C as a reference. Then, the operation-times associated with \( F_{\text{ref}} \) correspond to those with \( F_i \), based on their shape similarity:

\[
F_i(t_i) = F_{\text{ref}}(t_r),
\]

where \( t_i \) (\( t_r \)) is in the operation-time domain of \( F_i \) (\( F_{\text{ref}} \)). The reference operation-time can be solved offline by using simple linear curve fitting. On the other hand, the relationship between \( t_i \) and \( t_r \) is expressed as

\[
t_i = F_i^{-1}F_{\text{ref}}(t_r).
\]

The relationship \( F_i^{-1}F_{\text{ref}} \), as shown in Fig. 4, can be approximated to be a set of linear functions \( \Pi = \{ \pi_1, \pi_2, \ldots, \pi_{n} \} \), where \( \pi_i = a \times t + b \) is differentiated by the discharge rate. For instance, given two points in the operation-time domain with a discharge rate iC, two corresponding points in the operation-time domain are obtained via \( \pi_i \). These two points are then applied to \( F_{\text{ref}} \), yielding a voltage drop.

The reference, \( \Pi \) and \( F_{\text{ref}} \), can also be used to simulate or estimate the cell’s operation-time, as in [26]. The reference-based estimation, however, must consider the recovery efficiency to enhance the accuracy in estimation.
The recovery effect depends on the discharge rate, the discharge time, and the rest time. First, the recovery effect is proportional to the increase in the discharge rate. This relationship, however, is not linear, as shown in Fig. 5-(a), because the diffusion process occurs even during the course of the low-discharge activity. From the recovery-efficiency curve, one can find local optimal discharge rates that maximize the recovery efficiency. For instance, when the cell is discharged at 0.8261C or 2.0435C, its recovery effect is locally maximal. Second, the discharge time over which the cell is continuously discharged has a similar effect on the recovery efficiency as shown in Fig. 5-(b). For instance, when the cell is discharged for 5 or 13 minutes at 0.8261C, the recovery efficiency is locally maximal. Third, most of the charge current that has been temporarily unavailable is recovered after a short rest. Fig. 5-(c) shows the average cumulative recovery rate with respect to the rest time. 70% of the dropped voltage is recovered within one minute, and 85% within two minutes.

The correlation among these three elements indicates the monitoring interval at which the charge, discharge, and rest activities should be scheduled and the number of parallel-connected cells to be discharged, yielding the recovery-efficiency factor, v. For this purpose, we need to establish the correlation, denoted by the function \( F_r : c \times t_d \times t_r \rightarrow V \), where \( c \), \( t_d \), and \( t_r \) are the discharge rate, the discharge time, and the rest time, respectively. Solving for \( F_r \) can be facilitated by applying a multivariate linear regression for each given value of \( t_r \), yielding a set of functions. We compute the first-order derivative of \( F_r \) with respect to \( c \), and solve \( \frac{dF_r}{dc} = 0 \) for \( v \), as \( dc \rightarrow 0 \). Since we are interested in local maxima, \( v \) is filtered under the condition that \( \frac{d^2F_r}{dc^2} < 0 \). As a result, \( v \) is used to determine \( k \), the number of parallel-connected cells, which will be detailed next.

IV. Weighted-kRR Scheduling

This section first describes the architecture of the weighted-kRR scheduling framework, and then details its components.

A. The Architecture

The architecture of the weighted-kRR scheduling framework, as shown in Fig. 6, is made up of the adaptive filter, the kRR scheduler, and a battery pack built with a battery-cell array and switches. The input of the adaptive filter is a history of the loads measured at certain intervals, and an estimate of the upcoming load demand is provided as an output. The estimated load demand is passed to the kRR scheduler. The scheduler, using the feedback from cells, manages charge, discharge, and rest activities with the help of the battery reconfiguration system. The scheduler is responsible to solve for a group threshold, \( \delta_G \), with which to partition the cells into two groups, and determine \( k \), the number of cells in one group to be discharged within an interval. These two parameters are periodically updated and adapted to various loads. Particularly, the two groups partitioned via \( \delta_G \) can be charged and discharged simultaneously. At some point, however, either the charge or discharge process is allowed exclusively. In what follows, we give a detailed account of each of the above two components.
B. Adaptive Filter

The filter uses a recursive least-squares (RLS) algorithm [14], and measures & records loads including \( d_{n-1} \) at \( p + 1 \) intervals (i.e., \( (p + 1) \times \Delta t \)), yielding a history of the loads, \( u(n-1) = [d_{n-1}, \ldots, d_{n-p-1}]^T \). Since the importance of each element in \( u(n-1) \) is different, \( u(n-1) \) is weighted by a vector filter \( w_n = [w_{n-1}(0), w_{n-1}(1), \ldots, w_{n-1}(p)]^T \), estimating the load demand, \( d_n^* \) in the next \( \Delta t \) as:

\[
d_n^* = w_n^T \cdot u(n-1).
\]  

(3)

Next, we want to minimize the estimation error:

\[
e(n) = |d_n - d_n^*|.
\]  

(4)

Since the filter weights/coefficients dictate the estimation error, we solve for the filter coefficients to do so. We apply a weighted-least-squares error function as a function of \( e(n) \):

\[
f(w_n) = \sum_{i=0}^{n} \lambda^{n-i} |e(i)|^2,
\]  

(5)

where \( 0 < \lambda \leq 1 \) is an exponential forgetting factor that effectively limits the number of input samples; the smaller the value of \( \lambda \), the more sensitive to recent samples the filter becomes, causing more fluctuations in the filter coefficients.

The number of operations required for calculation of \( w_n \) grows proportionally to \( n \), as the number of columns in \( u(n) \) increases. To avoid the repetitive execution of the least squares algorithm in updating \( w_n \), we apply RLS as:

\[
w_n = w_{n-1} + \Delta w_{n-1},
\]  

(6)

where \( \Delta w_{n-1} \) is a correction factor expressed as \( e(n) \cdot g(n) \), where \( g(n) \) is a gain vector. \( g(n) \) is obtained by solving the following equation:

\[
g(n) = \frac{P(n-1)u(n)}{\lambda + u^T(n)P(n-1)u(n)},
\]  

(7)

where \( P(n) \) is a \( (p + 1) \times (p + 1) \) inverse correlation matrix. \( P(n) \) is solved recursively via

\[
P(n) = \lambda^{-1}P(n-1) - g(n)u^T(n)\lambda^{-1}P(n-1),
\]  

(8)

where \( P(0) = \eta^{-1}I \) where \( I \) is the \( (p + 1) \times (p + 1) \) identity matrix and \( \eta \) is a positive constant.

The RLS algorithm, thus, provides an effective way of updating the filter while reducing the computation overhead, yielding an estimate of the load demand, \( d^* \), that will be passed to the scheduler.

C. kRR Scheduler

A battery cell’s voltage is partly proportional to its SoC level that can be determined by counting coulombs. However, a lower-voltage cell can be at a higher-SoC level than a higher-voltage cell, due to the discharge and recovery efficiency. This discordance suggests that some cells whose voltages fall below the cutoff voltage still have a sufficient charge current to deliver, making it essential to schedule the charge, discharge, and rest activities.

1) Partitioning Battery-Cell Array: The scheduler manages all cells in the array in accordance with their SoC level, partitioning it into two groups: (1) \( G_H \) in which the member cells’ SoC level is higher than \( \delta_G \), and (2) \( G_L \) in which the member cells’ SoC level is lower than, or equal to, \( \delta_G \). Also, when an individual cell’s voltage is below the cutoff voltage, the cell is put into \( G_L \). There are two reasons for this partitioning. First, the scheduler prevents cells from being deep-discharged, which may otherwise cause an irreversible damage to the cells. If a high load on the cells with a very low SoC level causes their voltage to drop sharply below the cutoff voltage, \( G_L \) plays a buffering role to reserve energy, allowing only a low load demand to be met. Second, the scheduler allows cells to be charged and discharged simultaneously. Cells in \( G_L \) can be charged while those in \( G_H \) are being discharged. (We will describe the charging mechanism shortly.)

At a high discharge rate, a cell’s voltage drops rapidly, while it drops slowly at a low discharge rate. Thus, \( \delta_G \) must be dynamically adjusted to the load. This dynamic adjustment is made as

\[
\delta_G = d^* \times \frac{d_{\text{max}}}{NC},
\]  

(9)

where \( d_{\text{max}} \) is an estimated maximum load demand within a time interval \( \Delta t \), and \( NC \) is the cell’s nominal capacity.

Fig. 7 shows the pseudo-code of partitioning the cells in the array for charge and discharge activities. First, the scheduler sets \( \delta_G \) based on Eq. (9). The value of \( \delta_G \) is updated during the partitioning. When the SoC of any cell in \( G_H \) is lower than \( \delta_G \), the corresponding cells are put into \( G_L \). When \( G_H = \emptyset \), i.e., no cells available to be discharged, some cells in \( G_L \) are put back into \( G_H \) with a decreased value of \( \delta_G \). \( \delta_G \) is linearly decreased by \( \alpha < 1 \). A flag, \( F_S \), is turned on, indicating that the SoC level of all cells is too low to meet a high load demand. Thus, the scheduling mechanism is switched to \( nRR \) from \( kRR \). Conversely, when \( G_H \) is not empty, the scheduler checks whether there are any charged cells in \( G_L \) whose SoC is greater than, or equal to, the average SoC in \( G_H \). If there are, they are placed in \( G_H \). As a result, all the cells in the array are classified into the two groups in accordance with their SoC level.

<table>
<thead>
<tr>
<th>Partitioning:</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E_{\text{SoC}}(G) ): return the average SoC of elements in ( G );</td>
</tr>
<tr>
<td>( \alpha ): a linear decreasing factor;</td>
</tr>
<tr>
<td>( F_S \leftarrow \text{Off} );</td>
</tr>
<tr>
<td>( \delta_G \leftarrow \text{Eq. (9)}; ) /* updated for discharging */</td>
</tr>
<tr>
<td>move cells ((s \in G_H)) into ( G_L ), conditioning ( \text{SoC}_{\text{cell}} \leq \delta_G );</td>
</tr>
<tr>
<td>if ( G_H = \emptyset );</td>
</tr>
<tr>
<td>( \delta_G \leftarrow \delta_G \times \alpha ); /* decremented */</td>
</tr>
<tr>
<td>move cells ((s \in G_L)) into ( G_H ), conditioning ( \text{SoC}_{\text{cell}} &gt; \delta_G );</td>
</tr>
<tr>
<td>( F_S \leftarrow \text{On} ); /* switching kRR to nRR */</td>
</tr>
<tr>
<td>else /* balancing SoC for charging */</td>
</tr>
<tr>
<td>move cells ((s \in G_L)) into ( G_H ), conditioning ( \text{SoC}<em>{\text{cell}} \geq E</em>{\text{SoC}}(G_H) );</td>
</tr>
</tbody>
</table>

Fig. 7. Partitioning of battery cells

2) Determining \( k \): The load demand, \( d \), is to be shared primarily among the cells in \( G_H \). Selecting an appropriate value of \( k \) is crucial since the more cells become available,
the lower the per-cell load. For this purpose, based on the recovery-efficiency factor \( v \), and \( d^* \) from the adaptive filter, \( k \) is given by

\[
k = \frac{d^*}{v},
\]

where \( 1 \leq k \leq n \), and \( n \) is the total number of available cells in \( G_H \). Although \( k \) is within the range, when the per-cell load is too low (i.e., \( v > d/k \) and \( k > 2 \)), \( k \) is decreased by 1, thus increasing the per-cell load. When \( v \) is in the form of an array, \( k \) also becomes an array. This way, \( k \) is determined to exploit recovery efficiency and prevent each cell from excessive discharge.

The scheduler selects \( k \) cells with the highest SoC level in \( G_H \), referred to as the \( kRR \) scheduling, and then determines the per-cell load. Each cell \( i \) shares \( d \), weighted by

\[
d \times \frac{SoC_i}{SoC_1 + \ldots + SoC_i + \ldots + SoC_k}.
\]

This weighted load-sharing is effective for balancing voltages of the \( k \) cells, even when some of them become faulty.

3) Discharge Activity: A scheduling mechanism can vary with \( k \). First, when \( k = 1 \), at each interval, the cell with the highest SoC level is selected, referred to as round-robin (1RR) scheduling. The 1RR scheduling is effective for balancing SoC levels and voltages of the cells, while it may cause excessive per-cell discharge and the overhead for frequently switching from one cell to another. Second, when the cell remains in connection with the load until it gets fully discharged to the end of its operation-time, this mechanism is referred to as sequential (1+1RR) scheduling. Obviously, the 1+1RR scheduling has nothing to do with voltage-balancing, but it is vulnerable to excessive per-cell discharge, unable to exploit the battery characteristics. Third, when all \( n \) cells in \( G_H \) are discharged together until the end of their operation-time, this mechanism is referred to as parallel (nRR) scheduling. The nRR scheduling is robust against excessive per-cell discharge, although it fails to exploit the battery characteristics. Conversely, the \( kRR \) scheduling acts as 1RR at a low discharge rate and as nRR at a high discharge rate, achieving the merits of all. Note that the \( kRR \) scheduling subsumes all of these three scheduling mechanisms. Fig. 8 shows the pseudo-code of the \( kRR \) scheduling of the discharge activity.

4) Charge Activity: For the charging process, priority is given to the cells in \( G_L \). The cell with the lowest SoC level is selected and charged until its SoC level reaches the average SoC level of the cells in \( G_H \). The charged cell is then put into \( G_H \). This procedure, called quick average charging, repeats until no cell is left in \( G_L \). This way, undesirable cases, such as voltage-imbalances at any point in time, are avoided. After all cells in \( G_L \) are charged, the cells in \( G_H \) start to be charged. Unlike the way applied in \( G_L \), all the cells in \( G_H \) are charged at the same time. The reason for this is that despite the unpredictable disruption of the charging process, the cells should be available to be discharged without the concern of voltage-imbalances. To prevent any cells from being overcharged, the charging process stops whenever any of the cells in \( G_H \) is fully charged. Then, individual cells with a lower-SoC level are separately charged. Fig. 9 shows the pseudo code of the \( kRR \) scheduling of the charge activity.

**Discharging (method, \( F_S, G_H \)):**

- \( N(G) \): return the number of elements in \( G \);
- \( H(k, G) \): return \( k \) cell(s) with the highest SoC level in \( G \);
- \( A(G) \): return the first available cell in \( G \);

if \( G_H = \emptyset \)
return; /* fully-discharged */

switch method

    case ‘kRR’:
        \( k \leftarrow \text{Eq. (10)}; \)
        if \( k > N(G_H) \mid F_S \) is On
            \( k \leftarrow N(G_H); \) /* switching kRR to nRR */
            \( G_k \leftarrow H(k, G_H); \)
        case ‘1RR’:
            \( G_k \leftarrow H(1, G_H); \)
        case ‘1+1RR’:
            \( G_k \leftarrow A(G_H); \)
        case ‘nRR’:
            \( n \leftarrow N(G_H); \)
        \( G_k \leftarrow n \) cell(s) \( \in G_H \);
        Discharge \( G_k \) based on Eq. (11);
        
        **Fig. 8. kRR scheduling of discharge**

**Charging (\( G_H, G_L \)):**

- \( V_n \): nominal voltage;
- \( L(k, G) \): return \( k \) cell(s) with the lowest SoC level in \( G \);

if \( G_L = \emptyset \)
/* The discharge process is inactive */
\( G_c \leftarrow G_H; \)
else if \( G_H = \emptyset \)
\( G_c \leftarrow G_L; \)
else
/* The discharge processes can be active */
\( G_c \leftarrow L(1, G_L); \)
Charge cell(s) \( (\in G_c \& V_{cell} < V_n) \) /* Prevention of overcharge */

**Fig. 9. kRR scheduling of charge**

The quick average charging is resilient against overstresses that occur to the cells experiencing an intensive charge-and-discharge cycle. To illustrate this, suppose a simple policy is used where individual cells in \( G_L \) are deferred from their transfer to \( G_H \) until they are fully charged. In this policy, when the charging process is switched to the discharging process, a just-fully-charged, weak cell—note that a weak cell may be charged and/or discharged faster than healthy cells—is likely selected first for discharge due to its higher-SoC level, and then continues being discharged until its SoC level reaches the average SoC level of the cells in \( G_H \). At a high discharge rate, the weak cell’s SoC level, however, falls below \( \delta_c \), thus transferring the weak cell back to \( G_L \). Then, the weak cell is most likely selected for charge again. As a consequence, the cycle of charging and discharging the weak cell makes it weaker, eventually leading to its failure.

**D. Main Program**

A main scheduling program involves offline and online procedures. In the offline procedure, the reference to the battery characteristics is created via the charge-and-discharge cycle; this cycle may repeat several times. The reference is used to set
important system parameters including \( v \), the nominal voltage, and the cutoff voltage. In the online procedure, the reference is updated since the physical condition of the cells changes over time as long as the charge-and-discharge cycle is repeatedly scheduled. In particular, \( v \) is updated periodically, effectively offsetting the aging effect of the cells.

V. CELL ARRANGEMENT AND DISCHARGE PROFILE

This section describes the numerical evaluation of the arrangement of cells in a large-scale battery-cell array and specifies a discharge profile for the purpose of evaluation.

A. How Many Cells Do We Need?

Clearly, the more the cells included in the array, the higher the power it can deliver, but the costlier its manufacturing. Also, the efficiency of converting chemical energy to electrical energy is nonlinear, depending on applications including all components of an electric vehicle. This question must then be narrowed down to how many parallel-connected cells are necessary to allow a battery-cell array to be both schedulable and long-lasting.

Given the requirement of \( p \) watts where 745.70 watts is equivalent to 1 mechanical horsepower (HP), the cells are best to be connected in series and in parallel as follows:

\[
p = m \cdot V \times (1 + q) \cdot n \cdot v,
\]

where \( m \) and \( n \) are the numbers of series-connected cells and parallel groups, respectively; \( V \) is a cell’s discharge-varying terminal voltage; \( q \) is the redundancy factor, where \( q \geq 0 \). When \( q = 0 \), no redundant parallel groups are available for scheduling. Obviously, the higher the redundancy, the higher the schedulability.

Suppose that 35 HPs, on average, is required for an electric vehicle. Then, how many parallel groups should we have while exploiting \( v \)? This depends mainly upon the required supply voltage and \( v \). As shown in Fig. 10, the higher the supply voltage, the fewer the number of parallel groups, with the discharge rate adjusted to \( v \). Without redundancy, at a high discharge rate (the average electric current is set to \( v = 2.0435C \)) with the supply voltage of 615V \((m = 150)\), on average, 23 parallel groups are required. At a low discharge rate \((v = 0.8261C)\) with the same voltage, average 57 groups are required.

\[
\begin{align*}
&\text{Fig. 10. Battery-cell arrangement for the supply power of 35 HPs, with high} \\
&\text{recovery efficiency. The cell’s terminal voltage ranges from 3.5V to 4.1V} \\
&\text{during its operation-time. For simplicity, no redundancy is considered.}
\end{align*}
\]

B. Discharge Profile

A discharge profile is defined as workload, \( \rho \), over the operation-time, i.e., a sequence of variable loads, \( ds \), required. In the discharge profile, \( ds \) are approximated by piece-wise constant loads, i.e., represented by a set of \( M \) levels of the discharge rate, \((i_1, \ldots, i_M)\), where \( M \) is used to characterize \( d \), given \( \Delta t \) that is a fraction of the total operation-time, \( T \). As discussed earlier, the smaller the value of \( \Delta t \), the higher the accuracy in characterizing \( \rho \). In case \( \Delta t = T \), \( ds \) are considered constant. The load at time \( t \) is denoted as a stepwise function of time,

\[
d(t) = \sum_{j=1}^{T} \rho(j) \cdot 1(t_{j-1}, t_j)(t),
\]

where \( 1(t_{j-1}, t_j)(t) \) is an indicator function. \( \rho \) can be obtained from empirical measurements. For the purpose of evaluation, we generate a synthetic discharge profile, such that \( d(i) = d(i-1) + \gamma \), where \( \gamma \) is an increment or decrement that is expressed by a random discharge that follows a Gaussian distribution.

VI. EVALUATION

Our goal is to extend a battery pack’s operation-time and lifetime—defined as the duration during which the pack provides required (converted) electrical energy to the load while each of its cells repeats the charge-and-discharge cycle—by effectively scheduling charge, discharge, and rest activities. To evaluate the efficacy and efficiency of the weighted-\( kRR \) scheduling, the metrics we use include the battery pack’s operation-time and usability, and the reduction in voltageimbalance between parallel-connected cells in the pack.

We first describe the evaluation setup and then, based on the metrics, demonstrate the superior performance of the \( kRR \) scheduling over the other three scheduling mechanisms: \( 1RR \), \( 1+1RR \), and \( nRR \).

A. Evaluation Setup

We designed a battery management emulator that includes the four scheduling mechanisms, discharge profile, and battery-activity profiling. Two-step recovery effects are considered in the battery-activity profiling: cells are rested for \( \Delta t \) or \( 2 \times \Delta t \), where \( \Delta t = 1 \) (minute). To generate a reference including \( Fref \) and \( \Pi \) presented in Section III-A, we use Dualfoil [10, 12, 13]. Note, however, that the evaluation results in this section are not subject to Dualfoil’s accuracy in the emulation. Dualfoil simply offers a reference to Lithium-ion cell characteristics as most battery manufacturers do.

The battery management emulator uses the discharge profile, \( \rho \), specified in Eq. (13). In \( \rho \), the random discharge \( \gamma \) is normally distributed as

\[
\gamma \sim d_i \times N(0, 0.5),
\]

where \( d_i = 0.4C \times \Delta t \) which is a lower bound of \( \rho \). The upper bound, \( d_u \), of \( \rho \) is set to \( 4.3C \times \Delta t \). We also set \( d_{\text{max}} \) and \( d(0) \) to \( d_u \) and \( d_i \), respectively. Based on these parameter values, we specify the discharge profile. Given the discharge profile, the battery management emulator simulates the cell discharge according to each scheduling mechanism. The battery pack is assumed to contain 4 parallel-connected cells, but it can be extended in various ways, as discussed in Section II-A, e.g., \( n \).
parallel-connected battery packs of m-series-connected cells. In addition to the discharge profile, we set a cell’s nominal capacity (NC) to 3602.7mAh, assuming that all cells in the battery pack have the same characteristics, unless specified otherwise. Also, the cell’s terminal and cutoff voltages are set to 4.06267V and 2.00000V, respectively.

B. Evaluation Results

1) The battery characteristics are modeled effectively: Our battery-cell characteristics model is reference-based and captures the discharge and recovery efficiency. Based on the model, four scheduling mechanisms are comparatively evaluated. For this comparison to be effective, it is important to evaluate the accuracy in its simulation of battery activities. A discharge profile is synthetically specified such that the workload consists of charge at rate C for 10 minutes, rest for 2 minutes, charge at rate C for 10 minutes, and charge at rate 2C for 20 minutes; this is sufficient to show the discharge and recovery efficiency. The discharge profile is then ported to our battery management emulator and Dualfoil, thus yielding voltage curves over the operation-time as shown in Fig. 11-(a). The two voltage curves are almost identical except for the turning points at which the voltage drops steeply. To be more precise, we calculate $\chi^2$-distance [11] as:

$$\chi^2(X, Y) = \sum_{i=1}^{T} \frac{(X - Y)^2}{(X + Y)} \quad (15)$$

where $X$ and $Y$ are terminal voltage samples from Dualfoil and our model, respectively, and $T = 42$. Clearly, $\chi^2 = 0$ if and only if all $X$’s samples match $Y$’s. The lower the value of $\chi^2$, the more closely the two curves match. As shown in Fig. 11-(b), the difference between the two curves is negligibly small. Between the 10-th and the 12-th samples during which the recovery effect takes place, our model is shown to accurately simulate the recovery efficiency. Also, even when the discharge rate is abruptly changed at the 22nd minute, the distance value (i.e., difference), from that point, is still below 2.5E-4, and thus, the $\chi^2$ distance is as small as 0.0208 (V).

2) The adaptive filter operates effectively: The adaptive filter used in our scheduling framework is based on the RLS algorithm that recursively calculates an estimate from a given history of real outputs. The size of the history in this evaluation is set to 5, i.e., $p = 4$. Individual elements in the history are weighted separately; we $w$ with 50 samples, which converges to $[0.1012, 0.1916, 0.0708, 1.0377, -0.3881]^T$. In solving for the filter coefficients, the exponential forgetting factor $\lambda$ is set to 0.999, making it less sensitive to recent samples. Also, the gain factor $g(n)$ is obtained by recursively solving an inverse correlation matrix, $P(n)$ with $P(0) = \eta^{-1}I$, where $\eta = 100$. With these values applied, as shown in Fig. 12-(a), the adaptive filter is highly effective in estimating the outputs. The estimation errors are very low, as shown in Fig. 12-(b), and the average error ratio of estimated outputs to real outputs is as low as 0.0024.

3) $kRR$ outperforms the others: $kRR$ and $nRR$ are load-shared. That is, $d$ is distributed to parallel-connected cells available in proportion to their remaining SoC level. By contrast, individual cells are discharged, one at a time, in both $1RR$ and $1+1RR$. Thus, $1RR$ and $1+1RR$ are likely to make individual cells overloaded and hence exhausted. Fig. 13 compares the operation-time gains of $kRR$, $1RR$, $1+1RR$, and $nRR$ with the discharge profiles generated from Eqs. (13) and (14). Applying $kRR$, the battery pack lasts up to 44% longer than $1RR$, 56% longer than $1+1RR$, and 7% longer than $nRR$. $nRR$ outperforms $1RR$ and $1+1RR$ by 41% and 54%, respectively. $1+1RR$ performs as ineffectively as $1RR$.

4) $kRR$ and $nRR$ are suitable for heavy workloads: The results shown in Fig. 14 confirm these two mechanisms’ suitability for heavy workloads. By contrast, $1RR$ and $1+1RR$ make it overwhelming for individual cells to cope with the heavy workload, limiting their capability of effective scheduling. Although $1RR$ outperforms $1+1RR$, both $1RR$’s and $1+1RR$’s battery usability appears degraded significantly.

Fig. 11. Comparison of Dualfoil and our model w.r.t. voltage change

Fig. 12. An effective adaptive filter

Fig. 13. Comparison of the average operation-time gains with the four scheduling mechanisms from 100 runs of the charge-and-discharge cycle.
as the workload becomes heavier. Also, the difference in the battery usability between 1RR and 1+1RR, and kRR and nRR grows up to 19% larger.

kRR effectively adapts itself to different workloads developed in 4 phases. For workloads lighter than 1C × T (Phase 1 in the figure), kRR acts as 1RR which is most effective. This, compared to nRR, suggests that at a very low discharge rate, electrochemical dynamics inside the cell are inefficient in converting chemical energy to electrical energy although the effect is subject to individual cell characteristics. Compared to 1+1RR, 1RR is found to perform better at a low discharge rate, showing a 13% performance gain. As the workload gets heavier, however, 1RR loses its performance since the recovery efficiency lags behind the discharge rate. In Phase 2, kRR’s performance is excellent, because kRR effectively determines the right number of parallel-connected cells to accommodate the increasing workload while attempting to make best of their recovery efficiency. As a result, kRR outperforms 1RR and 1+1RR by up to 11% and 21%, respectively, and achieves a reasonable gain of 1.4% against nRR. In Phase 3, kRR’s scheduling performance is, on average, as good as nRR’s. In other words, as few parallel-connected cells as 4 are insufficient for kRR to exploit v. As the workload becomes even heavier (Phase 4), more parallel-connected cells are required since not enough time is spent on offsetting a significant voltage drop. Obviously, such a large voltage drop can be handled by using more parallel-connected cells. Thus, kRR acts as nRR. kRR maximizes the battery usability by scheduling the charge, discharge, and rest activities while adapting itself to the varying workloads.

Under heavy workloads, however, it appears difficult to adjust δG due mainly to unmanageable, abrupt, and steep voltage drops, implying that the high load for an individual cell is a dominating factor in the efficiency of scheduling the discharge and rest activities.

Fig. 15. Operation-time gain of kRR with dynamic δG over static δG; the same configuration as in Fig. 14 is applied.

6) kRR and 1RR are more tolerant of voltage-imbalance: Voltage-imbalance should be avoided in discharging parallel-connected cells. When voltages of individual cells drift apart—which is, to some degree, unavoidable—higher-voltage cells can charge lower-voltage cells, thereby causing the battery pack’s terminal voltage to drop. Thus, scheduling the discharge activity must cope with voltage-imbalance. As shown in Fig. 16-(a), kRR manages to keep the difference in voltage under 0.5%, while 1RR is subject to various loads and the cells’ SoC level since a single cell should accommodate the whole load at a time. For instance, at low SoC levels of the cells, the difference in their voltage goes up to 2.5% for the 1RR scheduling. Obviously, voltage-imbalance are rarely experienced under nRR as long as all the cells of the same characteristics are discharged simultaneously at the same rate. nRR is susceptible to cell failures or anomalous voltage variations (thus unbalancing voltages). As shown in Fig. 16-(b), despite the weighted discharge, nRR experiences voltage-imbalance. By contrast, kRR and 1RR quickly suppress it up to 50%.

5) Dynamic δG improves the schedulability: δG is a threshold according to which parallel-connected cells are partitioned into GH and GL. The cells whose SoC level is below δG are put into GL. This classification is to prevent the cells from suffering the entire terminal voltage drop of the battery pack due mainly to instantaneous high loads at their low SoC level. Moreover, GL serves as a buffer in which the charge is reserved and then delivered to accommodate light loads. The size of the buffer, however, should be adjusted to the load; the higher the load, the larger the value of δG. The dynamic adaptation of δG supports the schedulability and hence extends the battery pack’s operation-time. Fig. 15 shows the operation-time gains by applying the dynamic δG. Under wide-ranging workloads (below 3.3C × T), the dynamic δG proves to be effective; the maximum gain reaches 0.0179.

![Graph showing battery usability with respect to workload](image)

Fig. 14. Battery usability with respect to workload (C × T); the battery pack is discharged at a constant discharge rate over T. The discharge rate is increased by 0.1C from 0.4C to 4.3C. This procedure is repeated 100 times.

![Graph showing balancing voltages](image)

Fig. 16. Balancing voltages: The discharge profile shown in Fig. 12-(a) is applied. The difference in voltages is obtained by calculating (Vmax − Vmin)/Vmax × 100%.

**VII. Related Work**

Appropriate control of charge current and discharging behavior helps extend a battery’s operation and life times. For
instance, discharging cells completely one-by-one, in parallel, or in a round-robin fashion [15,16,25] makes a great impact on a battery’s operation and life times. A method for validating such policies usually relies on a stochastic model of battery behavior [7,19,20,24]. Using a stochastic model, Benini et al. [2] observed that virtual parallel discharge is more effective in maximizing a battery’s operation-time than sequential discharge. Similarly, Chiasserini and Rao [5] considered the recovery effect. Their models, however, capture only discharge behavior and ignore the dependency between the discharge rate and the recovery rate, thus limiting their scheduling efficacy. They need a comprehensive model for battery-cell characteristics to validate scheduling policies as well as task scheduling techniques such as those [21,22].

A synergetic battery pack [9] is a simple battery charger designed to charge 4 cells connected in series, with the accompanying control circuitry. In the pack, two switches at each cell are set at its boundary, making it connected to or disconnected from power buses. Likewise, a battery switch array system [1,23] presents an architecture for arranging micro-scale cells, allowing them to be bypassed. This system, however, similar to a configurable multi-cell battery [6], fails to select appropriate cells dynamically, based on battery-cell characteristics such as SoC, state-of-health, and load demands.

VIII. Conclusion

For longer operation and life times of a large-scale battery pack, one must effectively and efficiently schedule the charge, discharge, and rest activities for the cells in the pack. In this paper, we presented the weighted-\textit{kRR} scheduling framework to not only extend the battery pack’s operation-time, but also make it robust to anomalous voltage-imbbalances. We started by modeling the battery-cell characteristics, including the discharge and recovery efficiency. We found conditions under which the recovery efficiency is maximized. Based on this finding, we designed the \textit{kRR} scheduling framework to effectively schedule the charge, discharge, and rest activities. The framework is composed of the adaptive filter and the \textit{kRR} scheduler. The adaptive filter returns an estimate of the upcoming load demand. The scheduler determines \( k \), the number of parallel-connected cells to be discharged, while making best of recovery effects. The scheduler also balances the loads (and thus voltages) among \( k \) cells. Our evaluation results show that the \textit{kRR} scheduling framework allows the battery pack to last up to 56% longer than the \textit{1RR} scheduling, and be 50% more fault-tolerant of voltage imbalances than the \textit{nRR} scheduling. The \textit{kRR} scheduling framework effectively adapts itself to various workloads, improving battery usability. In summary, at a high discharge rate, the energy conversion in battery cells is inefficient, and their voltage-imbbalances must be minimized. The \textit{kRR} scheduling offers a unique solution to the inefficiency in charging and discharging battery cells, greatly improving the adaptability to various load conditions.

REFERENCES


