CHAPTER 16
ENERGY STORAGE PERFORMANCE TESTING

David Rosewater and David Schoenwald (Sandia National Laboratories)

Abstract

Fundamentally, energy storage (ES) technologies shift the availability of electrical energy through time and provide increased flexibility to grid operators. Specific ES devices are limited in their ability to provide this flexibility because of performance constraints on the rate of charge, rate of discharge, total energy they can hold, the efficiency of storage, and their operational cycle life. These performance constraints can be found experimentally through specific testing procedures. This chapter describes these tests and how they are applied differently at the battery cell and integrated system levels.

Key Terms

Beginning of life (BOL), capacity, capacity test, charge capacity, coulombic efficiency, depth of discharge (DOD), device under test (DUT), discharge capacity, electric power system (EPS), electric vehicle (EV), energy efficiency, hybrid pulsed power characterization (HPPC) test, open-circuit voltage (VOC), rate of charge, rate of discharge, rated capacity, reference performance test (RPT), state of charge (SOC), state of health (SOH)

1. Introduction

Battery energy storage systems (BESSs) are being installed in power systems around the world to improve efficiency, reliability, and resilience. This is driven in part by: engineers finding better ways to utilize battery storage, the falling cost of batteries, and improvements in BESS performance. Performance, in this context, can be defined as how well a BESS supplies a specific service. The various applications for energy storage systems (ESSs) on the grid are discussed in Chapter 23: Applications and Grid Services. A useful analogy of technical performance is miles per gallon (mpg) in internal combustion engine vehicles. A car's city mpg is a performance metric that gives a buyer proxy information about its efficiency, relative to other cars, and allows them to predict how much money they will spend on gas given their driving needs. This metric also allows regulators and policy makers to influence how efficient cars are. These factors are why performance is useful in the context of energy storage as well. Performance metrics in batteries, such as round-trip efficiency or degradation rate, allow customers, and regulators alike to make informed technical decisions. Utilities also use performance metrics in system planning to decide where to place energy storage on the power grid to maximize its impacts.

In addition to informing decision making, performance metrics can be used to automate charge/discharge decisions through controllers or energy management systems (EMSs). EMS controllers view the performance of ESSs as a set of equations governing the relationships among physical quantities such as power, state-of-charge, voltages, current, temperatures, and state-of-health. These models can be used to control ESSs to provide a range of grid services as described in *Battery Energy Storage Models for Optimal Control* [1].

Performance metrics are derived through regimented experimental analysis (i.e., testing). A performance testing procedure exposes the device under test (DUT) to a series of expected environmental and operational conditions to provide needed information about the device to a specific entity such as a company (e.g., the manufacturer) or another interested organization (e.g., the local utility).

> **Note 1**: Safety testing, which uses abnormal environmental and operational conditions, requires different methods and materials and is not discussed in this chapter. Also, testing on the materials and composites used to make energy storage components, while important in the research use to improve the technology, is out of the scope of this chapter. See Chapter 17: Safety of Electrochemical Energy Storage Devices for more information.

> **Note 2**: Performance is distinct from interconnection and interoperability, requirements for testing interconnection and interoperability can be found in *IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces* [2].

Performance tests can be categorized as:

- **Type tests** – Performed on a single unit as a representative sample of the design.
- **Unit tests** – Only concerned with the specific DUT.
- **Production tests** – Verifies if a unit functions as designed before it leaves the factory. Sometimes called factory acceptance testing. Performed at different times through a device's life.
- **Commissioning tests** – Verifies functionality after installation but before operation. Performed at different times through a device's life.
- **Periodic tests** – Checks if functionality has degraded over time or if it requires maintenance. Performed at different times through a device's life.
- **Reference performance tests (RPTs)** – A combination of commissioning and periodic tests used in long-term cycling or duty cycle testing programs.

The different test categories provide unique information needed in a specific circumstance. A production test may only check that a battery cell's voltage does not collapse on discharge, verifying that no manufacturing defects are present. A commissioning test may fully discharge the cell, verifying that the capacity required for the application is available. A periodic test may cycle the cell to calculate efficiency, which can be used to estimate when the cell may need to be replaced. The information needed about the DUT determines when the test should be performed, the category of the test, and the specific procedures.

This chapter reviews the methods and materials used to test energy storage components and integrated systems. While the emphasis is on battery-based ESSs, non-battery technologies such as flywheels and thermal storage are also discussed. Section 2 reviews the current state of energy storage performance testing and is divided into two main subsections: 2.1 on battery cell testing and 2.2 on integrated system testing. When reading procedures included in this chapter, keep in mind that they can be applied in any combination of testing categories depending on what information the entity needs and when it is needed. Section 3 discusses challenges in the existing state of the art and Section 4 discusses opportunities for future research and development.

2. State of Current Technology

2.1. Battery Cell Testing

The performance of battery cells is useful for: manufactures to benchmark technology improvements, system integrators to make supplier selection decisions, and battery management system (BMS) designers to develop state estimation models. Testing is performed on a sufficient number of cells in order to calculate statistical variations. A testing program is divided into reference performance tests and life-cycle testing.

2.1.1. Apparatus and Materials

Battery testing is performed in a controlled environment. As battery performance can be temperature dependent, temperature chambers are utilized to ensure experimental reproducibility. The measurement requirements for laboratory battery testing are outlined in the *United States Advanced Battery Consortium Battery Test Manual for Electric Vehicles* [3].

2.1.2. Reference Performance Testing Methods

Reference performance tests (RPTs), in the context of battery testing, are a combination of commissioning tests and periodic tests as they are performed at the beginning of life (BOL) and periodically throughout life cycle testing to get snapshots of a battery's capabilities and properties at a moment in time. This section describes two RPT procedures: the capacity test and the hybrid pulsed power characterization (HPPC) test. Similar test procedures for determining performance under a range of conditions are available in the *United States Advanced Battery Consortium Battery Test Manual for Electric Vehicles* [3].

2.1.2.1. Capacity Test

A battery's capacity is related to the energy that it can supply in a given application. Rated capacity, in the context of batteries, refers to the charge (in Ampere-hours) supplied by a battery at a C/3 rate over the full electrochemical range between Vmax100 and Vmin0, which are voltages defined by the manufacturer [3]. The rated capacity should be provided by the manufacturer, or it is established at the BOL and remains fixed during life aging.

> **Note:** The definitions of capacity and C rate are linked. Capacity is defined as the charge supplied at C/3, and C/3 is defined in terms of the rated capacity. In practice, batteries are designed to have a specified rated capacity, and this specification is what is used to first calculate C/3. The C/3 rate is then used to verify that the rated capacity has been achieved by a given design and from then on, the rated capacity and the associated C rates are defined together.

The terminology of "C rate" is used to scale the discharge rate according to the specified duration of discharge expected. For example: C/3 is the current at which battery is expected to take three hours to reach end of discharge. Hence, for a battery with a rated capacity of 3000 mAh (as specified by the manufacturer), C/3 would correspond to 1.0 A. The term capacity is often imprecisely used as if it were synonymous with rated energy [4], which is the energy (normally in Wh) supplied by an ESS at a specified rate over its operational range. These quantities are certainly related, but they very imply different assumptions about the tests used to calculate them.

Capacity testing is performed to understand how much charge / energy a battery can store and how efficient it is. In energy storage applications, it is often just as important how much energy a battery can absorb, hence we measure both charge and discharge capacities. Battery capacity is dependent on the discharge rate and temperature, so it is important to have multiple tests under a range of test conditions. To prepare the battery for discharge, it should be charged according to the manufacture's specifications. An example charging procedure is shown in Figure 1. The first stage is a constant current "bulk" charging stage. When the battery reaches its maximum operational voltage, the procedure switches to a constant voltage "taper" stage. The taper stage ends when the current required to hold the maximum operational voltage falls below a set threshold. At the end of the manufacturer's specified charging procedure the battery is defined to be at 100% state-of-charge (SOC). Note that a manufacturer can specify a voltage range that corresponds to a given design life, rather than the maximum and minimum safe voltages for the cell chemistry.

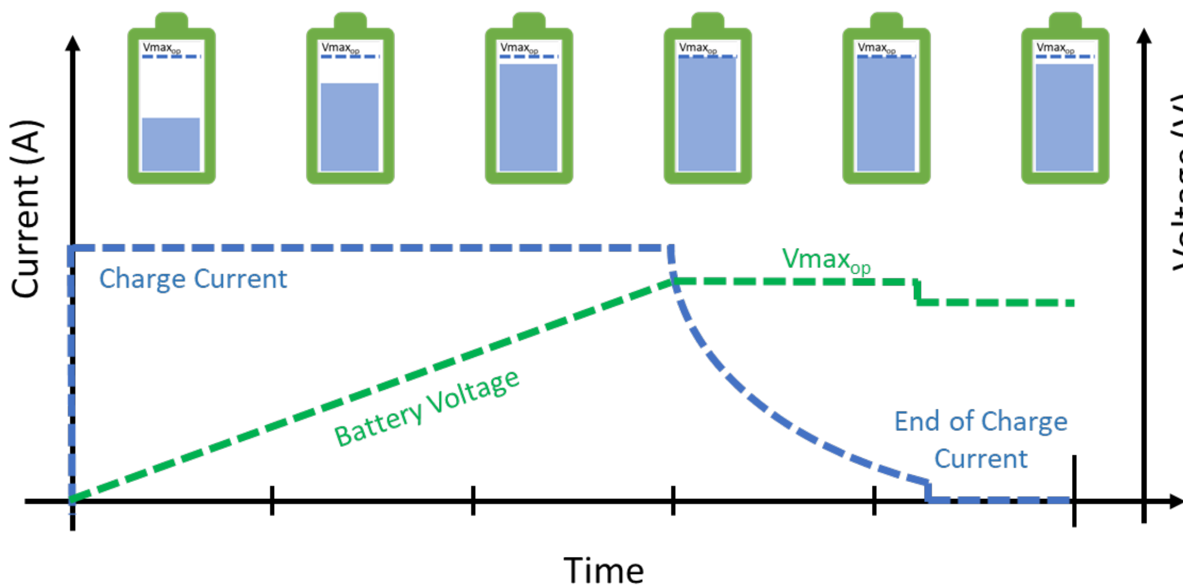

Figure 1. Example battery charging procedure

Once charged, a discharge capacity test proceeds in reverse of the charging procedure. An example discharge capacity test procedure is shown in Figure 2. Starting at 100% SOC, the discharge capacity test starts with a constant current "bulk discharge" stage. A battery should be tested at multiple constant currents over multiple tests. When the minimum operational voltage is reached, the procedure switches to a constant voltage "taper" stage. Note that this is different from procedures for mobile applications because a battery that is unable to sustain the current load required by a mobile system simply stops functioning. A grid-connected battery system simply reduces the maximum discharge power available and continues to supply what services it can. Once current needed to sustain the minimum operating voltage falls below the end of discharge current, discharge is stopped, and the battery is defined to be at 0% SOC

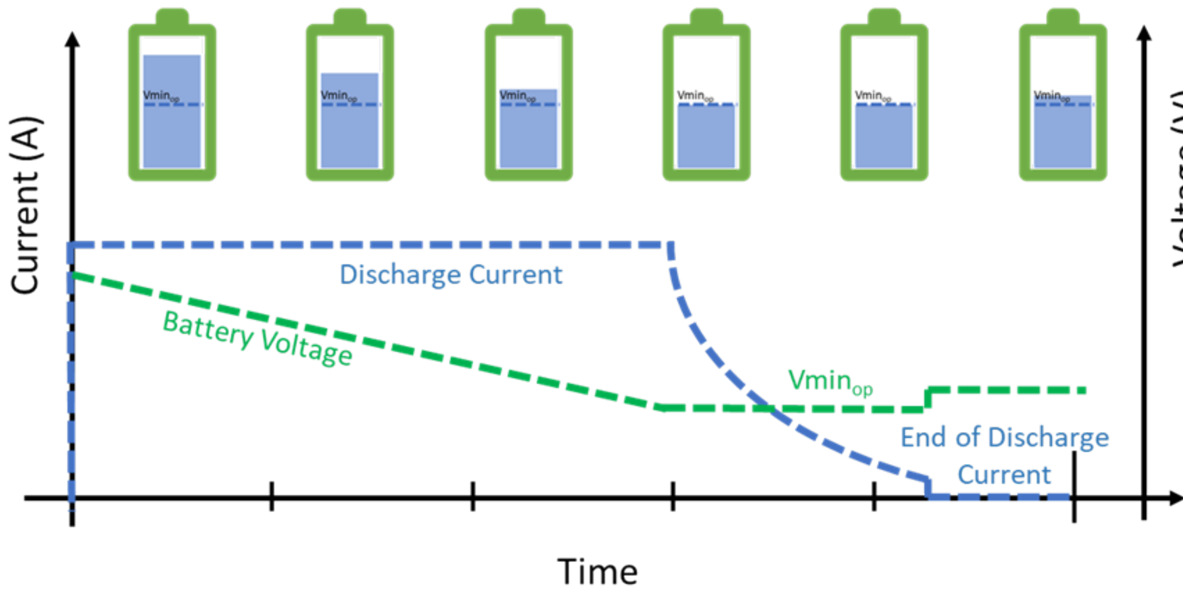

Figure 2. Example battery discharge capacity testing procedure

The battery's discharge capacity is calculated as the integral of current over time in Ampere-hours (Ah). Alternatively, the battery's discharge energy capacity is calculated as the integral of current multiplied by voltage over time in Watt-hours (Wh). Starting at 0% SOC, the charging capacity test is performed by executing the example battery charging procedure shown in Figure 1. The charging capacity, in both charge (Ah) and energy (Wh), are then calculated using the same method as the discharge capacity. The battery's coulombic efficiency is the ratio of charge capacity in Ah to discharge capacity in Ah, while the battery's energy efficiency is the ratio of charge capacity in Wh to discharge capacity in Wh. Note that discharge capacities, charge capacities, and both types of efficiency should be, and often is, reported in a battery's specification sheet along with the charge/discharge rates and environmental temperature at which they were calculated [4].

A battery will often retain a "memory effect" on performance for a short time. This means that a capacity test following a long rest will yield different results from a capacity test following a period of cycling. So, for replicability of results, the capacity tests should be performed three or more times in succession, with the performance metrics only being recorded starting in the third iteration. This practice allows the battery's performance to settle into capacity testing before the capacity is measured.

2.1.2.2. Pulsed Power Characterization

The hybrid pulsed power characterization (HPPC) test, widely used in EV battery testing [3], is designed to explore how much power is available over the operational range of SOC. This test procedure is modified for energy storage technologies to enable more precise model development. The overall test profile is shown in Figure 3. The profile starts with a capacity test to clear any memory from previous operation. Ten percent of the battery's capacity is then discharged, followed by a one-hour rest. The battery is then discharged for 10 seconds at whatever current is needed to reach the minimum operational voltage. The battery is then allowed to rest for 10 seconds, followed immediately by a charge pulse, where the battery is charged for 10 seconds at whatever current is needed to reach the maximum operational voltage. The power supplied during the discharge pulse and absorbed during the charge pulse are recorded. The process repeats by removing another 10% of the battery's capacity, followed by rest, followed by pulse testing (as shown in Figure 4), until the end of discharge current at the minimum operational voltage is reached in between pulse tests. At the end of the test the battery is restored to full charge.

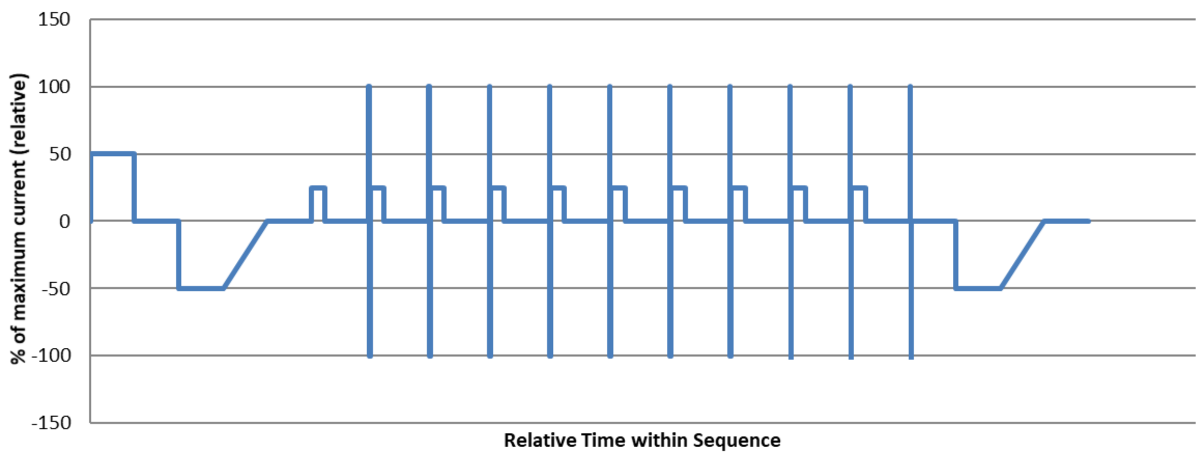

Figure 3. Pulse power testing profile (reproduced with data from [3])

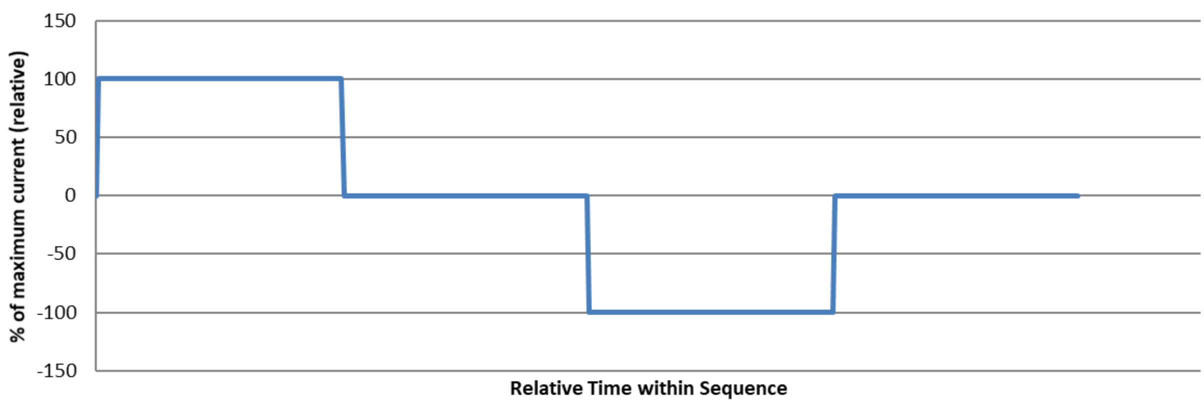

Figure 4. Discharge/Charge pulse test sequence (reproduced with data from [3])

Determining the current to use during pulse testing is not necessarily intuitive. An illustration of the maximum charge and discharge pulse power is shown in Figure 5. After an extended rest period, the battery voltage will be very near the open-circuit voltage (VOC) at the test SOC (a). The maximum discharge pulse power is the power at which the battery voltage reaches the minimum operational voltage (b), while the maximum charge power is the power at which the battery voltage reaches the maximum operational voltage (c).

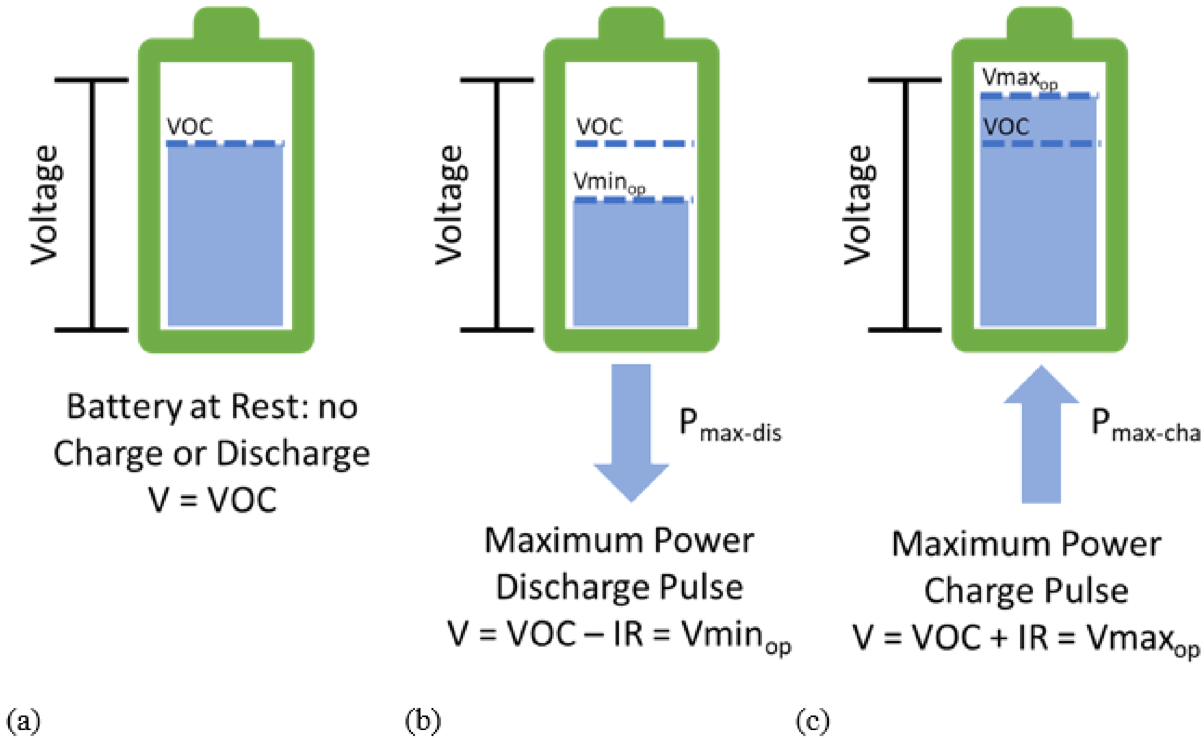

Figure 5. Illustration of maximum power pulse testing, (a) battery at rest, (b) discharge pulse, (c) charge pulse. V (voltage), VOC (open-circuit voltage), I (current), R (resistance), $Vmin_{op}$ (minimum operating voltage), $Vmax_{op}$ (maximum operating voltage), $P_{max\text{-}dis}$ (maximum discharge power at a given SOC), $P_{max\text{-}cha}$ (maximum charge power at a given SOC)

When these measurements are taken over the whole range of SOC, a plot such as the one shown in Figure 6 can be generated. In electric vehicles there are minimum performance targets for charge and discharge power that are used to determine the operational range of SOC available in that application. In grid connected systems there are not necessarily minimum performance targets, and the whole range of SOC could potentially be utilized. However, these curves will dictate the maximum power at different SOC. For example, a battery system with a 100-kW inverter, at 10% SOC may only be able to supply 20 kW on discharge but would likely be able to absorb the full 100 kW. Similarly, this system at 90% SOC may only be able to absorb 20 kW, while the full discharge power of 100 kW would be available. As a battery ages with time and use, its power capabilities and capacity tend to diminish. This has a net effect, illustrated in Figure 6, as reduced power HPPC test performance.

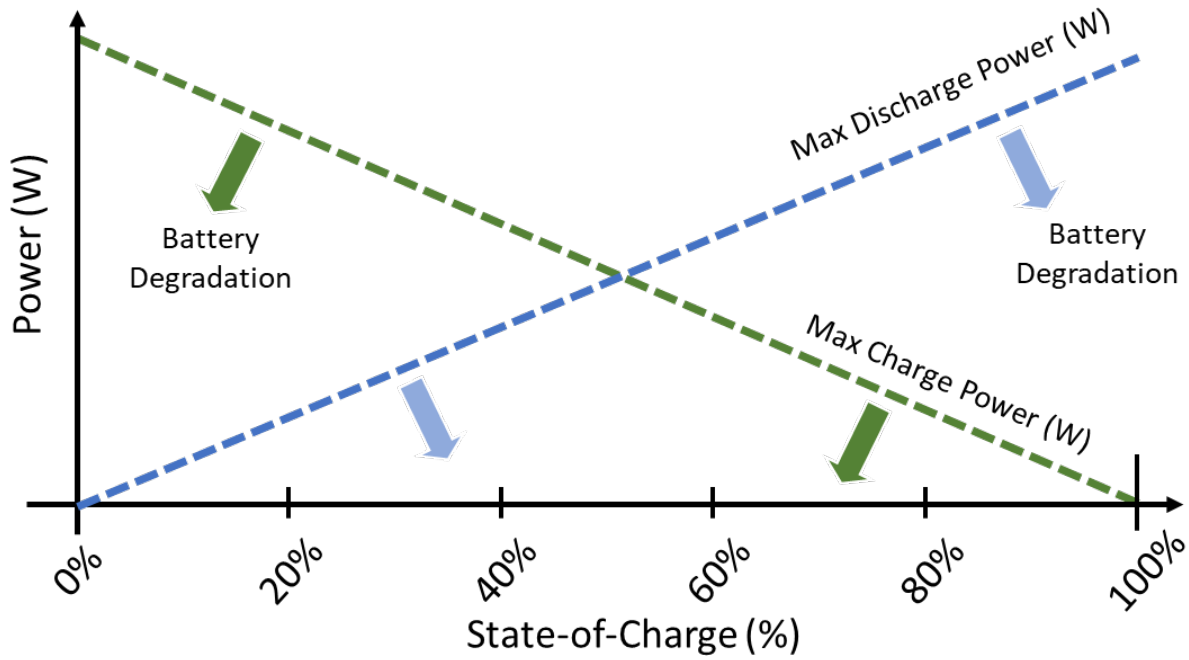

Figure 6. Illustrative example of HPPC test results and how they change with battery degradation (in real batteries, power limits are non-linear)

2.1.3. Life Cycle Testing Methods

Life cycle testing is meant to better understand how long specific battery cell type will last in each application or combination of applications. The term life cycle testing broadly applies to a range of specific tests. Life verification testing is a type of life cycle testing that establishes, with a defined level of statistical confidence, that a battery's life is at or above the expected life by using only a few years of accelerated ageing. Accelerated ageing is performed by defining and applying one or more elevated stress factors, such as temperature or depth-of-discharge (DOD), to induce a failure mode representative of actual use. By understanding the mechanisms of failure and designing tests to accelerate those mechanisms, the expected life under normal operating conditions can be modeled and predicted to a high degree of accuracy. Best practices for using accelerated ageing to predict and verify battery life come from the DOE-United States Advanced Battery Consortium, Electrochemical Energy Storage Technical Advisory Committee [5].

A simple example is included here to illustrate how an accelerated degradation testing program can achieve high confidence in design life. First, 210 cells randomly sampled from a production line are set aside then divided into six groups to be cycled under different total stress factors. If, under anticipated operation, the batteries undergo no more than two cycles per day on average, an accelerated degradation cycle to achieve a total stress factor of 10 may be designed with 20 cycles per day. Two cycling profile examples are shown in Figures 7 and 8. Stress factors are calculated by degradation models that are developed and verified by experimental analysis performed before the accelerated degradation testing program.

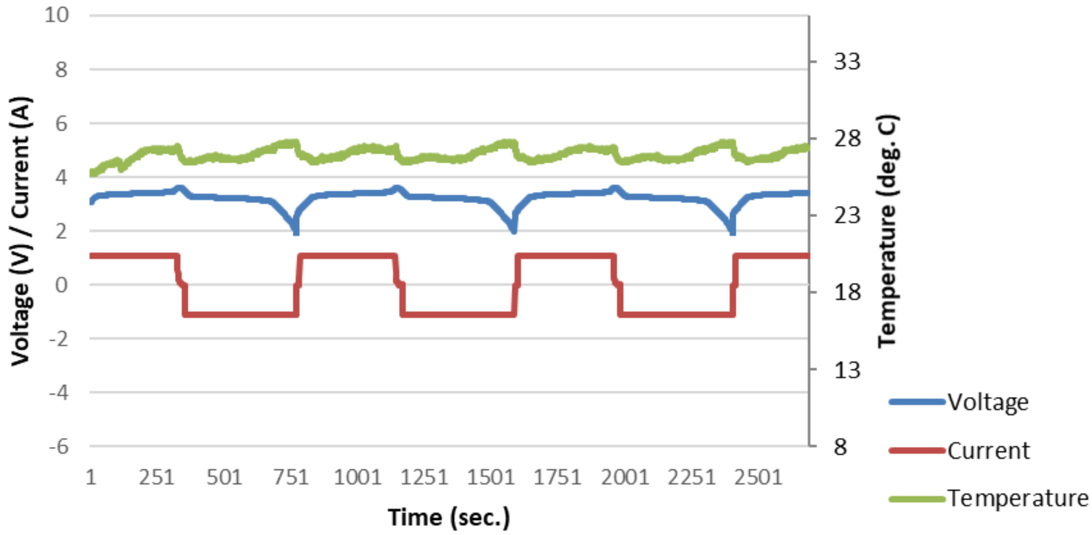

Figure 7. Example 1 – Accelerated Degradation Cycling (reproduced from cell testing data published in [6]) (assuming this profile has a total stress factor of 7.5)

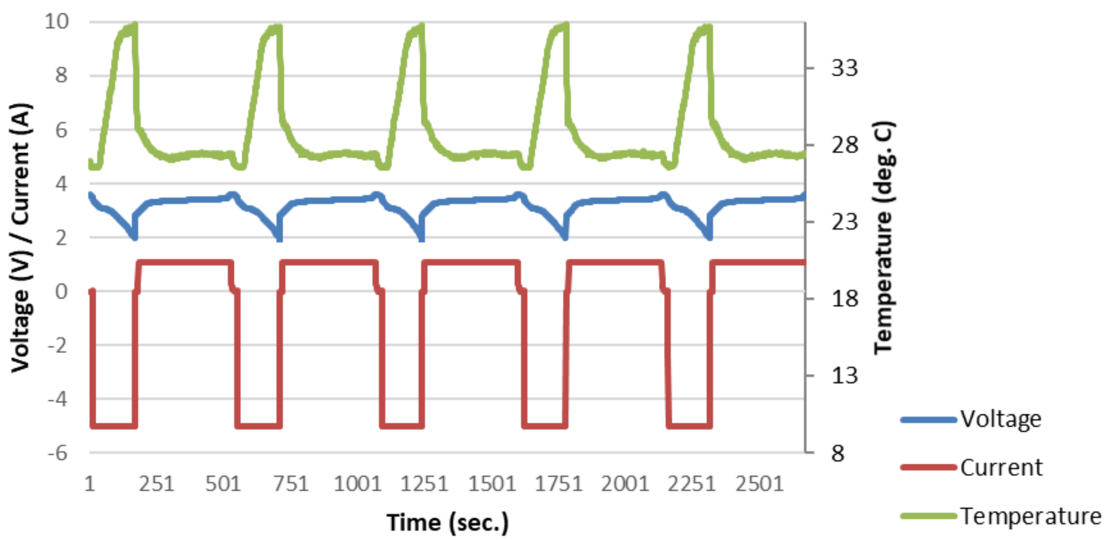

Figure 8. Example 2 – Accelerated Degradation Cycling (reproduced from cell testing data published in [6]) (assuming this profile has a total stress factor of 15)

The sample cells are cycled according to the accelerated degradation profiles and are regularly tested for their remaining capacity. While specific end of life (EOL) conditions vary by application and battery chemistry, EOL is often defined to be the threshold when remaining capacity falls below 80% of initial capacity. This is a convention based on lead-acid battery degradation mechanisms that make them less reliable in backup power applications after passing this threshold. The example testing data is included in Table 1 with the number of cells cycled under each stress factor and the resulting mean time to EOL in years. The standard deviation of time to EOL is also included to capture the statistical uncertainty of cell degradation rates.

Table 1. Example Accelerated Life Testing Data*

Group #	Number of Cells	Total Stress Factor	Mean Time to EOL Conditions (Y)	Standard Deviation (Y)
1	50	7.5	2.071	0.035
2	30	10	1.594	0.03
3	30	12.5	1.268	0.027
4	40	15	1.054	0.025
5	30	17.5	0.87	0.023
6	30	20	0.796	0.022

*These data have been contrived for our example and did not come from testing

Once these data have been collected, the predicted life of each group can be calculated by multiplying the mean time to EOL by the total stress factor. Assuming the cell degradation rates follow a standard normal distribution, the mean time to EOL minus two standard deviations will yield the projected minimum life (lower bound) of a cell with 97.2% confidence. In other words, under these testing conditions a cell has a 97.2% probability of lasting longer than this lower bound. The results calculating the projected life and lower bound are listed in Table 2.

Table 2. Example Projected Life and Lower Bound by Group

Group #	Number of Cells	Mean Projected Life (Y)	97.2% Confidence Lower Bound (Y)
1	50	15.5325	15.0075
2	30	15.94	15.34
3	30	15.85	15.175
4	40	15.81	15.06
5	30	15.225	14.42
6	30	15.92	15.04

A weighted average is then taken to calculate the final projected life and lower bound. The results of life verification testing are listed in Table 3 and shown in Figure 9.

Table 3. Example Results from Life Verification Testing

Weighted Average Projected Life (Y)	Weighted Average 97.2% Confidence for Lower Bound (Y)
15.70	15.01

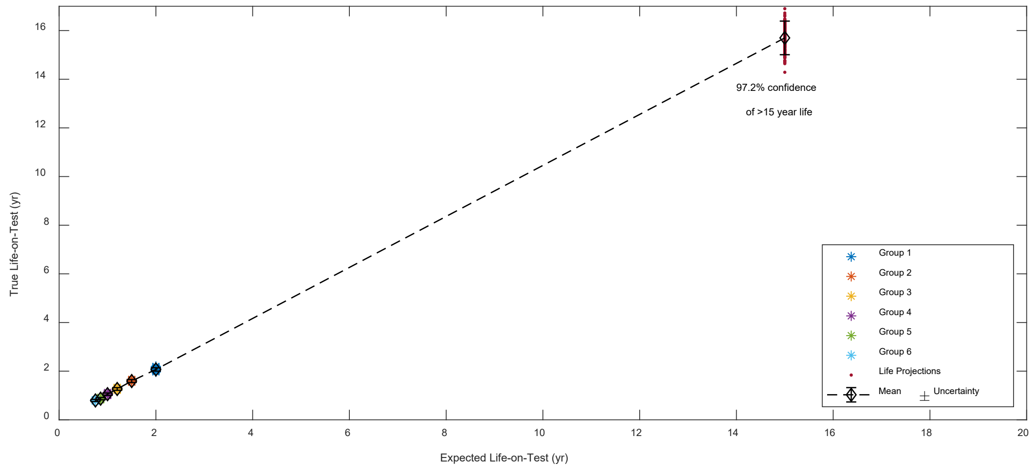

Figure 9. Example Data and Results from Life Verification Testing

This simple example yields 97.2% confidence that the cells coming off the production line will last at least 15 years in the anticipated application. However, there are many assumptions that lead to this conclusion and, therefore many opportunities for misinterpretation. If given a battery module with 100 cells in series, on average two or three cells in this module will reach EOL conditions before 15 years, which means the whole module could fail early. As another example, cells used in a cold environment to avoid high temperature acceleration of degradation may not exhibit the expected results because the accelerated degradation testing program is based on an imperfect degradation model that may not account for mechanisms that occur at low temperature. Misinterpretation of results can be avoided through understanding the process of life testing. Some common misconceptions about battery life are:

- Misconception: a battery will last its design life no matter how it is used
    - Truth: Battery life is highly dependent on both its use and its environment
- Misconception: life cycle test results on a specification sheet are specific to the cell make and model they are listed for.
    - Truth: As life cycle testing takes years, newly developed batteries will often reuse old data until data on the new batteries is available. Also, manufacturing processes and material qualities can change leading to better or worse performance than expected.
- Misconception: life cycle testing reflects a cycle test over the whole life of a battery
    - Truth: Accelerated degradation testing is used to extrapolate battery life beyond the duration of any actual testing. This extrapolation is done based on models for how a battery degrades and may not reflect all degradation mechanisms.
- Misconception: In a large battery system, if all cells are used equally and in the same environment, then all cells will degrade at roughly the same rate.
    - Truth: A statistical average, or even a confidence interval, is not a guarantee. Even with a 97.2% confidence interval, a system with 1000 cells will likely have 28 reach EOL conditions before their expected life. As battery strings are limited by their

lowest capacity cell, normal uneven aging can lead to operational problems. Battery management systems and maintenance technicians try and identify these outlier cells so that they can be replaced or removed before any issues arise.

2.2. Integrated System Testing

2.2.1. Apparatus and Materials

The materials needed to perform tests on an integrated ESS are an electrical connection to the electric power system (EPS), metering to collect accurate data, and a control system to implement user commands. Additionally, many services require access to specific information such as wholesale energy price. This access is critical to performing application specific tests in those areas. Testing performed on installed and operational ESSs should have data collection systems as specified in *Electrical Energy Storage Data Submission Guidelines* [7].

2.2.2. Reference Performance Testing Methods

As discussed for battery cells, an RPT is a combination of commissioning test and periodic test performed to get a snapshot of an ESS's performance, independent of application. An RPT can be used to get a baseline performance at BOL prior to duty-cycle testing. RPTs are also used periodically to estimate state-of-health as performance degrades over time and use. RPTs include tests for stored energy, efficiency, response-rate, and stand by energy loss rate. Other RPTs can be added as needed according to the technology and the application. The procedures for many RPTs are recorded explicitly in "Protocol for Uniformly Measuring and Expressing the Performance of Energy Storage Systems" [8] and *Energy Storage Integration Council (ESIC) 2016 Energy Storage Test Manual* [9]. Two RPT procedures are discussed in this section – the stored energy test and the energy storage pulsed power characterization (ESPPC) test.

2.2.2.1. Stored Energy Test Routine

The stored energy test is a system level corollary to the capacity test described in Section 2.1.2.1. The goal of the stored energy test is to calculate how much energy *can* be supplied discharging, how much energy *must* be supplied recharging, and how efficient this cycle is. The test procedure applied to the DUT is as follows:

Specify charge power P_{cha} and discharge power P_{dis}

Preconditioning (only performed before testing starts):

1. The DUT is charged to its upper SOC limit by charging at P_{cha} in accordance with the system manufacturer's specifications and operating instructions.
2. The system is left at rest in an active state for 5 minutes.

Test:

1. The DUT is discharged at P_{dis} to its lower SOC limit in accordance with the ESS manufacturer's specifications and operating instructions.
2. The system is left at rest in an active state for 5 minutes.

3. The DUT is charged to its upper SOC limit by charging at P_{cha} in accordance with the system manufacturer's specifications and operating instructions.
4. The system is left at rest in an active state for 5 minutes.

The test is then repeated several times to verify consistency and then repeated at several specified P_{cha} and P_{dis} settings. Just as with battery cells, many ESSs have reduced available discharge power at low SOC and reduced available charge power at high SOC. When the DUT encounters these limits, it should continue at reduced power until the SOC limit is reached. The total energy supplied on discharge during step 1 of the test is recorded as the discharge energy capacity. The total energy absorbed on charge during step 3 of the test is recorded as the charge energy capacity. The ratio of discharge energy capacity to charge energy capacity is the round-trip energy efficiency expressed as a percentage.

2.2.2.2. Energy Storage Pulsed Power Testing

The energy storage pulsed power characterization (ESPPC) test is a system-level corollary to the HPPC test described in Section 2.1.2.2. The goal of ESPPC testing is to define the bounds of the region shown in Figure 10**.**. A secondary goal of this procedure is to determine the AC/DC energy conversion efficiency of battery-based ESSs at different power levels and over the operational range of SOC. At high SOC, charge power is often limited, just as at low SOC, discharge power is often limited, and it is critical to know these operational limits.

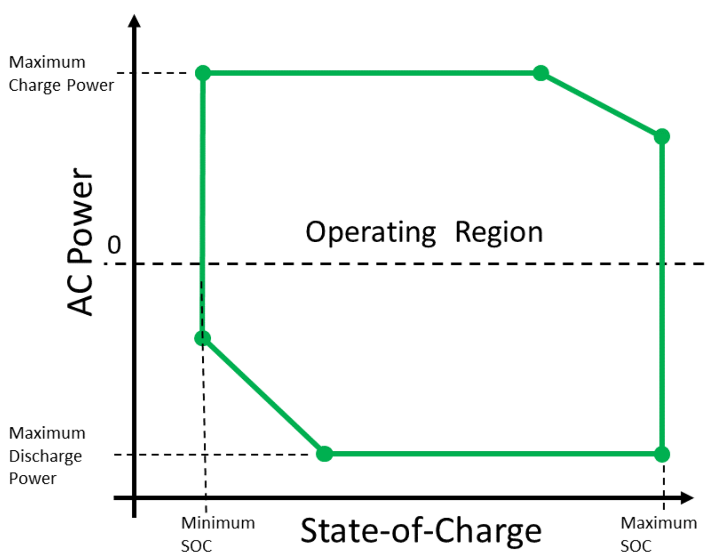

Figure 10. Feasible Power vs. SOC Region. The values p_{max} and p_{min} are the maximum charge power and maximum discharge power respectively. The values ς_{max} and ς_{min} are the maximum and minimum operational SOC respectively (in real systems, limits may be non-linear)

To calculate the bounds of this region, pulse testing is performed at a distribution of SOCs, as reported by the DUT. The profile starts with a capacity test to clear any memory effects from previous operation. Ten percent of the DUT's capacity is then discharged, followed by a one-hour rest. The DUT is then commanded to 60 second pulses/rest stages (as shown in Figure 12) where 100% discharge is the rated discharge power and 100% charge is the rated charge power. A single charge pulse and discharge pulse can be used for systems that do not have AC/DC inverter systems. The power supplied during the discharge pulse and absorbed during the charge pulse are recorded. The process repeats, by removing 10% SOC, followed by rest, followed by pulse testing (as shown in Figure 11), until the end-of-discharge conditions are met. At the end of the test the DUT is restored to full charge.

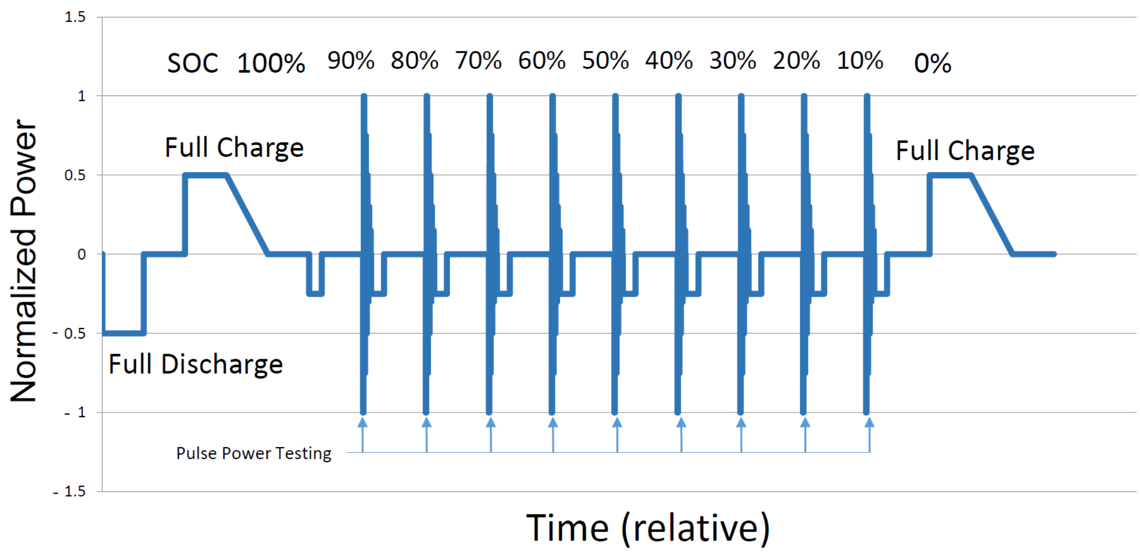

Figure 11. ESPPC testing profile

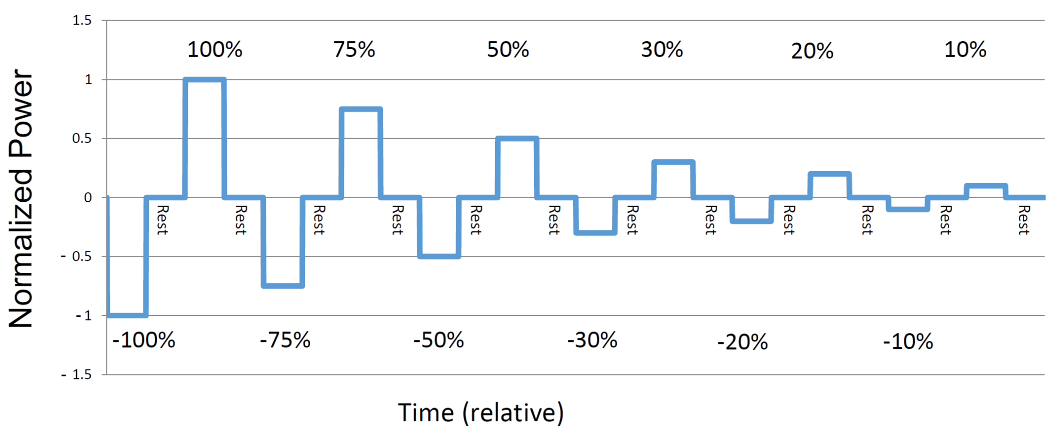

Figure 12. ESPPC Power Pulse Testing (negative corresponds to discharging)

The results are then plotted on the SOC/AC Power axes in Figure 9. Each maximum discharge at a specified SOC is a single point on the lower half of the plot. Similarly, each maximum charge at a specified SOC is a single point on the upper half of the plot. Together, the points outline the operational range of the DUT. In battery systems, the AC/DC conversion efficiency of the inverter can also be measured at each power level and each SOC. On discharge, the AC power divided by the DC power is the conversion efficiency. On charge, the DC power divided by the AC power is the conversion efficiency.

2.2.3. Duty Cycle Testing Methods

The goal of duty cycle testing is to calculate the ability of a particular device to perform in a given application [8]. The performance of a device in duty cycle testing is expressed using a common set of measurement requirements and test procedures. The resulting performance metrics enable a potential end-user to do an apples-to-apples comparison among different devices for a given application. There are no pass/fail criteria for any of the duty cycle tests because each end-user is likely to have different criteria by which they would select a particular device.

This section first discusses how to adjust a generic duty cycle to a specific DUT. It then covers four example duty cycles intended to give an understanding of their structure and how they can be applied. Similar duty cycles are available for a range of energy storage applications and services [8, 9].

2.2.3.1. Duty Cycle Test Scaling

Testing on integrated systems is often specified in normalized quantities and must be scaled to the specific DUT. In testing for use in electric vehicles, this is accomplished by establishing performance targets and scaling the tests to the minimum number of cells that would meet the targets [3]. In ESSs this process is not as simple, as the application space is diverse and there is no minimum performance metric. Instead the tests are scaled by energy or power depending on the limiting factor for a given test. The illustrations in Figure 13 (a) and (b) show how different tests are scaled differently on the same device.

A duty cycle is first scaled according to the rated power of the DUT. If the DUT has surplus or unused energy under this power scaled duty cycle, as shown in Figure 13 (a), then the test can be applied to the DUT successfully. This may be the case for a power intensive service; however, this may not be the case for an energy intensive service. If the DUT has insufficient rated energy under the power scaled duty cycle, then the duty cycle is scaled according to its rated energy. In this case, the DUT will have surplus or unused power capacity as shown in Figure 13 (b). Note that scaling a duty cycle to the rated energy of a DUT involves either trial-and-error testing or simulated testing using a model of the DUT, or some of both.

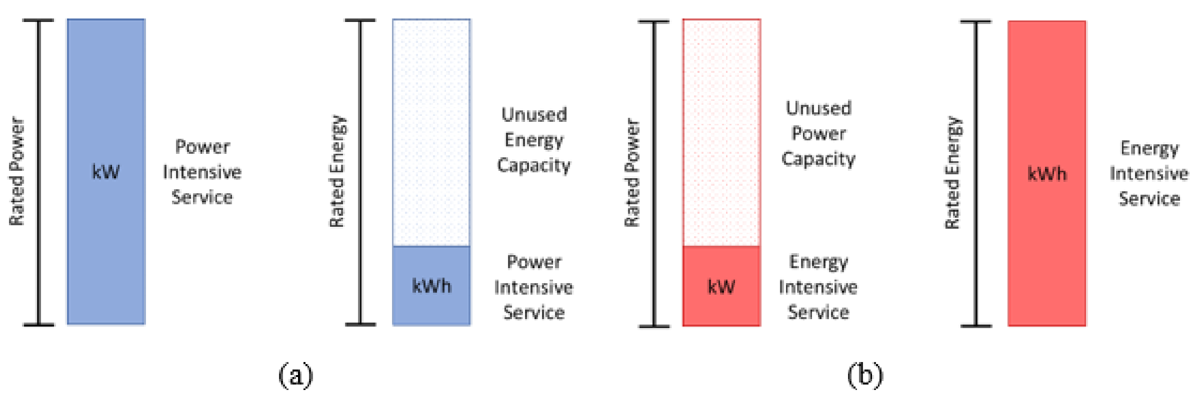

Figure 13. Test scaling based on service type and ESS ratings, (a) for a power intensive service, and (b) for an energy intensive service

For example, if a device has ratings of 1 MW/1 MWh, testing its ability to perform frequency regulation (a power intensive service) may only make use of 250 kWh, or one quarter of its energy capacity. Similarly, when testing its performance for peak shaving services, the duty-cycle is scaled to make sure the DUT has enough energy to discharge for 2, 4, or 6 hours. A device is well-suited for a specific service if there is very little unused power or energy capacity. However, some technologies, such as batteries, have nonlinear degradation stress factors that can make full utilization of energy capacity non-viable for long-term operation. For these devices, the scaling should account for these stress factors, constrained by a minimum expected operational life provided by the manufacturer.

2.2.3.2. Peak Shaving

Peak shaving refers to the capability of lowering peak demand on an EPS by charging an ESS when power demands are low so that its energy is available to discharge when power demands are high. The peak shaving duty cycles, illustrated in Figure 14, use charge and discharge time windows instead of normalized power levels or discharge rates [8]. In applying the duty cycles, the discharge power and SOC range for each duty cycle are selected so that the power remains constant throughout the required discharge period (6 , 4 , and 2 hours for duty cycles A, B, and C, respectively). The charging time window is 12 hours (10:00 pm to 10:00 am in simulated or real time) for each duty cycle, within which the DUT can follow any charging procedure.

The full duty cycle test consists of bringing the ESS to the desired initial SOC before applying duty cycle A, then B, then C. Following the application of duty cycle C, the ESS is brought back to its initial SOC. The peak shaving duty cycle performance is then calculated as the constant power supplied over each discharge window respectively. The round-trip duty cycle efficiency should also be reported as described in the *Protocol for Uniformly Measuring and Expressing the Performance of Energy Storage Systems*.

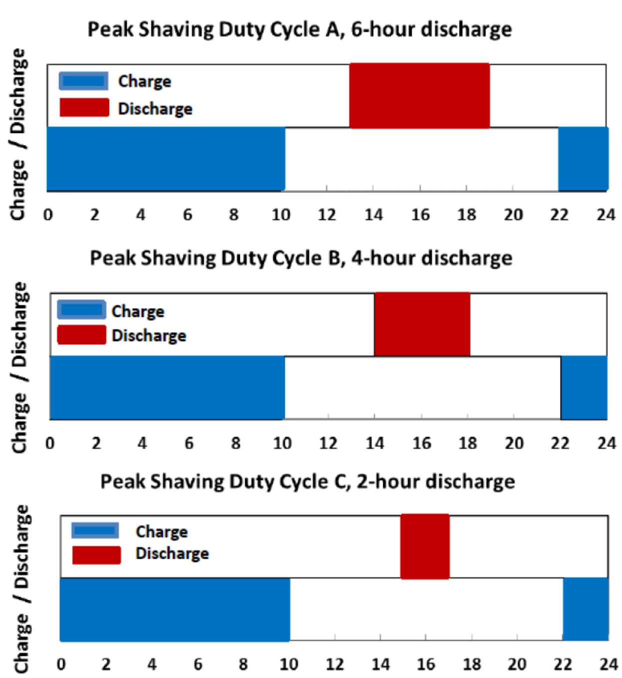

Figure 14. Peak Shaving Duty Cycles, X-axis is Time in hours [8]

2.2.3.3. Frequency Regulation

Frequency regulation refers to the capability to maintain system frequency (e.g., 60 Hz) in response to changes caused by imbalance between generation and load on an EPS [10]. In practice, the local balancing authority supplies a signal to follow according to its calculated imbalance or area control error (ACE). The goal of a frequency regulation duty-cycle test is to understand a DUT's efficiency and performance when providing frequency regulation to an EPS.

Development of a standard duty cycle started with a detailed analysis of publicly available PJM[1] frequency regulation control signal data from April 1, 2011 to March 31, 2012 [11]. This analysis determined that the dispatch signal's standard deviation could be used as a metric for aggressiveness or rigor of operation, and that 99.9% of the power in the daily signal is represented in periods shorter than 2 hours. This enabled the development of a 24-hour duty cycle from a series of 2-hour representative ''average'' and representative ''aggressive'' signals. The representative average signal was selected by finding the most energy neutral 2 hours within the three days determined to have standard deviations closest to the yearly average signals' standard deviation. The representative aggressive control signal was selected by finding the most energy neutral 2 hours within the day with the highest standard deviation. Both signals were checked to ensure their properties adequately represented the days from which they were selected. The final duty cycle was composed of representative control signals performed in the following order:

1. Preconditioning
2. Three average signals
3. One aggressive signal
4. Three average signals
5. One aggressive signal
6. Four average signals
7. Post-conditioning

The derivation of this duty cycle from PJM frequency regulation data was to enable laboratory testing analogous to real-world use and it accomplished that goal. However, the performance specifications for frequency regulation differ by region and will likely change over time. Indeed, PJM has since changed their ACE algorithm so that modern frequency regulation signals have much higher standard deviations then even the most strenuous days in 2012.

The full duty cycle image shown in Figure 15 is color coded by average signal (red) and aggressive signal (green). The required pre-/post-conditioning steps that ensure the DUT starts and ends the test at a designated SOC are not shown. An example of this procedure would be to fully charge the system (100% SOC), then discharge the system to prepare for the duty cycle (e.g. estimated 50% SOC), and then after the duty-cycle is completed, fully recharge the system (100% SOC). These additional steps enable the calculation of round-trip duty cycle efficiency as described in the *Protocol for Uniformly Measuring and Expressing the Performance of Energy Storage Systems*.

[1] PJM coordinates and directs the operation of a regional transmission grid for 13 states and the District of Columbia, administers a wholesale electricity market, and plans regional transmission expansion improvements to maintain grid reliability and relieve congestion.

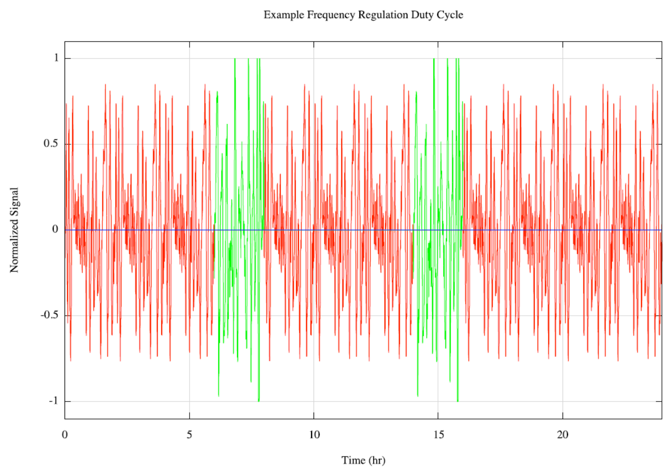

Figure 15. Frequency-Regulation Duty Cycle [10, 11]

2.2.3.4. PV Smoothing

PV smoothing is the use of an ESS to mitigate rapid fluctuations in variable PV power output. The purpose of PV smoothing is to mitigate frequency variation and stability issues that can arise at both distribution and transmission levels in high penetration PV scenarios and to help meet ramp-rate requirements [12, 13]. In distribution systems, PV smoothing is implemented to mitigate voltage flicker and voltage excursions outside safe limits. At the transmission level, PV variability can require additional operating reserve to be set aside and increase traditional generation cycling. The method by which the ESS can provide smoothing of PV output power is to absorb or supply power at appropriate times as determined by a control system, resulting in a less variable composite power on the EPS.

PV power output and battery power output (both expressed in kW) from the Public Service Company of New Mexico (PNM) Prosperity Energy Storage Project was used for construction of the PV smoothing duty cycle [14].

The duty cycle is constructed by capturing one-hour “slices” of PV generation from different days and splicing these slices together into a composite signal of 10 hours. Most of these slices represent moderate to very high levels of PV variability. Thus, the composite signal will lead to an aggressive tracking signal. This splicing process also enables different times of the day and times of the year to be captured by a single duty cycle. The tracking signal is then computed by subtracting the 30-minute moving average of the composite “day” from the composite signal itself. The duty cycle is obtained by normalizing the tracking signal to the rated power of the DUT. The full “day” signal as well as each hour of that day must be sufficiently close to net energy neutral. Some of the relevant metrics for PV smoothing include:

- Reference signal tracking accuracy
- SOC extrusion
- Duty cycle round-trip efficiency

The duty cycle is illustrated in Figure 16. A plot showing the smoothed PV power after application of a 30-minute rolling average is shown with the raw PV power in Figure 17.

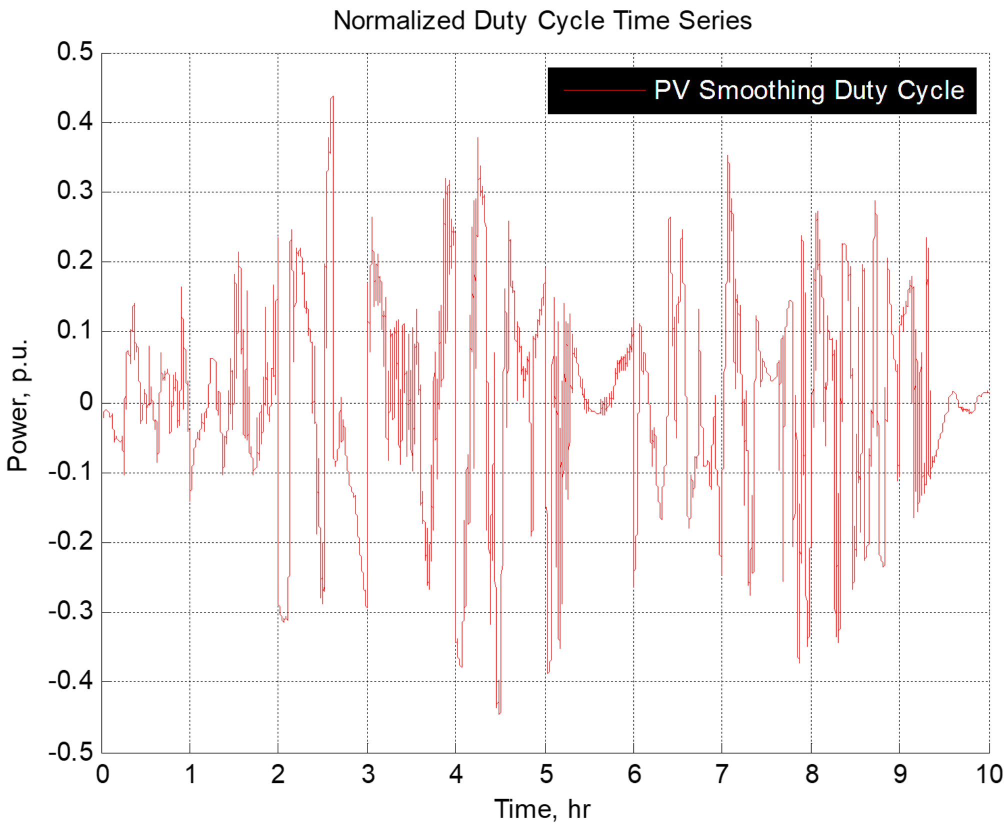

Figure 16. Duty Cycle – PV Smoothing [12, 13]

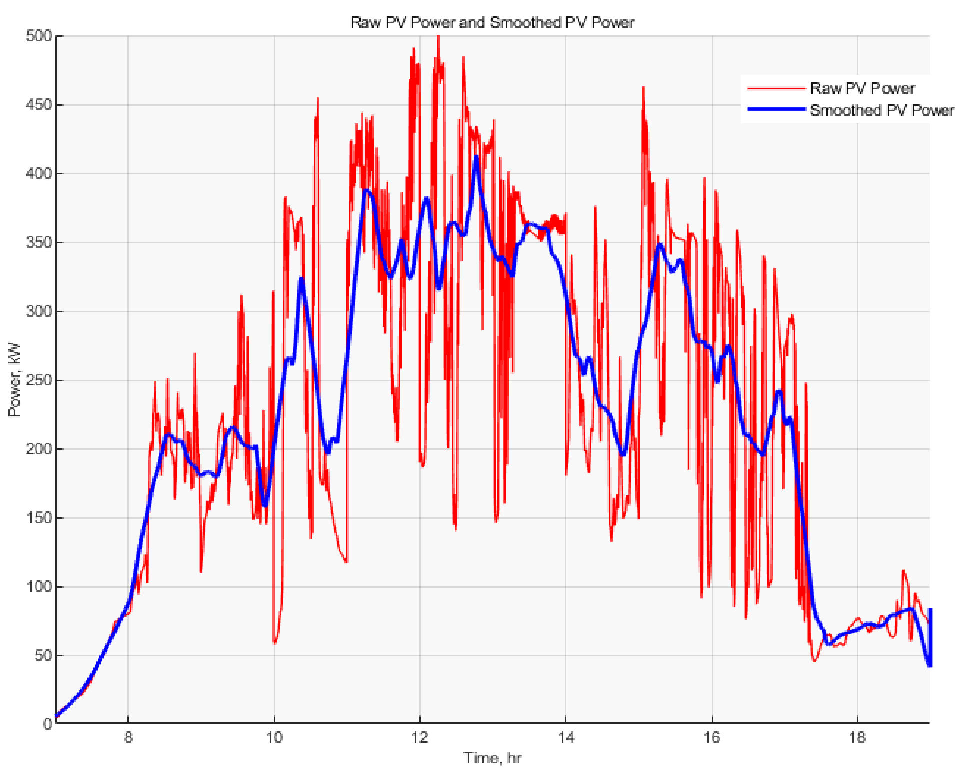

Figure 17. Resultant smoothed PV power after application of energy storage [12, 13]

2.2.3.5. Wind Firming

Wind firming is the use of an ESS to provide energy to supplement wind power generation so that their combination (ESS output + wind power generation) produces steady power output over a desired timeframe. Firming provides a means of overcoming wind-based generation intermittency, thus increasing the amount of renewable energy that can be provided to the grid [15].

This control strategy has the ESS inject energy (discharge) when wind power generation falls below a certain threshold. Likewise, the ESS absorbs energy (charge) when wind power generation exceeds either the same or a different threshold. Generally, smoothing attempts to limit ramp rates over short time intervals (e.g., one minute) while firming is more concerned with creating predictable supply periods for longer time intervals (e.g., one hour). Therefore, smoothing is considered to be a *power* intensive application whereas firming is considered to be an *energy* intensive application.

Wind power data supplied by Southern California Edison was used to help construct the duty cycle. Two 24-hour duty cycles were developed for wind firming because they each represent important yet distinct methods of controlling a battery for the wind firming application that would be difficult to capture in one duty cycle. The stairstep control method is used to develop the more aggressive duty cycle (higher ramp rates), and the moving average control method is used to developed the nominal (less aggressive) duty cycle. Both control methods start by calculating the smoothed power signal from the raw wind power data. The difference between the smooth power signal and the raw power signal then becomes the duty cycle's reference power for the DUT. The duty cycles are illustrated in Figure 18. Figure 19 and Figure 20 show the firmed wind power that results from applying the duty cycles in Figure 18 on the same axes as the raw wind power to illustrate the effects of using the nominal and aggressive duty cycles, respectively. Some of the relevant metrics for wind firming include:

- Reference signal tracking accuracy
- SOC extrusion
- Duty cycle round-trip efficiency

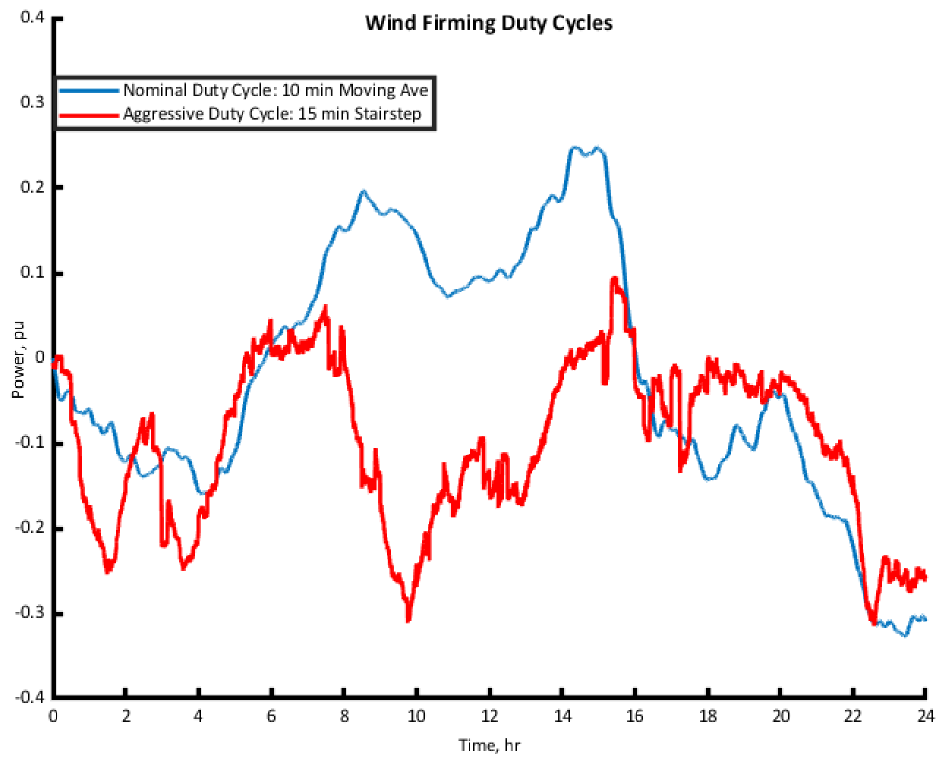

Figure 18. Normalized time series plots of wind firming duty cycles [15]

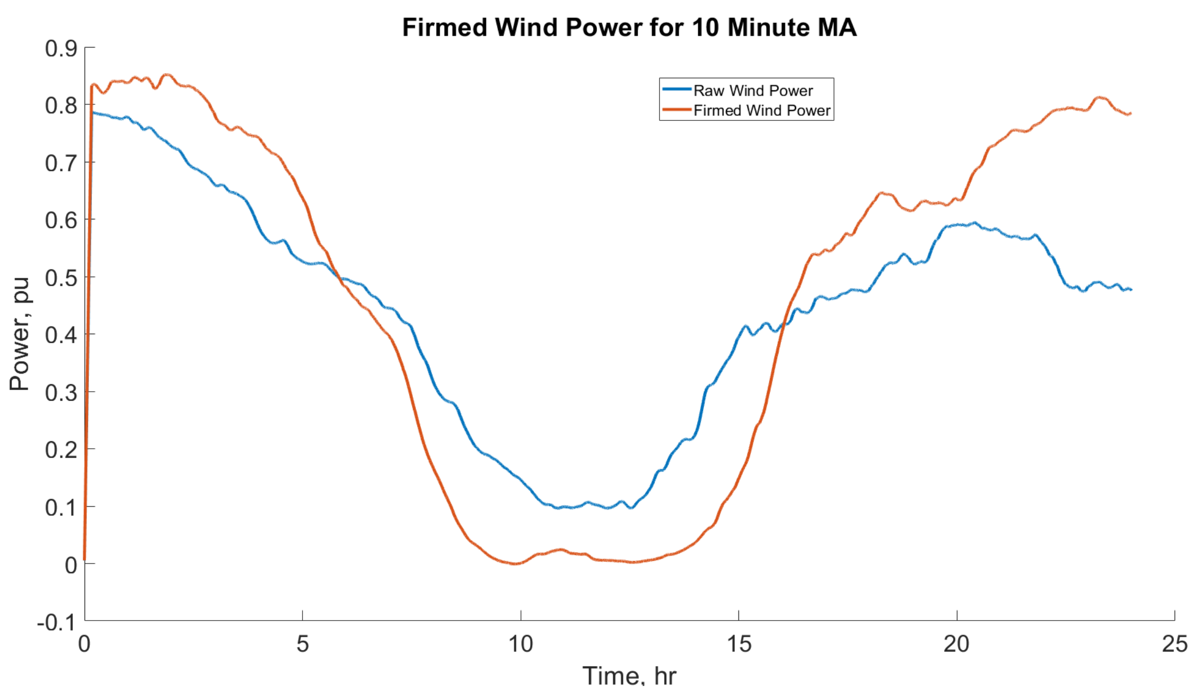

Figure 19. Firmed wind power (red curve) after application of nominal ES duty cycle to raw wind power (blue curve) [15]

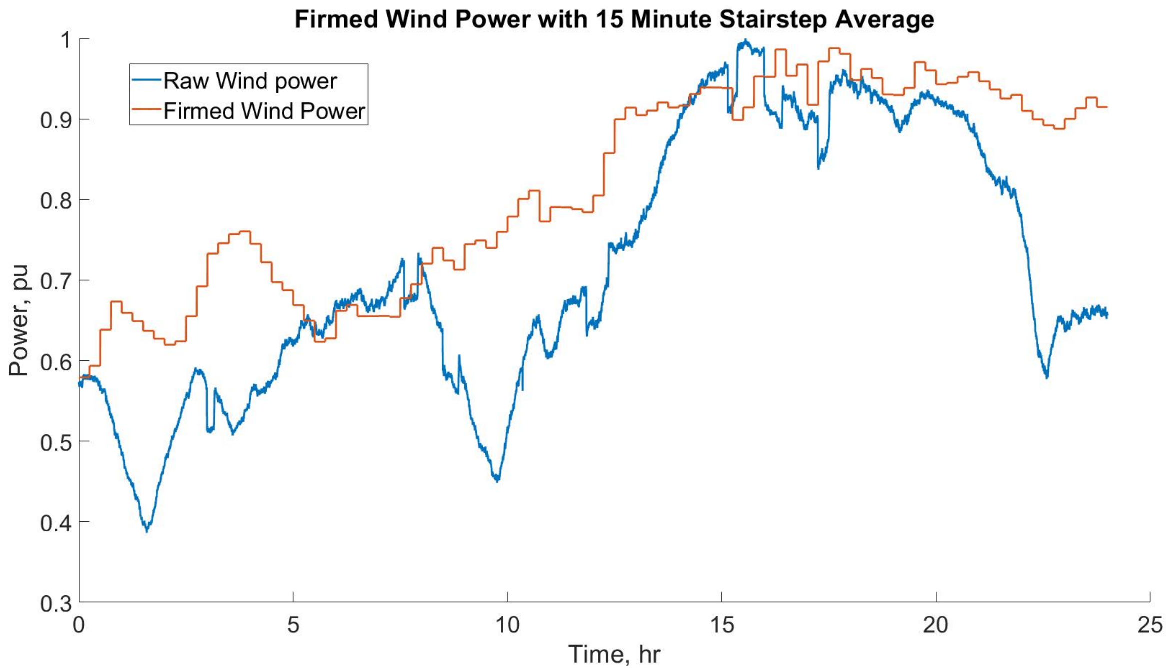

Figure 20. Firmed wind power (red curve) after application of aggressive ES duty cycle to raw wind power (blue curve) [15]

3. Challenges in Energy Storage Performance Testing

Battery cell performance testing is well developed for use in personal devices, automotive applications, and even backup power supply applications; however, it is not as developed for grid supportive applications. The reason for this deficit is that the connections between grid service requirements and cell performance requirements are complex, uncertain, and in constant flux. The broad variety of energy storage grid services means that cells have many sets of performance targets, rather than just one set. Each set of performance requirements for a specific grid service only has a loose correlation with the actual requirements of a real application environment. This loose correlation means that if a cell performs well in one application environment (e.g. smoothing wind power in Texas) then it is still not certain to perform well in another application environment (e.g. smoothing wind power in Hawaii). While some limited testing has been done applying duty cycles at the cell level [16], correlating this result to system-level performance has high uncertainty due to the many assumptions required in the calculation. Lastly, the grid service requirements themselves are changing over time as power system operators adapt their practices to incorporate energy storage. Different approaches to cell performance testing for grid services are needed to reduce the complexity and uncertainty of mapping cell to grid service performance requirements.

Integrated system performance—how well an ESS provides a certain service—has proven difficult to measure because of the wide variety of applications and the ways in which battery usage within those applications varies substantially between power systems. The state-of-the-art approach to measure energy storage performance, discussed throughout this chapter, is a rough allegory to the solution for miles per gallon (mpg) in cars – a consensus is reached through stakeholder meetings to decide on a duty cycle testing protocol that uniformly measures the ESS performance. This duty-cycle approach is motivated by the balance of ease (cost of testing to manufacturers), and salience (usefulness of the derived metrics to decision makers).

To point out the deficiencies of the duty-cycle approach we will return to the automotive analogy. For cars, the applications are few—city and highway—and static in that the roads have not changed appreciably in the 70 years. Therefore, an easily applied, representative duty-cycle for city driving can reasonably expect to be salient to how consumers will use it across the United States. In contrast, the applications for ESSs are many and changing. Changes to state and federal policies, and local energy resource mixes cause usage profiles in some applications to change over time. Another complication is the relationship between battery operation and life. Cars, as they are mechanical systems, have a close to linear relationship between use and life; whereas batteries can have very complex relationships between use and life [17]. This provides an opportunity for manufacturers to exploit the duty-cycle approach by maximizing test performance at the expense of battery life.

4. Opportunities for Advanced Energy Storage Performance Testing

There are several directions to develop the state-of-the-art in performance testing that can be categorized into the following groups:

- Improved duty-cycles and testing procedures
- Model-based testing
- Using operational data to calculate standardized performance

By improving the duty cycles, test engineers can be more confident about installed performance. The frequency regulation duty cycle in Section 2.2.3.3, for example, was developed before PJM changed how it calculated its dynamic frequency regulation signal. While a system tested using the current duty-cycle would have more performance certainty then an untested system, the test would provide little confidence that a system could perform in that application. A few stacked duty-cycles currently exist but more combinations would be helpful in specific situations. Research is needed to both update existing duty cycles and develop more duty cycles for applications that have not yet been performed.

Moving past duty-cycle testing, there is an alternative approach that leverages the low cost of computation and advances in machine learning to improve both the ease and salience of testing. Rather than testing ESSs with duty cycles relevant to different applications, it is possible to test an ESS to develop a highly accurate computer model and then expose the model to a wide range of real world use relevant to different applications. This “model-based testing” approach solves many of the challenges with the duty-cycle approach. First, it limits the number of tests that must be performed. Rather than needing a test for both the PJM and California markets, the model could be applied to both at no marginal cost. Second, it allows for a more represented use profile than is possible in physical testing as years of simulated operation in each application can be accomplished, again at no marginal cost. Third, battery degradation models can be included in simulated operational decisions by constraining performance to achieve a desired operational life. These benefits come at the cost of model uncertainty. No model perfectly represents a physical system and this discrepancy generates errors between the model predicted performance and what the system is actually capable of. However, this uncertainty can be calculated and bounded as it can be a part of the model itself, rather than hidden within the derivation of the duty cycle. More research is needed to develop methods for model-based testing.

Because any testing regimen is expensive it would be greatly beneficial to calculate standard performance metrics using only operational data produced during the service of an application. This would be useful not only in verifying warrantied performance but also for calculating the degradation rate and updating the controller to optimize performance. Where performance metrics such as round-trip energy efficiency are normally calculated using a specific charge/discharge profile, they can be estimated after a grid service causes the ESS to discharge and then recharge. This kind of in situ performance testing can reduce cost and improve salience of the information produced.

5. Concluding Remarks

Performance testing is a critical component of safe and reliable deployment of energy storage systems on the electric power grid. Specific performance tests can be applied to individual battery cells or to integrated energy storage systems. Battery cells can be tested for both reference performance (e.g., capacity and efficiency) and for life-cycle performance (e.g., cycle-life for a specific intended use). This chapter has provided an overview of the various types of methods and procedures that are used in battery cell testing. Integrated energy storage systems can include batteries, or non-battery technologies such as flywheels, capacitors, or compressed air. Integrated system tests are applied uniformly across energy storage technologies to yield performance data. Duty-cycle testing can produce data on application-specific performance of energy storage systems. This chapter reviewed a range of duty-cycle tests intended to measure performance of energy storage supplying grid services. Understanding the motivation behind testing, and how each test is constructed and applied, avoids many common misconceptions associated with interpreting test results. The shortcomings of the conventional approach to performance testing also become apparent. Performance metrics such as capacity or efficiency are built on layers of assumptions and caveats. Testing programs can be prohibitively expensive and often only produce information relevant to one or two applications. To improve the ease and salience of performance test results future research should focus on improved duty-cycles and testing procedures, developing methods for model-based testing, and the use of operational data to calculate standardized performance metrics.

David Rosewater is a Senior Member of the Technical Staff at the Sandia National Laboratories. He received a BS and an MS in electrical engineering from Montana Tech of the University of Montana, as well as a Ph.D. in electrical and computer engineering from the University of Texas at Austin. His research interests include modeling and simulation, performance testing, safety, and standardization of battery energy storage systems. From 2009 to 2011, he worked at the Idaho National Laboratory developing advanced spectral impedance measurement techniques for hybrid vehicle battery cells. He is a Senior Member in the IEEE and currently chairs the IEEE P2686 working group developing a recommended practice for design and configuration of battery management systems in energy storage applications. David holds a professional engineering license in the state of New Mexico.

David Schoenwald is a Principal Member of the Technical Staff in the Electric Power Systems Research Department at Sandia National Laboratories. In his current work, Dr. Schoenwald focuses on control system design for damping inter-area power system oscillations for which he received the 2017 R&D 100 Award. In addition, his interests include mitigation of network-induced issues in feedback control systems using wide-area measurements, and development of performance standards for grid-scale energy storage. Before joining Sandia, he was with Oak Ridge National Laboratory where he developed models and controls for manufacturing applications. He was also an adjunct assistant professor in the Electrical Engineering Department, University of Tennessee, Knoxville, where he taught a graduate level course in nonlinear control systems. He is a Senior Member of the IEEE. Dr. Schoenwald received his BS, MS, and PhD degrees in electrical engineering, respectively, from the University of Iowa, the University of Illinois, Urbana-Champaign, and The Ohio State University.

References

[1] D. M. Rosewater, D. A. Copp, T. A. Nguyen, R. H. Byrne and S. Santoso, "Battery Energy Storage Models for Optimal Control," in IEEE Access, vol. 7, pp. 178357-178391, 2019.doi: 10.1109/ACCESS.2019.2957698 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8922750&isnumber=8600701

[2] IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces, IEEE Std 1547-2018 (Revision of IEEE Std 1547-2003) , vol., no., pp.1-138, 6 April 2018 doi: 10.1109/IEEESTD.2018.8332112 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8332112&isnumber=8332111

[3] United States Advanced Battery Consortium Battery Test Manual for Electric Vehicles Rev. 3, June 2015 https://www.uscar.org/commands/files_download.php?files_id=405

[4] 1679-2010 - IEEE Recommended Practice for the Characterization and Evaluation of Emerging Energy Storage Technologies in Stationary Applications

[5] J. Christopherson "Battery Technology Life Verification Test Manual Revision 1" Idaho National Laboratory, December 2012, https://inldigitallibrary.inl.gov/sites/sti/sti/5605669.pdf

[6] H. Barkholtz et al. "A Database for Comparative Electrochemical Performance of Commercial 18650-Format Lithium-Ion Cells," in J. Electrochem. Soc. 2017, 164, A2697-A2706

[7] B. Schenkman, D. Rosewater, S. Willard "Electrical Energy Storage Data Submission Guidelines" Sandia National Laboratories and the Electric Power Research Institute, technical update, February 2020, SAND2020-2569 R, EPRI pt. 3002016277

[8] D. Conover, A. Crawford, J. Fuller, S. Gourisetti, V. Viswanathan, S. Ferreira, D. Schoenwald, D. Rosewater, "Protocol for Uniformly Measuring and Expressing the Performance of Energy Storage Systems," Pacific Northwest National Laboratory and Sandia National Laboratories Technical Report, PNNL-22010 Rev 2, SAND2016-3078 R, Richland WA and Albuquerque NM, April 2016. https://energystorage.pnnl.gov/pdf/PNNL-22010Rev2.pdf

[9] Energy Storage Integration Council (ESIC) Energy Storage Test Manual 2016 https://www.epri.com/#/pages/product/3002009313/?lang=en-US

[10] S. Ferreira and D. Schoenwald, "Duty-Cycle Signal for Frequency Regulation Applications of ESSs," Sandia National Laboratories Technical Report, SAND2013-7315P, Albuquerque, NM, August 2013. http://www.sandia.gov/ess/publications/SAND2013-7315P.xlsx

[11] D. Rosewater and S. Ferreira, "Development of a frequency regulation duty-cycle for standardized energy storage performance testing," *Journal of Energy Storage*, Vol. 7, pp. 286-294, August 2016. http://dx.doi.org/10.1016/j.est.2016.04.004

[12] D. Schoenwald and J. Ellison, "Determination of Duty Cycle for Energy Storage Systems in a PV Smoothing Application," Sandia National Laboratories Technical Report, SAND2016-3474, Albuquerque, NM, April 2016. http://www.sandia.gov/ess/publications/SAND2016-3474.pdf

[13] D. Schoenwald, "PV Smoothing Duty Cycle," Sandia National Laboratories Technical Report, SAND2016-2543R, Albuquerque, NM, April 2016. http://www.sandia.gov/ess/publications/SAND2016-2543R.xlsx

[14] D. Roberson, J. Ellison, D. Bhatnagar, and D. Schoenwald, "Performance Assessment of the PNM Prosperity Electricity Storage Project," Sandia National Laboratories Technical Report, SAND2014-2883, Albuquerque, NM, May 2014. https://www.sandia.gov/ess/publications/SAND2014-2883.pdf

[15] D. Schoenwald, "Determination of Duty Cycles for Energy Storage Systems Providing Wind Firming and Combination of Frequency Regulation and Peak Shaving Applications," Sandia National Laboratories Technical Report, Albuquerque, NM, February 2020.

[16] S. Ferreira, "Multiple Uses in Storage: Results of Lithium-Ion Cells Tested under Stacked Cycling Profiles," 2015 EESAT Conference, Portland, OR, September 2015.

[17] B. Xu, A. Oudalov, A. Ulbig, G. Andersson and D. S. Kirschen, "Modeling of Lithium-Ion Battery Degradation for Cell Life Assessment," in IEEE Transactions on Smart Grid, vol. 9, no. 2, pp. 1131-1140, March 2018, doi: 10.1109/TSG.2016.2578950.