LESSON BOOK



Battery Energy Storage Systems

From one cell to what a telemetry record can and cannot determine

Nine lessons · 181 pages

Ricardo Marquez, Ph.D.

ricardo-marquez.com

Battery Energy Storage Systems

From one cell to what a telemetry record can and cannot determine

A nine-lesson tutorial, written up as a book

Ricardo Marquez

October 5, 2026 (one chapter per lesson, nine lessons)

How to read this book

This book assumes no prior knowledge of batteries. Each chapter expands one lesson of a tutorial that starts from a single lithium-ion cell and ends at the question an engineer must answer before fitting any model to field data: given the sensors installed, what can actually be known? The progression is

$$\text{the machine} \to \text{electrical quantities} \to \text{the cell} \to \text{the circuit model} \to \text{the hardware} \to \text{heat} \to \text{aging} \to \text{faults} \to \text{information}.$$

One system is used as the running example throughout. The *reference container* is a grid-scale lithium iron phosphate battery of 3,564 cells arranged as nine racks of 396 cells, storing about 3.42 MWh, and the *reference record* is its field telemetry: rack currents and voltages, a reported state of charge, 396 cell voltages and 216 temperatures per rack. Every number attributed to the reference container or its record is a property of that system and is used here to make the general ideas concrete; numbers introduced only to illustrate an equation are labeled illustrative.

Where the field uses a shorthand term, a *Jargon* box expands it once and the text then uses the expanded meaning. Every symbol is defined in the notation chapter that follows and is used with the same meaning in every chapter. Each chapter closes with a statement of what to retain, exercises, and a short section connecting the chapter to the TERRA framework, a graph-based way of modeling physical systems as conservation laws on nodes with flows on edges, in which the question of what a set of measurements can determine is asked before any parameter is estimated.

Contents

How to read this book **ii**

Notation and vocabulary **vi**

1 What a battery energy storage system is **1**
1.1 The machine and its purpose . . . 1
1.2 One cell and its energy . . . 3
1.3 Series strings and the rack . . . 6
1.4 Parallel racks and the container . . . 8
1.5 Current, charge and state of charge . . . 11
1.6 Supervision, heat and the grid interface . . . 13
1.7 What to retain . . . 17
1.8 Exercises . . . 17
1.9 Connection to the TERRA framework . . . 18

2 Electrical quantities **20**
2.1 Three quantities at two terminals . . . 20
2.2 Charge and energy . . . 22
2.3 Power and its relation to energy . . . 24
2.4 State of charge and coulomb counting . . . 26
2.5 C-rate, parallel racks and the five quantities . . . 29
2.6 Worked examples and the reference record . . . 31
2.7 What to retain . . . 34
2.8 Exercises . . . 34
2.9 Connection to the TERRA framework . . . 36

3 Inside a lithium-ion cell **38**
3.1 What moves inside a cell . . . 38
3.2 Discharge, charge and the origin of voltage . . . 40
3.3 Open-circuit voltage and the LFP plateau . . . 42
3.4 Terminal voltage and internal resistance . . . 45
3.5 Heat and the series string . . . 47
3.6 Aging, the weakest cell and the per-cell model . . . 50
3.7 What to retain . . . 52
3.8 Exercises . . . 53
3.9 Connection to the TERRA framework . . . 54

4 The equivalent-circuit model **55**
4.1 Three voltages . . . 55
4.2 Internal resistance as a lumped parameter . . . 57
4.3 The first-order equivalent circuit . . . 59

4.4 States, measurements and identifiability . . . 63
4.5 Flat curve; resistance and capacity . . . 66
4.6 The hierarchy of models and model choice . . . 69
4.7 What to retain . . . 71
4.8 Exercises . . . 71
4.9 Connection to the TERRA framework . . . 72

5 The hardware around the cells 74
5.1 From cell to grid asset . . . 74
5.2 Conductors, protection and measurement . . . 77
5.3 The string voltage identity . . . 80
5.4 The common bus and the missing sensor . . . 83
5.5 The battery management system . . . 85
5.6 Power conversion and control . . . 89
5.7 What to retain . . . 92
5.8 Exercises . . . 92
5.9 Connection to the TERRA framework . . . 93

6 Thermal behavior and cooling 95
6.1 Heat as a conserved quantity . . . 95
6.2 Storage, cooling and the heat balance . . . 98
6.3 Thermal dynamics and time scales . . . 100
6.4 Sensors and the rack-mean boundary . . . 104
6.5 The reference thermal equation . . . 105
6.6 Faults, coupling and the system so far . . . 108
6.7 What to retain . . . 112
6.8 Exercises . . . 113
6.9 Connection to the TERRA framework . . . 114

7 Degradation and state of health 115
7.1 Two kinds of degradation . . . 115
7.2 Health as a vector . . . 117
7.3 Aging mechanisms and stress factors . . . 119
7.4 Imbalance and the weakest cell . . . 121
7.5 Estimating health . . . 125
7.6 SOH, SOP, RUL and the time-varying model . . . 127
7.7 What to retain . . . 130
7.8 Exercises . . . 130
7.9 Connection to the TERRA framework . . . 131

8 Faults, detection and diagnosis 133
8.1 Five terms . . . 133
8.2 Residuals and thresholds . . . 136
8.3 Fault types and their signatures . . . 139
8.4 Diagnosis as competing hypotheses . . . 144
8.5 Faults the model can absorb . . . 147
8.6 Instrumentation and visibility . . . 149
8.7 What to retain . . . 152
8.8 Exercises . . . 152
8.9 Connection to the TERRA framework . . . 153

9 Observability and identifiability **155**
9.1 What can be known . . . 155
9.2 Excitation and redundancy . . . 158
9.3 Structural and practical identifiability . . . 161
9.4 What the sensors resolve . . . 163
9.5 Boundaries, degrees of freedom, trade-offs . . . 166
9.6 Matching the dataset to the question . . . 168
9.7 What to retain . . . 171
9.8 Exercises . . . 171
9.9 Connection to the TERRA framework . . . 173

Notation and vocabulary

Every symbol below is used with the same meaning in every chapter. Current is positive on discharge and negative on charge throughout.

Symbol	Unit	Meaning
V	V	Terminal voltage: the voltage measured across a cell, rack or bus while it operates.
$U(\mathrm{SOC})$	V	Open-circuit voltage: the voltage of a rested cell with no current, a function of state of charge.
I	A	Current; positive on discharge, negative on charge.
$P = V\,I$	W	Electrical power, the rate of energy transfer.
Q	Ah	Electric charge moved or stored; $Q = \int I \,\mathrm{d}t$.
C	Ah	Charge capacity of a cell or a series string.
E	Wh	Energy; $E = \int P \,\mathrm{d}t$, nominally $E \approx V\,C$.
C-rate	h^{-1}	Current divided by capacity, I/C.
SOC	–	State of charge: charge remaining divided by available capacity. An estimate, not a sensor reading.
C_{assumed}	Ah	The capacity a battery management system divides by when it computes state of charge.
$\mathrm{SOC}_{\mathrm{BMS}}$	–	The state of charge a battery management system reports.
I_{side}	A	Lumped unmeasured side current of a rack: balancing, self-discharge and controller consumption.
R	Ω	Lumped internal resistance of a cell.
R_0	Ω	Ohmic resistance: the instantaneous part of the voltage drop.
R_1, C_1	Ω, F	Resistance and capacitance of the polarization branch; time constant $\tau = R_1 C_1$.
V_1	V	Polarization voltage: the slow, history-dependent part of the voltage drop.
V_{cell}, V_i	V	Voltage of one cell; V_i is cell i of a string.
V_{rack}, V_r	V	Terminal voltage of a rack; V_r is rack r.
V_{bus}, I_{bus}	V, A	Voltage and total current of the common DC bus.
R_{series}	Ω	Lumped resistance of the connection hardware between the cell string and the rack terminals.
V_{off}	V	Constant offset between two voltage measurements of the same quantity.
R_{shunt}	Ω	Resistance of a current shunt.
P_{heat}	W	Heat generated by resistive loss, I^2R.
T, T_{env}, T_{ref}	K	Temperature of the battery, of its surroundings, and a reference temperature.
C_{th}	J/K	Thermal capacitance.

Symbol	Unit	Meaning
h	W/K	Effective cooling conductance.
k, q_0	W/A^2, W	Lumped heating coefficient and steady unaccounted heat term of the rack thermal balance.
SOH	–	State of health; here most often the capacity ratio C/C_{new}.
SOP	W	State of power: the power that can safely be delivered now.
RUL	time	Remaining useful life: a forecast of the time to an end-of-life criterion.
DoD	–	Depth of discharge: the state-of-charge swing of a cycle.
N_{EFC}	–	Number of equivalent full cycles: total charge discharged divided by capacity.
r	varies	Residual: measured value minus the value a model predicts.
H_0, H_i	–	The normal hypothesis and a fault hypothesis.
θ	varies	Vector of unknown model parameters.

Vocabulary

Cell, module, rack, container The physical hierarchy: a cell is one electrochemical unit; a module is a mechanical group of cells; a rack is a series string of modules; a container holds racks in parallel.

BESS Battery energy storage system: the battery, its management and protection hardware, its cooling and its power conversion system.

BMS Battery management system: the supervisory system that measures, estimates, protects and balances.

PCS Power conversion system: the inverter that connects the DC battery to the AC grid.

LFP Lithium iron phosphate, the cell chemistry of the reference container.

Coulomb counting Estimating state of charge by integrating measured current and dividing by an assumed capacity.

Reference container, reference record The running example of this book and its telemetry, described in “How to read this book”.

Chapter 1

What a battery energy storage system is: cells, series strings, parallel racks and the container

1.1 The machine and its purpose

This book assumes no background in batteries, and so it does not begin with a battery model. It begins with the machine: what is physically inside a grid-scale battery, where its energy goes, what its voltages and currents mean, and why thousands of small cells are arranged the way they are. The running example is the reference container, a grid-scale lithium-ion battery of 3,564 cells arranged in nine racks and storing about 3.42 MWh. Its telemetry, the reference record, contains rack currents and rack voltages, the voltage of every cell, temperatures, the outputs of a battery management system and the operation of a grid inverter. Each of those channels is a measurement taken somewhere in the machine described in this chapter, and none of them can be interpreted until the machine is understood. The question of the chapter is

what a battery energy storage system is physically, and how 3,564 *cells of about 3.2 volts each become one machine that exchanges megawatts with the grid.*

Jargon: BESS, PCS, BMS, cell, module, rack, container, series, parallel, SOC

A *battery energy storage system* (BESS) is a machine that takes electrical energy from the grid, stores it chemically in battery cells and returns it to the grid later; the term covers the cells, their management and protection hardware, their cooling and their power conversion system. The *power conversion system* (PCS), commonly called the *inverter*, converts between the direct current (DC) of the battery and the alternating current (AC) of the grid, and sets how much power flows in or out. The *battery management system* (BMS) is the computer and sensor network that measures the battery, estimates its state, enforces its limits and opens or closes its switches. A *cell* is the smallest electrochemical unit: a sealed device with two terminals and a voltage of a few volts. A *module* is a mechanical group of cells, 44 in the reference container. A *rack* is a string of modules connected end to end, here nine modules or 396 cells. A *container* is the enclosure that holds several racks, here nine, together with their cooling and protection. Components are in *series* when they are connected end to end, so that the same current passes through each of them; they are in *parallel* when they are connected across the same two conductors, so that they share the same voltage. The *state of charge* (SOC) is the fraction of the usable charge that remains, from 0 to 100 %.

Key idea

A battery energy storage system is a very large number of low-voltage cells, connected in series to reach a usable voltage and in parallel to reach a usable current and energy, together with a converter that exchanges that energy with the AC grid and a supervisory system that keeps every cell within its limits. Every channel of the reference record is a voltage, a current or a temperature somewhere in that arrangement, or a number computed from them.

This section describes the machine at its coarsest level: its three blocks and its two modes of operation.

1.1.1 Grid, power conversion system and battery

At the coarsest level a BESS is three blocks in a chain,

$$\text{grid} \overset{\text{AC}}{\longleftrightarrow} \text{PCS} \overset{\text{DC}}{\longleftrightarrow} \text{battery system},$$

and Figure 1.1 draws it. The three boxes are, from left to right, the AC grid, the power conversion system and the battery system. The short double-headed arrows between them are the two electrical connections, labeled with the kind of current each carries. The long blue arrow above the chain is the path of energy during charging and the long green arrow below it the path during discharging. The dashed box at the bottom lists the supporting systems, which act on the battery and on the converter but do not carry the main energy flow.

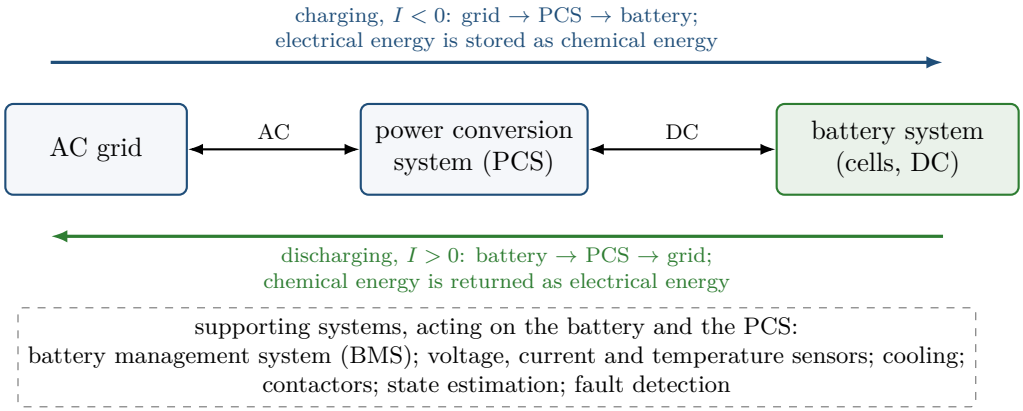


Figure 1.1: The three blocks of a battery energy storage system and the two directions of energy flow. The battery is a DC device and the grid an AC system; the power conversion system sits between them and carries energy in either direction. Current I is measured on the DC side and is positive on discharge.

The battery stores energy chemically and delivers it as *direct current*: current that flows in one direction at a time, between terminals whose polarity does not change. The grid is an *alternating current* system: its voltage and current reverse direction 50 or 60 times per second. The two cannot be connected directly. The power conversion system is the power-electronic device between them; it switches semiconductor devices thousands of times per second to synthesize the grid's alternating waveform from the battery's direct voltage, or to rectify the grid's waveform into a direct current that charges the battery. Because it controls that switching, the PCS also decides how much power flows and in which direction: the battery does not push energy into the grid on its own, the converter draws it. Section 1.6.4 returns to the PCS once the battery side of the chain has been built up.

1.1.2 Charging and discharging as energy conversion

A BESS has two modes of operation, and each is an energy conversion. In *charging*, energy flows from the grid through the PCS into the battery,

$$\text{grid} \rightarrow \text{PCS} \rightarrow \text{battery}: \qquad \text{electrical energy} \rightarrow \text{chemical energy},$$

and the chemical reactions inside every cell are driven in the direction that stores energy. In *discharging* the flow reverses,

$$\text{battery} \rightarrow \text{PCS} \rightarrow \text{grid}: \qquad \text{chemical energy} \rightarrow \text{electrical energy},$$

and the reactions run in the direction that releases the stored energy as a current through the external circuit. Neither conversion is lossless. Part of the energy in transit becomes heat in the cells, in the conductors and in the converter, so the energy returned to the grid in a full cycle is less than the energy taken from it. The losses are small enough that the round trip is worthwhile and large enough that they heat the battery, which Section 1.6.3 takes up.

The two modes are told apart by the sign of the current. Throughout this book the current I on the DC side is positive on discharge and negative on charge, so a rack current of $+100\,\text{A}$ means the rack is discharging and a rack current of $-100\,\text{A}$ means it is charging. The electrical power delivered by the battery, $P = VI$, follows the same sign: positive when the battery supplies the grid, negative when the grid supplies the battery.

Everything else in a BESS exists to make this conversion safe, controllable, measurable and reliable:

Battery management system measures the cells and the racks, estimates quantities that cannot be measured directly, and stops operation when a limit is reached (safe, measurable).

Sensors for voltage, current and temperature supply every number the BMS and the operator see (measurable).

Cooling removes the heat generated by the losses and keeps the cells inside their temperature range (safe, reliable).

Contactors are heavy-duty switches, operated by the BMS, that connect each rack to the rest of the system or isolate it (safe, controllable).

State estimation computes quantities such as the state of charge that no sensor reads (controllable, measurable).

Fault detection looks for cells, sensors or connections that depart from normal behavior before they cause damage (reliable).

The rest of this chapter builds the battery side of Figure 1.1 from its smallest unit upward, and then returns to these supporting systems.

1.2 One cell and its energy

The cell is the unit from which the whole machine is built. This section describes it as a two-terminal device, derives its stored energy from its voltage and its capacity, separates energy from power, and shows why a single cell cannot serve the grid.

1.2.1 The cell as a two-terminal device

A cell has two terminals, a positive one (+) and a negative one (−), and a voltage between them. The reference container uses lithium iron phosphate (LFP) cells with a nominal voltage

of $V_{\text{cell}} \approx 3.2\,\text{V}$ and a charge capacity of $C \approx 300\,\text{Ah}$. *Nominal* means a representative value: the actual voltage of a cell changes with how full it is, with the current flowing through it and with its temperature, and the chapters on the cell and on its circuit model describe those changes. The *capacity* is the amount of electric charge the cell delivers between full and empty under defined conditions, measured in ampere-hours: 300 Ah means, in the ideal case, 300 A for one hour, 100 A for three hours or 30 A for ten hours.

Figure 1.2 draws the cell beside a water-tank analogy. On the left is one prismatic cell with its two terminals on top and the voltage V_{cell} between them; its capacity and energy are written beneath it. On the right is an open water tank with an outlet near the bottom. The labels map each part of the tank onto an electrical quantity: the height of the water, which sets the pressure at the outlet, corresponds to the voltage; the size of the tank corresponds to the capacity; the water stored corresponds to the stored charge; and the flow rate through the outlet corresponds to the current.

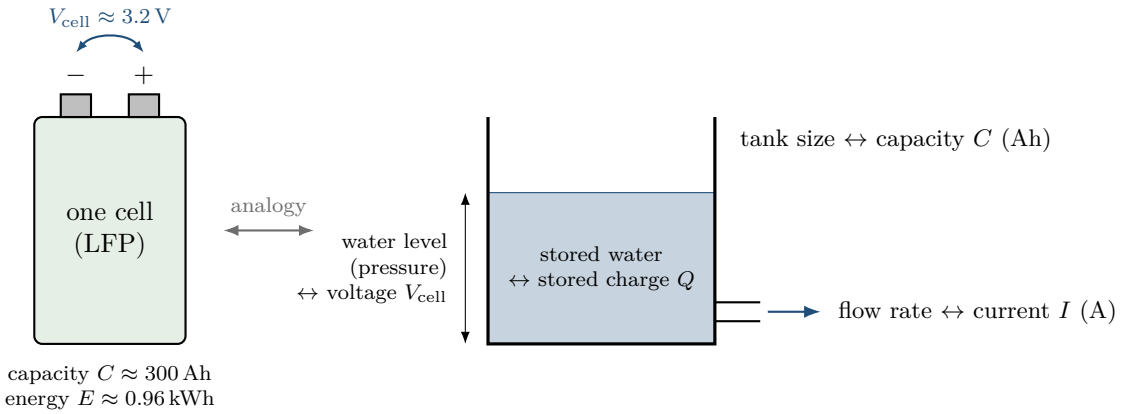


Figure 1.2: One cell as a two-terminal device (left) and the water-tank analogy (right). The nominal values are those of the reference container's cells. The analogy is imperfect: in a tank the pressure falls steadily as the tank drains, whereas the voltage of an LFP cell stays nearly constant over most of its range.

The analogy is useful because it separates two ideas that are easy to confuse. Voltage is a kind of electrical pressure: it is the energy each unit of charge carries, and it determines how hard the cell pushes charge through an external circuit. Capacity is the size of the electrical tank: it says how much charge the cell can move before it is empty. A cell with a high voltage and a small capacity pushes hard for a short time; a cell with a low voltage and a large capacity pushes gently for a long time. The analogy is also imperfect, and the imperfection matters later. In a water tank the pressure at the outlet falls in proportion to the level, so reading the pressure tells how full the tank is. In an LFP cell the voltage stays nearly flat over most of the range between full and empty, so the voltage says little about how full the cell is; the chapter on the cell develops this.

1.2.2 Energy as voltage times capacity

Electrical power is the product of voltage and current, $P = VI$, and energy is power accumulated over time,

$$E = \int P\,\mathrm{d}t = \int V\,I\,\mathrm{d}t. \tag{1.1}$$

If the voltage is held at a constant value V, it comes out of the integral, and the remaining integral of current over time is the charge moved, $\int I\,\mathrm{d}t = Q$. Over a full discharge the charge

moved is the capacity, so the energy stored in a cell is approximately

$$E \approx V\,C. \tag{1.2}$$

The units confirm it: volts times ampere-hours is $\mathrm{V} \cdot \mathrm{A} \cdot \mathrm{h} = \mathrm{W} \cdot \mathrm{h}$, a watt-hour. For one cell of the reference container,

$$E = 3.2\,\mathrm{V} \times 300\,\mathrm{Ah} = 960\,\mathrm{Wh} \approx 0.96\,\mathrm{kWh}.$$

That is about one kilowatt for about one hour. Delivering 1 kW at 3.2 V takes a current of $1000/3.2 \approx 312$ A, which is close to the current that empties a 300 Ah cell in one hour, so the two statements agree. Equation (1.2) ignores two things. The voltage is not exactly constant during a discharge, so the integral in equation (1.1) is the exact statement and $V\,C$ the nominal one; and part of the energy is lost as heat inside the cell, so less than 960 Wh reaches the terminals at high current. It also ignores limits: whether the cell may in fact be discharged at 312 A is set by the manufacturer and enforced by the BMS.

1.2.3 Power and energy

Power and energy are different quantities, and a battery has a limit on each. Power, in kilowatts (kW) or megawatts (MW), is how fast energy moves. Energy, in kilowatt-hours (kWh) or megawatt-hours (MWh), is how much is available. In the water-tank picture, energy is the water in the tank and power is the flow rate through the outlet. They are linked by time: at a constant power P held for a duration t, the energy delivered is

$$E = P\,t. \tag{1.3}$$

A battery that stores 4 MWh can therefore provide 1 MW for four hours, 2 MW for two hours or 4 MW for one hour, as Table 1.1 lists; the product is the same in each row. Which of these it can actually do depends on its power limit, set by the rating of the PCS and by the current the cells tolerate, which is a separate number from its energy. A statement of a battery's size is therefore incomplete unless it gives both: megawatts for how fast and megawatt-hours for how much.

Table 1.1: Ways of delivering the energy of an illustrative 4 MWh battery, equation (1.3), ignoring losses and limits.

Power P	Duration t	Energy $E = P\,t$
0.5 MW	8 h	4 MWh
1 MW	4 h	4 MWh
2 MW	2 h	4 MWh
4 MW	1 h	4 MWh

1.2.4 Why one cell is not enough

A single cell is far too small for the grid in both voltage and energy. Consider voltage first. A power conversion system connected to a medium-voltage grid works with a DC voltage of the order of a thousand volts. At 3.2 V, the current needed to deliver 1.3 MW would be

$$I = \frac{P}{V} = \frac{1.3 \times 10^{6}\,\mathrm{W}}{3.2\,\mathrm{V}} \approx 406{,}000\,\mathrm{A},$$

a current no practical conductor or converter carries. Raising the voltage lowers the current needed for the same power, and the losses in the conductors scale with the square of the current (the heat I^2R of Section 1.6.3), so a high voltage is not a convenience but a necessity.

Energy is the second shortfall. One cell stores about 0.96 kWh; a container that stores 3.42 MWh needs

$$\frac{3{,}420\,\text{kWh}}{0.96\,\text{kWh}} \approx 3{,}560$$

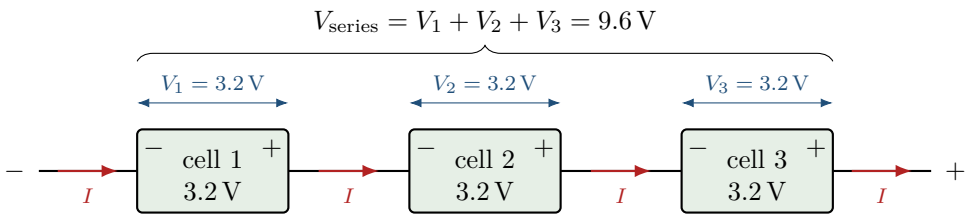
cells. Cells must therefore be connected together, and there are two ways of doing it. Connecting them in series and connecting them in parallel have different effects on voltage, current and capacity, and the arrangement of the reference container uses both. The next two sections take them in turn.

1.3 Series strings and the rack

Connecting cells end to end is how a battery reaches a voltage the power conversion system can use. This section states the two rules of a series connection and applies them to the rack of the reference container.

1.3.1 Voltages add in series

In a series connection the positive terminal of one cell is joined to the negative terminal of the next. Figure 1.3 shows three cells connected this way. Each box is a cell with its negative terminal on the left and its positive terminal on the right; the short blue arrow above each cell is that cell's voltage V_i, and the brace above all three is the voltage across the whole chain. The red arrows on the wires are the current, drawn in the direction it flows during discharge.



the same current I enters and leaves every cell:
if $I = 100\,\text{A}$ at the terminals, each of the three cells carries 100 A

Figure 1.3: Three cells in series during discharge. The voltages add; the current is common to all three cells.

The voltages add. Three cells of 3.2 V in series give

$$3.2\,\text{V} + 3.2\,\text{V} + 3.2\,\text{V} = 9.6\,\text{V},$$

and in general a string of N cells in series has the voltage

$$V_{\text{series}} = \sum_{i=1}^{N} V_i. \tag{1.4}$$

The reason is that voltage is energy per unit charge. A unit of charge that passes through the whole string passes through every cell in turn, and each cell adds its own energy to it, so the energy per unit charge across the string is the sum of the energies per unit charge across the cells. The same statement in circuit language is Kirchhoff's voltage law: around any closed path, the voltages sum to zero, so the voltage between the two ends of the string equals the sum of the voltages between the intermediate junctions. Equation (1.4) holds for unequal cells as well: if one cell sits at 3.1 V and the others at 3.2 V, the string voltage is 9.5 V.

1.3.2 The common current

The second rule of a series connection is that the same current flows through every cell. In Figure 1.3 there is only one path for charge: whatever enters cell 1 at its negative terminal leaves it at its positive terminal, enters cell 2, and so on. Charge does not accumulate at the junction between two cells, because charge is conserved and the junction stores none, so the current leaving one cell equals the current entering the next. If the current at the terminals of a rack is 100 A, then every cell in that rack carries 100 A. The current is not divided by the number of cells in series: it is not 100/3 in a string of three nor 100/396 in a string of 396.

An analogy is a line of runners roped together at the waist. Whatever the speed of the line, every runner moves at that speed; adding runners to the line does not make each of them slower or faster. The rope also makes the line only as fast as its slowest runner, and a series string has the corresponding property: because every cell carries the same current, the string must stop discharging when any one cell reaches its lower limit, whatever the state of the others. Section 1.6.2 returns to that consequence.

The common current has a second consequence that runs through the whole book. A single current sensor at the end of a string measures the current of every cell in the string. One measured number drives the charge balance of hundreds of cells at once, which makes the current measurement the most consequential sensor in the battery.

1.3.3 The rack of the reference container

The rack of the reference container applies these two rules at scale. Each module holds 44 cells and each rack holds nine modules, so a rack holds

$$44 \times 9 = 396$$

cells, all connected in series. The module is a mechanical and electrical grouping for assembly and maintenance; electrically the rack is one string of 396 cells. Figure 1.4 draws it. The nine rounded boxes are the modules, numbered along the string; the small green rectangles and the dots inside each module stand for its 44 cells; the short wires between modules join the positive end of one module to the negative end of the next; and the two leads at the ends are the rack's terminals, with the rack current I entering at the left. The brace over the first module gives the module's numbers and the brace under the whole row gives the rack's.

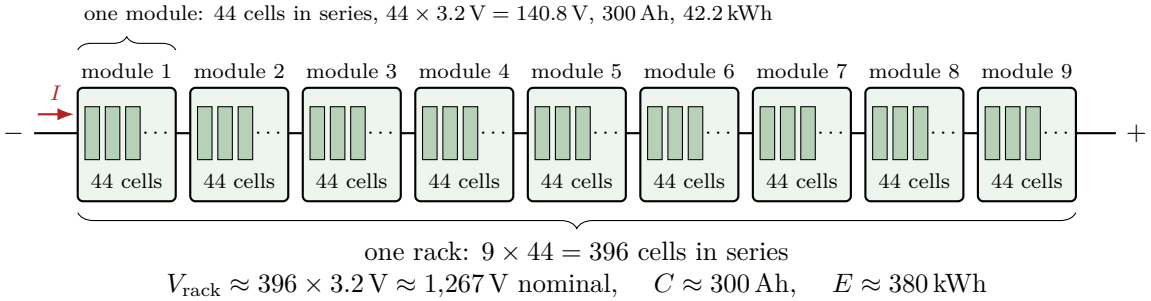


Figure 1.4: One rack of the reference container: nine modules of 44 cells, all 396 cells in one series string. The rack current I passes through every cell; the rack voltage is the sum of the 396 cell voltages.

By equation (1.4), a module has a nominal voltage of $44 \times 3.2 = 140.8$ V and the rack a nominal voltage of

$$V_{\text{rack}} \approx 396 \times 3.2\,\text{V} \approx 1{,}267\,\text{V}. \tag{1.5}$$

The rack voltages measured in the reference record lie between about 1,250 V and 1,350 V. Dividing by 396, that range corresponds to an average cell voltage between $1{,}250/396 \approx 3.16$ V

and 1,350/396 ≈ 3.41 V; the variation comes from the state of charge and from the current, both described in later chapters. What is sold or described as a "1,300 V battery" is therefore 396 small cells of about 3.2 V stacked in series.

1.3.4 Capacity and energy of the rack

A series connection increases the voltage but not the charge capacity. Every cell in the string carries the same current for the same length of time, so every cell passes the same charge. When 300 Ah has flowed through the rack terminals, 300 Ah has flowed through each of the 396 cells, and each of them is empty. The capacity of the string is therefore the capacity of one cell,

$$C_{\text{rack}} = C_{\text{cell}} \approx 300\,\text{Ah},$$

for identical cells. If the cells differ, the string's usable capacity is set by the cell that empties first, which is the property of the roped runners in Section 1.3.2. The energy of the rack follows from equation (1.2) with the rack voltage,

$$E_{\text{rack}} \approx V_{\text{rack}}\, C \approx 1{,}267\,\text{V} \times 300\,\text{Ah} \approx 380{,}000\,\text{Wh} \approx 380\,\text{kWh}. \tag{1.6}$$

The same number comes from the cells: $396 \times 960\,\text{Wh} = 380{,}160\,\text{Wh}$. The two routes agree because series connection adds energy as well as voltage: each cell contributes its own 960 Wh, and the string merely delivers them all at a higher voltage through the same 300 Ah of charge. Series connection thus multiplies the voltage and the energy by the number of cells and leaves the capacity in ampere-hours unchanged.

1.4 Parallel racks and the container

A rack has the voltage the converter needs but only a ninth of the container's energy. This section connects nine racks in parallel, states the rules of a parallel connection beside those of a series one, and assembles the whole container.

1.4.1 Racks in parallel on a common DC bus

The nine racks of the reference container are connected in parallel to a common *DC bus*, a pair of heavy conductors to which every rack and the PCS connect. In a parallel connection every rack is connected across the same two conductors, so every rack has the same voltage, and the currents of the racks add.

Figure 1.5 draws the arrangement. The thick blue line at the top is the positive conductor of the DC bus (the return conductor is not drawn); the nine green boxes below it are the racks, each a string of 396 cells; and the vertical line from each rack to the bus carries that rack's current I_r, drawn as a red arrow in the direction of discharge. On the right the bus enters the PCS, and the red arrow on the bus is the total current I_{bus}. The note beneath the racks states the two rules. In this section the subscript of V and I indexes racks, $r = 1, \ldots, 9$.

The two rules are

$$V_1 \approx V_2 \approx \cdots \approx V_9 \approx V_{\text{bus}}, \qquad I_{\text{bus}} = I_1 + I_2 + \cdots + I_9 = \sum_{r=1}^{9} I_r. \tag{1.7}$$

The current rule is conservation of charge at the bus: the bus stores no charge, so the current leaving it toward the PCS equals the sum of the currents arriving from the racks. This is Kirchhoff's current law. The voltage rule is approximate because each rack reaches the bus through its own cables, fuses and contactors, which have a small resistance; the voltage at a rack's terminals differs from the bus voltage by the product of that resistance and the rack's

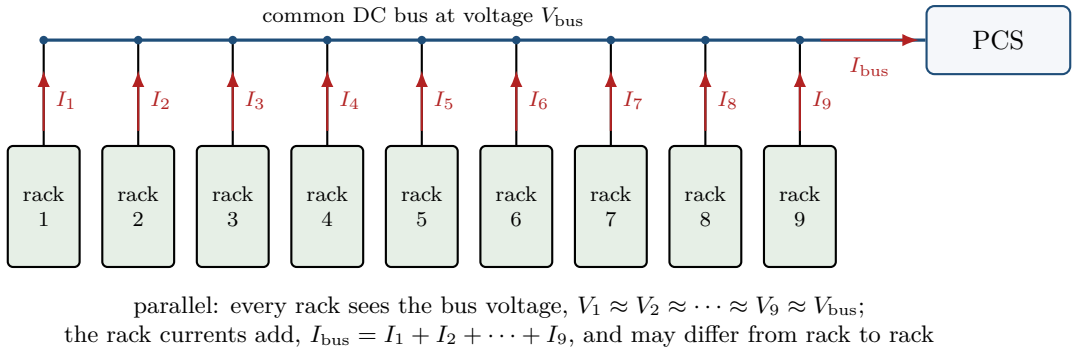


Figure 1.5: The nine racks of the reference container in parallel on a common DC bus, during discharge. Each rack is a series string of 396 cells; the bus current into the PCS is the sum of the rack currents.

current, a fraction of a volt to a few volts. The rack currents need not be equal. A rack whose cells are at a slightly different state of charge, or whose internal resistance is slightly higher, settles at a different current while sharing the common voltage. If the bus current is 900 A and the racks were identical, each would carry 100 A; in practice the nine currents spread around that value, for instance between 95 and 105 A (illustrative), and that spread is itself informative about the racks.

Capacity adds in parallel. Each rack supplies its own part of the bus current and drains its own 300 Ah, so nine racks in parallel can supply

$$9 \times 300\,\text{Ah} = 2{,}700\,\text{Ah}$$

to the bus, at the rack voltage. The voltage does not increase: the container has the voltage of one rack.

1.4.2 Series and parallel compared

Table 1.2 sets the two kinds of connection side by side. Each row is one electrical quantity; the two middle columns say what happens to that quantity when N identical units are connected in series or in parallel, and the last column gives the reference container's instance of each.

Table 1.2: Series and parallel connection of N identical units.

Quantity	Series	Parallel	Reference container
Voltage	adds: $N\,V$	same: V	396 cells in series give $\approx$ 1,267 V; nine racks in parallel stay at that voltage
Current	same through every unit	adds: $\sum I$	100 A in a rack is 100 A in each of its cells; the bus current is the sum of nine rack currents
Capacity (Ah)	same: C	adds: $N\,C$	a rack has 300 Ah; the container 2,700 Ah
Energy (Wh)	adds: $N\,E$	adds: $N\,E$	380 kWh per rack; 3.42 MWh per container

The table is worth memorizing, because most of the reasoning about measurements in the later chapters starts from it. A cell voltage is one term of a sum; a rack current is the current

of 396 cells; a bus current is a sum of nine rack currents; and only energy adds under both connections.

Principle 1.1. *In series, voltages add and the current is common; in parallel, currents and capacities add and the voltage is common. Energy adds in both.*

1.4.3 The whole container

The container is nine racks of 396 cells, so it holds

$$9 \times 396 = 3{,}564$$

cells. Its energy is nine times the energy of one rack,

$$E_{\text{container}} \approx 9 \times 380\,\text{kWh} \approx 3{,}420\,\text{kWh} \approx 3.42\,\text{MWh}, \tag{1.8}$$

and the same number follows from the cells, $3{,}564 \times 0.96\,\text{kWh} \approx 3{,}421\,\text{kWh}$, or from the bus, $1{,}267\,\text{V} \times 2{,}700\,\text{Ah} \approx 3{,}421\,\text{kWh}$. The three routes agree because they are the same product of a voltage and a charge, grouped differently.

Figure 1.6 draws the architecture as a ladder. Each box in the top row is one level of the hierarchy with its count of cells, its nominal voltage, its capacity and its energy; the label on each arrow says how many units of the level on the left make one unit of the level on the right, and how they are connected. The bottom row is the end of the chain, the PCS and the grid. Reading from left to right, the voltage grows only through the series steps and the capacity only through the parallel step, while the energy grows at every step.

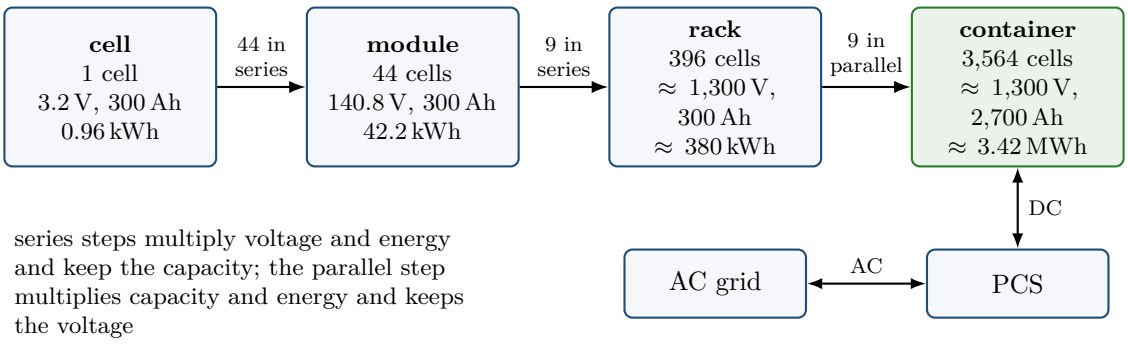


Figure 1.6: The architecture of the reference container, from one cell to the grid. Voltages are rounded: the nominal rack voltage is $396 \times 3.2 = 1{,}267\,\text{V}$ and measured rack voltages lie between about 1,250 and 1,350 V.

Written as a chain, the architecture is

$$\text{cell } (\approx 3.2\,\text{V}, \approx 300\,\text{Ah}) \xrightarrow{44 \text{ in series}} \text{module} \xrightarrow{9 \text{ in series}} \text{rack } (396 \text{ cells}, \approx 1{,}300\,\text{V}, \approx 300\,\text{Ah}, \approx 380\,\text{kWh})$$
$$\xrightarrow{9 \text{ in parallel}} \text{container } (3{,}564 \text{ cells}, \approx 1{,}300\,\text{V}, \approx 2{,}700\,\text{Ah}, \approx 3.42\,\text{MWh}) \to \text{PCS} \to \text{grid}.$$

Every cell voltage in the reference record belongs to one position in this ladder, every rack voltage and rack current to one of the nine strings, and every bus or PCS quantity to the top of it.

1.5 Current, charge and state of charge

The current flowing through a string says how much charge has left or entered every one of its cells. This section defines charge and its units, writes the charge balance that turns a current record into a charge record, defines the state of charge, and explains how the reported state of charge of the reference record is produced.

1.5.1 Current as a flow of charge

Electric charge is measured in coulombs (unit symbol C, set upright, not to be confused with the capacity C in italic). Current is the rate at which charge flows past a point, and one ampere is one coulomb per second,

$$1\,\mathrm{A} = 1\,\mathrm{C/s}.$$

The coulomb is a small unit for a battery, so engineers count charge in ampere-hours: one ampere flowing for one hour, or $1\,\mathrm{Ah} = 1\,\mathrm{A} \times 3{,}600\,\mathrm{s} = 3{,}600\,\mathrm{C}$. The capacity of one cell of the reference container is then $300 \times 3{,}600 = 1.08 \times 10^6\,\mathrm{C}$.

Because the 396 cells of a rack are in series, a rack current of 100 A means that 100 C of charge passes through each of the 396 cells every second. The rack-current channel of the reference record is therefore not the current of one component but the common current of 396 cells, and it is the only direct record of how much charge each of them has moved.

1.5.2 Counting charge

Charge moved is current accumulated over time. A rack of 300 Ah cells delivering 100 A for one hour has delivered 100 Ah, which is

$$\frac{100\,\mathrm{Ah}}{300\,\mathrm{Ah}} = \frac{1}{3} \approx 33\,\%$$

of its nominal capacity. For a current that varies, the accumulation is an integral. With $Q(t)$ the charge stored in a cell at time t and the current positive on discharge,

$$Q(t) = Q(0) - \int_0^t I(\tau)\,\mathrm{d}\tau. \tag{1.9}$$

Discharge ($I > 0$) lowers the stored charge and charge ($I < 0$) raises it. A telemetry record does not contain the continuous current but samples of it, I_k at intervals Δt, and the integral becomes a running sum,

$$Q_k = Q_{k-1} - I_k\,\Delta t. \tag{1.10}$$

With one-second samples, $\Delta t = 1/3{,}600\,\mathrm{h}$, so a sample of 100 A removes $100/3{,}600 \approx 0.028\,\mathrm{Ah}$. As an illustrative sequence, a rack that discharges at 150 A for 30 minutes and then charges at 60 A ($I = -60\,\mathrm{A}$) for 20 minutes changes its stored charge by

$$-150\,\mathrm{A} \times 0.5\,\mathrm{h} + 60\,\mathrm{A} \times \tfrac{1}{3}\,\mathrm{h} = -75\,\mathrm{Ah} + 20\,\mathrm{Ah} = -55\,\mathrm{Ah}.$$

Equation (1.9) is the *charge balance*: the charge stored changes only by the charge that crosses the terminals. It is one of the four relations on which the rest of the book is built.

1.5.3 State of charge

The *state of charge* expresses the stored charge as a fraction of the usable capacity,

$$\mathrm{SOC} = \frac{\text{charge currently available}}{\text{usable charge capacity}} = \frac{Q}{C}. \tag{1.11}$$

An SOC of 100 % means full and 50 % half full. An SOC of 0 % means the defined lower operating limit, not a cell emptied of all its chemical content: the cell is never discharged that far, and 0 % marks the point at which operation stops. A rack with 150 Ah remaining of 300 Ah is at

$$\mathrm{SOC} = \frac{150\,\mathrm{Ah}}{300\,\mathrm{Ah}} = 50\,\%,$$

and a full rack that delivers 100 A for one hour ends at $(300 - 100)/300 \approx 66.7\,\%$.

Dividing the charge balance (1.9) by the capacity gives the state of charge as a function of time,

$$\mathrm{SOC}(t) = \mathrm{SOC}(t_0) - \frac{1}{C}\int_{t_0}^{t} I(\tau)\,\mathrm{d}\tau, \tag{1.12}$$

which says that the state of charge at any time is its value at a starting time minus the charge delivered since then, expressed as a fraction of the capacity.

1.5.4 Coulomb counting and the reported state of charge

The state of charge is generally not measured. There is no SOC sensor in a cell: no instrument reads the charge stored inside it. The sensors on a battery measure voltage, current and temperature, and the state of charge must be estimated from them. The BMS estimates it, and the most common method is *coulomb counting*, which is equation (1.12) computed in real time: start from a known value, integrate the measured current, and divide by an assumed capacity. Many BMSs also reset the count when the battery reaches a recognizable condition, such as full charge or a long rest; between resets the reported value is a function of the measured current alone.

In the reference record the reported state of charge is produced essentially this way: the BMS integrates the same rack-current measurement that appears in the record and divides by an assumed capacity. The SOC channel is therefore not an independent measurement of the true stored charge. It is partly a calculation based on another sensor, and it inherits that sensor's errors and the assumptions of the calculation:

- *A current offset accumulates.* If the current sensor reads 1 A high (illustrative), the computed charge drifts by 24 Ah per day, which is $24/300 = 8\,\%$ of state of charge per day.
- *A capacity error scales every change.* If the true capacity has fallen to 285 Ah (illustrative) and the BMS still assumes 300 Ah, a 100 Ah discharge is reported as a drop of 33.3 % when the true drop is $100/285 = 35.1\,\%$.
- *An initial error persists.* Equation (1.12) carries $\mathrm{SOC}(t_0)$ forward unchanged; an error in it is never corrected by the integral.

The practical consequence is a rule for analysis. Comparing the reported SOC with a charge balance computed from the same rack current tests only whether the arithmetic was done; agreement is guaranteed by construction and says nothing about the battery or the current sensor. To learn something about the true stored charge, the analysis needs a measurement that the integral did not use, such as the cell voltage at rest.

Principle 1.2. *A reported state of charge is a derived channel, not a measurement: it is an integral of the measured current divided by an assumed capacity, and it cannot independently confirm a model driven by the same current.*

1.6 Supervision, heat and the grid interface

The cells, strings and racks store the energy; three further parts make the machine work. This section describes the battery management system and why it measures every cell, the battery as a thermal system, and the power conversion system, and then assembles the complete picture and outlines where the following chapters go.

1.6.1 The battery management system

The *battery management system* is the computer and control system that supervises the battery. It works as a loop: its inputs are the readings of the cell-voltage, current and temperature sensors; it compares them with limits and computes estimates; and its outputs are alarms, operating limits sent to the PCS, and commands that open or close the contactors. In the reference container the BMS reads 396 cell voltages and 216 temperatures in each rack, together with the rack voltage and rack current. It monitors cell voltages, rack voltage, current, temperatures and the state of charge, manages *balancing* (equalizing the charge of the cells in a string, typically by bleeding a small current from the cells that are fuller than the rest), and protects against overvoltage, undervoltage, overheating and overcurrent. Table 1.3 lists the typical protective functions; each row gives the condition guarded against, the measurement that reveals it, and the typical response.

Table 1.3: Typical protective functions of a BMS.

Condition	Measurement	Typical response
Overvoltage	any cell voltage above its upper limit	reduce or stop charging; open the rack contactor if it persists
Undervoltage	any cell voltage below its lower limit	reduce or stop discharging; open the contactor if it persists
Overheating	any temperature above its limit	reduce the allowed current; stop operation at a higher threshold
Overcurrent	rack current above its rating	limit the current through the PCS; open the contactor on a fault
Imbalance	spread of cell voltages in a string	balance the cells

Two features of the table matter for what follows. Every protective function acts on the most extreme cell or sensor, not on an average; and the BMS's outputs (limits, contactor states, the reported state of charge) appear in the reference record beside the measurements, so an analysis of the record must keep track of which channels were measured and which were computed or decided by the BMS.

1.6.2 Why every cell is measured

A rack is only as healthy as its weakest cells, and an aggregate measurement hides a weak cell. Consider a rack in which 395 cells sit at 3.35 V and one cell sits at 2.80 V. The rack voltage is

$$395 \times 3.35\,\mathrm{V} + 2.80\,\mathrm{V} = 1{,}323.25\,\mathrm{V} + 2.80\,\mathrm{V} = 1{,}326.05\,\mathrm{V},$$

while a rack of 396 cells all at 3.35 V would read $396 \times 3.35 = 1{,}326.60\,\mathrm{V}$. The low cell lowers the rack voltage by 0.55 V, or 0.04 %, and the average cell voltage from 3.3500 V to $1{,}326.05/396 \approx 3.3486\,\mathrm{V}$. A rack voltage of 1,326 V lies in the middle of the normal range of 1,250 to 1,350 V, and the 0.55 V shift is far smaller than the change that the state of charge produces in the course of a day.

Figure 1.7 shows the example. The left panel plots the voltage of each cell against its position in the string, from 1 to 396: the blue points form a flat line at 3.35 V and the single red point, placed at position 200 for illustration, sits at 2.80 V. The right panel plots the string voltage, the sum of the 396 cell voltages, for a rack with all cells normal (blue bar) and for the rack with the one low cell (red bar), on an axis that starts at zero. The deviation that is obvious in the left panel is invisible in the right one.

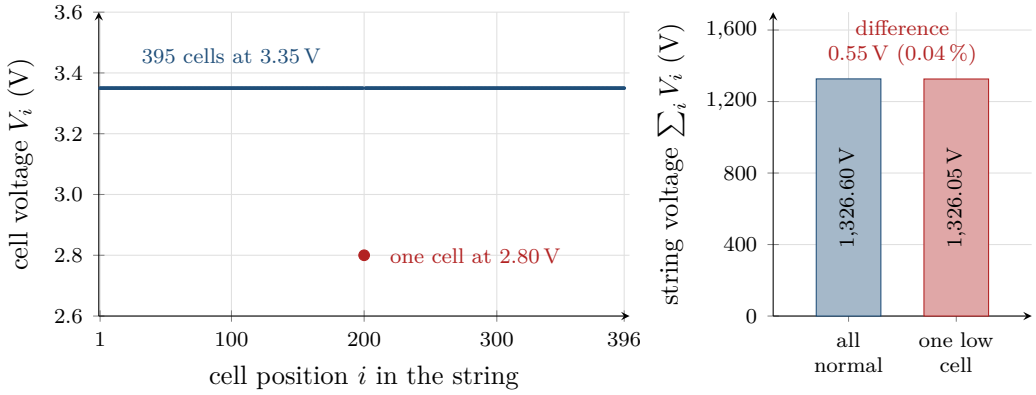


Figure 1.7: The dilution of one abnormal cell in a string sum. Left: cell voltages of a 396-cell rack with 395 cells at 3.35 V and one at 2.80 V. Right: the string voltage, equation (1.4), with and without the low cell. Values are illustrative; the position of the low cell does not affect the sum.

The reason is arithmetic. In a sum of 396 comparable terms, any one term contributes about $1/396 \approx 0.25\,\%$ of the total, so even a large change in one cell is a small change in the sum. The problem disappears in the aggregate and in the average. Yet the low cell is the one that determines what the rack can do. Because all 396 cells carry the same current (Section 1.3.2), the BMS must stop the discharge of the whole rack when the lowest cell reaches its lower voltage limit, even though the other 395 still hold usable charge. The usable capacity of the rack is set by its weakest cell, while the rack voltage and the rack average look normal. This is why the BMS measures every cell, and why an analysis that uses only rack-level channels can miss the condition that limits the rack.

Principle 1.3. *An aggregate can look normal while one cell limits the system: a single cell contributes about* $1/396$ *of a rack's voltage, so its deviation is diluted in the rack sum and in the rack average, but it alone can end a discharge.*

1.6.3 The battery as a thermal system

A battery is also a thermal system. Current flowing through a cell meets an internal resistance R, and the electrical energy lost in that resistance becomes heat at the rate

$$P_{\text{heat}} = I^2 R. \tag{1.13}$$

The heat depends on the square of the current, so it is the same on charge and on discharge, and doubling the current quadruples it. The chain of cause and effect is

$$\text{current} \rightarrow \text{resistive loss} \rightarrow \text{heat} \rightarrow \text{temperature rise} \rightarrow \text{cooling removes heat.}$$

The temperature is set by the balance between the heat generated and the heat removed. In its most primitive form, the heat stored in the battery changes at the rate at which heat is

generated minus the rate at which it is removed,

$$C_{\text{th}} \frac{\mathrm{d}T}{\mathrm{d}t} = I^2 R - \dot{Q}_{\text{cool}}, \tag{1.14}$$

where C_{th} is the thermal capacitance (the heat needed to raise the temperature by one kelvin) and $\dot{Q}_{\text{cool}}$ the rate of heat removal by the cooling system. When the two terms on the right are equal the temperature is steady; when generation exceeds removal it rises.

As an illustration, take a rack whose 396 cells and connections add up to a resistance of $R = 0.2\,\Omega$ (illustrative; about $0.5\,\text{m}\Omega$ per cell). At a rack current of 111 A the heat generated is $111^2 \times 0.2 \approx 2{,}460\,\text{W}$, about 2.5 kW; at 222 A it is about 9.9 kW. With an illustrative thermal capacitance of $C_{\text{th}} = 2 \times 10^6\,\text{J/K}$ and no cooling, 2.46 kW raises the rack temperature at $2{,}460/(2 \times 10^6) \approx 1.2 \times 10^{-3}\,\text{K/s}$, about 4.4 K per hour. The analysis of the reference record uses a rack-level balance of exactly the form of equation (1.14), with the heat generation tied to the measured rack current and the rack temperatures as the measured response; the chapter on thermal behavior develops it.

1.6.4 The power conversion system

The battery is DC and the grid is AC, and the PCS converts in both directions: from DC to AC during discharge and from AC to DC during charge. It also controls how much power goes in or out, by setting the current it draws from or pushes into the DC bus. On the DC side the power is the product of the bus voltage and the bus current,

$$P_{\text{DC}} \approx V_{\text{bus}}\, I_{\text{bus}}. \tag{1.15}$$

With $V_{\text{bus}} = 1{,}300\,\text{V}$ and $I_{\text{bus}} = 1{,}000\,\text{A}$,

$$P_{\text{DC}} \approx 1{,}300\,\text{V} \times 1{,}000\,\text{A} = 1.3 \times 10^6\,\text{W} = 1.3\,\text{MW}.$$

If the nine racks share that current equally, each carries about $1{,}000/9 \approx 111\,\text{A}$ and delivers about $1{,}300 \times 111 \approx 144\,\text{kW}$. At 1.3 MW the container's 3.42 MWh would last about $3.42/1.3 \approx 2.6$ hours, ignoring losses and state-of-charge limits. The AC power delivered to the grid is slightly less than P_{DC}, by the losses in the converter.

Equation (1.15) is how an ordinary current measurement becomes a grid-scale power flow. A thousand amperes on the bus at 1,300 V is more than a megawatt, and a 1 % error in the current measurement, 10 A, is an error of 13 kW in power and, by equation (1.9), of 10 Ah per hour in the charge balance of the bus.

1.6.5 The complete picture

Figure 1.8 assembles the machine. The top row is the chain grid, PCS and DC bus, with the AC and DC sides of the converter labeled. Below the bus hang the nine racks, each connected through its own line carrying its current I_r; the bus current I_{bus} flows into the PCS. The dashed lines from rack 1 open it up at the bottom of the figure: a row of cells, of which five are drawn (V_1, V_2, V_3, V_{395} and V_{396}, here indexing cells), joined in series and carrying the single rack current I_1. The brace under the cells gives the rack voltage as their sum, and the note on the right states the two rules that hold inside and between racks.

Read from top to bottom, the figure states the facts to keep:

1. The grid is AC; the PCS converts between AC and DC and sets the power.
2. The DC side is a common bus to which racks 1 to 9 connect in parallel.
3. Inside a rack, 396 cells are connected in series.

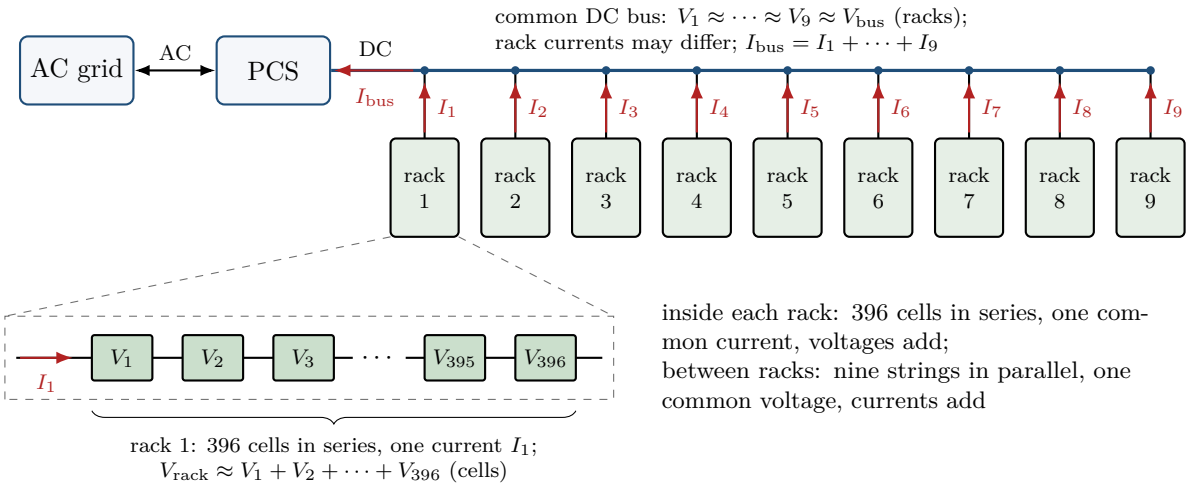


Figure 1.8: The complete picture of the reference container: grid, PCS, common DC bus, nine racks in parallel, and inside each rack 396 cells in series. Arrows show the direction of current during discharge.

4. The same current flows through every cell of a rack.
5. The rack voltage is approximately the sum of its cell voltages, $V_1 + \cdots + V_{396}$.
6. The rack voltages are approximately the same, because the racks share the bus.
7. The rack currents can differ from one another.
8. The total current is the sum of the rack currents, $I_{\text{bus}} = I_1 + \cdots + I_9$.

Every channel of the reference record attaches to one place in this picture: a cell voltage to one box in the bottom row, a rack current to one vertical line, a rack voltage to one rack, a temperature to a location inside a rack, the reported state of charge to the BMS's integral of a rack current, and the power to the PCS.

1.6.6 What the following chapters add

This chapter has described the machine; the following chapters describe what happens inside it and what its measurements can reveal. The electrical quantities used here (voltage, current, charge, power, energy) and their units are defined carefully in Chapter 2. Chapter 3 opens the cell: its chemistry, the relation between open-circuit voltage and state of charge, and why that relation is nearly flat for LFP, which is the imperfection of the water-tank analogy. Chapter 4 introduces internal resistance and the circuit model of a cell, which explains why the terminal voltage V differs from the open-circuit voltage U whenever current flows. Chapter 5 describes the BMS and the hardware in detail, Chapter 6 the thermal behavior, Chapter 7 degradation, Chapter 8 faults, and Chapter 9 what the sensors measure directly versus what must be inferred.

Four relations organize that work, and each has appeared in this chapter in its simplest form: the charge balance, equation (1.9); the rack or string voltage identity, equation (1.4); a per-cell electrical model relating each cell's voltage to its current and state, of which equation (1.2) is the crudest version; and the heat balance, equation (1.14). The question the book works toward is: given these measurements and these physical laws, what can be known about the battery, and what is fundamentally unidentifiable?

1.7 What to retain

Definition 1.1 (Battery energy storage system)**.** A battery energy storage system is a power converter and a supervisory system around a large array of low-voltage cells, in which series connection builds the voltage (396 cells, about 1,300 V per rack) and parallel connection builds the current, capacity and energy (nine racks, 2,700 Ah, 3.42 MWh). Within a series string every cell carries the same current, so one current measurement and one integral describe the charge of 396 cells, and the reported state of charge is that integral, not an independent measurement.

Table 1.4 lists the chapter's objects, the relation or value that defines each in the reference container, and the subsection in which it is introduced.

Table 1.4: The objects of this chapter.

Object	Relation or value	Subsection
BESS	grid $\leftrightarrow$ PCS $\leftrightarrow$ battery; charging and discharging as energy conversion	1.1.1, 1.1.2
Cell	$V_{\text{cell}} \approx 3.2\,\text{V}$, $C \approx 300\,\text{Ah}$	1.2.1
Energy	$E \approx V\,C = 960\,\text{Wh}$ per cell	1.2.2
Power and energy	$E = P\,t$; MW for how fast, MWh for how much	1.2.3
Series string	$V_{\text{series}} = \sum_i V_i$; common current	1.3.1, 1.3.2
Rack	396 cells; $\approx 1{,}267\,\text{V}$ nominal, 300 Ah, 380 kWh	1.3.3, 1.3.4
Parallel bus	$V_r \approx V_{\text{bus}}$; $I_{\text{bus}} = \sum_r I_r$	1.4.1
Container	3,564 cells, 2,700 Ah, 3.42 MWh	1.4.3
Charge balance	$Q(t) = Q(0) - \int I\,\mathrm{d}t$	1.5.2
State of charge	$\text{SOC} = Q/C$; coulomb counting	1.5.3, 1.5.4
BMS	measures, estimates, protects, balances	1.6.1
Dilution of one cell	one cell is $\approx 1/396$ of the rack voltage	1.6.2
Heat balance	$C_{\text{th}}\,\mathrm{d}T/\mathrm{d}t = I^2R - \dot{Q}_{\text{cool}}$	1.6.3
PCS power	$P_{\text{DC}} \approx V_{\text{bus}}\,I_{\text{bus}} = 1.3\,\text{MW}$ at 1,300 V, 1,000 A	1.6.4

1.8 Exercises

Exercise 1.1. An illustrative container uses cells of 3.2 V and 280 Ah, with 52 cells per module, eight modules per rack connected in series, and ten racks in parallel. Compute (a) the energy of one cell; (b) the number of cells in a rack, and the rack's nominal voltage, capacity and energy; (c) the number of cells in the container, the capacity available at the DC bus, and the container's energy. Check the last answer by two of the three routes of Section 1.4.3.

Exercise 1.2. The reference container discharges at 1.3 MW with a bus voltage of 1,300 V. (a) Compute the bus current and, assuming the racks share it equally, the current in one rack and in one cell. (b) What fraction of a rack's 300 Ah is delivered in 30 minutes? (c) Ignoring losses, how long can the container sustain 1.3 MW from an SOC of 90 % down to 10 %? (d) With the illustrative rack resistance $R = 0.2\,\Omega$ of Section 1.6.3, compute the heat generated in one rack at this operating point, and at twice the power with the same bus voltage.

Exercise 1.3. A rack with $C = 300\,\text{Ah}$ starts at $\text{SOC} = 80\,\%$. It discharges at 120 A for 45 minutes, rests for 15 minutes, and then charges at 90 A ($I = -90\,\text{A}$) for 40 minutes. (a) Use equation (1.12) to compute the state of charge at the end of each phase. (b) The current sensor reads 1.5 A too high throughout the 100 minutes, including the rest. What error does this

introduce in the reported state of charge at the end? (c) The true capacity is 270 Ah while the BMS assumes 300 Ah. Starting from a true SOC of 80 %, compute the true SOC at the end of the discharge phase and compare it with the reported value. (d) Explain why comparing the reported SOC with an independent analyst's integral of the same rack current would reveal neither error.

Exercise 1.4. A rack of 396 cells has 395 cells at 3.30 V and one cell at a voltage V_{low}. (a) Compute the rack voltage for $V_{\text{low}} = 3.30\,\text{V}$ and for $V_{\text{low}} = 2.70\,\text{V}$, and the percentage difference. (b) Show that no single cell, at any voltage between 0 and 3.30 V, can lower the rack voltage by more than about 0.25 %. (c) If the rack-voltage sensor has an illustrative accuracy of ± 0.5 %, can the rack voltage alone reveal the low cell? (d) The BMS stops discharge when any cell reaches 2.50 V. If the low cell stays 0.60 V below the others throughout a discharge (a simplification), at what voltage are the other 395 cells when the discharge stops, and what does this imply for the usable capacity of the rack?

1.9 Connection to the TERRA framework

The architecture of this chapter maps directly onto a conservation graph, and the mapping already shows which questions the reference record can answer.

- **Charge conservation on a rack node.** The charge balance, equation (1.9), is an accumulator: the rack is a node whose stored charge changes only by the current on the edge that connects it to the bus. The DC bus is a node with no storage, and the current rule of equation (1.7), $I_{\text{bus}} = \sum_r I_r$, is its conservation law.

- **The series string is an edge with one flow.** The common current of Section 1.3.2 means that a rack is an edge whose single flow, I_r, passes through 396 cell components. One measured flow drives 396 component models at once, which is why the rack-current sensor is the most consequential measurement in the graph.

- **The string voltage identity is a redundancy relation.** Equation (1.4) relates two independent sets of sensors, the rack-voltage sensor and the 396 cell-voltage sensors, through a physical identity. Because both sides are measured, its residual is a test of the sensors and of the connections between cells and terminals, not a definition.

- **Measured versus derived channels.** The reported state of charge of Section 1.5.4 is computed from the same rack current that drives the charge balance. A conservation boundary with no independent measurement crossing it gives an identity, not a test: fitting the charge balance to the reported SOC confirms the BMS's arithmetic and leaves the true capacity undetermined.

- **The heat balance is an energy-conservation node.** Equation (1.14) is a thermal node with storage C_{th}, an input edge I^2R tied to the electrical flow, and an output edge $\dot{Q}_{\text{cool}}$ to the cooling system; the rack temperatures are the measured states.

- **A fault is a hypothesis with a signature.** The low cell of Figure 1.7 changes the rack voltage by 0.04 %, which a normal model of the rack can absorb into its state of charge or its sensor noise. At the cell node the same fault is a 0.55 V departure. A fault is detectable only where its signature lies outside what the normal model can explain, which is why the graph must carry the cells and not just the racks.

In summary, a battery energy storage system is a graph of a few thousand cell components on nine series edges joined at one bus node, with a converter at its boundary and a thermal

node beside each rack. Its two connection rules are conservation laws, its reported state of charge is a derived channel rather than a measurement, and the dilution of one cell in a sum of 396 is the first instance of the book's central question: which properties of the battery the available measurements can and cannot determine. The next chapter, Chapter 2, defines the electrical quantities used here with their units and the relations between them.

Chapter 2

Electrical quantities: voltage, current, power, energy, ampere-hours, C-rate and state of charge

2.1 Three quantities at two terminals

Seen from outside, a battery is first a two-terminal device. Whatever happens inside the cells, the rest of the system meets the battery at a positive and a negative terminal, and at those terminals three quantities can be asked about: the voltage across them, the current through them, and the power that the two together carry. Everything else this book discusses (charge, energy, capacity, state of charge, heat, aging) is either built from these three or inferred from them. The question of this chapter is

which electrical quantities describe a battery at its terminals, how they are related, and which of them are measured rather than computed.

The chapter defines each quantity in turn, works every relation through with the numbers of the reference container, and ends by reading the channels of the reference record with those definitions in hand.

Jargon: volt, ampere, coulomb, ampere-hour, watt, watt-hour, C-rate, coulomb counting

The *coulomb* (symbol C) is the SI unit of electric charge. The *volt* (V) is the unit of voltage: one volt is one joule of energy per coulomb of charge. The *ampere* (A) is the unit of current: one ampere is one coulomb per second. The *ampere-hour* (Ah) is the charge moved by a current of one ampere flowing for one hour, so 1 Ah = 3600 coulombs; battery capacities are stated in ampere-hours. The *watt* (W) is the unit of power, one joule per second, and the *watt-hour* (Wh) is the energy transferred by one watt in one hour, 1 Wh = 3600 joules; the kilowatt-hour (kWh) and megawatt-hour (MWh) are 10^3 and 10^6 watt-hours. The *C-rate* is a current divided by the battery's capacity in ampere-hours, written with a suffix C: a 300 Ah cell discharged at 150 A runs at "0.5C". *Coulomb counting* is the estimation of state of charge by integrating the measured current over time and dividing by an assumed capacity. The letter C therefore does three jobs in this field: the unit coulomb (upright C), the capacity C (italic, in Ah), and the C-rate suffix. The context always says which is meant.

Key idea

A battery is described at its terminals by voltage, current and their product, power. Charge and energy are the time integrals of current and power, and state of charge is a charge integral divided by an assumed capacity: it is computed from the current, not measured.

This section defines voltage and current, the two quantities a sensor can read directly at the terminals of a rack.

2.1.1 Voltage

Voltage, measured in volts, is the electrical potential difference between two points. Physically it says how strongly charge is pushed through a circuit connected between those points: one volt means that each coulomb moved from the higher to the lower potential gives up one joule of energy. That definition, energy per unit charge, is the one to keep, because it is what later turns ampere-hours into watt-hours. A single cell of the reference container has a voltage of about 3.2 V. The value is not fixed: it depends on how full the cell is and on the current flowing, which are the subjects of Chapters 3 and 4. For this chapter 3.2 V is the nominal value.

Cells connected in series, positive terminal of one to the negative of the next, add their voltages. A rack of the reference container is a series string of 396 cells, so its voltage is the sum

$$V_{\text{rack}} = V_1 + V_2 + \cdots + V_{396} = \sum_{i=1}^{396} V_i, \tag{2.1}$$

where V_i is the voltage of cell i. With every cell at its nominal value this gives

$$V_{\text{rack}} \approx 396 \times 3.2\,\text{V} = 1267.2\,\text{V} \approx 1267\,\text{V}.$$

The reference record shows rack voltages between about 1250 V and 1350 V, which brackets the nominal value: the cells sit above 3.2 V when the rack is full or being charged and below it when the rack is nearly empty or delivering a large current. Equation (2.1) is an identity of the circuit, not an approximation, and it will matter later that the record contains both sides of it: a rack voltage channel and 396 cell voltage channels.

2.1.2 Current

Current, measured in amperes, is the rate at which charge flows past a point of the circuit. If Q is the charge that has passed,

$$I = \frac{\mathrm{d}Q}{\mathrm{d}t}, \qquad 1\,\text{A} = 1 \text{ coulomb per second}. \tag{2.2}$$

Read backwards, equation (2.2) says that charge is accumulated current. With a constant current the charge moved in an interval Δt is simply

$$\Delta Q = I\,\Delta t, \tag{2.3}$$

so a current of 100 A flowing for one hour moves 100 Ah, and for two hours 200 Ah. When the current varies, the product is replaced by an integral,

$$\Delta Q = \int_{t_0}^{t_1} I(t)\,\mathrm{d}t, \tag{2.4}$$

which is the area under the current trace between t_0 and t_1. With I in amperes and t in hours the area is in ampere-hours; with t in seconds it is in coulombs, and 1 Ah = 3600 coulombs converts between the two.

Figure 2.1 draws both cases. In each panel the horizontal axis is time in hours and the vertical axis is current in amperes; the shaded region is the charge moved. Panel (a) is the constant 100 A of the example: the darker rectangle is the first hour (100 Ah) and the lighter one the second (200 Ah in total). Panel (b) is a current that swings between 60 A and 140 A; the shaded area is again 200 Ah, because the swings above and below 100 A cancel over two hours. The narrow dark strip in panel (b) is one term $I_k\,\Delta t$ of the sum that a digital system actually computes: a sensor reports the current at discrete instants, and the integral of equation (2.4) becomes

$$\Delta Q \approx \sum_k I_k\,\Delta t_k, \tag{2.5}$$

with I_k the k-th reading and Δt_k the interval it stands for. This integral is the reason current matters so much for state of charge: Section 2.4.1 shows that the state of charge a battery management system reports is equation (2.5) divided by a capacity.

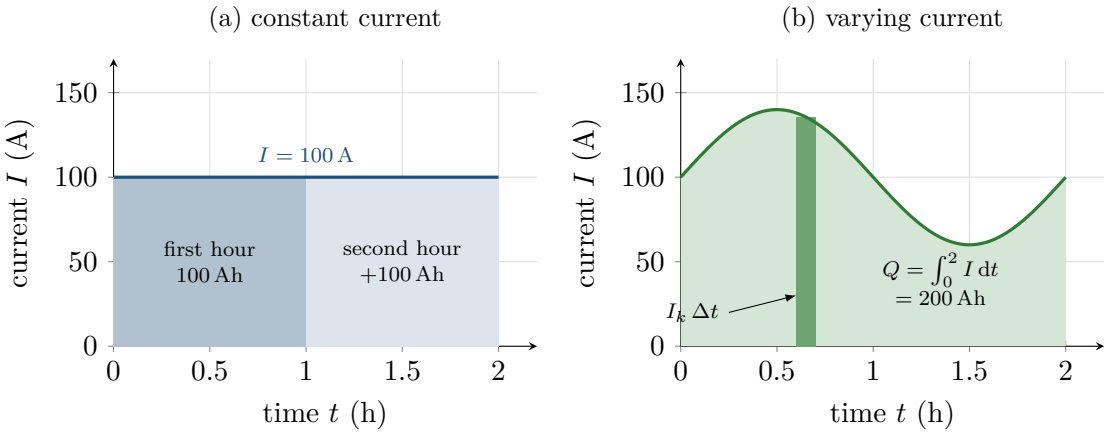


Figure 2.1: Charge moved as the area under a current trace, equations (2.3) and (2.4). (a) A constant 100 A moves 100 Ah in one hour and 200 Ah in two. (b) The current $I(t) = 100 + 40\sin(\pi t)$ A (t in hours, illustrative) moves the same 200 Ah in two hours; the dark strip is one term $I_k\,\Delta t$ of the sampled sum, equation (2.5).

2.2 Charge and energy

This section separates the two stored quantities that battery specifications both call "capacity": charge, in ampere-hours, and energy, in watt-hours, and shows how energy is obtained from voltage and current over time.

2.2.1 Ampere-hours as stored charge

A cell of the reference container is rated 300 Ah. The rating does not mean that the cell produces 300 A. It means that the cell holds enough charge to sustain a current–time product of about 300 Ah:

$$300\,\text{A} \times 1\,\text{h}, \qquad 150\,\text{A} \times 2\,\text{h}, \qquad 75\,\text{A} \times 4\,\text{h}$$

all move the same 300 Ah. This is equation (2.3) solved for the time, $\Delta t = Q/I$. The list ignores practical limits: the cell may not be permitted to carry 300 A, and the charge a real cell delivers depends somewhat on the current, the temperature and the cell's age. As a first statement, though, the ampere-hour rating measures stored charge capacity, written C in this book.

The same reading applies to a whole rack. The 396 cells of a rack are in series, so the same current passes through every one of them and every cell gives up the same charge. A rack

therefore has the capacity of one cell, $C \approx 300\,\text{Ah}$, not 396 times that. Putting cells in series raises the voltage; it does not raise the ampere-hours. Section 2.5.2 shows the opposite case, racks in parallel, which raises the ampere-hours and leaves the voltage unchanged.

2.2.2 Ampere-hours are not energy

A 300 Ah battery at 3.2 V and a 300 Ah battery at 1300 V hold very different amounts of energy. The reason is the definition of voltage as energy per unit charge: moving a charge Q across a voltage V transfers the energy

$$E = V\,Q, \qquad E[\text{Wh}] = V[\text{V}] \times C[\text{Ah}], \tag{2.6}$$

where the second form is the nominal energy content of a battery of capacity C at voltage V. Volts times ampere-hours are watt-hours, because a volt is a joule per coulomb and an ampere-hour is 3600 coulombs.

Table 2.1 applies equation (2.6) at the three levels of the reference container. Its columns are the level, its nominal voltage, its charge capacity, the product and the energy in the customary unit. A cell stores $3.2\,\text{V} \times 300\,\text{Ah} = 960\,\text{Wh} \approx 0.96\,\text{kWh}$. A rack has the same 300 Ah at 1267 V and stores $1267.2 \times 300 = 380{,}160\,\text{Wh} \approx 380\,\text{kWh}$, which is 396 times the cell's energy because energy, unlike charge, adds in series. Nine racks store $9 \times 380.16 \approx 3421\,\text{kWh} = 3.42\,\text{MWh}$, the documented energy of the reference container. Reading down the capacity column shows the point of the section: the charge capacity is 300 Ah for both the cell and the rack, while their energies differ by a factor of 396.

Table 2.1: Nominal charge and energy at the three levels of the reference container, equation (2.6), with a nominal cell voltage of 3.2 V.

Level	Voltage	Capacity	$V \times C$	Energy
Cell	3.2 V	300 Ah	960 Wh	0.96 kWh
Rack (396 cells in series)	1267 V	300 Ah	380,160 Wh	380 kWh
Container (9 racks in parallel)	1267 V	2700 Ah	3,421,440 Wh	3.42 MWh

Principle 2.1. *Ampere-hours measure charge, not energy. The energy a battery stores is its charge capacity multiplied by its voltage, so two batteries with the same ampere-hour rating can differ in energy by any factor.*

2.2.3 Energy as the integral of power

Equation (2.6) assumes a fixed voltage, and a battery's voltage is not fixed. The general statement follows from the same definition applied to a small amount of charge: moving $\mathrm{d}Q$ across the voltage V transfers $\mathrm{d}E = V\,\mathrm{d}Q$, and since $\mathrm{d}Q = I\,\mathrm{d}t$ by equation (2.2),

$$\mathrm{d}E = V(t)\,I(t)\,\mathrm{d}t \qquad \Longrightarrow \qquad E = \int_{t_0}^{t_1} P(t)\,\mathrm{d}t = \int_{t_0}^{t_1} V(t)\,I(t)\,\mathrm{d}t, \tag{2.7}$$

where $P = VI$ is the power, the subject of Section 2.3.1. When the power is constant the integral reduces to $E = P\,\Delta t$. In sampled form, with the voltage and current read at the same instants, $E \approx \sum_k V_k I_k\,\Delta t_k$.

Because voltage and current both vary during operation, the energy a battery actually delivers comes from integrating their product, and the nominal $V\,C$ is only an approximation to

it. An illustrative case shows the size of the difference. A rack discharged at a constant 100 A for two hours moves 200 Ah. If its voltage falls linearly from 1320 V to 1260 V over those two hours, the average voltage is 1290 V and equation (2.7) gives

$$E = 100\,\text{A} \times 1290\,\text{V} \times 2\ \text{h} = 258{,}000\,\text{Wh} = 258\,\text{kWh},$$

whereas multiplying the 200 Ah by the starting voltage would claim 264 kWh. The 6 kWh difference is the cost of treating a varying voltage as a constant; it is the reason a nominal energy figure depends on which voltage is chosen, a point that returns in Section 2.5.2.

2.3 Power and its relation to energy

This section defines power, computes it for one rack and for the reference container, and separates the power a battery can deliver from the energy it holds, the distinction a two-number rating such as "1 MW / 4 MWh" expresses.

2.3.1 Power

Power is the product of voltage and current,

$$P = V\,I, \tag{2.8}$$

measured in watts. It says how fast energy is being transferred at this moment; energy says how much has been transferred in total, and equation (2.7) links the two. The units check: a volt is a joule per coulomb and an ampere is a coulomb per second, so their product is a joule per second, a watt.

A rack at 1300 V carrying 100 A transfers

$$P = 1300\,\text{V} \times 100\,\text{A} = 130{,}000\,\text{W} = 130\,\text{kW}.$$

The nine racks of the reference container are in parallel on one bus at the same voltage. If each carries 100 A, the bus current is the sum, $I_{\text{bus}} = 9 \times 100 = 900$ A, and the container transfers

$$P \approx 1300\,\text{V} \times 900\,\text{A} = 1{,}170{,}000\,\text{W} \approx 1.17\,\text{MW}.$$

Equivalently, it is nine times the 130 kW of one rack. Power, like energy, adds across parallel racks, because the currents add and the voltage is shared.

2.3.2 Power and energy are different quantities

Power and energy are easily confused because both are spoken of as the "size" of a battery. A water tank separates them. Figure 2.2 draws two tanks holding the same volume of water; the shaded region is the water and the pipe at the bottom right is the outlet, drawn narrow on the left and wide on the right. The water in the tank stands for energy, how much is stored; the rate at which water leaves through the outlet stands for power, how fast it is being delivered. The width of the outlet is the power capability, which in a battery is set by the power conversion equipment and by the current the cells may carry. Both tanks hold the same amount; the wide outlet empties its tank four times sooner.

The analogy translates into one formula. A battery holding energy E and delivering a constant power P runs for

$$t = \frac{E}{P}. \tag{2.9}$$

A 1 MW / 4 MWh battery delivers full power for $t = 4/1 = 4$ h, and a 4 MW / 4 MWh battery for $t = 4/4 = 1$ h. The two store the same energy and have very different power capabilities;

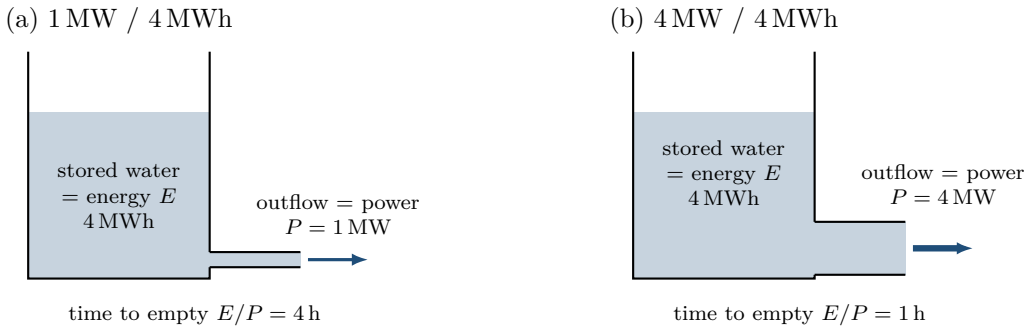


Figure 2.2: The water-tank analogy for power and energy. Both tanks hold the same volume (energy, 4 MWh); the outlet width is the power capability. The narrow outlet (a) delivers 1 MW and empties the tank in four hours; the wide outlet (b) delivers 4 MW and empties it in one.

they are different machines for different jobs, the first for shifting energy over an evening and the second for delivering a large amount quickly. Figure 2.3 draws the same two cases as rectangles. In each panel the horizontal axis is time in hours and the vertical axis is power in megawatts, so the area of each rectangle, power times duration, is the energy. Both rectangles have area 4 MWh; one is low and long, the other tall and short.

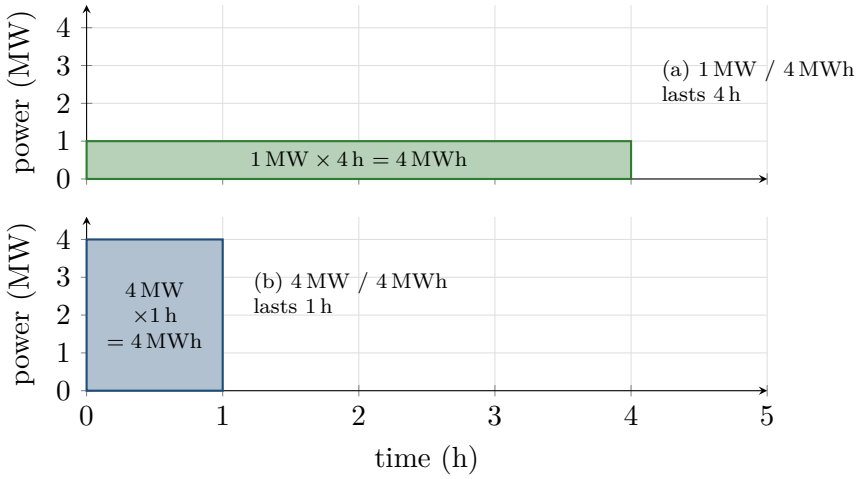


Figure 2.3: The same stored energy at two power levels, equation (2.9). Each rectangle has height equal to the delivered power and width equal to the duration, so its area is the energy, 4 MWh in both panels.

Principle 2.2. *A power rating says how fast a battery can deliver energy; an energy rating says how much it holds. Their ratio, E/P, is the duration at full power, and neither number describes the battery without the other.*

2.3.3 A discharge at 195 kW

Take one rack of the reference container at 1300 V discharging at 150 A. Its power is

$$P = 1300\,\text{V} \times 150\,\text{A} = 195{,}000\,\text{W} = 195\,\text{kW},$$

and if both values held for one hour, equation (2.7) with a constant integrand gives

$$E = 195\,\text{kW} \times 1\ \text{h} = 195\,\text{kWh},$$

about half of the rack's nominal 380 kWh ($195/380 \approx 0.51$).

The charge view gives the same answer from the other side: 150 A for one hour removes 150 Ah, exactly half of the rack's 300 Ah. The energy fraction comes out slightly above one half only because the example uses 1300 V while the nominal 380 kWh was computed at 1267 V; with a common voltage the two fractions agree. In practice neither number would hold for the full hour: as charge leaves, the rack voltage drifts down, so the power at constant current falls and the energy delivered over the hour is a little less than 195 kWh, by the mechanism of the illustrative example of Section 2.2.3.

2.4 State of charge and coulomb counting

This section defines state of charge, derives how it changes with current, shows that the value a battery management system reports is a computation that depends on an assumed capacity, and fixes the sign convention under which the computation is written.

2.4.1 State of charge

The state of charge, SOC, is the fraction of a battery's available charge capacity that remains in it,

$$\mathrm{SOC} = \frac{\text{charge remaining}}{\text{available charge capacity}} = \frac{Q_{\mathrm{rem}}}{C}, \tag{2.10}$$

expressed either as a fraction between 0 and 1 or as a percentage. For a cell or rack with $C = 300$ Ah, a state of charge of 100 % corresponds to 300 Ah remaining, 50 % to 150 Ah and 20 % to 60 Ah. The state of charge is the battery's fuel gauge: it says how much of the charge the battery could deliver is still there.

The definition is conceptual, and the word "available" carries a good deal. The charge a real battery can deliver depends on its temperature, on its degradation (an aged cell holds less), on the operating limits the battery management system enforces at the top and bottom of the range, and on the cell chemistry. A state of charge is therefore always relative to some value of C, and two systems that assume different values of C will report different states of charge for the same battery. The next three subsections make that dependence explicit.

2.4.2 Coulomb counting

Take a rack at $\mathrm{SOC}_0 = 80\,\%$ with $C = 300$ Ah that discharges at 60 A for one hour. The charge removed is $\Delta Q = 60\ \mathrm{A} \times 1\ \mathrm{h} = 60$ Ah, which is $60/300 = 0.20$ of the capacity, so the state of charge falls to about $80\,\% - 20\,\% = 60\,\%$. Written generally, the charge remaining falls by the integral of the current, and dividing by C gives

$$\mathrm{SOC}(t) = \mathrm{SOC}(t_0) - \frac{1}{C}\int_{t_0}^{t} I(\tau)\,\mathrm{d}\tau, \tag{2.11}$$

with I in amperes, τ in hours and C in ampere-hours. This is *coulomb counting*: the state of charge is updated by counting the coulombs that cross the terminals. A battery management system (BMS) evaluates it from sampled current with equation (2.5),

$$\mathrm{SOC}_k = \mathrm{SOC}_{k-1} - \frac{I_k\,\Delta t_k}{3600\,C}, \tag{2.12}$$

with Δt_k in seconds. At the sampling interval of the reference record's current channel, about 1.03 s, a current of 120 A moves $120 \times 1.03/3600 \approx 0.034$ Ah per sample, a change in state of charge of about 0.011 percentage points.

Under a constant current the integral is $I\,(t - t_0)$ and the state of charge is a straight line in time,

$$\mathrm{SOC}(t) = \mathrm{SOC}_0 - \frac{I}{C}\,(t - t_0), \tag{2.13}$$

with slope $-I/C$. Figure 2.4 plots it for the 300 Ah rack. The horizontal axis is time in hours and the vertical axis is state of charge in percent. The green line is the discharge of the example, $I = +60$ A from 80 %, which falls by 20 percentage points per hour and passes through 60 % at one hour (the marked point). The blue line is a charge at $I = -60$ A from 20 %, which rises at the same rate. The two lines are mirror images: the magnitude of the current sets the steepness, its sign the direction.

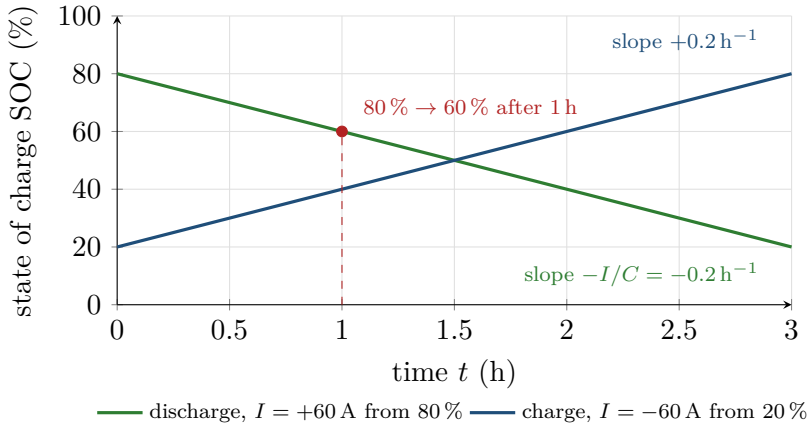


Figure 2.4: State of charge under constant current, equation (2.13), for $C = 300$ Ah. Discharging at $+60$ A from $\mathrm{SOC}_0 = 80$ % (green) and charging at -60 A from $\mathrm{SOC}_0 = 20$ % (blue) both change the state of charge by 20 percentage points per hour; the marked point is the worked example of Section 2.4.2.

2.4.3 Dependence on the assumed capacity

Equation (2.11) needs two inputs besides the current: the starting value $\mathrm{SOC}(t_0)$ and the capacity C. The BMS does not measure C; it assumes it. If aging has reduced the true capacity of a rack to 270 Ah while the BMS still divides by 300 Ah, the estimate drifts away from the truth. An illustrative calculation shows how fast. Start both the rack and the BMS at 80 % and remove 60 Ah. The BMS reports $80 - 60/300 \times 100 = 60$ %. The true charge remaining is $0.80 \times 270 - 60 = 156$ Ah, which is $156/270 = 57.8$ % of the true capacity. After one hour the estimate is 2.2 percentage points high, and the error grows with every ampere-hour counted: removing the full 216 Ah that the rack actually held empties it while the BMS still reports $80 - 216/300 \times 100 = 8$ %. Nothing in the count itself reveals the problem, because the count contains no information except the current.

Figure 2.5 draws the computation as it is performed. The boxes are operations, read left to right: the rack, a current sensor that produces the measured current $I(t)$, an integrator, a division by the assumed capacity, and a subtraction from the starting value, producing the estimate $\widehat{\mathrm{SOC}}(t)$ (the hat marks an estimate). The two inputs entering from below, the assumed capacity and the initial value, are the only places where anything other than the current enters. The dashed box at the top is the reading the chain is sometimes mistaken for: a direct instrument from battery to state of charge. No such instrument exists. A diagram labeled "current sensor → SOC" means measured $I(t)$, then integration, then division by C, then an estimated state of charge; it does not mean a battery connected to a state-of-charge sensor.

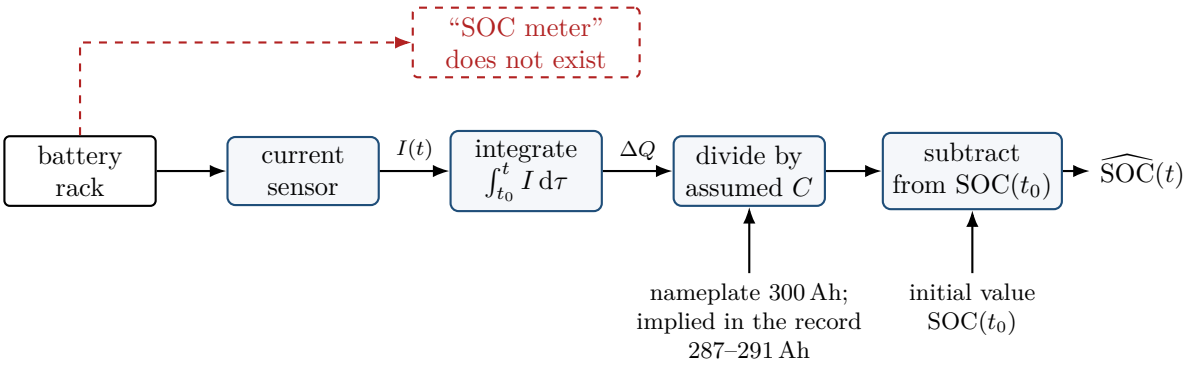


Figure 2.5: The coulomb-counting chain, equation (2.11). The only measured input is the current; the capacity and the initial value are assumptions. The dashed box is the instrument the chain is sometimes mistaken for; there is no sensor that reads state of charge directly.

The reference record shows this structure directly. An analysis of the record finds that the reported state of charge behaves like a coulomb count on the same current telemetry, with an implied capacity of about 287 to 291 Ah. The implied capacity is found by inverting equation (2.11) over intervals in which the reported value changes,

$$C_{\text{implied}} = -\frac{\int_{t_0}^{t_1} I\,\mathrm{d}t}{\mathrm{SOC}(t_1) - \mathrm{SOC}(t_0)}, \tag{2.14}$$

with the current integrated from the record's own current channel. For illustration, a charging interval whose integrated current is -43.35 Ah and over which the reported state of charge rises by 15 percentage points gives $C_{\text{implied}} = 43.35/0.15 = 289$ Ah, the middle of the range. The implied value is a property of the BMS's calculation, the capacity it divides by. Whether it equals the rack's true present capacity cannot be read from the reported state of charge, because that channel was computed from the current and carries no information the current does not.

Principle 2.3. *State of charge is inferred, not sensed. A reported state of charge is the integral of measured current divided by an assumed capacity, and it inherits every error of the current sensor, of the starting value and of the assumed capacity.*

2.4.4 The sign of the current

Charging reverses the direction of the current, and equation (2.11) needs a convention for which direction is positive. This book uses

$$I > 0 \text{ on discharge}, \qquad I < 0 \text{ on charge}, \tag{2.15}$$

the convention of the reference record. A current of $+100$ A means charge leaving the battery at 100 coulombs per second; -100 A means charge entering at the same rate. With this sign the minus in equation (2.11) makes the state of charge fall on discharge and rise on charge, as in Figure 2.4.

The convention carries through to power. Because $P = VI$ and the voltage of a battery is always positive, $P > 0$ is power delivered by the battery and $P < 0$ is power absorbed by it. An example operation taken from the reference record has a negative current because the rack is charging, and the worked reading of Section 2.6.3 interprets its sign. Other systems use the opposite convention (positive on charge), and the first thing to establish about any battery dataset is which one it follows: a sign error turns every charge into a discharge and makes a coulomb count run backwards.

2.5 C-rate, racks in parallel and the five quantities

This section normalizes current by capacity, assembles the nine racks of the reference container in parallel, explains why their currents differ, and collects the chapter's quantities and relations in one place.

2.5.1 C-rate

The C-rate expresses a current relative to the battery's capacity,

$$\text{C-rate} = \frac{I}{C}, \tag{2.16}$$

with I in amperes and C in ampere-hours, so the unit is h^{-1}. For a 300 Ah cell, $150\,\mathrm{A}/300\,\mathrm{Ah} = 0.5\,\mathrm{h}^{-1}$, spoken as "0.5C". Its reciprocal is the time a full battery would last at that current, ignoring practical limits: at 1C a full battery empties in about one hour. For the 300 Ah cell the scale is

$$0.25\mathrm{C} = 75\,\mathrm{A}\ (4\,\mathrm{h}), \quad 0.5\mathrm{C} = 150\,\mathrm{A}\ (2\,\mathrm{h}), \quad 1\mathrm{C} = 300\,\mathrm{A}\ (1\,\mathrm{h}), \quad 2\mathrm{C} = 600\,\mathrm{A}\ (0.5\,\mathrm{h}).$$

Figure 2.6 plots this scale. The horizontal axis is the C-rate and the vertical axis is the time to empty a full battery, 1/C-rate, in hours; the curve is a hyperbola, and the four marked points are the four rates above, each labeled with its current for a 300 Ah cell. Doubling the C-rate halves the duration, so the curve is steep at low rates: going from 0.25C to 0.5C saves two hours, going from 1C to 2C saves half an hour.

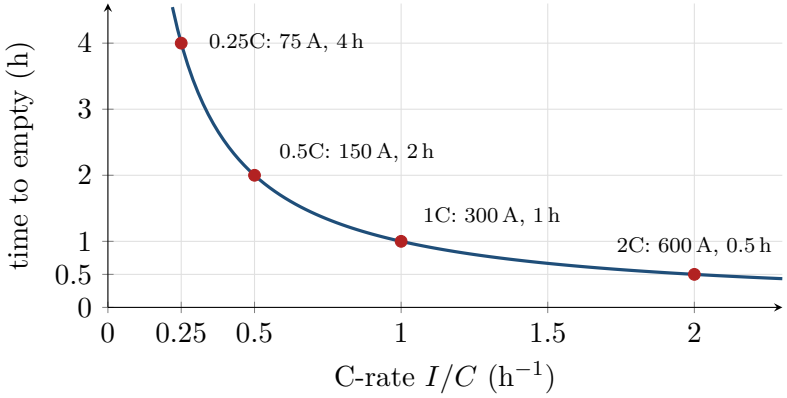


Figure 2.6: The C-rate scale. The curve is the time to empty a full battery, $t = 1/(\text{C-rate})$, which ignores practical limits; the labeled currents are for a 300 Ah cell, $I = \text{C-rate} \times 300\,\mathrm{Ah}$.

Why the C-rate is useful. A statement such as "charging at 150 A" says nothing about how hard a battery is being worked until its capacity is known. For a 300 Ah cell it is 0.5C, a moderate rate; for a 50 Ah cell it is $150/50 = 3$C, a fast rate that empties or fills the cell in twenty minutes. The C-rate normalizes current by battery size, so that heating, voltage drop and wear can be compared across cells of different capacity on one scale.

2.5.2 Nine racks in parallel

Each rack of the reference container has a capacity of about 300 Ah. Nine racks connected in parallel to one bus share its voltage and add their currents, so the container has

$$9 \times 300\,\mathrm{Ah} = 2700\,\mathrm{Ah}$$

at about 1300 V. Equation (2.6) then gives $1300\,\text{V} \times 2700\,\text{Ah} = 3{,}510{,}000\,\text{Wh} \approx 3.51\,\text{MWh}$. The documented figure for the container is about 3.42 MWh, and the difference is entirely the choice of voltage: at the nominal 1267.2 V of Section 2.1.1, $1267.2 \times 2700 = 3{,}421{,}440\,\text{Wh} \approx 3.42\,\text{MWh}$, the last row of Table 2.1. A nominal energy is only as definite as the voltage chosen to compute it, which is the point of Section 2.2.3 in another form.

The structure of the reference container is now complete in electrical terms. Putting 396 cells in series produces a high voltage, about 1267 V, at the capacity of one cell. Putting nine racks in parallel produces a high current capability and a large ampere-hour capacity, 2700 Ah, at the voltage of one rack. Series raises voltage; parallel raises current and charge capacity; both raise energy.

2.5.3 Unequal currents in parallel racks

Racks in parallel share a common bus voltage, but their currents need not match. Three racks may carry, at the same instant, $I_1 = 118\,\text{A}$, $I_2 = 123\,\text{A}$ and $I_3 = 115\,\text{A}$ (illustrative values, with a spread of a few percent). The parallel connection forces approximately equal voltages at the rack terminals; the current then distributes according to the characteristics of each branch. A rack whose internal voltage is a little higher, or whose resistance is a little lower, carries a little more of the load.

Figure 2.7 draws the arrangement for three racks. The two horizontal lines are the positive and negative conductors of the common DC bus; each dashed box is a rack, drawn as a source (the two plates, long plate positive) in series with a lumped resistance $R + R_{\text{series}}$ that combines the cells' internal resistance with the connection hardware. Every rack sees the same bus voltage V_{bus} between the lines, and each carries its own current upward into the bus, where the currents add to $I_{\text{bus}} = 118 + 123 + 115 = 356\,\text{A}$ toward the power conversion system. The differences between the branches come from five sources: internal resistance, state of charge (which sets the source voltage), temperature (which changes both), connection resistance, and cell-to-cell differences in the characteristics of the 396 cells of each string.

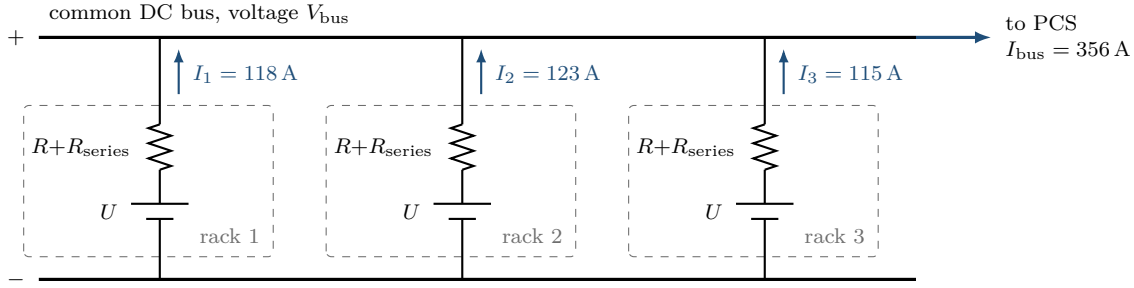


Figure 2.7: Unequal currents in parallel racks on a common bus. All three racks see the same bus voltage V_{bus}; each branch is a source U in series with a lumped resistance $R + R_{\text{series}}$ whose values differ slightly from rack to rack, so the currents split unequally. The currents are illustrative; the branch values are not shown because they are not measured.

This matters later, when the question becomes whether an unusual rack current indicates a problem. A rack that carries 123 A while its neighbors carry 115 to 118 A is not by that fact faulty: a spread of a few percent is the normal consequence of small differences in the branches. Deciding when a difference stops being normal requires a model of how much spread the branch characteristics can produce, which Chapters 4 and 8 develop.

2.5.4 Five quantities and four relations

The chapter has now defined five quantities. Table 2.2 lists them with their symbols, units and physical meanings: voltage, the potential difference; current, the rate of charge flow; power, the rate of energy transfer; charge capacity, in ampere-hours; and energy capacity, in kilowatt-hours or megawatt-hours. The first three are instantaneous quantities at the terminals; the last two are amounts stored.

Table 2.2: The five electrical quantities of this chapter.

Quantity	Symbol	Unit	Meaning
Voltage	V	V	potential difference; energy per unit charge
Current	I	A	rate of charge flow; positive on discharge
Power	P	W, kW, MW	rate of energy transfer at this moment
Charge capacity	C	Ah	charge a full battery can deliver
Energy capacity	E	kWh, MWh	energy a full battery can deliver

Four relations connect them:

$$P = V\,I, \qquad E = \int P\,\mathrm{d}t, \qquad Q = \int I\,\mathrm{d}t, \qquad E \approx V\,C \ \text{(nominal)}. \tag{2.17}$$

The first is instantaneous, the middle two accumulate over time, and the last is the approximation that holds when the voltage is treated as constant. The state of charge of equation (2.11) and the C-rate of equation (2.16) are both ratios built on these: charge accumulated divided by C, and current divided by C.

2.6 Worked examples and the reference record

This section applies the five quantities to one rack in operation, reads the channels of the reference record in their terms, interprets one recorded instant, and solves the first forecasting problem a battery poses: when a target state of charge will be reached.

2.6.1 A rack discharging at 120 A

Example 2.1. A rack has $V = 1280\,\mathrm{V}$, $C = 300\,\mathrm{Ah}$ and $\mathrm{SOC} = 80\,\%$, and discharges at $120\,\mathrm{A}$. Its power is

$$P = 1280\,\mathrm{V} \times 120\,\mathrm{A} = 153{,}600\,\mathrm{W} = 153.6\,\mathrm{kW},$$

and its C-rate is $120/300 = 0.4\mathrm{C}$. After one hour it has delivered $120\,\mathrm{Ah}$, which is $120/300 = 40\,\%$ of the nominal capacity, so its state of charge has fallen from $80\,\%$ to $40\,\%$. If the voltage held at $1280\,\mathrm{V}$, the energy delivered would be

$$E \approx 153.6\,\mathrm{kW} \times 1\ \mathrm{h} = 153.6\,\mathrm{kWh}.$$

Every relation of equation (2.17) appears in Example 2.1: power from voltage and current, charge from current and time, state of charge from charge and capacity, energy from power and time. The one approximation is in the last line. Over an hour in which $40\,\%$ of the charge leaves, the rack voltage drifts downward, so the true energy is the integral of equation (2.7) and is somewhat below $153.6\,\mathrm{kWh}$; the charge, $120\,\mathrm{Ah}$, is exact for a constant current regardless of the voltage.

The example also shows that the charge and energy bookkeeping run in parallel at different levels of certainty. The charge removed follows from the current alone. The energy removed needs the voltage too. The state of charge needs the capacity, which is assumed. Each additional quantity brings an additional source of error.

2.6.2 The channels of the reference record

The reference record contains five kinds of channel per rack. Table 2.3 lists them with their symbols, their approximate sampling intervals, their physical reading, and whether each is measured directly or derived from other measurements. The rack current I_r arrives about every 1.03 s and the rack terminal voltage V_r at a similar rate; the reported state of charge S_r about every 78 s; the 396 cell voltages V_i about once per minute; and the 216 temperatures T_i about every 213 s.

Table 2.3: The channels of the reference record, per rack.

Channel	Symbol	Interval	Physical reading	Kind
Rack current	I_r	≈ 1.03 s	how much charge is moving, and in which direction	measured
Rack terminal voltage	V_r	≈ 1 s	potential across the rack	measured
Reported state of charge	S_r	≈ 78 s	the BMS estimate of charge remaining	derived from I_r
Cell voltages (396)	V_i	≈ 60 s	individual cell voltages	measured
Temperatures (216)	T_i	≈ 213 s	thermal state at many locations	measured

Read physically, the current says how much charge is moving, the voltage gives the potential across the rack, the reported state of charge is the BMS's estimate of the charge remaining, the cell voltages resolve the rack voltage into its 396 parts, and the temperatures give the thermal state at many points. The channels are not equally independent. Four are direct readings of a sensor. The fifth, S_r, is computed: Section 2.4.3 established that it behaves like a coulomb count of I_r. It therefore adds an estimate of the starting value and of the assumed capacity, but no new information about the charge that moved. Between two reports, about 78 s apart, a current of 120 A moves $120 \times 78/3600 = 2.6$ Ah, so each report steps the state of charge by about 0.87 percentage points at that current. There is also a redundancy among the measured channels: equation (2.1) says V_r should equal the sum of the 396 V_i, so the rack voltage and the cell voltages measure the same quantity twice.

Which channels are independent is the question on which every later attempt to infer the health of the battery turns. A model that treats S_r as an independent check on I_r will find perfect agreement and learn nothing from it; a model that uses V_r and the V_i together gains a genuine consistency test. Chapter 9 returns to this distinction in full.

> **Principle 2.4.** *A telemetry record mixes measurements and derived values. Before a channel is used as evidence, it must be established whether it was read from a sensor or computed from other channels, because a derived channel cannot test the channels it was derived from.*

2.6.3 Reading one instant

Suppose one instant of the reference record shows a rack with $I = -120$ A, $V = 1300$ V and SOC = 65 %. Under the sign convention of equation (2.15), the negative current means the rack is charging. The power is

$$P = V\,I = 1300\,\mathrm{V} \times (-120\,\mathrm{A}) = -156{,}000\,\mathrm{W} = -156\,\mathrm{kW},$$

and the negative sign means power flowing into the battery: about 156 kW is being absorbed. The 65 % is the BMS's estimate of the charge remaining, computed as in Figure 2.5.

The three numbers have different standing. The current and voltage are sensor readings at that instant. The power is computed from them by equation (2.8) and is exact to the extent that they are. The state of charge is the running result of every current reading since the BMS last set its starting value, divided by its assumed capacity. The charging rate is $120/300 = 0.4$C, and that rate, with the state of charge, is enough to say how long the rack will take to reach a given level, the problem of the next subsection.

2.6.4 Time to reach a target state of charge

Under a constant current, equation (2.13) with $t_0 = 0$ is $\mathrm{SOC}(t) = \mathrm{SOC}_0 - (I/C)\,t$. Setting $\mathrm{SOC}(t) = \mathrm{SOC}^*$, a target, and solving for t gives

$$t = \frac{C\,(\mathrm{SOC}_0 - \mathrm{SOC}^*)}{I}. \tag{2.18}$$

The sign convention makes the formula work in both directions. On charge, $I < 0$ and the target lies above the start, $\mathrm{SOC}^* > \mathrm{SOC}_0$, so numerator and denominator are both negative and t is positive. On discharge both are positive. If the signs disagree (a discharging rack and a target above the start), t comes out negative, which says correctly that the target is never reached under that current.

Example 2.2. The rack of Section 2.6.3 charges at $I = -120$ A from $\mathrm{SOC}_0 = 65\,\%$, with $C = 300$ Ah. To reach $\mathrm{SOC}^* = 80\,\%$ the state of charge must rise by 0.15, which is $0.15 \times 300 = 45$ Ah of charge. Equation (2.18) gives

$$t = \frac{300 \times (0.65 - 0.80)}{-120}\ \mathrm{h} = \frac{-45}{-120}\ \mathrm{h} = 0.375\ \mathrm{h} = 22.5\ \mathrm{min}.$$

A shortcut gives the same result. The rate of change is $|I|/C = 120/300 = 0.4$ per hour, which is 40 percentage points per hour, or one percentage point every 1.5 minutes. Fifteen points therefore take $15 \times 1.5 = 22.5$ minutes.

The shortcut gives every other target at once. Table 2.4 lists the targets 70, 80, 90 and 100 % with the percentage points to be gained, the charge required and the time at 1.5 minutes per point. Figure 2.8 draws the same calculation: the horizontal axis is time in minutes, the vertical axis is state of charge, the line is the constant-current ramp from 65 %, and each dashed corner marks a target on the vertical axis and the time it is reached on the horizontal axis. The equal spacing of the last three times, 15 minutes per 10 points, is the linearity of equation (2.13).

Table 2.4: Time for the rack of Example 2.2 to reach each target from 65 % at a constant -120 A, $C = 300$ Ah.

Target SOC*	Points to gain	Charge required	Time
70 %	5	15 Ah	7.5 min
80 %	15	45 Ah	22.5 min
90 %	25	75 Ah	37.5 min
100 %	35	105 Ah	52.5 min

Two caveats qualify the table. The first is the constant current. A real charging current does not stay constant all the way to 100 %: as the cells approach their upper voltage limit, the charger reduces the current to keep the voltage in bounds, so the last part of the ramp bends

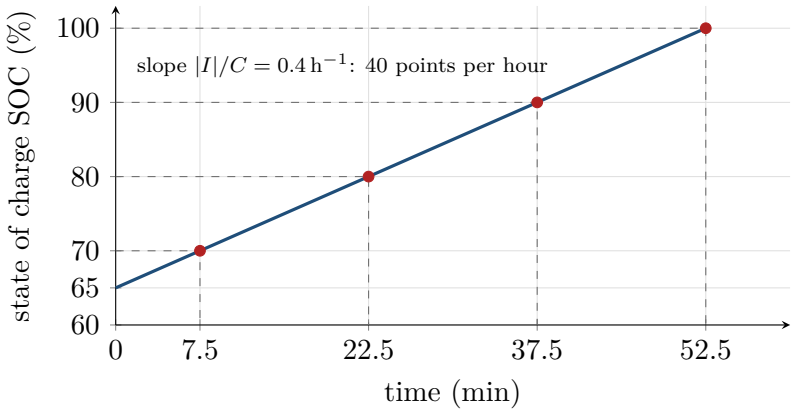


Figure 2.8: The state-of-charge ramp from 65 % under a constant charging current, $\mathrm{SOC}(t) = 0.65 + (|I|/C)\,t$ with $|I| = 120\,\mathrm{A}$ and $C = 300\,\mathrm{Ah}$, equation (2.13). The marked points are the targets 70, 80, 90 and 100 %, reached at 7.5, 22.5, 37.5 and 52.5 minutes.

over and takes longer. The 52.5 minutes to full is therefore a lower bound, and the times to 70 and 80 %, well below the upper limit, are the reliable ones.

The second caveat is the capacity. The calculation divides by the nameplate 300 Ah, while the BMS of the reference record behaves as if the capacity were about 289 Ah. With that value the time from 65 to 80 % at 120 A is

$$t = \frac{289 \times 0.15}{120}\ \mathrm{h} = 0.361\ \mathrm{h} \approx 21.7\ \mathrm{min},$$

rather than 22.5; across the implied range of 287 to 291 Ah it is 21.5 to 21.8 minutes. The reported state of charge will reach 80 % about 0.8 minutes sooner than the nameplate calculation predicts, because it is counting against a smaller denominator.

This is a first, small battery forecasting problem: given the starting state of charge SOC_0, the current $I(t)$ over the coming interval and the capacity C, predict when the target SOC^* is reached. Each of the three inputs is uncertain: the starting value is itself an estimate, the future current is a forecast of how the system will be operated, and the capacity is an assumption. The answer inherits all three uncertainties, and the two caveats above are the first two of them made concrete.

2.7 What to retain

Definition 2.1 (The electrical description of a battery)**.** A battery is described at its terminals by voltage, current and power, $P = VI$. Charge and energy are the time integrals of current and power, and state of charge is integrated current divided by an assumed capacity: a computed estimate, never a direct measurement.

Table 2.5 lists the objects the chapter introduced, each with its defining relation and the subsection where it is defined.

2.8 Exercises

Exercise 2.1. A rack of 396 cells has a nominal cell voltage of 3.25 V and a capacity of 300 Ah. (a) Compute the nominal rack voltage and the nominal rack energy. (b) The rack discharges at 90 A at that voltage: compute the power and the C-rate. (c) Compute the time to go from 90 % to 20 % at that constant current, and the energy delivered if the voltage stayed constant. (Answers: 1287 V and 386.1 kWh; 115.8 kW and 0.3C; 2.33 h and about 270 kWh.)

Table 2.5: The objects of this chapter.

Object	Definition	Subsection
Voltage V, series sum	energy per unit charge; $V_{\mathrm{rack}} = \sum_i V_i$	2.1.1
Current I	$I = \mathrm{d}Q/\mathrm{d}t$; charge is the area under the current trace	2.1.2
Charge capacity C	charge a full battery can deliver, in Ah; unchanged by series connection	2.2.1
Energy E	$E = V\,C$ nominally; Ah are not energy	2.2.2
Energy integral	$E = \int V I \,\mathrm{d}t$; depends on the varying voltage	2.2.3
Power P	$P = V I$, rate of energy transfer	2.3.1
Duration E/P	time at full power; 1 MW/4 MWh lasts 4 h, 4 MW/4 MWh lasts 1 h	2.3.2
State of charge SOC	charge remaining divided by available capacity	2.4.1
Coulomb counting	$\mathrm{SOC}(t) = \mathrm{SOC}(t_0) - C^{-1} \int I \,\mathrm{d}\tau$	2.4.2
Implied capacity	the C a reported SOC divides by; 287–291 Ah in the reference record	2.4.3
Sign convention	$I > 0$ discharge, $I < 0$ charge; $P < 0$ is power into the battery	2.4.4
C-rate	I/C in h^{-1}; normalizes current by battery size	2.5.1
Parallel racks	capacities and currents add, voltage shared; 2700 Ah, 3.42 MWh	2.5.2
Unequal rack currents	set by branch characteristics under a common V_{bus}	2.5.3
Measured and derived channels	S_r is derived from I_r; V_r and the V_i are redundant	2.6.2
Time to target	$t = C(\mathrm{SOC}_0 - \mathrm{SOC}^*)/I$ under constant current	2.6.4

Exercise 2.2. A BMS assumes $C = 300\,\text{Ah}$ for a rack whose true capacity has faded to 280 Ah. Both the rack and the BMS start at 90 %, and the rack discharges at 140 A for 1.5 hours. (a) Compute the reported and the true state of charge at the end. (b) How much charge does the BMS believe remains, and how much actually remains? (c) At what reported state of charge is the rack actually empty? (d) Over the same discharge, use equation (2.14) to show what capacity an analyst would infer from the reported state of charge and the current, and explain why it is 300 Ah and not 280 Ah.

Exercise 2.3. Nine racks on a common bus at 1310 V carry, at one instant, 118, 123, 115, 120, 117, 121, 119, 116 and 122 A. (a) Compute the bus current, the container power and the mean rack current. (b) Compute the C-rate of the most heavily loaded rack, taking $C = 300\,\text{Ah}$. (c) The spread of the currents is 8 A. Express it as a percentage of the mean, and list the branch characteristics of Section 2.5.3 that could produce it without a fault.

Exercise 2.4. A rack reads $V = 1295\,\text{V}$, $I = -90\,\text{A}$ and $\text{SOC} = 40\,\%$. (a) State whether it is charging or discharging and compute its power with the sign convention of this chapter. (b) With $C = 300\,\text{Ah}$, compute the time to reach 85 % at constant current, and the energy absorbed over that time if the voltage stayed constant. (c) Repeat the time with the implied capacity of 289 Ah. (d) Compute the time to reach 100 % and explain why it is a lower bound.

2.9 Connection to the TERRA framework

The quantities of this chapter are the variables of a conservation graph for a battery, and the chapter's distinctions (measured versus computed, charge versus energy) are the ones such a model makes explicit.

- **Coulomb counting is the charge balance on a rack node.** Equation (2.11) is charge conservation for a rack: the stored charge changes by exactly the charge that crosses the terminals, $\mathrm{d}Q_{\text{rem}}/\mathrm{d}t = -I$. In TERRA the rack is a node with a stored charge as its state, the current is the flow on its edge, and the state of charge is that stored charge divided by the parameter C.
- **The bus is a node where currents add.** The nine racks of Section 2.5.2 are nine edges into one bus node at the common potential V_{bus}, and conservation of charge at that node is $I_{\text{bus}} = \sum_r I_r$. The unequal split of Figure 2.7 is not part of the balance; it is set by each rack's component relation, the branch characteristics that later chapters model.
- **The series string is one edge with a common flow.** The 396 cells of a rack carry the same current, so the string is a single edge whose flow is shared by 396 components, and equation (2.1), $V_{\text{rack}} = \sum_i V_i$, is a redundancy relation between the rack voltage channel and the 396 cell voltage channels: two measurements of one quantity, and therefore a test.
- **The reported state of charge is a derived channel.** The record's S_r is computed from I_r by the chain of Figure 2.5. Across the rack's terminals the only independent measurement is the current, so a model that checks S_r against I_r obtains an identity, not a test. TERRA asks first whether the measurements can represent the balance, and for the charge balance alone the answer is that they represent it exactly and only once.
- **Capacity is not identifiable from current and reported state of charge.** The capacity enters equation (2.11) only as the divisor of the integrated current. From I_r and S_r alone the posterior over the true capacity is as wide as its prior: the data determine the BMS's assumed value, the 287 to 291 Ah of equation (2.14), and nothing about the rack's actual one. Information about the true capacity has to come from a channel the BMS did not compute, which is the role the voltage will play.

- **Power and energy are derived edge quantities.** $P = VI$ combines two measured channels on the same edge, and energy is its time integral. Both are exact functions of the measurements, so they add no information of their own; their uncertainty is that of the voltage and current they are computed from.

In summary, a battery at its terminals is a node with a stored charge, an edge carrying a current, and a voltage across that edge. Charge and energy are integrals of the edge flow and of its product with the voltage, and the state of charge is the stored charge divided by a capacity that the measurements alone do not determine. What the reference record measures directly is current, voltage and temperature; everything else in this chapter is computed from them.

Chapter 3 takes up the quantity this chapter treated as a nominal 3.2 V: why a cell has a voltage at all, and why the voltage at its terminals differs from its open-circuit voltage, leading to $V \approx U(\mathrm{SOC}) - IR$.

Chapter 3

Inside a lithium-ion cell: ions, electrons, open-circuit voltage and internal resistance

3.1 What moves inside a cell

The reference container is built from one repeated part. Each of its 3,564 cells is a lithium iron phosphate cell rated at about 3.2 V and 300 Ah, and everything the container does, from storing 3.42 MWh to warming its racks, is the sum of what those cells do. Chapter 2 defined voltage, current, charge and energy as quantities measured at terminals. This chapter opens the cell and asks

where the voltage of a cell comes from, why it changes with the state of charge and with the current, and what that implies for inferring the state and health of a cell from its measured voltage.

A cell stores energy chemically. Charging and discharging move lithium ions from one host material to another, and the energy released or absorbed by that transfer appears at the terminals as a voltage. The answer to the question therefore has three layers: the chemistry that creates a voltage, the dependence of that voltage on how much lithium has been moved, and the losses that appear as soon as current flows.

Jargon: anode, cathode, separator, electrolyte, LFP, open-circuit voltage, terminal voltage, internal resistance, plateau

The *anode* is the negative electrode of a cell; in the reference cells it is graphite. The *cathode* is the positive electrode. The battery industry keeps these names fixed by the discharge direction, although in strict electrochemistry the two swap roles during charge. The *separator* is a thin porous polymer film between the two electrodes that lets ions pass and keeps the electrodes from touching. The *electrolyte* is a lithium salt dissolved in an organic solvent; it fills the pores of both electrodes and of the separator and conducts lithium ions but not electrons. *LFP* is lithium iron phosphate, $LiFePO_4$, the cathode material of the reference cells. The *open-circuit voltage* U is the voltage of a cell that has rested with no current long enough to reach equilibrium; it depends on the state of charge. The *terminal voltage* V is the voltage actually measured across the cell terminals, with or without current. The *internal resistance* R is the resistance the cell itself presents to current; it makes the terminal voltage differ from the open-circuit voltage under load. A *plateau* is a range of state of charge over which the open-circuit voltage changes very little.

Key idea

A cell behaves, to first order, as a voltage source $U(\mathrm{SOC})$ set by where its lithium sits, in series with a resistance R; current moves the state of charge, drops the voltage by IR and makes heat I^2R. For LFP the source voltage is nearly flat over most of the state-of-charge range, so the voltage that is easiest to measure says least about the state that matters most.

The two subsections below describe the parts of a cell and the one rule that makes it useful: electrons are not allowed to take the short path.

3.1.1 Anatomy of a cell

A lithium-ion cell is a layered sandwich. From the negative terminal to the positive terminal it contains a copper current collector, the negative electrode, the separator, the positive electrode and an aluminum current collector. Figure 3.1 draws a cross section in that order, from left to right. The gray block is the graphite negative electrode, drawn as stacked carbon layers with lithium (green dots) stored between them. The light blue block with dashed lines is the separator, soaked in electrolyte; the electrolyte also fills the pores of both electrodes, which the drawing does not show. The purple block is the LFP positive electrode, drawn as a lattice of phosphate cells, some of which hold lithium. Above the cell, the external circuit joins the two current collectors through a load, which for the reference container is the power conversion system (PCS) and, through it, the grid.

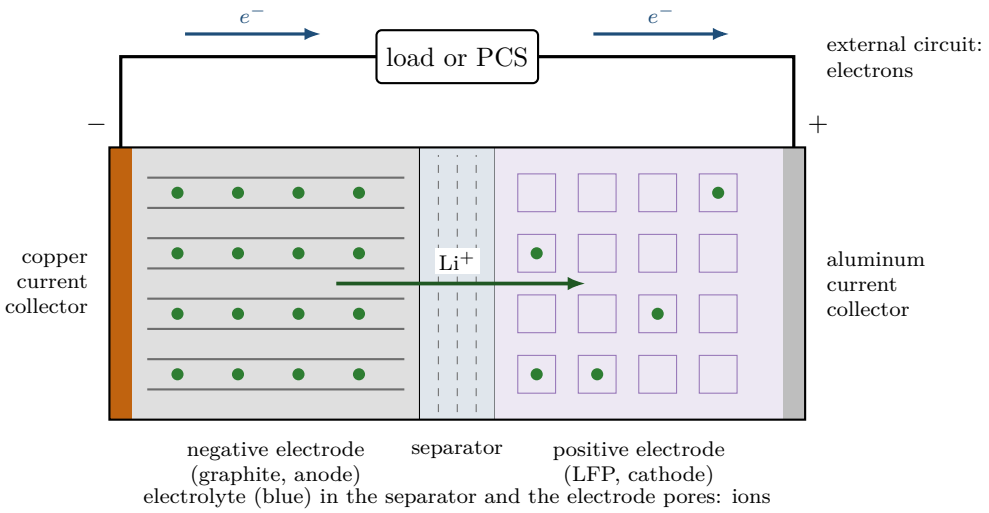


Figure 3.1: Cross section of a lithium-ion cell of the reference type, shown during discharge. Lithium ions (green arrow) cross from the graphite negative electrode through the separator to the LFP positive electrode inside the cell; electrons (blue arrows) travel from the negative terminal through the external circuit to the positive terminal. Layer thicknesses are not to scale: in a real cell the separator is typically 10 to 25 micrometers thick.

Two different things move when the cell operates, and the figure keeps them on separate paths. Lithium ions, Li^+, move *inside* the cell, through the electrolyte and across the separator. Electrons, e^-, move *outside* the cell, through the current collectors, the terminals and the external circuit. The electrodes are called host materials because lithium does not form a new compound with them in the ordinary sense: it slips into spaces in their structure (between the carbon layers of graphite, into the phosphate lattice of LFP) and slips out again. This insertion and removal, called intercalation, is reversible, and its reversibility is why the cell can be cycled thousands of times.

The amount of lithium involved is modest. One ampere-hour is 3,600 coulombs, and one mole of electrons carries Faraday's constant $F = 96{,}485$ coulombs. A 300 Ah cell therefore moves

$$\frac{300 \times 3{,}600 \text{ C}}{96{,}485 \text{ C/mol}} \approx 11.2 \text{ mol} \tag{3.1}$$

of lithium ions from one electrode to the other over a full discharge, about 78 grams of lithium at 6.94 grams per mole. Every ion that crosses inside is matched by exactly one electron that crosses outside, which is why counting electrons in the external circuit (measuring current) is a way of counting lithium inside the cell.

3.1.2 Why electrons take the external path

The separator and the electrolyte together enforce a single rule: ions may cross the cell, electrons may not. The electrolyte conducts ions and is an electrical insulator for electrons, and the separator is a porous insulating film that keeps the two electrodes from touching. When a lithium atom leaves the graphite during discharge, it splits into an ion and an electron. The ion can travel the short distance, tens of micrometers, through the electrolyte to the positive electrode. The electron cannot follow it; its only route to the positive electrode is the long way around, out of the negative terminal, through the wires and the load, and back in through the positive terminal. That external flow of electrons is the current the load receives.

This arrangement is how electrical work is extracted from a chemical reaction. The reaction wants to happen; the cell is built so that it can happen only if the electrons pass through a circuit outside, where the energy they release can drive a motor, charge a capacitor or feed an inverter. If the electrons could cross internally, the same reaction would run inside the cell and its energy would be released there as heat. That is exactly what an internal short circuit does when a damaged separator lets the electrodes touch, and it is why separator integrity is a safety matter rather than a performance detail. Faults of this kind are treated in Chapter 8.

3.2 Discharge, charge and the origin of voltage

This section follows the ions and the electrons through one discharge and one charge, explains why the two electrodes produce a voltage at all, and connects that voltage to energy through the definition of the volt.

3.2.1 Discharge

During discharge the cell delivers energy. At the negative electrode lithium leaves its graphite host,

$$\mathrm{Li} \longrightarrow \mathrm{Li^+} + e^-, \qquad \text{written for graphite as} \qquad \mathrm{LiC_6} \longrightarrow \mathrm{C_6} + \mathrm{Li^+} + e^-. \tag{3.2}$$

The ion crosses the electrolyte and the separator. The electron goes out of the negative terminal, through the wire, the load or the grid converter, and into the positive terminal. At the positive electrode the ion and the electron meet again and the lithium enters the phosphate lattice,

$$\mathrm{FePO_4} + \mathrm{Li^+} + e^- \longrightarrow \mathrm{LiFePO_4}. \tag{3.3}$$

Both carriers therefore travel from the negative side to the positive side: the electrons from negative to positive through the external circuit, the lithium ions from negative to positive through the cell.

Figure 3.2 shows the two directions side by side. Each panel is a simplified cell with the graphite electrode on the left, the separator (blue, soaked in electrolyte) in the middle and

the LFP electrode on the right; the green arrow inside is the ion flow and the blue arrow on the wire above is the electron flow. The left panel is discharge. Electric current is defined by convention as the flow of positive charge, so it runs opposite to the electrons: during discharge conventional current leaves the positive terminal, passes through the load and returns to the negative terminal. This book counts that current as positive, $I > 0$ on discharge, as stated in the notation chapter.

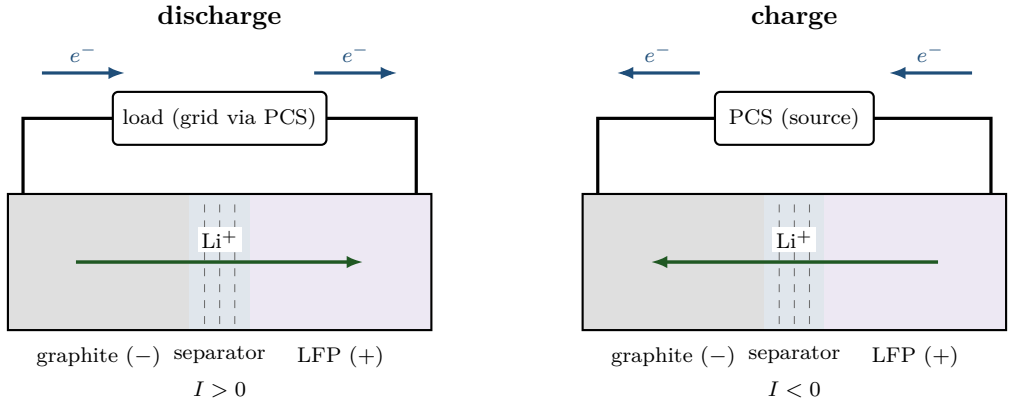


Figure 3.2: Directions of the two carriers in discharge (left) and charge (right). In discharge both the lithium ions inside the cell and the electrons in the external circuit move from the negative side to the positive side, and the load absorbs energy. In charge the PCS acts as the source and both directions reverse. Conventional current runs opposite to the electrons and is counted positive on discharge.

3.2.2 Charge

During charge the PCS pushes energy back into the cell, and every direction reverses. The PCS drives electrons out of the positive terminal, through the external circuit, and into the negative terminal. At the positive electrode lithium is pulled out of the phosphate lattice, $LiFePO_4 \rightarrow FePO_4 + Li^+ + e^-$; the ion crosses the electrolyte to the graphite, where it meets the electron that came around the outside and is stored between the carbon layers, $C_6 + Li^+ + e^- \rightarrow LiC_6$. The right panel of Figure 3.2 shows both arrows pointing from positive to negative: the ions inside the cell and the electrons outside it. Conventional current now enters the positive terminal, and by the sign convention of this book $I < 0$.

The charge direction makes clear what the state of charge is physically. A full LFP cell has most of its movable lithium stored in the graphite; an empty one has most of it back in the phosphate lattice. The state of charge is ultimately a statement about where the lithium is distributed between the two electrodes: the fraction of the cyclable lithium that currently resides in the negative electrode, measured within the window the cell is allowed to use. No sensor sees that distribution directly. What can be measured is the current, which counts the ions as they cross, and the voltage, which, as the next two subsections explain, reflects the energy of the lithium in each electrode.

3.2.3 The electrochemical potential difference

The two electrode materials hold lithium with different strengths. Lithium stored in graphite is in a high-energy state: it is only weakly bound, and it would release energy by moving somewhere it is held more tightly. The iron phosphate lattice is such a place. The difference in how strongly the two materials accept or release lithium and electrons is an *electrochemical potential difference*, and it is observed at the terminals as a voltage. Expressed per mole of

lithium, the open-circuit voltage is the difference in the chemical potential μ of lithium in the two electrodes divided by the charge carried by one mole of electrons,

$$U = \frac{\mu_{\mathrm{Li}}^{\mathrm{neg}} - \mu_{\mathrm{Li}}^{\mathrm{pos}}}{F}. \tag{3.4}$$

For a cell near 3.2 V the numerator is about $3.2 \times 96{,}485 \approx 309$ kilojoules per mole of lithium, or 3.2 electron-volts per transferred ion.

A useful picture is two reservoirs at different heights. The electrons in the negative electrode sit at a higher energy level than those in the positive electrode, and they lower their energy by moving from one to the other. The separator blocks the direct route, so the only way down is through the external circuit, and the energy they give up on the way is delivered to whatever is connected there. Charging lifts them back up, which requires an external source at a higher voltage than the cell.

Equation (3.4) also explains why the voltage depends on the state of charge. The chemical potential of lithium in each electrode depends on how full that electrode already is: lithium added to an almost empty host is held more tightly than lithium added to an almost full one. As the cell discharges, the graphite empties and the phosphate fills, the two potentials move, and the voltage between them changes. Section 3.3 develops that dependence.

3.2.4 The volt as a joule per coulomb

The voltage is energy per unit charge. By definition

$$1\,\mathrm{V} = 1\ \mathrm{J/C}, \tag{3.5}$$

so a 3.2 V cell exchanges about 3.2 joules for every coulomb of charge that passes through it. The energy moved with a charge Q at voltage V is therefore

$$E = V\,Q. \tag{3.6}$$

For the reference cell, $Q = 300\,\mathrm{Ah} = 300 \times 3{,}600 = 1{,}080{,}000$ coulombs, and

$$E = 3.2\ \mathrm{J/C} \times 1{,}080{,}000\ \mathrm{C} = 3{,}456{,}000\ \mathrm{J} = 960\,\mathrm{Wh},$$

using $1\,\mathrm{Wh} = 3{,}600$ joules. The shortcut of multiplying volts by ampere-hours gives the same result directly, $3.2\,\mathrm{V} \times 300\,\mathrm{Ah} = 960\,\mathrm{Wh}$. Multiplied by the 3,564 cells of the reference container, $960\,\mathrm{Wh} \times 3{,}564 \approx 3.42\,\mathrm{MWh}$, which is the container's nominal energy.

Equation (3.6) holds exactly only when the voltage is constant while the charge moves. The cell voltage varies with the state of charge, so the exact energy of a discharge is the integral $E = \int U\,\mathrm{d}Q$ over the charge moved, minus resistive losses. The nominal 3.2 V is a representative value for that integral, and for LFP, whose voltage is nearly constant over most of the range, the approximation is close.

3.3 Open-circuit voltage and the LFP plateau

The voltage of a rested cell is a function of its state of charge, and the shape of that function decides how much the voltage says about the state. This section introduces the function, shows how flat it is for LFP, and states the estimation problem that the flatness creates.

3.3.1 State of charge sets the open-circuit voltage

When a cell has rested with no current long enough for the lithium inside each electrode to even out, its terminal voltage settles to an equilibrium value that depends only on the state of charge and, weakly, on temperature. That equilibrium value is the open-circuit voltage,

$$U = U(\mathrm{SOC}), \tag{3.7}$$

abbreviated OCV. It is a rising curve: as the state of charge goes from 0 to 100 % more lithium sits in the high-energy graphite, the potential difference of equation (3.4) grows, and the voltage rises. The curve is a property of the electrode pair, so each chemistry has its own. A cell with a nickel-manganese-cobalt cathode, for example, has a curve that climbs steadily across the whole range, while LFP has the nearly flat curve described next.

The open-circuit voltage curve is measured in the laboratory by stepping a cell through its state-of-charge range in small increments and letting it rest after each step until the voltage stops changing, which can take hours. In the field such long rests are rare, and the curve is taken from the cell manufacturer or from a laboratory test of a sample cell. It is the bridge between the quantity the battery management system (BMS) measures, voltage, and the quantity it reports, state of charge.

3.3.2 The flat open-circuit curve of LFP

For an LFP cell the curve has a characteristic shape. It starts near 3.2 V at low state of charge and rises quickly; it then runs along a long plateau near 3.3 to 3.4 V across the middle of the range, roughly from 20 to 80 %; and it climbs toward 3.6 V as the cell approaches full. Figure 3.3 draws an illustrative smooth curve with that shape. The horizontal axis is the state of charge in percent and the vertical axis is the open-circuit voltage in volts; the shaded band marks the plateau. Over the shaded 60 percentage points the voltage rises by only about 40 millivolts, whereas the first 10 points and the last 10 points each account for a change of 75 millivolts or more.

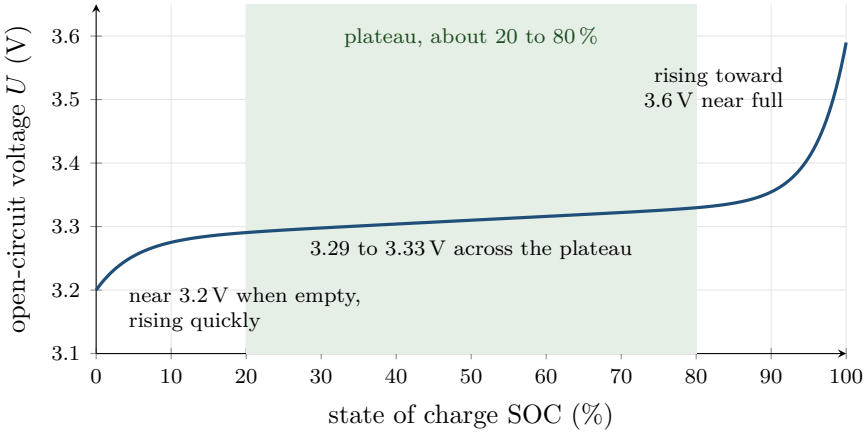


Figure 3.3: Illustrative open-circuit voltage curve of an LFP cell, $U(s) = 3.31 + 0.06\,(s - 0.5) - 0.08\,e^{-s/0.05} + 0.25\,e^{-(1-s)/0.04}$ volts with $s = \mathrm{SOC}/100$. The parameters were chosen to reproduce the typical shape (about 3.20 V empty, a plateau with slope 0.06 V per unit state of charge, about 3.59 V full); a measured curve also shows small steps on the plateau and a difference of a few millivolts between charge and discharge, both omitted here.

The plateau is where an LFP cell spends most of its working life, because grid operation keeps cells away from the extremes. An analysis of the reference record finds that 59 % of the analyzed time windows lie on the flat part of the curve. The consequence for estimation is

direct: in the middle of the range, voltage alone is a poor indicator of state of charge. The slope of the illustrative curve makes the point quantitatively. On the plateau $\mathrm{d}U/\mathrm{dSOC} \approx 0.06\,\mathrm{V}$ per unit state of charge, which is 0.6 millivolts per percentage point. At 5 % the slope is about 6.5 millivolts per point and at 95 % about 18.5 millivolts per point, so the same change in state of charge moves the voltage ten to thirty times more at the ends than in the middle.

Principle 3.1. *On the LFP plateau the open-circuit voltage changes by less than one millivolt per percentage point of state of charge, so a voltage reading in the middle of the range carries little information about the state of charge.*

3.3.3 Voltage alone does not determine state of charge

Take two cells on the plateau, one at 45 % and one at 60 %. Their equilibrium voltages on the curve of Figure 3.3 are

$$U(45\,\%) \approx 3.307\,\mathrm{V}, \qquad U(60\,\%) \approx 3.316\,\mathrm{V},$$

a difference of 9 millivolts for a difference of 15 percentage points of state of charge, or 45 Ah in a 300 Ah cell. Read in the other direction, a measured voltage of 3.30 V does not determine the state of charge. If the reading is uncertain by ±10 millivolts, an illustrative figure that covers sensor error, temperature and the residual effect of recent current, the reading is consistent with every state of charge whose equilibrium voltage lies between 3.29 and 3.31 V, which on the illustrative curve is the whole interval from about 19 to 50 %.

Figure 3.4 shows both readings of the plateau on an enlarged voltage axis. In both panels the horizontal axis is the state of charge and the vertical axis is the open-circuit voltage from 3.26 to 3.38 V, about one fifth of the range of Figure 3.3. The left panel marks the two cells: the dashed lines from 45 % and 60 % meet the curve at two voltages that are barely separated. The right panel starts from the voltage instead: the horizontal band is a reading of $3.30 \pm 0.01\,\mathrm{V}$, and the vertical band is the set of states of charge consistent with it.

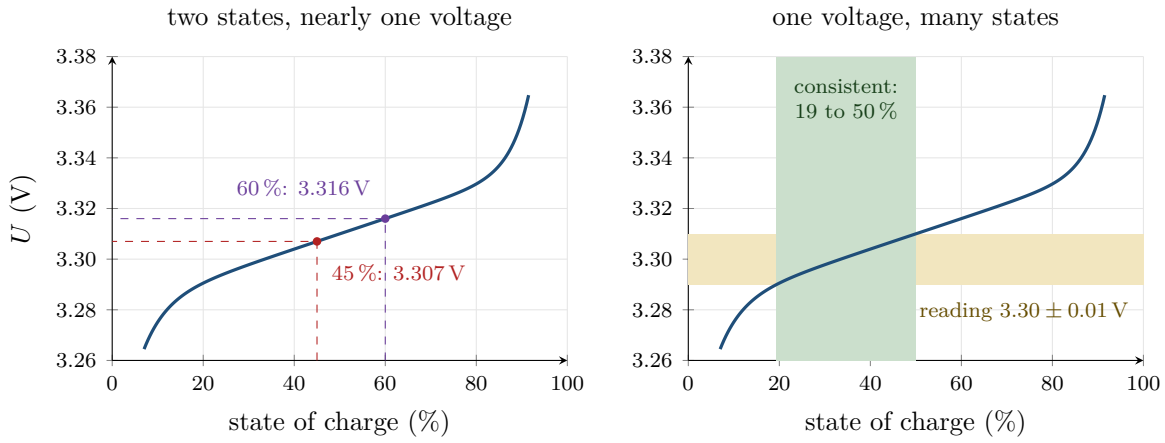


Figure 3.4: The plateau of Figure 3.3 on an enlarged voltage axis (same illustrative curve). Left: cells at 45 % and 60 % differ by 9 millivolts at equilibrium. Right: an illustrative reading of $3.30 \pm 0.01\,\mathrm{V}$ is consistent with any state of charge from about 19 to 50 %.

This is an *identifiability* problem. A quantity is identifiable from a measurement when different values of the quantity produce measurements that can be told apart. On the plateau, one measured voltage corresponds to a broad range of states, so the state of charge is only weakly identifiable from voltage there. The problem is not solved by a better voltmeter alone:

the sensor would have to resolve a fraction of a millivolt, and the voltage under operation also contains the resistive and slow relaxation effects of current, which are larger than that. It is solved, where it is solved at all, by bringing in other information: counting the charge that has passed since the last point at which the state was known, which is what the BMS does, or waiting for the cell to visit a steep part of the curve. Chapter 9 returns to the question of what a set of measurements can determine.

3.4 Terminal voltage and internal resistance

The open-circuit voltage is the voltage of a resting cell. This section describes what happens to the measured voltage when current flows, why the change has opposite signs in discharge and charge, and why the same cell at the same state of charge can show three different voltages.

3.4.1 Open-circuit voltage and terminal voltage

At rest the terminal voltage equals the open-circuit voltage, $V = U(\mathrm{SOC})$. With current flowing the measured terminal voltage differs from it, because the cell resists the passage of current. The simplest model that captures the difference is

$$V = U(\mathrm{SOC}) - I\,R, \tag{3.8}$$

where V is the terminal voltage, U the equilibrium (open-circuit) voltage, I the current, positive on discharge, and R the internal resistance. The resistance lumps several physical effects: the electronic resistance of the current collectors and electrode materials, the ionic resistance of the electrolyte in the separator and in the electrode pores, and the resistance of the reactions at the electrode surfaces.

Figure 3.5 draws equation (3.8) as a circuit. Inside the dashed box, which is the cell, an ideal voltage source of value $U(\mathrm{SOC})$ sits in series with a resistor R. The two open circles on the right are the terminals, and the voltage between them, marked V, is what a voltmeter reads. The arrow on the upper wire is the current in the discharge direction. Only the terminals are accessible: the source voltage and the resistor voltage cannot be measured separately, which is why the model is needed to interpret a terminal reading.

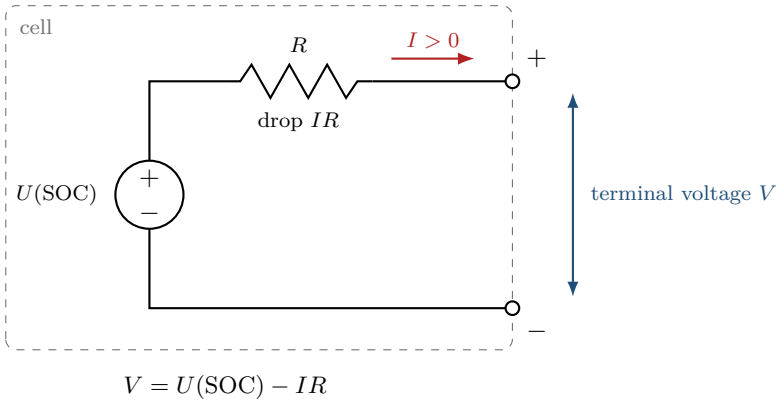


Figure 3.5: The simplest circuit model of a cell: an ideal source whose voltage $U(\mathrm{SOC})$ depends on the state of charge, in series with an internal resistance R. Only the two terminals are accessible, so only V can be measured.

3.4.2 The resistive drop

The cell resists current internally, and the voltage lost across that resistance is

$$\Delta V = I\,R. \tag{3.9}$$

With illustrative values $U = 3.30\,\mathrm{V}$, $I = 100\,\mathrm{A}$ and $R = 1\,\mathrm{m\Omega} = 0.001\,\Omega$, the drop is $IR = 100 \times 0.001 = 0.1\,\mathrm{V}$, and the terminal voltage under discharge is

$$V = 3.30 - 0.10 = 3.20\,\mathrm{V}.$$

A current of 100 A in a 300 Ah cell is a rate of one third of the capacity per hour, a moderate rate in the C-rate notation of Chapter 2. The resistance is small, a thousandth of an ohm, but the current is large, and their product is a tenth of a volt. The open-circuit voltage changes by only about 40 millivolts across the whole LFP plateau, so the resistive drop is not a small correction to be ignored: in the middle of the range it is larger than the entire state-of-charge signal.

The drop also costs energy. At 100 A discharge the source delivers $UI = 330$ watts, the terminals deliver $VI = 320$ watts, and the difference of 10 watts is dissipated inside the cell. That loss is the heat of Section 3.5.1.

3.4.3 The sign of the drop under charge

During charge the current is negative by the sign convention of this book. Equation (3.8) needs no change: with $I = -100\,\mathrm{A}$ it gives

$$V = U - (-100)\,R = U + 100\,R, \tag{3.10}$$

so the terminal voltage now lies *above* the open-circuit voltage. Physically, the PCS must push the current into the cell against both the cell's own voltage and its internal resistance, so it has to apply more than U. The sign reverses naturally once the current carries its sign.

The three regimes follow from one equation:

$$\text{discharge: } V < U; \qquad \text{rest: } V \approx U; \qquad \text{charge: } V > U.$$

The symbol $\approx$ at rest is deliberate. After a period of current the terminal voltage takes minutes to hours to relax fully to the open-circuit value, because part of the voltage drop builds up and decays slowly rather than appearing instantly. The single resistor of Figure 3.5 ignores that slow part; Chapter 4 adds it.

3.4.4 Three terminal voltages of one cell

Example 3.1. A cell at a fixed state of charge has $U = 3.30\,\mathrm{V}$ and an illustrative internal resistance $R = 0.8\,\mathrm{m\Omega}$. At 100 A the resistive drop is $100 \times 0.0008 = 0.08\,\mathrm{V}$. The cell's terminal voltage is

$$\begin{aligned} \text{discharge at } 100\,\mathrm{A}: &\quad V = 3.30 - 0.08 = 3.22\,\mathrm{V}, \\ \text{rest}: &\quad V = 3.30\,\mathrm{V}, \\ \text{charge at } 100\,\mathrm{A}: &\quad V = 3.30 + 0.08 = 3.38\,\mathrm{V}. \end{aligned}$$

Figure 3.6 plots Example 3.1 as a three-minute test. The upper panel is the current: +100 A for the first minute, zero for the second and −100 A for the third. The lower panel is the resulting terminal voltage, with the open-circuit voltage $U = 3.30\,\mathrm{V}$ as a dashed line. Each minute moves 1.7 Ah, about 0.6 percentage points of state of charge, so on the plateau the

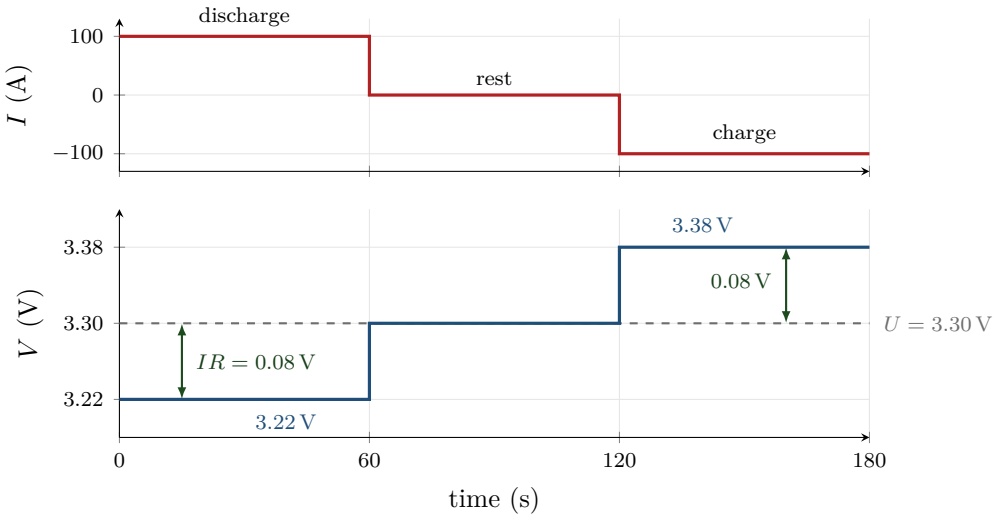


Figure 3.6: Terminal voltage of one cell at one state of charge under discharge, rest and charge (Example 3.1): the model $V = U - IR$ with $U = 3.30\,\mathrm{V}$ and $R = 0.8\,\mathrm{m\Omega}$ (illustrative). The slow relaxation that a real cell shows after each step is omitted.

open-circuit voltage changes by less than half a millivolt and is drawn constant. The terminal voltage steps from 3.22 to 3.30 to 3.38 V while nothing about the cell's state has changed.

The span between the charging and the discharging voltage, 0.16 V, is eighteen times the 9-millivolt difference between 45 and 60 % state of charge found in Section 3.3.3. A voltage reading cannot be interpreted without the current at which it was taken: with the cell of Example 3.1 at 100 A, the same 3.30 V can be a resting cell on the plateau, a discharging cell near 3.38 V open-circuit, or a charging cell near 3.22 V open-circuit, which on Figure 3.3 is a cell below 10 %.

Principle 3.2. *Terminal voltage cannot be read without current. The same cell at the same state of charge shows a lower voltage on discharge, the open-circuit voltage at rest and a higher voltage on charge, separated by IR.*

3.5 Heat and the series string

The resistance that lowers the terminal voltage also turns electrical energy into heat, and in a rack the resistances of hundreds of cells and of the connection hardware add. This section quantifies the heat of one cell and then assembles the string of 396 cells that forms a rack of the reference container.

3.5.1 Resistive heating

The power lost inside the cell is the difference between the power of the source and the power at the terminals,

$$P_{\mathrm{heat}} = UI - VI = (U - V)\,I = (IR)\,I = I^2 R, \tag{3.11}$$

using equation (3.8). The same expression holds on charge, since I^2 is positive whatever the sign of I: resistive heating occurs in both directions. With $R = 1\,\mathrm{m\Omega}$,

$$I = 100\,\mathrm{A}:\ P_{\mathrm{heat}} = 100^2 \times 0.001 = 10\ \mathrm{W}; \qquad I = 200\,\mathrm{A}:\ P_{\mathrm{heat}} = 200^2 \times 0.001 = 40\ \mathrm{W}.$$

Doubling the current quadruples the heat. The delivered power, by contrast, only roughly doubles, so the fraction of the energy lost to heat grows in proportion to the current: $10/330 \approx 3\,\%$ of the source power at 100 A and $40/660 \approx 6\,\%$ at 200 A.

Figure 3.7 plots equation (3.11). The horizontal axis is the current magnitude from 0 to 250 A and the vertical axis is the heat generated in one cell in watts. The red curve is a cell with $R = 1.0\,\mathrm{m\Omega}$, with the two values computed above marked; the green curve is a cell with half that resistance, $0.5\,\mathrm{m\Omega}$, which Section 3.6.1 uses as the resistance of a new cell. Both curves are parabolas: flat at low current and steep at high current.

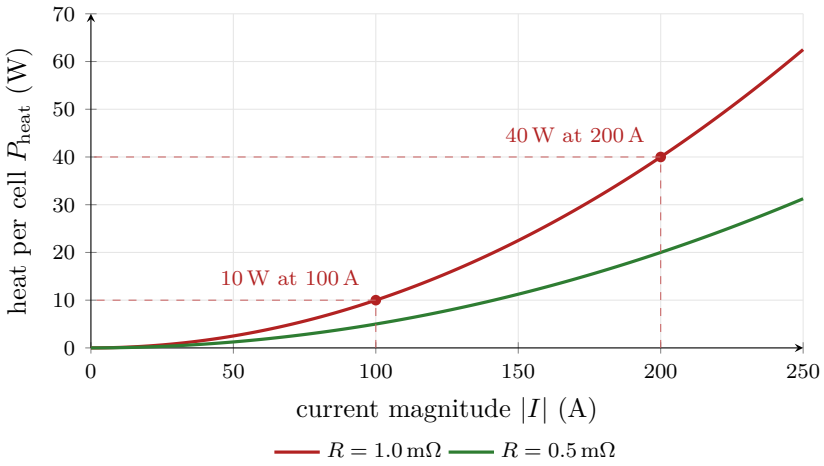


Figure 3.7: Resistive heat generated in one cell against current, $P_{\mathrm{heat}} = I^2 R$, for $R = 1.0\,\mathrm{m\Omega}$ and $R = 0.5\,\mathrm{m\Omega}$ (illustrative values). Doubling the current from 100 to 200 A multiplies the heat by four.

High-power operation is thermally demanding for this reason. A duty that alternates between rest and full power generates more heat than one that delivers the same charge at a steady half power, because heat follows the square of the current: one hour at 200 A and one hour at rest generate 40 watt-hours in a $1\,\mathrm{m\Omega}$ cell, while two hours at 100 A move the same charge for 20 watt-hours. The heat must be removed by the cooling system, and the temperature it produces in turn affects resistance and aging. Chapter 6 builds the heat balance of a rack on this term.

Principle 3.3. *Resistive heat grows with the square of the current. Doubling the current quadruples the heat, in charge and discharge alike.*

3.5.2 The series string of 396 cells

A rack of the reference container is a string of 396 cells connected in series: the positive terminal of each cell is joined to the negative terminal of the next. In a series connection there is no branch, so the same current I flows through every cell. The cell voltages add,

$$\sum_{i=1}^{396} V_i = \sum_{i=1}^{396} U(\mathrm{SOC}_i) - I \sum_{i=1}^{396} R_i, \tag{3.12}$$

where V_i, SOC_i and R_i are the terminal voltage, state of charge and internal resistance of cell i. Each cell produces heat $I^2 R_i$, so the total heat of the string is $I^2 \sum_i R_i$ and the total resistive drop is $I \sum_i R_i$. Series resistances add.

With an illustrative uniform $R_i = 0.8\,\mathrm{m\Omega}$, the string resistance is $396 \times 0.0008 = 0.317\,\Omega$. At 100 A the string loses $100 \times 0.317 = 31.7\,\mathrm{V}$ to its internal resistance and generates $100^2 \times 0.317 \approx$

3,170 watts of heat. With every cell at $U = 3.30\,\mathrm{V}$ the open-circuit string voltage is $396 \times 3.30 = 1{,}306.8\,\mathrm{V}$, and under 100 A discharge the cell voltages sum to $1{,}306.8 - 31.7 = 1{,}275.1\,\mathrm{V}$. The rack stores $396 \times 960\,\mathrm{Wh} \approx 380\,\mathrm{kWh}$, and the nine racks together give the container's 3.42 MWh.

3.5.3 Connection hardware between the string and the rack terminals

The cells are not the only resistance in the current path. Between the ends of the cell string and the rack terminals the current passes through busbars and inter-module links, a fuse that opens on overcurrent, contactors that connect and disconnect the rack, and the current shunt, a calibrated resistor whose voltage drop is how the rack current is measured. Each of these has a resistance. Figure 3.8 draws the path. The row of boxes across the middle is the current path from the negative rack terminal on the left to the positive rack terminal on the right: a contactor, the 396 cells (only four drawn), busbars, the fuse, the shunt and a second contactor. The small tick marks between the cells are the voltage taps at which the BMS measures each cell voltage. The upper brace spans the part of the path that the sum of the cell measurements covers; the lower brace spans the whole path between the rack terminals, which is what the rack voltage sensor reads.

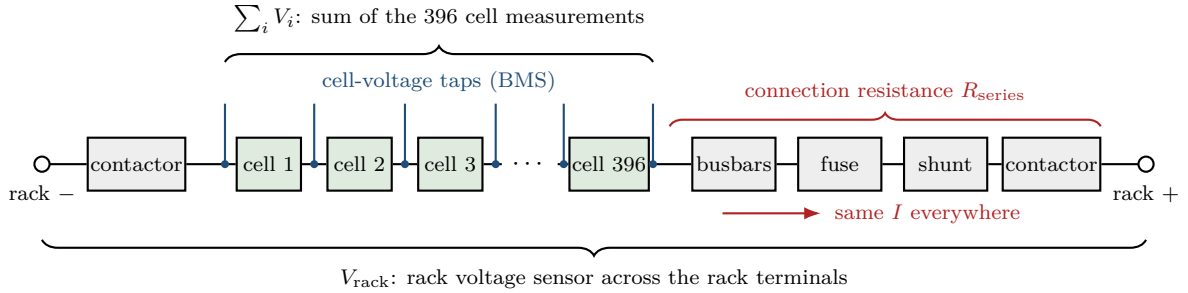


Figure 3.8: One rack of the reference container as a current path. The same current flows through the contactors, the 396 series cells, the busbars, the fuse and the shunt. The BMS measures each cell between its voltage taps (blue); the rack sensor measures across the rack terminals, which also includes the connection hardware. The first contactor and the inter-module links among the cells also carry connection resistance; the red brace marks the largest group.

Because the connection hardware sits outside the cell taps, the rack terminal voltage is not exactly the sum of the cell voltages. Under current the hardware drops a further IR_{series}, where R_{series} is the lumped resistance of everything between the taps and the terminals. The two voltage measurements also come from different sensors with different calibrations, so they can disagree by a constant offset V_{off} even at rest. The relation is

$$V_{\mathrm{rack}} \approx \sum_{i=1}^{396} V_i - I\,R_{\mathrm{series}} + V_{\mathrm{off}}, \tag{3.13}$$

and the analysis of the reference record models the difference $V_r - \sum_i V_{r,i}$ for each rack r in exactly this form, as a straight-line function of current with an effective series resistance as its slope and a voltage offset as its intercept. The two parameters are separable because one multiplies the current and the other does not: at rest the difference is V_{off}, and its change with current is $-R_{\mathrm{series}}$.

Table 3.1 gives illustrative resistances for the elements of Figure 3.8. Their total is $R_{\mathrm{series}} = 1.5\,\mathrm{m\Omega}$, which at 100 A drops 0.15 V and dissipates $100^2 \times 0.0015 = 15$ watts. Against a rack voltage of 1,275 V the drop is about one part in ten thousand, but it is twice the drop of an individual cell, and it is a resistance the cell measurements cannot see. Any comparison between the rack voltage and the sum of the cells has to account for it before a difference can be read as a fault.

Table 3.1: Illustrative resistances of the connection hardware of one rack.

Element	Resistance	Role
Busbars and inter-module links	0.8 mΩ	carry current between modules and to the rack terminals
Fuse	0.3 mΩ	opens on overcurrent
Contactors (two)	0.3 mΩ	connect and isolate the rack
Current shunt	0.1 mΩ	calibrated resistor; its voltage drop measures I
Total, R_{series}	1.5 mΩ	drop 0.15 V and heat 15 W at 100 A

3.6 Aging, the weakest cell and the per-cell model

The quantities C and R are not fixed for the life of a cell, and in a series string they are not equal from cell to cell. This section describes how they change with age, why the cell with the least capacity sets the capacity of the whole string, and how the chapter's pieces combine into a per-cell model whose parameters one would like to infer from telemetry.

3.6.1 Capacity loss and resistance growth

Aging changes a cell in two separate ways. *Capacity loss* reduces the charge the cell can store, for example from 300 Ah new to 270 Ah after years of service, as lithium becomes trapped in side reactions and parts of the electrode stop taking part. *Resistance growth* increases the internal resistance, for example from 0.5 mΩ to 1.0 mΩ, as surface films thicken and contact within the electrodes degrades. Both values are illustrative.

Table 3.2 works through what the two changes mean for the quantities of this chapter. The capacity loss shrinks the stored energy by 10 %. The resistance growth doubles the resistive drop and the heat at any current: at 100 A the drop rises from 0.05 to 0.10 V, and the heat from 5 to 10 watts, the green and red curves of Figure 3.7. The aged cell therefore stores less, sags more under load and runs hotter. These are different dimensions of degradation, and they need not progress together: a cell can lose capacity with little change in resistance, or the reverse. A single health number cannot represent both. Chapter 7 treats the mechanisms.

Table 3.2: A new and an aged cell compared (illustrative values, $U = 3.30$ V, nominal 3.2 V for energy).

Quantity	New	Aged	Effect
Capacity C	300 Ah	270 Ah	stores less
Nominal energy 3.2 V × C	960 Wh	864 Wh	−10 %
Internal resistance R	0.5 mΩ	1.0 mΩ	doubles
Drop IR at 100 A	0.05 V	0.10 V	sags more
Terminal voltage at 100 A discharge	3.25 V	3.20 V	reaches the lower limit sooner
Heat I^2R at 100 A	5 W	10 W	runs hotter
Heat I^2R at 200 A	20 W	40 W	runs hotter

3.6.2 The weakest cell

In a series string every cell carries the same current, so every cell gives up the same charge. Cells with different capacities therefore move through their state-of-charge range at different speeds: the state of charge of cell i changes at the rate $-I/C_i$, faster for a smaller C_i.

Example 3.2. A rack has 395 healthy cells of 300 Ah and one cell of 270 Ah, all fully charged. On discharge every cell gives up the same charge. After 270 Ah the weak cell is empty while each healthy cell still holds 30 Ah. The BMS must stop the rack at that point, because discharging further would drive the weak cell below its lower limit and damage it. The rack delivers 270 Ah, about $396 \times 270 \times 3.2 \approx 342$ kWh at nominal voltage, instead of 380 kWh; the 395 healthy cells still hold $395 \times 30 \times 3.2 \approx 37.9$ kWh that cannot be used. One cell in 396 has cost 10 % of the rack's usable capacity.

The same limitation appears on charge. If the cells are balanced at the bottom instead, all empty together, the weak cell fills first and reaches its upper voltage limit while the others are still at 90 %. A cell with normal capacity but a higher state of charge than its neighbors, an imbalanced cell, behaves the same way: it reaches the upper limit first and ends the charge. In either direction the usable capacity of the rack can be limited by a single cell, which is why the BMS measures every cell voltage rather than only the rack voltage and why the 396 cell voltages per rack in the reference record exist.

Runners chained together. A string of 396 cells is like 396 runners chained together at the waist. They must all move at the same pace, as the cells all carry the same current. If one runner tires first, the whole group stops, however fresh the others are. The pace of the group over a long run is set by its weakest member, not by its average.

> **Principle 3.4.** *In a series string every cell carries the same current, so the string stops when its weakest or most imbalanced cell reaches a limit. The usable capacity of a rack is set by its worst cell, not its average cell.*

3.6.3 The per-cell voltage model

The pieces of this chapter combine into a model of each cell. The terminal voltage of cell i is

$$V_i = U(\mathrm{SOC}_i) - I\,R_i, \tag{3.14}$$

and its state of charge evolves by counting the charge that has passed,

$$\mathrm{SOC}_i(t) = \mathrm{SOC}_i(0) - \frac{1}{C_i}\int_0^t I(\tau)\,\mathrm{d}\tau, \tag{3.15}$$

with C_i in ampere-hours and the integral in ampere-hours. Substituting the second into the first gives

$$V_i(t) = U\!\left(\mathrm{SOC}_i(0) - \frac{1}{C_i}\int_0^t I(\tau)\,\mathrm{d}\tau\right) - I(t)\,R_i. \tag{3.16}$$

Equation (3.16) relates two measured signals, the cell voltage $V_i(t)$ and the common current $I(t)$, to three unknowns per cell: the initial state of charge $\mathrm{SOC}_i(0)$, the capacity C_i and the resistance R_i, given the shared curve U. For a rack that is $3 \times 396 = 1{,}188$ unknowns. In principle the model can be fitted to the telemetry to infer each cell's capacity and resistance, which is exactly the estimation problem the analysis of the reference record studies.

The two parameters enter the model differently, and this decides how well each can be estimated. The resistance multiplies the current directly, so every change in current produces an immediate change $\Delta V_i = -R_i\,\Delta I$; a step from rest to 100 A in a 0.8 mΩ cell moves the voltage by 80 millivolts at once. The capacity enters only through the open-circuit curve, after the current has been integrated, so it shows up only to the extent that U changes with the state of charge.

3.6.4 Weak identifiability on the plateau

For a small change in state of charge, the change in open-circuit voltage is approximately

$$\Delta U \approx \frac{\mathrm{d}U}{\mathrm{dSOC}}\,\Delta\mathrm{SOC}. \tag{3.17}$$

On the LFP plateau $\mathrm{d}U/\mathrm{dSOC} \approx 0$: in the illustrative curve of Figure 3.3 it is 0.06 V per unit state of charge. A meaningful error in state of charge therefore produces a tiny voltage effect: an error of 10 percentage points changes the voltage by 6 millivolts.

Example 3.3. Two cells on the plateau, one of 300 Ah and one of 270 Ah, are discharged together by 100 Ah. The first moves $100/300 = 0.333$ in state of charge and the second $100/270 = 0.370$. With $\mathrm{d}U/\mathrm{dSOC} = 0.06$ V, their open-circuit voltages fall by $0.06 \times 0.333 = 20.0$ and $0.06 \times 0.370 = 22.2$ millivolts. A 10 % difference in capacity appears as a 2.2-millivolt difference in voltage after an hour of discharge at 100 A. If the same 100 Ah ended near 95 % state of charge instead, where the slope of the illustrative curve is about 1.85 V per unit, the difference of 0.037 in state of charge would appear as about 69 millivolts.

Example 3.3 shows why capacity is hard to infer from voltage on the plateau: the information about C_i arrives almost entirely when the cell visits the steep ends of the curve, and in the reference record more than half of the analyzed windows lie on the plateau. Resistance, by contrast, is visible at every current change. Some parameters are therefore only weakly identifiable from the telemetry, and some can be determined only relative to one another rather than absolutely. If the current sensor reads 2 % high, every integrated charge is 2 % high and every fitted capacity absorbs the same factor; the ratio of two cells' capacities is unaffected, because the same current flows through both. Comparing cells with their neighbors in the same string is therefore more reliable than stating the capacity of any one of them. Chapter 9 makes these distinctions precise.

3.6.5 Three equations to keep

The chapter reduces to three equations,

$$\mathrm{SOC}(t) = \mathrm{SOC}_0 - \frac{1}{C}\int_0^t I\,\mathrm{d}\tau, \qquad V \approx U(\mathrm{SOC}) - IR, \qquad P_{\mathrm{heat}} = I^2R, \tag{3.18}$$

each a consequence of current. Current changes the state of charge, by moving lithium from one electrode to the other; current causes a voltage drop, through the internal resistance; and current causes heating, through the same resistance. Together they are the physical skeleton of a cell: the charge balance, the voltage relation and the heat source. Every later chapter adds detail to one of the three, and none replaces them.

The order in which the equations are used for inference matters. The first needs only current and an initial state, and it accumulates any error in either. The second ties the state to the one quantity that can be measured at every cell, but only through the curve U, which is flat where the cells spend most of their time. The third links the electrical side to the thermal measurements, the subject of Chapter 6.

3.7 What to retain

Definition 3.1 (The cell as a source and a resistance)**.** A lithium-ion cell stores energy as lithium in a high-energy host and behaves as an open-circuit voltage $U(\mathrm{SOC})$ in series with an internal resistance R: current moves its state of charge, lowers its terminal voltage by IR and generates heat I^2R. For LFP the open-circuit voltage is nearly flat over most of the range, so voltage determines neither state of charge nor capacity well there.

Table 3.3 lists the chapter's objects with the subsection that defines each.

Table 3.3: The objects of the cell model.

Object	Definition	Section
electrodes, separator, electrolyte	graphite and LFP hosts; ion-only path between them	3.1.1
external path	electrons must cross through the circuit; that flow is the current	3.1.2
discharge, charge	both carriers negative to positive, and the reverse	3.2.1, 3.2.2
state of charge	where the cyclable lithium is distributed	3.2.2
$U = (\mu_{\mathrm{Li}}^{\mathrm{neg}} - \mu_{\mathrm{Li}}^{\mathrm{pos}})/F$	electrochemical potential difference	3.2.3
$E = VQ$	energy as voltage times charge; $1\,\mathrm{V} = 1\,\mathrm{J/C}$	3.2.4
$U(\mathrm{SOC})$	open-circuit voltage curve	3.3.1
plateau	LFP range, about 20 to 80 %, with slope near 0.6 mV per point	3.3.2
identifiability	distinct values give distinguishable measurements	3.3.3
$V = U(\mathrm{SOC}) - IR$	terminal voltage under current	3.4.1
IR	resistive drop; negative current raises V	3.4.2, 3.4.3
$P_{\mathrm{heat}} = I^2R$	resistive heat, quadratic in current	3.5.1
$\sum_i V_i$, $I\sum_i R_i$	series string: voltages and resistances add	3.5.2
$V_{\mathrm{rack}} \approx \sum_i V_i - IR_{\mathrm{series}} + V_{\mathrm{off}}$	rack voltage with connection hardware and offset	3.5.3
capacity loss, resistance growth	two separate dimensions of aging	3.6.1
weakest cell	the string stops when one cell reaches a limit	3.6.2
$V_i(t)$ with $\mathrm{SOC}_i(0)$, C_i, R_i	per-cell voltage model	3.6.3
$\Delta U \approx (\mathrm{d}U/\mathrm{dSOC})\,\Delta\mathrm{SOC}$	sensitivity; weak and relative identifiability	3.6.4

3.8 Exercises

Exercise 3.1. A cell has a nominal voltage of 3.25 V and a capacity of 280 Ah. (a) Express its capacity in coulombs and its nominal energy in joules and in watt-hours. (b) How many moles of lithium cross between the electrodes in a full discharge, using $F = 96{,}485$ coulombs per mole? (c) How many such cells are needed for 1 MWh of nominal energy, and what is the energy per transferred lithium ion in electron-volts?

Exercise 3.2. A cell has $U = 3.32\,\mathrm{V}$ and $R = 0.6\,\mathrm{m\Omega}$. (a) Compute its terminal voltage at 150 A discharge, at rest and at 150 A charge, and the heat it generates in each case. (b) Compute the power delivered at the terminals during the discharge and the fraction lost as heat. (c) After aging the resistance has doubled. Repeat (a), and find the charging current at which the terminal voltage reaches an upper limit of 3.65 V.

Exercise 3.3. A rack has 396 cells in series, each at $U = 3.31\,\mathrm{V}$ with $R_i = 0.7\,\mathrm{m\Omega}$. (a) Compute the sum of the cell voltages and the total string heat at 150 A discharge. (b) The difference $V_{\mathrm{rack}} - \sum_i V_i$ reads $+0.25\,\mathrm{V}$ at rest and $+0.04\,\mathrm{V}$ at 150 A discharge. Using equation (3.13), find R_{series} and V_{off}, and the heat dissipated in the connection hardware at 150 A. (c) One cell of the rack has a capacity of 255 Ah and the others 300 Ah, all starting full. What fraction of the rack's nominal capacity can be discharged, and how much energy, at a nominal 3.2 V, is left in the healthy cells when the rack stops?

Exercise 3.4. Use the plateau slope $\mathrm{d}U/\mathrm{dSOC} = 0.06\,\mathrm{V}$ per unit state of charge. (a) Two cells of 300 Ah and 280 Ah are discharged together by 60 Ah on the plateau. Compute the change in each cell's open-circuit voltage and the difference between them, and compare it with a voltage resolution of 1 millivolt. (b) Repeat the comparison for a discharge that ends near 95 % state of charge, where the slope is 1.85 V per unit. (c) The current sensor reads 2 % low. What capacities would a fit of equation (3.16) report for the two cells, and what ratio?

3.9 Connection to the TERRA framework

The cell model of this chapter maps onto a conservation graph directly, and the identifiability questions it raises are the ones the framework asks before it estimates anything.

- **The state-of-charge equation is a charge balance.** Equation (3.15) is conservation of charge on a cell node: the stored charge $C_i\,\mathrm{SOC}_i$ changes by exactly the charge carried by the current. The reported state of charge of the reference record is this integral evaluated with the measured current, so it is a derived channel, not an independent measurement of the balance; comparing the two tests nothing.
- **The series string is an edge with one flow.** The current I is the flow on the edge that passes through all 396 cells, the connection hardware and the shunt of Figure 3.8. One measured flow drives 396 component states, which is why the string is informative about differences between cells and why the weakest cell of Section 3.6.2 is a bound on the shared flow.
- **The per-cell voltage model is a component relation.** Equation (3.16) relates each cell's measured voltage to its state and to its parameters C_i and R_i. In the framework these are component parameters with posterior distributions, inferred from $V_i(t)$ and $I(t)$.
- **The string voltage identity is a redundancy relation.** Equation (3.13) links two independently measured quantities, the rack voltage and the sum of the cell voltages, through R_{series} and V_{off}. Because both sides are measured, its residual is a genuine test of the sensors and the connection path, unlike the comparison of reported and counted state of charge.
- **The heat term is a source on a thermal node.** $I^2 \sum_i R_i + I^2 R_{\mathrm{series}}$ is the electrical heat entering the rack's energy balance, which couples the resistance parameters to the temperature measurements.
- **The plateau is a practical identifiability limit.** Where $\mathrm{d}U/\mathrm{dSOC} \approx 0$, the data constrain C_i weakly and its posterior stays wide; a current-sensor gain error makes absolute capacity structurally confounded with that gain, while capacity ratios remain determined. The framework reports this as a wide or correlated posterior rather than as a point estimate.

In summary, a cell contributes one charge balance, one voltage relation and one heat source to the graph; a rack contributes a shared current edge, a redundancy relation between two voltage measurements and the constraint that the weakest cell bounds the whole; and the flat LFP curve determines which of the resulting parameters the measurements can actually pin down. The next chapter refines the single resistor into an equivalent circuit whose slow polarization branch explains why a resting cell does not return to its open-circuit voltage at once.

Chapter 4

Open-circuit voltage, internal resistance and the equivalent-circuit model

4.1 Three voltages

Chapter 3 introduced a relation that looks too simple to need a chapter of its own: the voltage measured at the terminals of a cell equals an equilibrium voltage set by the chemistry, minus a loss proportional to the current,

$$V = U(\mathrm{SOC}) - IR. \tag{4.1}$$

Almost all basic battery estimation sits inside this line. State of charge, capacity, resistance, power capability and heating are each read through it, and each of the difficulties met later in this book (the flat voltage curve, the missing rest periods, the parameters that cannot be separated) is already visible in it. The question of this chapter is

how the voltage measured at a cell's terminals is related to the state of charge inside it, and what a model of that relation must contain before it can be used to estimate anything.

Jargon: open-circuit voltage, terminal voltage, ohmic drop, polarization, equivalent circuit, time constant, state-space model

The *open-circuit voltage* U is the voltage of a cell that carries no current and has rested long enough for every effect of past current to fade; it reflects the cell's electrochemical state and is a function of the state of charge. The *terminal voltage* V is the voltage actually measured between the cell's terminals, with or without current. The *ohmic drop* is the part of the difference $U - V$ that appears the instant current flows and disappears the instant it stops, written IR_0. *Polarization* is the part that builds up and decays slowly, written V_1: the temporary departure of the electrochemical system from equilibrium caused by current. An *equivalent circuit* is a network of ideal circuit elements (voltage sources, resistors, capacitors) whose voltage response to current reproduces that of the cell; its elements are modeling devices, not physical parts. A *time constant* τ is the time a first-order system takes to cover about 63 % of the way to its new steady value after a step. A *state-space model* writes a dynamic system as a set of first-order differential equations for a vector of internal states, driven by an input, plus an equation giving the measured output as a function of the states and the input.

Key idea

The terminal voltage of a cell is its open-circuit voltage minus an instantaneous ohmic drop and minus a slow polarization voltage. One sensor reads the sum; which term changed must be inferred, and on the flat open-circuit curve of a lithium iron phosphate cell the term that carries the state of charge is the one that changes least.

This section names the three voltages, says precisely what "open circuit" means, and shows with one number why a terminal voltage cannot be converted to a state of charge without the current.

4.1.1 Open-circuit voltage, terminal voltage and the internal drop

Three voltages appear in equation (4.1). The open-circuit voltage U is the chemical or equilibrium voltage of the cell: the voltage its electrode materials establish when no charge is being pushed through it. The internal drop IR is the voltage lost inside the cell in driving the current I through it; R is the cell's internal resistance. The terminal voltage V is what remains at the terminals, and it is the only one of the three that a voltmeter reads. Equation (4.1) therefore says: the equilibrium voltage minus the internal loss gives the measured terminal voltage.

The sign convention of this book makes current positive on discharge. On discharge $I > 0$, the drop IR is positive, and $V < U$: the cell delivers less voltage than its chemistry establishes. On charge $I < 0$, the term $-IR$ is positive, and $V > U$: the charger must push harder than the chemistry pushes back. At rest $I = 0$ and $V = U$. One equation covers all three cases because the sign of I carries the direction.

4.1.2 What open circuit means

Setting $I = 0$ in equation (4.1) gives $V = U$, and this is the operational definition of the open-circuit voltage: the terminal voltage after the cell has rested long enough for every load-related effect to settle. The qualification "long enough" matters. Interrupting the current removes the ohmic drop at once, but Section 4.3.1 will show that part of the drop decays only slowly; a voltage read a few seconds after the current stops is not yet the open-circuit voltage.

Once the cell has rested, its voltage represents the underlying electrochemical state, which for a given cell and temperature is determined by how much lithium sits in each electrode, that is, by the state of charge. The open-circuit voltage is therefore a function,

$$U = U(\mathrm{SOC}), \tag{4.2}$$

the open-circuit voltage curve of Chapter 3. For the lithium iron phosphate (LFP) cells of the reference container the curve rises steeply at low state of charge, runs almost flat through the middle of the range, and rises steeply again near full. The curve is a property of the chemistry; the cell's position on it is the state.

4.1.3 Why terminal voltage alone does not give the state of charge

Take a cell with open-circuit voltage $U = 3.30\,\mathrm{V}$ and internal resistance $R = 0.8\,\mathrm{m\Omega}$. At rest the terminal voltage is 3.30 V. At a discharge current of 100 A the internal drop is

$$IR = 100\,\mathrm{A} \times 0.0008\,\Omega = 0.08\,\mathrm{V},$$

and the terminal voltage is $V = 3.30 - 0.08 = 3.22\,\mathrm{V}$. Charging at the same current gives $3.30 + 0.08 = 3.38\,\mathrm{V}$. One cell at one state of charge shows three terminal voltages, 3.22, 3.30 and 3.38 V, depending only on the current.

Figure 4.1 draws the consequence. The horizontal axis is the state of charge and the vertical axis is voltage. The solid green curve is an illustrative LFP open-circuit voltage curve $U(\mathrm{SOC})$; the blue curve below it is the terminal voltage at 100 A discharge, $U - 0.08\,\mathrm{V}$, and the gray curve above it the terminal voltage at 100 A charge. At $\mathrm{SOC} = 0.50$ the three dots mark 3.38, 3.30 and 3.22 V, and the red double arrow is the 0.08 V drop. The dashed path shows what happens if the discharge reading of 3.22 V is treated as if it were an open-circuit voltage: it meets the open-circuit curve at $\mathrm{SOC} \approx 0.08$. A reading of 3.22 V taken alone suggests a nearly empty cell; the cell is in fact half full. The voltage and the current must be interpreted together.

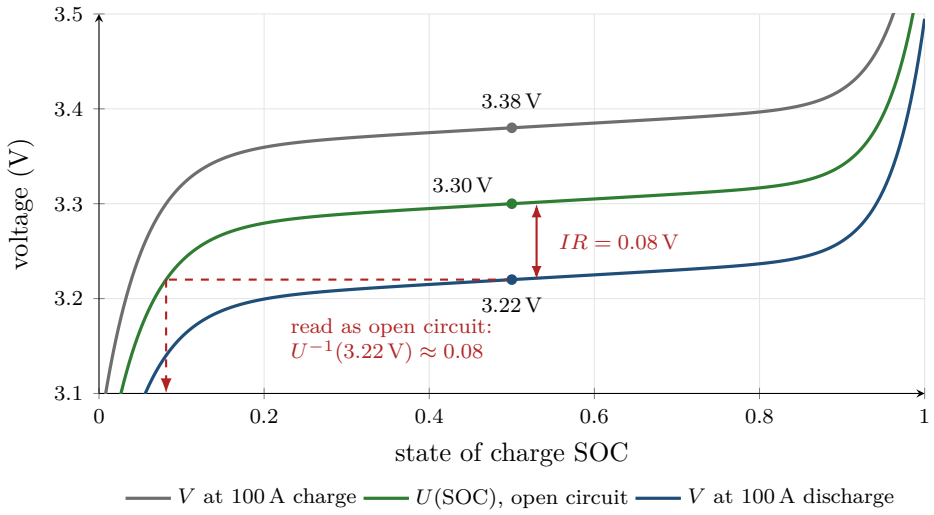


Figure 4.1: The three voltages of one cell. The open-circuit curve is the illustrative LFP-like function $U(\mathrm{SOC}) = 3.275 + 0.05\,\mathrm{SOC} - 0.30\,e^{-\mathrm{SOC}/0.05} + 0.25\,e^{-(1-\mathrm{SOC})/0.04}$ V; the terminal voltages at 100 A are $U \mp IR$ with $R = 0.8\,\mathrm{m\Omega}$. Reading the discharge voltage 3.22 V as an open-circuit voltage places the cell at $\mathrm{SOC} \approx 0.08$ instead of 0.50.

Principle 4.1. *A terminal voltage is interpretable only together with the current that was flowing when it was read. The same cell at the same state of charge reads* 3.22, 3.30 *or* 3.38 V *at* 100 A *discharge, rest, or* 100 A *charge.*

4.2 Internal resistance as a lumped parameter

The internal resistance R of equation (4.1) is drawn as one resistor, but it stands for several physical processes and it changes with the operating conditions. This section says what it lumps together, what it depends on, and how it is estimated from a current step.

4.2.1 Internal resistance is not one resistor

The model behind equation (4.1) draws an ideal voltage source U in series with a resistor R. Inside the cell there is no such resistor. The voltage lost under current is the sum of several distinct effects:

- *ionic resistance*: the resistance of the electrolyte to the motion of lithium ions between the electrodes;
- *electronic resistance*: the resistance of the electrode materials and the metal current collectors to the motion of electrons;
- *electrode effects*: resistive surface films and interfaces on the electrode particles;
- *contact resistance*: the welds, tabs and terminals that join the internal stack to the outside;

- *reaction kinetics*: the extra voltage needed to drive the charge-transfer reaction at the electrode surfaces at a given rate;
- *transport limitations*: the concentration differences that build up when lithium must diffuse through the electrolyte and into the electrode particles faster than it would at rest.

The first four behave much like ordinary resistors: their voltage appears and vanishes with the current. The last two do not; they build up and relax over time, and Section 4.3.5 gives them their own term.

R is therefore a *lumped* parameter: one number standing in for a collection of physics, chosen so that the model reproduces the cell's terminal behavior in the conditions of interest. This move, replacing complicated physics by a few parameters that act on the measured quantities, recurs throughout the book: the thermal capacitance and cooling conductance of Chapter 6 and the capacity-fade parameters of Chapter 7 are lumped in the same way. A lumped parameter is useful exactly to the extent that it is stable over the data it is fitted to and predicts the behavior it is used for.

4.2.2 Resistance depends on state of charge, temperature and age

A lumped parameter inherits the dependences of the processes it lumps, so in general

$$R = R(\mathrm{SOC}, T, \mathrm{age}, \ldots). \tag{4.3}$$

Three dependences dominate. *Temperature*: ionic conduction and reaction kinetics slow down in the cold, so a cold cell shows a higher resistance. A common empirical form is $R(T) = R(T_{\mathrm{ref}})\exp\!\left[\beta\,(1/T - 1/T_{\mathrm{ref}})\right]$ with temperatures in kelvin. With the illustrative value $\beta = 3000\,\mathrm{K}$, cooling a cell from 25°C (298 K) to 0°C (273 K) multiplies its resistance by $\exp[3000 \times (1/273 - 1/298)] = \exp(0.92) \approx 2.5$, so a $0.5\,\mathrm{m\Omega}$ cell at 25°C has about $1.26\,\mathrm{m\Omega}$ at 0°C. *Age*: surface films thicken and contacts degrade as the cell is used, so an aged cell shows a higher resistance; Chapter 3 used the example of growth from $0.5\,\mathrm{m\Omega}$ to $1.0\,\mathrm{m\Omega}$, which doubles the drop at 100 A from 50 to 100 mV. *State of charge*: resistance is roughly constant through the middle of the range and rises toward the ends, where one electrode is nearly full or nearly empty of lithium and the reaction slows.

Treating R as a constant is therefore an approximation that holds over a limited operating window: a range of temperature, of state of charge, and of time short enough that aging is negligible. Outside that window a constant-resistance model is still a model, but the constant it fits is an average over conditions and should be reported with the conditions it was fitted in.

4.2.3 Estimating resistance from a voltage step

Equation (4.1) suggests how to measure R. If the current changes abruptly from I_1 to I_2, the state of charge has no time to move, so $U(\mathrm{SOC})$ is the same immediately before and immediately after the step. Subtracting the two readings eliminates it:

$$\Delta V = V_{\mathrm{after}} - V_{\mathrm{before}} = -(I_2 - I_1)\,R \quad\Longrightarrow\quad R \approx -\frac{\Delta V}{\Delta I}. \tag{4.4}$$

For a step from rest to 100 A discharge in which the voltage falls from 3.30 V to 3.22 V,

$$R \approx \frac{0.08\,\mathrm{V}}{100\,\mathrm{A}} = 0.0008\,\Omega = 0.8\,\mathrm{m\Omega}.$$

Voltage sag under a current step reveals resistance. The estimate needs no knowledge of the open-circuit curve or of the state of charge, because both cancel in the difference, which is why current steps are the standard resistance test.

The estimate has one hidden choice: the time at which the second voltage is read. If the cell were exactly equation (4.1), it would not matter. Real cells show that it does, and that

observation is the starting point of the next section: in Figure 4.2, the 0.08 V drop of this example is reached only after the voltage has finished relaxing, and the part that appears at the instant of the step is smaller.

4.3 The first-order equivalent circuit

The constant-resistance model explains a voltage drop that appears with the current; it cannot explain a voltage that keeps moving while the current is constant. This section adds the smallest structure that can, a resistor–capacitor branch, and writes the resulting model as three equations.

4.3.1 An instantaneous drop and a slow relaxation

Figure 4.2 shows the measured shape that motivates the model. The upper panel is the current: zero until $t = 0$, a constant 100 A discharge for five minutes, zero again afterwards. The lower panel is the terminal voltage over the same time axis, with the open-circuit voltage as a dashed line. At $t = 0$ the voltage falls vertically by 50 mV; it then continues to fall, more and more slowly, by a further 30 mV, and settles near 3.22 V. When the current stops at $t = 300$ s the voltage jumps up by the same 50 mV and then creeps back toward 3.30 V over the following minutes.

A single resistance explains the vertical segments: a drop that appears and disappears with the current is exactly IR. It cannot explain the curved segments, during which the current is constant (or zero) and the voltage still moves. Some part of the cell's response depends not on the present current but on the recent history of the current. A model that is to reproduce Figure 4.2 needs a term with memory, and that is the motivation for the equivalent-circuit model. It also resolves the hidden choice of Section 4.2.3: reading the voltage within a fraction of a second of the step gives $50\,\text{mV}/100\,\text{A} = 0.5\,\text{m}\Omega$, reading it after the relaxation has settled gives the $0.8\,\text{m}\Omega$ of that section, and a reading in between gives a number in between. A resistance quoted without the time at which it was read is ambiguous.

4.3.2 The simplest equivalent circuit

Figure 4.3 draws the circuit that reproduces Figure 4.2. On the left is an ideal voltage source whose voltage is the open-circuit voltage $U(\text{SOC})$. In series with it is a resistor R_0, and in series with that a parallel pair: a resistor R_1 and a capacitor C_1 connected across the same two nodes. The terminals of the cell are on the right, and the current I flows out of the positive terminal on discharge. The labels below the elements are the voltages across them: IR_0 across the series resistor and V_1 across the parallel pair.

Walking around the circuit from the negative terminal through the source and the two elements to the positive terminal gives the terminal voltage,

$$V = U(\text{SOC}) - IR_0 - V_1. \tag{4.5}$$

The two loss terms have different characters. R_0, the *ohmic resistance*, gives the fast drop: its voltage is proportional to the present current, so it produces the vertical segments of Figure 4.2. V_1, the voltage across the parallel branch, gives the slower polarization effect: a capacitor's voltage cannot jump, so V_1 moves only gradually, and it produces the curved segments. With $V_1 = 0$ equation (4.5) reduces to equation (4.1) with $R = R_0$; after a long constant current it reduces to equation (4.1) with $R = R_0 + R_1$. The constant-resistance model is the first-order model observed at one of its two time scales.

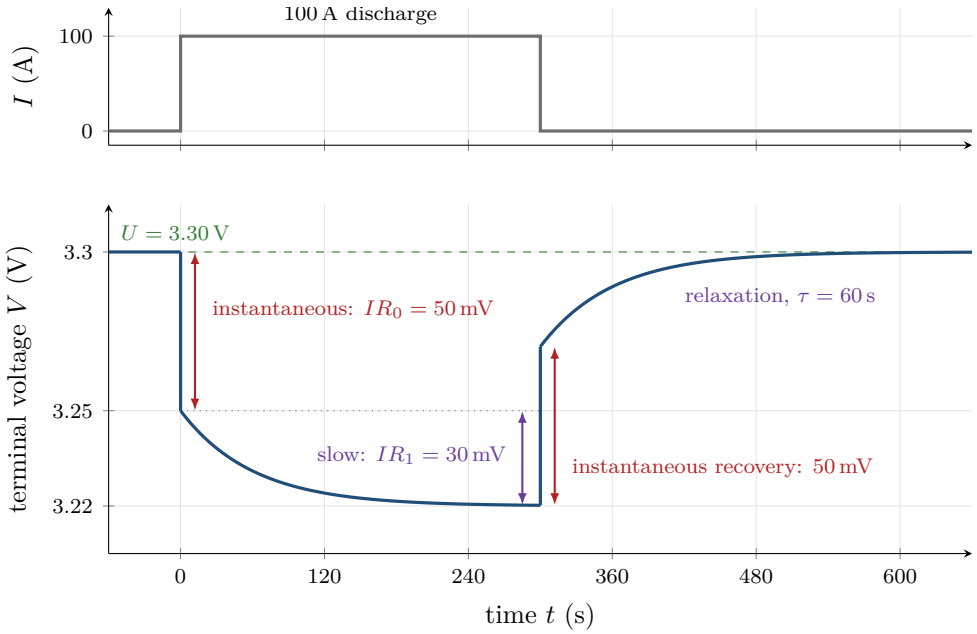


Figure 4.2: Terminal-voltage response of the first-order equivalent circuit to a five-minute current pulse. For $0 \le t \le 300\,\text{s}$, $V(t) = U - IR_0 - IR_1\,(1 - e^{-t/\tau})$; after the pulse, $V(t) = U - V_1(300)\,e^{-(t-300)/\tau}$, with illustrative $U = 3.30\,\text{V}$ (held constant), $I = 100\,\text{A}$, $R_0 = 0.5\,\text{m}\Omega$, $R_1 = 0.3\,\text{m}\Omega$ and $\tau = R_1C_1 = 60\,\text{s}$. The vertical segments are the ohmic drop; the curved segments are the polarization voltage building and decaying.

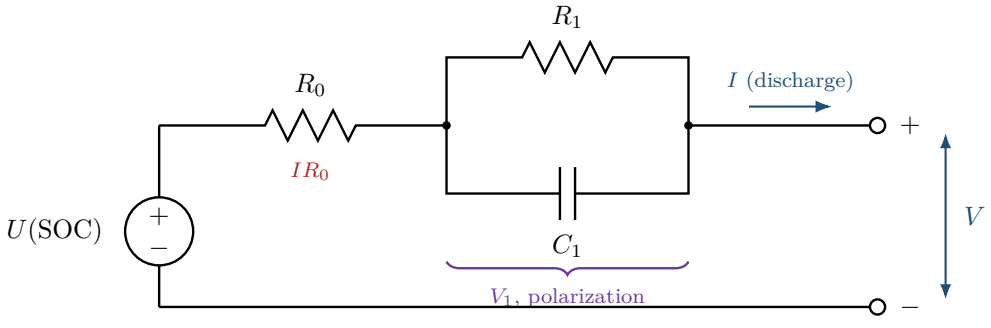


Figure 4.3: The first-order equivalent circuit of a cell: an ideal source $U(\text{SOC})$, a series ohmic resistance R_0, and a parallel R_1–C_1 branch whose voltage V_1 is the polarization voltage. The terminal voltage is $V = U(\text{SOC}) - IR_0 - V_1$.

4.3.3 Polarization

Current pushes the electrochemical system away from equilibrium. Lithium must be removed from one electrode and inserted into the other at a rate set by the current; ions pile up near one electrode surface and are depleted near the other; the reaction at each surface runs faster than it would at rest and needs an extra driving voltage to do so. Some of these processes respond almost instantly to a change of current: the conduction losses that R_0 represents. Others respond slowly, because they involve concentrations that take time to build up and time to even out again by diffusion. The slow, temporary departure of the voltage from its equilibrium value is the *polarization voltage* V_1.

Polarization is temporary: when the current stops, the concentration differences relax and V_1 returns to zero, which is the creep back toward 3.30 V after $t = 300$ s in Figure 4.2. It is also history-dependent: its value at any moment depends on how much current flowed in the recent past, not only on the current now. The R_1–C_1 branch is what gives the model that memory. A model with only R_0 is static, its voltage a function of the present current alone; a model with the branch is dynamic.

4.3.4 The polarization state and its time constant

The voltage across the parallel branch obeys one differential equation. The current I entering the branch splits: V_1/R_1 flows through the resistor, and the remainder charges the capacitor, whose voltage changes at the rate (current into it)$/C_1$. Hence

$$\frac{\mathrm{d}V_1}{\mathrm{d}t} = -\frac{V_1}{R_1 C_1} + \frac{I}{C_1}, \qquad \tau = R_1 C_1. \tag{4.6}$$

The first term pulls V_1 back toward zero at a rate set by the time constant τ; the second pushes it in the direction of the current. For a constant current I switched on at $t = 0$ with $V_1(0) = 0$, the solution is

$$V_1(t) = I R_1 \left(1 - e^{-t/\tau}\right), \tag{4.7}$$

which rises toward the steady value IR_1; after the current stops at time t_{off} it decays as $V_1(t) = V_1(t_{\text{off}})\, e^{-(t-t_{\text{off}})/\tau}$. At $t = \tau$ the branch has reached $1 - e^{-1} = 63\,\%$ of its steady value; at 3τ, 95 %; at 5τ, 99.3 %.

With the illustrative values of Figure 4.2, $R_1 = 0.3\,\mathrm{m\Omega}$ and $\tau = 60$ s, the capacitance is $C_1 = \tau/R_1 = 60/0.0003 = 200{,}000$ F, the steady polarization at 100 A is $IR_1 = 30$ mV, and after one minute of current it has reached $30 \times 0.63 = 19$ mV. The same numbers make "rested long enough" of Section 4.1.2 precise. Starting from 30 mV, the polarization falls below 1 mV after $\tau \ln 30 = 60 \times 3.40 \approx 204$ s; a voltage read one minute into the rest is still 11 mV away from the open-circuit value. Real cells contain slower processes as well, with time constants of many minutes to hours, so in practice a rest of tens of minutes or more is used before a terminal voltage is accepted as an open-circuit voltage.

Figure 4.4 shows the role of τ. The horizontal axis is time; the vertical axis is the polarization voltage in millivolts during and after the same five-minute 100 A pulse (gray band), for $R_1 = 0.3\,\mathrm{m\Omega}$ and two time constants. With $\tau = 20$ s (blue) the polarization reaches its steady value within a minute and has fallen to 1.5 mV a minute after the pulse: a fast response and a short memory. With $\tau = 120$ s (purple) it has reached only 92 % of its steady value when the pulse ends, and it is still above 1 mV six minutes later: a slow response and a long memory. The idea is the same as thermal inertia in a building, where a large thermal time constant means the indoor temperature responds slowly to a heater and remembers it for a long time after it is switched off.

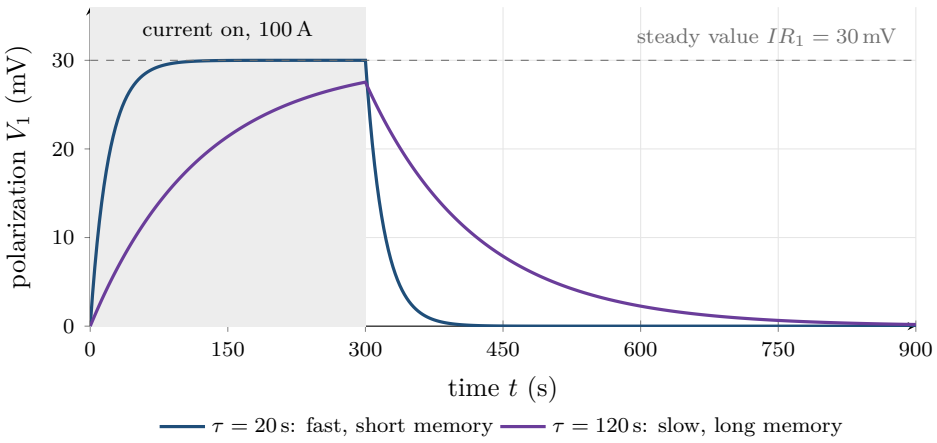


Figure 4.4: Polarization voltage during and after a five-minute 100 A pulse (gray band), from equation (4.6): $V_1(t) = IR_1(1 - e^{-t/\tau})$ for $t \le 300$ s and $V_1(300)\, e^{-(t-300)/\tau}$ afterwards, with illustrative $R_1 = 0.3\,\text{m}\Omega$ and $\tau = 20$ or 120 s. The time constant sets both how fast polarization builds and how long it is remembered.

Why a capacitor appears. There is no literal capacitor inside the cell. No pair of plates holds 200,000 F of charge separation; the capacitance is a mathematical element chosen because a resistor and a capacitor in parallel produce exactly the exponential build-up and decay of equation (4.7), which is the simplest description of a delayed electrochemical response. The same is true of every element of the equivalent circuit: the source, R_0, R_1 and C_1 are model abstractions that reproduce the cell's terminal behavior, in the way that the thermal resistance and thermal capacitance of a building model reproduce its temperature response without corresponding to any single wall or mass. Their values are fitted, and they are judged by whether the fitted circuit predicts the measured voltage, not by whether they can be found in a teardown.

4.3.5 The first-order model in three equations

Collecting the pieces, the first-order equivalent-circuit model of a cell is

$$\frac{\mathrm{dSOC}}{\mathrm{d}t} = -\frac{I}{C}, \tag{4.8}$$

$$\frac{\mathrm{d}V_1}{\mathrm{d}t} = -\frac{V_1}{R_1 C_1} + \frac{I}{C_1}, \tag{4.9}$$

$$V = U(\mathrm{SOC}) - IR_0 - V_1. \tag{4.10}$$

The first equation is coulomb counting from Chapter 2: the state of charge falls at the rate the current removes charge, divided by the capacity C. With C in ampere-hours, time is in hours; with time in seconds, the right-hand side is $-I/(3600\,C)$. The second is the polarization branch, equation (4.6). The third is the circuit relation, equation (4.5). Two equations describe how the cell's internal condition evolves; one says what the voltmeter sees.

For the reference cell of 300 Ah, one second at 100 A removes $100/3600 = 0.028$ Ah, which lowers the state of charge by $0.028/300 = 9.3 \times 10^{-5}$, about one hundredth of a percentage point. The whole five-minute pulse of Figure 4.2 removes 8.3 Ah, or 2.8 percentage points. On the flat part of the illustrative curve, whose slope is 0.05 V per unit of state of charge, that lowers U by only $0.05 \times 0.028 = 1.4$ mV, which is why the figure could hold U constant. The two dynamic equations act on very different time scales: the polarization state moves within minutes, the state of charge within hours.

4.4 States, measurements and identifiability

Equations (4.8)–(4.10) have the form of a standard dynamic system, and that form makes explicit which quantities are measured and which must be inferred. This section writes the model in state-space form, separates what the reference record observes from what it hides, and shows why the hidden quantities are hard to recover on an LFP cell.

4.4.1 State-space form

Define the state vector, the input and the output as

$$x = \begin{bmatrix} \mathrm{SOC} \\ V_1 \end{bmatrix}, \qquad u = I, \qquad y = V. \tag{4.11}$$

The model becomes

$$\dot{x} = \begin{bmatrix} 0 & 0 \\ 0 & -1/\tau \end{bmatrix} x + \begin{bmatrix} -1/C \\ 1/C_1 \end{bmatrix} u, \qquad y = U(x_1) - R_0\, u - x_2, \tag{4.12}$$

a linear state equation with a nonlinear output equation, because the open-circuit curve $U(\mathrm{SOC})$ is nonlinear. Figure 4.5 draws it as a block diagram. The measured input I enters on the left and feeds three paths: the charge integrator, whose output SOC passes through the open-circuit curve; the polarization branch, whose output is V_1; and the ohmic resistance. The three contributions are summed with the signs of equation (4.10) to give the measured output V on the right. The dashed box encloses the two states, which no sensor reads.

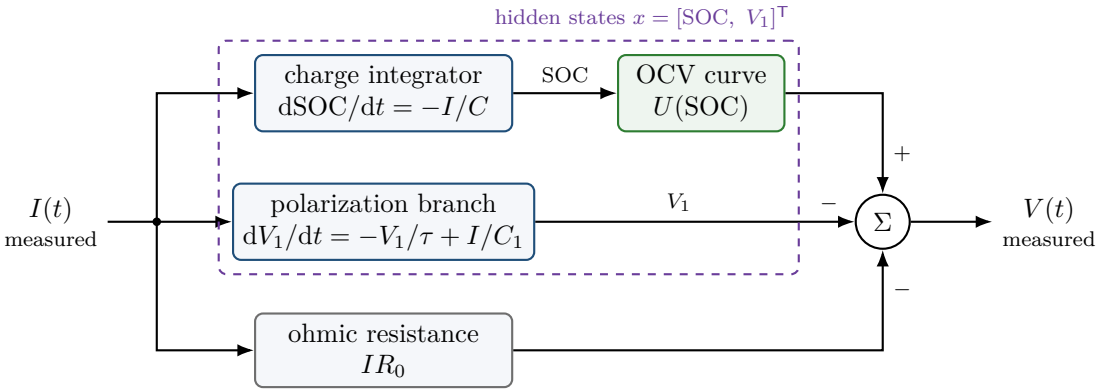


Figure 4.5: The first-order equivalent-circuit model, equation (4.12), as a block diagram. The input current and the output terminal voltage are measured; the state of charge and the polarization voltage inside the dashed box are not. The parameters C, R_0, R_1 and C_1 are not measured either.

This form is what connects battery modeling to a large body of general methods. A *Kalman filter* runs equation (4.12) forward from the measured current to predict the state, compares the predicted voltage with the measured one, and corrects the state in proportion to the difference; it is the standard state-of-charge estimator in battery management systems. *System identification* fits the parameters C, R_0, R_1 and C_1 from records of input and output. *Machine-learning* models of batteries either learn the output map directly or learn corrections to a model of this form. All of them start from the same structure: a hidden state driven by a known input and seen only through an output.

For computation the model is used in discrete time. Over a sampling interval Δt in seconds with the current held constant, the exact update is

$$\mathrm{SOC}[k+1] = \mathrm{SOC}[k] - \frac{\Delta t}{3600\,C}\, I[k], \qquad V_1[k+1] = e^{-\Delta t/\tau}\, V_1[k] + R_1\bigl(1 - e^{-\Delta t/\tau}\bigr)\, I[k], \tag{4.13}$$

where $[k]$ denotes the value at the k-th sample. With $\Delta t = 1\,\mathrm{s}$ and $\tau = 60\,\mathrm{s}$ the polarization keeps $e^{-1/60} = 0.9835$ of its value each second, which is the memory of Section 4.3.3 written as one number.

4.4.2 Observed and hidden quantities

Table 4.1 sorts the quantities of the model into those a sensor reads and those that must be inferred, and says how each appears in the reference record. The measured quantities are the current $I(t)$, the terminal voltage $V(t)$ and the temperature T. Everything else is hidden: the state of charge, the polarization state, and all four parameters C, R_0, R_1, C_1. The reported state of charge in the reference record is not an exception. It behaves like a coulomb count on the same current telemetry, so it is a derived channel, equation (4.8) evaluated with an assumed capacity, and not an independent measurement of the state.

Table 4.1: Measured and hidden quantities of the first-order model.

Quantity	Status	In the reference record
Current $I(t)$	measured	rack current, common to the 396 cells of a rack
Terminal voltage $V(t)$	measured	396 cell voltages per rack, at 1 mV resolution, and the rack voltage
Temperature T	measured	216 temperature sensors per rack
State of charge SOC	hidden	reported value is a coulomb count of the measured current, a derived channel
Polarization V_1	hidden	not reported
Capacity C	hidden	nominal 300 Ah; actual value per cell unknown
R_0, R_1, C_1	hidden	not reported; to be estimated

Estimating a battery's condition from its telemetry is therefore an estimation problem in the strict sense: two hidden states and four hidden parameters are to be recovered from two measured signals and a temperature, through a model. Whether that is possible depends on the model structure and on what the signals contain, and the rest of this section is about that dependence.

4.4.3 Competing explanations of a voltage drop

Suppose the terminal voltage of a cell is observed to drop. In equation (4.10), $V = U(\mathrm{SOC}) - IR_0 - V_1$, at least five changes produce that observation: the state of charge decreased, so U fell; the current increased, so IR_0 grew; the resistance increased, so IR_0 grew at the same current; the temperature fell, which raises R_0 and R_1; or the polarization increased, so V_1 grew. The voltmeter reports the sum. Which term is responsible is not immediately known.

Table 4.2 makes the competition concrete for a drop of 20 mV at 100 A, using the illustrative values of this chapter. Each row lists the explanation, the term of the equation it acts through, how large a change it needs to produce the 20 mV, and the time signature that could distinguish it. The sizes are all plausible: a 25 A current change, a 0.2 mΩ resistance increase, a cooling of about 7 K, or a state-of-charge change that is 40 points on the plateau but only 2.3 points on the steep part near SOC = 0.10.

The last column is what makes separation possible at all. Each explanation leaves a different signature in time and in its relation to the current: an ohmic change scales with the current and appears at steps; a polarization change relaxes with a characteristic time constant; a state-of-charge change accumulates with the integral of the current. Whether a given record lets those signatures be told apart is the question of *identifiability*: can the data determine the unknowns of the model uniquely? It depends on two things. *Model structure*: whether two

Table 4.2: Five explanations of a 20 mV drop in terminal voltage at 100 A (illustrative values: $R = 0.8\,\mathrm{m\Omega}$, curve slope 0.05 V per unit SOC on the plateau and 0.86 V at SOC = 0.10, temperature law of Section 4.2.2).

Explanation	Term	Change needed for 20 mV	Time signature
SOC decreased	$U(\mathrm{SOC})$	0.40 on the plateau; 0.023 at SOC = 0.10	slow; follows the integral of current
Current increased	IR_0	$\Delta I = 0.020/0.0008 = 25\,\mathrm{A}$	instantaneous; coincides with a current step
Resistance increased	IR_0	$\Delta R = 0.020/100 = 0.2\,\mathrm{m\Omega}$	persistent; scales with $\lvert I\rvert$
Temperature fell	R_0, R_1	factor 1.25 on R: about $25 \to 18\,^{\circ}\mathrm{C}$	slow; follows the temperature channel
Polarization grew	V_1	two thirds of the 30 mV steady value	relaxes with τ after a current change

different sets of states and parameters can produce exactly the same output for every input, in which case no data separates them. *Data richness*: whether the input actually contained the steps, pulses, rests and swings of state of charge that excite each signature. A record with constant current, for example, contains no step from which to read R_0 separately from U. Chapter 9 treats identifiability systematically; this chapter needs only the observation that the voltage equation mixes its causes.

4.4.4 The flat LFP curve

Lithium iron phosphate makes the problem harder than other chemistries. Its open-circuit curve has $\mathrm{d}U/\mathrm{dSOC} \approx 0$ over much of the middle range: on the illustrative curve of Figure 4.1 the slope between SOC = 0.3 and 0.8 is about 0.05 V per unit of state of charge, or 0.5 mV per percentage point. A ten-point change of state of charge moves the open-circuit voltage by about 5 mV, and the terminal voltage hardly moves at all relative to the 50–80 mV of the ohmic and polarization terms at 100 A. State-of-charge estimation from voltage is weak on the plateau. The reference record sits largely on that flat range: in the analysis of the record, 59 % of the analyzed windows lie on the flat part of the curve.

The state-space form of Section 4.4.1 shows the weakness exactly. Linearize the output equation (4.12) around an operating point: $y \approx U' x_1 - x_2 - R_0 u + \mathrm{const}$, where $U' = \mathrm{d}U/\mathrm{dSOC}$. The output row is $H = [\,U',\ -1\,]$ and the state matrix is $A = \mathrm{diag}(0, -1/\tau)$. A linear system's initial state can be recovered from its output if and only if the *observability matrix*, here H stacked on HA, has full rank:

$$\mathcal{O} = \begin{bmatrix} H \\ HA \end{bmatrix} = \begin{bmatrix} U' & -1 \\ 0 & 1/\tau \end{bmatrix}, \qquad \det \mathcal{O} = \frac{U'}{\tau}. \tag{4.14}$$

The determinant is proportional to the slope of the open-circuit curve. Where the curve is steep the state of charge is observable from voltage; where the curve is flat the determinant approaches zero and the state of charge becomes unobservable, however long the record. A Kalman filter reflects the same fact: the correction it applies to the state of charge is proportional to U', so on the plateau the filter corrects almost nothing and the estimate is carried by the coulomb count of equation (4.8), errors included.

4.4.5 Rest periods

A rest period is the one condition under which the voltage equation simplifies by itself. During rest $I \approx 0$, so $IR_0 \approx 0$; after enough time, $V_1 \to 0$ by equation (4.6); and equation (4.10) reduces to

$$V \approx U(\mathrm{SOC}). \tag{4.15}$$

Two of the three terms vanish without being estimated. A rested voltage gives clean access to the open-circuit curve: it places the cell on the curve, and two rests separated by a known amount of charge anchor the capacity, as Section 4.5.4 works out. The relaxation into the rest is also informative, since its shape is equation (4.7) in reverse and gives τ and R_1 directly. Laboratory characterization relies heavily on rests for these reasons.

The reference record archives charge and discharge operations but not the rest periods between them. A clean open-circuit reference is therefore unavailable: every voltage in the record carries an ohmic and a polarization term that must be modeled, and no reading places any cell on the open-circuit curve without them. Combined with the flat curve of Section 4.4.4, this limits what can be inferred from the record, and the limit is a property of what was archived, not of the method applied to it.

4.5 The flat curve and the separation of resistance from capacity

The flat open-circuit curve affects state of charge and capacity, but it leaves resistance largely untouched. This section quantifies the first effect, explains the difference through the two routes by which current reaches the voltage, and uses both to read the per-cell model of the reference record.

4.5.1 A thought experiment

Suppose a cell reads $V = 3.30\,\mathrm{V}$, and consider what that reading alone determines. By itself, not the state of charge: Figure 4.1 showed that the same reading arises at different states of charge depending on the current. Knowing that $I = 0$ helps, since it removes the ohmic term. Knowing that the cell has rested for a long time helps more, since it removes the polarization term and makes equation (4.15) hold. With the open-circuit curve in hand, the state of charge then follows by inverting it,

$$\mathrm{SOC} \approx U^{-1}(V). \tag{4.16}$$

For an LFP cell the inversion is poorly conditioned: on the plateau a tiny voltage error becomes a large state-of-charge uncertainty. Each piece of information in the sequence (current, rest, curve) removed one ambiguity; the last ambiguity is a property of the chemistry and no additional knowledge of the operating condition removes it.

4.5.2 The mathematics of the flat curve

A first-order expansion of the open-circuit curve gives

$$\Delta U \approx \frac{\mathrm{d}U}{\mathrm{dSOC}}\,\Delta\mathrm{SOC} \qquad \Longrightarrow \qquad \Delta\mathrm{SOC} \approx \frac{\Delta U}{\mathrm{d}U/\mathrm{dSOC}}. \tag{4.17}$$

For a given voltage uncertainty ΔU, the state-of-charge uncertainty is inversely proportional to the slope: a small slope means a large uncertainty. With an illustrative voltage uncertainty of $\pm 5\,\mathrm{mV}$, on the plateau where the slope is $0.05\,\mathrm{V}$ per unit,

$$\Delta\mathrm{SOC} \approx \frac{0.005}{0.05} = 0.10,$$

ten percentage points either way. On the steep part near SOC = 0.08, where the slope is 1.26 V per unit, the same 5 mV gives 0.005/1.26 = 0.004, under half a point.

Figure 4.6 draws both statements. The left panel is a close-up of the illustrative open-circuit curve, voltage against state of charge. Each shaded horizontal band is a ±5 mV uncertainty around a reading, 3.22 V (blue) on the steep part and 3.30 V (red) on the plateau, and each vertical band below it is the range of states of charge consistent with that reading. The blue band maps to a sliver; the red band maps to the whole interval from 0.40 to 0.60. The right panel plots equation (4.17) over the whole range: the state-of-charge uncertainty in percentage points, on a logarithmic scale, for a fixed 5 mV voltage uncertainty. It is below one point near the ends and close to ten points across the plateau.

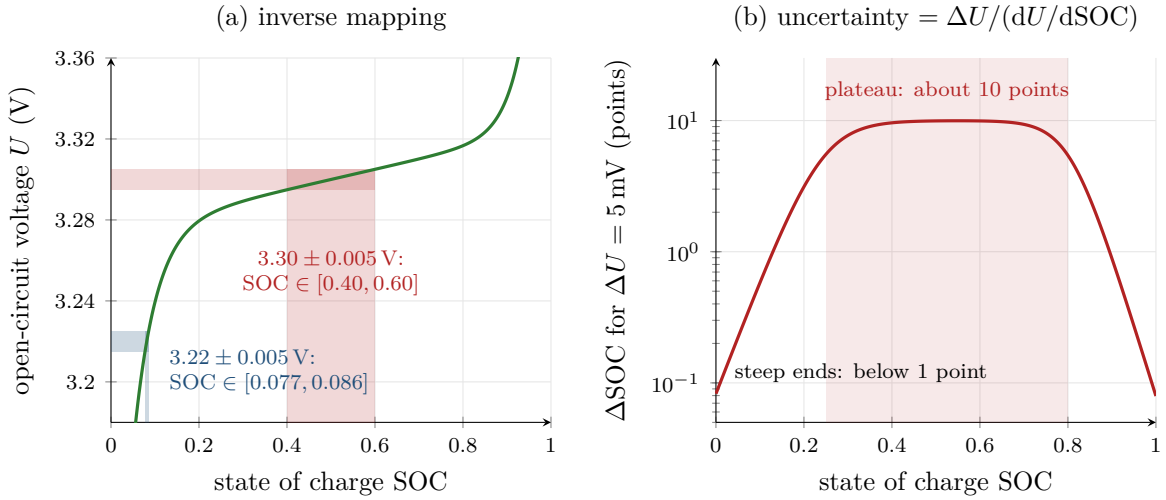


Figure 4.6: State-of-charge uncertainty on a flat and a steep part of the illustrative open-circuit curve of Figure 4.1. (a) A ±5 mV band around 3.22 V and around 3.30 V, projected through U^{-1}. (b) Equation (4.17), $100 \times 0.005/U'(\mathrm{SOC})$ percentage points with $U'(\mathrm{SOC}) = 0.05 + 6\,e^{-\mathrm{SOC}/0.05} + 6.25\,e^{-(1-\mathrm{SOC})/0.04}$ V per unit; the shaded interval is the plateau.

The voltage uncertainty in equation (4.17) has several sources, and on the plateau each of them is amplified twenty times. Quantization alone is one: the reference record reports cell voltages at 1 mV resolution, which on the plateau is already 0.001/0.05 = 0.02, two points of state of charge. Under current the ohmic term adds more: if R_0 is known only to within ±0.1 mΩ, the predicted drop at 100 A is uncertain by ±10 mV, which becomes ±20 points. This is the quantitative reason why rests matter (Section 4.4.5) and why their absence from the reference record matters.

Principle 4.2. *On an open-circuit curve of slope* $\mathrm{d}U/\mathrm{dSOC}$*, a voltage uncertainty* ΔU *becomes a state-of-charge uncertainty* $\Delta U/(\mathrm{d}U/\mathrm{dSOC})$*. On the LFP plateau, where the slope is about* 0.05 V *per unit, every millivolt of voltage uncertainty is two points of state of charge.*

4.5.3 Resistance and capacity act through different routes

Resistance and capacity both change the terminal voltage, but they act through different routes, and the difference is what allows them to be separated. Resistance acts through the term IR_0: its effect appears the instant the current changes and is proportional to the change. Capacity acts through the state of charge,

$$\mathrm{SOC}(t) = \mathrm{SOC}_0 - \frac{1}{C}\int_0^t I\,\mathrm{d}t', \tag{4.18}$$

and then through the open-circuit curve $U(\mathrm{SOC})$: its effect is tied to the accumulated charge history and grows only as charge is moved.

Figure 4.7 draws the two routes. The measured current on the left feeds an upper route, in which it is multiplied by R_0 and appears at once as a voltage drop, and a lower route, in which it is integrated, divided by C, turned into a state of charge, and passed through the open-circuit curve. Both routes end at the terminal voltage. The annotations under each route state the time scale on which it acts and the parameter it reveals: current, then ohmic drop, then resistance; current, then integration, then state of charge, then open-circuit change, then capacity.

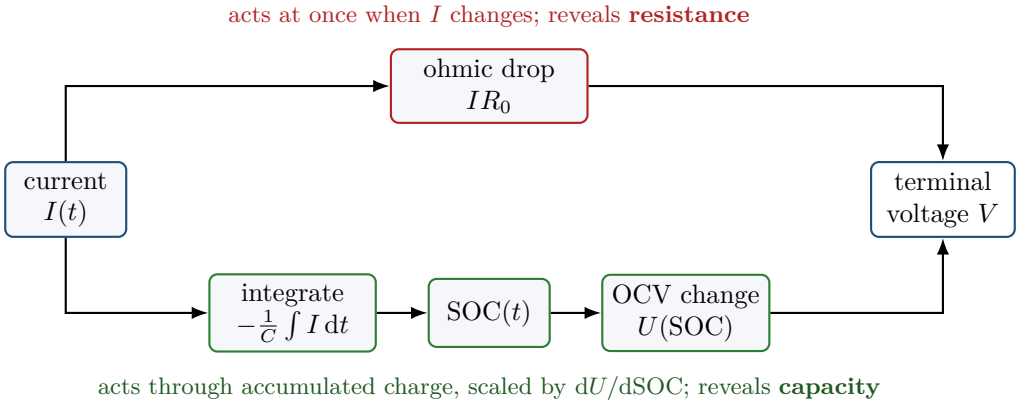


Figure 4.7: Two routes from current to terminal voltage. Upper: the ohmic route, instantaneous, whose size reveals the resistance. Lower: the charge route, cumulative, whose size reveals the capacity but is multiplied by the slope of the open-circuit curve and therefore nearly vanishes on the LFP plateau.

A worked comparison shows how unequal the two routes are on the plateau. Take two cells of the reference type, one at the nominal 300 Ah and one faded to 270 Ah, discharged together at 100 A for one hour. The charge removed is 100 Ah for both; by equation (4.18) their states of charge fall by $100/300 = 0.333$ and $100/270 = 0.370$. The difference, 0.037, reaches the voltage through the plateau slope as $0.05 \times 0.037 = 1.9\,\mathrm{mV}$, after a full hour. Now take two cells of equal capacity whose resistances are 0.5 and 1.0 mΩ: at the first instant of the same 100 A discharge their voltages differ by $100 \times 0.0005 = 50\,\mathrm{mV}$. A ten-percent capacity difference produces a two-millivolt signal that takes an hour to build; a resistance difference produces a fifty-millivolt signal at every current step. Resistance is the easier parameter to estimate on an LFP cell, and capacity is the harder one, for this reason.

Principle 4.3. *Resistance acts instantaneously, through IR_0, and is revealed by changes of current. Capacity acts cumulatively, through the integral of current and the open-circuit curve, and is revealed only by moving enough charge across a part of the curve with enough slope.*

4.5.4 The per-cell model of the reference record

The per-cell model used in the analysis of the reference record is more elaborate than $V = U - IR$, but it is grown from it. It has five ingredients: the cell voltage being explained, an open-circuit voltage term, a cell resistance, a capacity-related term, and an operation-specific offset. One way to write a model with these ingredients, for cell i during operation k (one archived charge or discharge), is

$$V_i(t) = U\big(\mathrm{SOC}_i(t)\big) - I(t)\,R_i + b_{i,k}, \qquad \mathrm{SOC}_i(t) = \mathrm{SOC}_{i,k} - \frac{1}{C_i}\int_{t_k}^{t} I\,\mathrm{d}t', \tag{4.19}$$

where R_i is the cell's resistance, C_i its capacity, $\mathrm{SOC}_{i,k}$ its state of charge at the start t_k of the operation, and $b_{i,k}$ a constant offset for that cell in that operation. The first four ingredients are the chapter's model with the polarization folded into the resistance; the offset needs explanation.

Every operation in the record starts at a state of charge that no rested voltage pins down, because the rests are not archived (Section 4.4.5). On the plateau, an error δ in the starting state of charge shifts the open-circuit term by $U(\mathrm{SOC}_{i,k} + \delta) - U(\mathrm{SOC}_{i,k}) \approx U' \delta$, a constant throughout the operation: an error of $\delta = 0.05$ shifts the voltage by $0.05 \times 0.05 = 2.5\,\mathrm{mV}$ from the first sample to the last. A constant voltage shift is exactly what the offset $b_{i,k}$ represents, so the offset absorbs the unknown starting point, and the model needs it because the absolute starting point on the open-circuit curve is not available from rest data. What remains visible is how each cell *departs* from the common trajectory of its rack: all 396 cells carry the same current, so differences between cells come from differences in R_i, in C_i through the integral, and in the offsets. What is not available is the independent information that would anchor absolute cell capacity.

With rests, that anchor would be direct. Two rests, one near full and one near empty on the steep parts of the curve, would give the two states of charge from equation (4.16); the charge moved between them is measured; and the capacity follows as charge over state-of-charge change. For example, rests placing a cell at 0.95 and 0.10 with 255 Ah moved between them give $C = 255/(0.95 - 0.10) = 300\,\mathrm{Ah}$. Without rests, the per-cell model can rank cells by how they depart from the rack, but it cannot convert that ranking into absolute ampere-hours with confidence. That is a statement about the information in the record, which is a deeper statement than "the fit was bad": a better optimizer or more samples of the same operations would not change it.

4.6 The hierarchy of models and model choice

The first-order equivalent circuit is one rung of a ladder of battery models of increasing detail. This section lays out the ladder, states which question each rung answers, and closes the chapter's body with the picture that connects the pieces.

4.6.1 A hierarchy of models

Figure 4.8 draws five levels as the rungs of a ladder, from the simplest at the bottom to the most detailed at the top. Each rung gives the level, its name, its governing equation and what it captures; the arrow on the left marks the direction in which physics, parameters and data requirements all increase.

Level 0, the energy bucket, is coulomb counting, equation (4.18): a tank whose level falls with the current drawn. It does energy and charge accounting and knows nothing about voltage. *Level 1, static resistance*, is equation (4.1), $V = U(\mathrm{SOC}) - IR$: the terminal voltage follows the current instantly, with no memory. *Level 2, the first-order equivalent circuit*, is equations (4.8)–(4.10): the RC state adds one memory with one time constant. *Level 3, higher-order equivalent circuits*, adds further RC branches,

$$V = U(\mathrm{SOC}) - IR_0 - V_1 - V_2 - \cdots , \tag{4.20}$$

each with its own state equation and time constant, to represent several time scales at once, for example a fast charge-transfer process of seconds and a slow diffusion process of many minutes. *Level 4, electrochemical models*, abandons the circuit analogy and solves partial differential equations for lithium diffusion in the electrode particles, reaction kinetics at their surfaces, the lithium concentration profiles through the electrodes, and transport in the electrolyte. Level 4 is the most faithful and needs geometry, material properties and laboratory data that field telemetry does not supply; it is not needed for anything in this book.

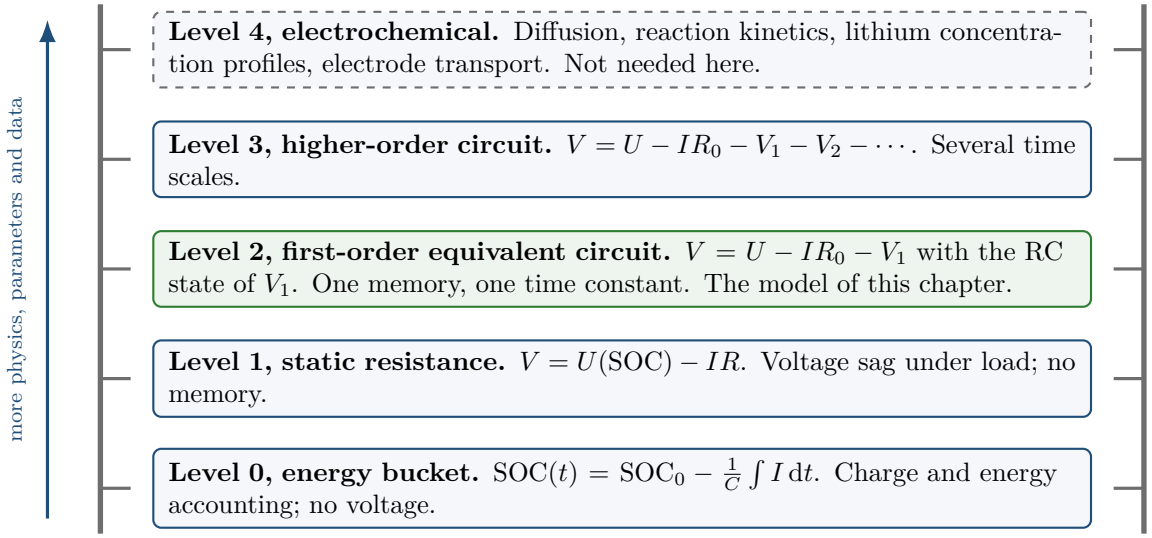


Figure 4.8: The hierarchy of battery models as a ladder. Each rung adds physics to the one below it and requires more parameters and richer data to fit. The first-order equivalent circuit (green) is the model of this chapter; the electrochemical level (dashed) is beyond what field telemetry supports.

Each level contains the one below it as a special case. Level 2 with $R_1 = 0$ is level 1; level 1 with the voltage equation discarded is level 0. Moving up the ladder never contradicts a lower level; it adds terms that the lower level absorbed into its parameters or ignored.

4.6.2 Matching the model to the question

The right level is set by the question, not by a preference for detail. Table 4.3 lists four questions and the smallest model that answers each. How long until the state of charge reaches 80 % needs only level 0: a 300 Ah cell at 95 % discharging at 150 A must deliver $0.15 \times 300 = 45$ Ah, which takes $45/150 = 0.3$ h, or 18 minutes. The expected terminal voltage during a current pulse needs level 1: at $U = 3.30$ V, 200 A and $R = 0.8$ mΩ it is $3.30 - 0.16 = 3.14$ V, which can be compared with the cell's lower voltage limit. Why the voltage relaxes after a current change needs level 2: only the RC state produces a voltage that moves at constant current, recovering 63 % of IR_1 in one time constant. What happens inside the electrode particles needs level 4, and no field record answers it.

Table 4.3: Questions and the smallest model that answers each.

Question	Model	Level
Time until SOC reaches 80 %	$\mathrm{dSOC}/\mathrm{d}t = -I/C$	0, energy bucket
Terminal voltage during a current pulse	$V = U(\mathrm{SOC}) - IR$	1, static resistance
Why the voltage relaxes after a current change	RC state, equation (4.6)	2, first-order circuit
What happens inside the electrode particles	diffusion and kinetics equations	4, electrochemical

A higher level is not free. Each rung adds parameters, and each parameter must be identified from data that excite it: an RC branch whose time constant is never excited by a current step is a parameter the data cannot determine, and fitting it anyway produces a number without

meaning. The model should be as detailed as the question requires and no more detailed than the data can support.

Principle 4.4. *Choose the lowest level of the model hierarchy that answers the question. Each additional level adds parameters that the data must be rich enough to determine.*

4.6.3 The picture

The chapter's content can be gathered into one picture of what happens between the chemistry and the voltmeter. The hidden chemistry sets the state of charge, and the state of charge sets the open-circuit voltage $U(\mathrm{SOC})$. The current I then acts in three ways: through the ohmic drop IR_0, through the polarization V_1, and through heating, I^2R, which at 100 A and 0.8 mΩ is $100^2 \times 0.0008 = 8$ W per cell and $8 \times 396 \approx 3.2$ kW per rack. The first two subtract from U to give the terminal voltage V; the third is the subject of Chapter 6, and it feeds back on the voltage because temperature changes the resistances. Schematically,

$$\text{hidden chemistry} \longrightarrow \mathrm{SOC} \longrightarrow U(\mathrm{SOC}), \qquad I \longrightarrow \{\, IR_0,\ V_1,\ I^2R \,\} \longrightarrow V,$$

and the core equations are the three of Section 4.3.5:

$$\frac{\mathrm{dSOC}}{\mathrm{d}t} = -\frac{I}{C}, \qquad V = U(\mathrm{SOC}) - IR_0 - V_1, \qquad \frac{\mathrm{d}V_1}{\mathrm{d}t} = -\frac{V_1}{R_1C_1} + \frac{I}{C_1}.$$

Figure 4.5 is the same picture drawn as a block diagram, with the hidden states marked.

4.7 What to retain

Definition 4.1 (The equivalent-circuit model)**.** The terminal voltage of a cell is $V = U(\mathrm{SOC}) - IR_0 - V_1$: the open-circuit voltage set by the state of charge, minus an instantaneous ohmic drop, minus a polarization voltage with memory. Measured voltage is a mixture of these causes, so the state of charge, resistance and capacity are inferred, not read, and how well they can be inferred depends on the slope of the open-circuit curve and on whether the data contain steps and rests.

Table 4.4 lists the chapter's objects with the subsection that defines each.

4.8 Exercises

Exercise 4.1. A cell at rest reads 3.310 V. The current steps from 0 to 150 A discharge; within a fraction of a second the voltage is 3.235 V, and after several minutes at constant current it has settled at 3.190 V. (a) Estimate R_0, R_1 and the lumped R of equation (4.4). (b) If $\tau = 60$ s, what is the terminal voltage one minute after the step? (c) What is the heating I^2R of the cell at this current, using the settled resistance?

Exercise 4.2. An RC branch has $R_1 = 0.4$ mΩ and $C_1 = 150{,}000$ F. (a) Compute τ. (b) A 200 A discharge pulse lasts 120 s from $V_1 = 0$; compute V_1 at the end of the pulse. (c) How long must the cell then rest before V_1 is below 2 mV? (d) Write the discrete-time coefficient $e^{-\Delta t/\tau}$ for $\Delta t = 1$ s.

Exercise 4.3. An LFP cell's open-circuit curve has slope 0.04 V per unit of state of charge on its plateau and 1.5 V per unit near full. (a) For a voltage uncertainty of ±2 mV, compute the state-of-charge uncertainty in each region with equation (4.17). (b) The cell is read at 100 A on the plateau with R_0 known to ±0.05 mΩ; what state-of-charge uncertainty does the resistance alone contribute? (c) State what measurement condition would remove that contribution.

Table 4.4: The objects of the chapter.

Object	Expression or role	Defined in
Open-circuit voltage	$U(\mathrm{SOC})$, the rested voltage	Section 4.1.2
Static model	$V = U(\mathrm{SOC}) - IR$	Section 4.1.1
Lumped resistance	$R(\mathrm{SOC}, T, \mathrm{age})$	Section 4.2.2
Step estimate of resistance	$R \approx -\Delta V/\Delta I$; depends on read time	Section 4.2.3
Polarization state	$\mathrm{d}V_1/\mathrm{d}t = -V_1/\tau + I/C_1$, $\tau = R_1C_1$	Section 4.3.4
First-order model	equations (4.8)–(4.10)	Section 4.3.5
State-space form	$x = [\mathrm{SOC}, V_1]^\mathsf{T}$, $u = I$, $y = V$	Section 4.4.1
Observability on the plateau	$\det \mathcal{O} = U'/\tau \to 0$	Section 4.4.4
Rest condition	$V \approx U(\mathrm{SOC})$	Section 4.4.5
Flat-curve uncertainty	$\Delta\mathrm{SOC} \approx \Delta U/(\mathrm{d}U/\mathrm{dSOC})$	Section 4.5.2
Two routes	ohmic (resistance) and cumulative (capacity)	Section 4.5.3
Operation-specific offset	$b_{i,k}$ absorbs the unknown starting point	Section 4.5.4
Model hierarchy	levels 0 to 4	Section 4.6.1

Exercise 4.4. Two cells in the same series string, of 300 Ah and 285 Ah, are discharged together at 120 A for 1.5 h on the plateau, where the slope is 0.05 V per unit. (a) Compute the difference in their state-of-charge changes and the resulting difference in open-circuit voltage at the end. (b) Compare it with the voltage difference produced at the first instant by a 0.1 mΩ difference in resistance at the same current. (c) With cell voltages reported at 1 mV resolution, which of the two differences is easier to detect, and why does Section 4.5.3 predict this?

4.9 Connection to the TERRA framework

The equivalent-circuit model is the per-cell component relation that every later chapter attaches to the graph.

- **Coulomb counting is a charge balance on a storage node.** Equation (4.8) states that stored charge changes by the current crossing the cell's boundary. In TERRA it is a conservation node with one storage state, SOC, and the capacity C as its parameter. The reported state of charge of the reference record is the same equation evaluated on the same current, so comparing it with the node's state is an identity, not a test.

- **The series string is one edge feeding 396 component relations.** The current I is the flow on the string edge and is common to every cell. Each cell's equation (4.10) is a component relation driven by that one flow, so the 396 cell voltages are 396 views of the same input through 396 sets of states and parameters $(\mathrm{SOC}_i, V_1, R_i, C_i)$, one per cell, which is what makes the cell-to-cell departures of Section 4.5.4 visible.

- **The polarization voltage is a component state with its own dynamics.** Equation (4.6) gives the cell component a second state beyond SOC. Forward simulation in time integrates both, in the discrete form of equation (4.13), and reproduces the step response of Figure 4.2.

- **Hidden states and parameters are posterior quantities.** SOC, V_1, R_0, R_1, C_1 and C are inferred from the measured I and V. The determinant U'/τ of equation (4.14) is the structural reason a posterior over SOC stays wide on the plateau, and the missing rests

are the practical reason the absolute capacity stays wide; TERRA asks both questions before estimating.

- **An offset parameter limits what can be detected.** The operation-specific offset $b_{i,k}$ absorbs any constant voltage shift. A fault whose signature is also a constant shift, such as a sensor bias, would be absorbed by it; a fault hypothesis is detectable only if its signature lies outside what the normal model can absorb, which is the subject of Chapter 8.

In summary, the first-order equivalent circuit turns one cell into a component with two states, four parameters, one measured input and one measured output, and the chapter's limits (the flat curve, the absent rests, the offset) are statements about which of those six unknowns the measurements can determine. The next chapter moves from the single cell to the hardware that assembles 3,564 of them into the reference container and measures them.

Chapter 5

The hardware around the cells: modules, racks, busbars, fuses, contactors, shunts, the BMS and the PCS

5.1 From cell to grid asset

Chapter 4 described one cell by the relation $V \approx U(\mathrm{SOC}) - IR$: an open-circuit voltage that depends on the state of charge, minus a resistive drop proportional to the current. A grid battery is not one cell. The reference container holds 3,564 cells, and between those cells and the grid sit conductors, protective devices, switches, sensors, a supervisory computer and a power converter. Each of these pieces adds a resistance, a measurement, a limit or a decision, and each changes what the telemetry of the system can say about the cells inside it. The question of this chapter is

How do thousands of cells become one controllable grid asset, and which of the quantities that describe that asset are measured and which are computed?

The answer is organized around the hardware. The cells are grouped into modules and racks; the racks are joined by heavy conductors, protected by fuses, switched by contactors and metered by shunts; a battery management system watches every cell; and a power conversion system ties the whole assembly to the alternating-current grid. Three relations come out of that hardware and are used in every later chapter: a voltage identity along each rack, a voltage equality across the racks, and a current balance at the point where the racks meet.

Jargon: module, rack, busbar, fuse, contactor, shunt, DC bus, balancing, redundancy, EMS

A *module* is a mechanical and electrical group of cells connected in series inside one enclosure. A *rack* is a cabinet of modules connected in series, forming one long string of cells that carries a single current. A *busbar* is a heavy metal bar or plate that carries large current between cells, modules and racks with very small resistance. A *fuse* is a protective element that melts and opens the circuit permanently when the current is dangerously large for long enough. A *contactor* is an electrically commanded high-current switch that connects or isolates a rack and can be opened and closed many times. A *shunt* is a precisely known small resistance placed in the current path so that the current can be computed from the small voltage across it. The *DC bus* (direct-current bus) is the common pair of conductors to which all racks are connected in parallel and from which the power converter draws or to which it delivers current. *Balancing* is the reduction of state-of-charge differences among the cells of a series string. *Redundancy* means that the same physical quantity is determined by two independent routes, so that a disagreement between them reveals an error. The *EMS* (energy management system), also called the site controller, decides what the battery as a whole should do: when to charge, when to discharge and at what power.

Key idea

A megawatt-hour battery is a hierarchy of small, manageable units joined by hardware that has its own resistance, its own sensors and its own limits, and the value of each sensor depends on whether another, independent measurement constrains the same quantity.

This section builds the hierarchy from the cell upward and describes the module and the rack, the two levels at which the cells of the reference container are physically organized.

5.1.1 The hierarchy

A battery energy storage system is assembled in levels. The chain from the smallest unit to the grid is

$$\text{cell} \rightarrow \text{module} \rightarrow \text{rack} \rightarrow \text{battery container (the DC system)} \rightarrow \text{power conversion system} \rightarrow \text{grid}.$$

For the reference container the counts at each level are fixed by its design: 44 cells are connected in series to form one module, nine modules are connected in series to form one rack of $9 \times 44 = 396$ cells, and nine racks are connected in parallel to form the container, for a total of $9 \times 396 = 3{,}564$ cells and an energy of about 3.42 MWh.

Figure 5.1 draws this chain. Each box is one level; the label on each arrow says how many units of the level on the left make one unit of the level on the right and whether they are joined in series or in parallel; the lower row continues from the container to the power conversion system and the grid. The numbers in the boxes follow from two rules. In series, voltages add and the capacity (the charge in ampere-hours that the string can deliver) stays that of one cell, because the same current passes through every cell. In parallel, capacities add and the voltage stays that of one branch, because every branch sits between the same two conductors. The nominal cell voltage of the reference container's lithium iron phosphate cells is 3.2 V, and dividing the container energy by the cell count and that voltage gives the nominal cell capacity,

$$\frac{3.42 \times 10^6\,\text{Wh}}{3{,}564 \times 3.2\,\text{V}} \approx 300\,\text{Ah},$$

so each cell stores about $3.2 \times 300 = 960$ Wh, each module $44 \times 960 \approx 42.2$ kWh, each rack $9 \times 42.2 \approx 380$ kWh, and the container $9 \times 380 \approx 3.42$ MWh.

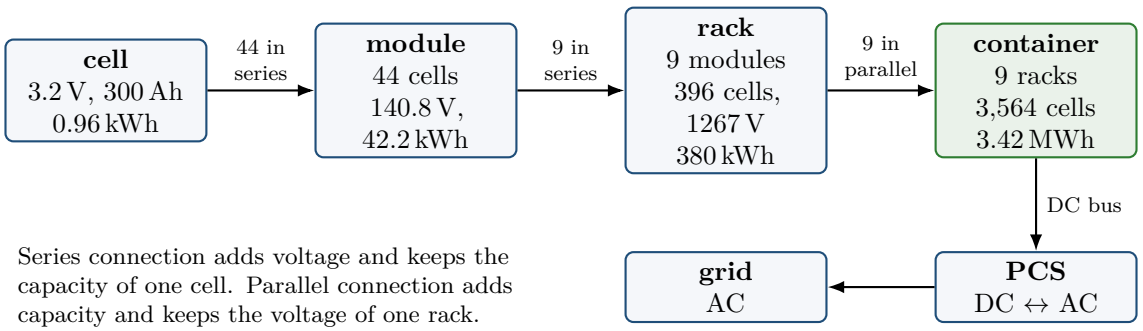


Figure 5.1: The hierarchy of the reference container, from one cell to the grid. Voltages and energies are nominal: 3.2 V and 300 Ah per cell, so $44 \times 3.2 = 140.8$ V per module, $396 \times 3.2 \approx 1267$ V per rack, and $3{,}564 \times 960$ Wh ≈ 3.42 MWh for the container.

The reason for the hierarchy is practical. No single electrochemical cell can be built to store megawatt-hours; a cell the size of a container would be impossible to manufacture, cool, monitor or repair. A megawatt-hour battery is therefore built from many manageable units, each small enough to be produced in volume, tested, monitored individually and replaced. The price of that choice is the subject of this chapter: every level of the hierarchy adds connections, protection and measurement, and every one of those has physical consequences.

5.1.2 The module

A module is the first grouping above the cell. Mechanically it is an enclosure that holds the cells in compression, carries their connections and gives the cooling air or liquid a defined path. Electrically, in the reference container, it is 44 cells in series, so its nominal voltage is

$$44 \times 3.2\,\text{V} = 140.8\,\text{V}. \tag{5.1}$$

A single module is therefore already a high-voltage object, well above the level at which direct-current voltage is treated as hazardous to touch. A module is handled, shipped and replaced as one unit, and it usually carries its own sensing board, which measures the voltage of each of its cells and the temperature at several points.

Nine modules in series give a nominal rack voltage of $9 \times 140.8 = 1267.2$ V, about 1267 V. The working voltage moves around this value as the state of charge and the current change: a lithium iron phosphate cell rests near 3.3 V over most of its range, so the rack rests near $396 \times 3.3 \approx 1300$ V in the middle of its state-of-charge range, which is the value used in the examples of this chapter. The module is a convenient unit for manufacturing and maintenance, but it is not the unit that decides the electrical behavior: that is the rack.

5.1.3 The rack

The rack is the major DC branch of the system. It is a single series string of 396 cells from its positive terminal, through modules 1 to 9, to its negative terminal, and it carries one current. A useful picture is a line of runners chained together: every runner covers the same distance at the same pace, and the slowest one sets the pace of all. In the same way every cell of a rack passes the same current, and the weakest cell sets the limits of the whole string.

Because voltages add in series, the rack voltage is the sum of the 396 cell voltages, less whatever voltage is lost in the connections that join them,

$$V_{\text{rack}} = \sum_{i=1}^{396} V_i - (\text{connection drops}). \tag{5.2}$$

Equation (5.2) is the seed of the string voltage identity of Section 5.3. The connection drops are not abstract: they are the voltages lost across specific pieces of hardware that the current must pass through on its way out of the rack.

Figure 5.2 draws that path. The nine modules are stacked vertically, as they are in the cabinet, with module 1 at the top; the thick gray segments between them are the busbars that join the negative terminal of one module to the positive terminal of the next. From module 1 the path runs through a fuse and a contactor to the positive rack terminal; from module 9 it runs through the shunt and a second contactor to the negative rack terminal. On the left, the cell voltage taps of every module feed the battery management system; on the right, a separate sensor measures the voltage between the two rack terminals. The figure shows that the rack current passes every cell, every busbar, the fuse, both contactors and the shunt in turn: one current, one path, many resistances.

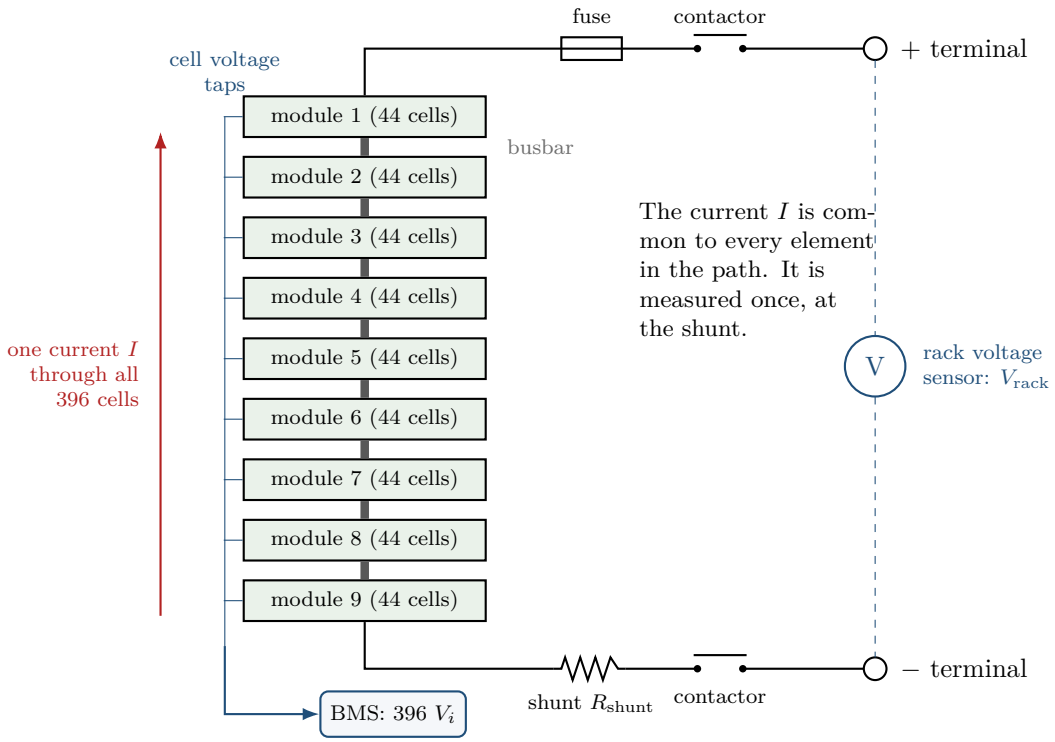


Figure 5.2: The series path of one rack of the reference container. On discharge the current enters at the negative terminal, passes a contactor and the shunt, rises through modules 9 to 1 joined by busbars, and leaves through a fuse and a second contactor at the positive terminal. The cell voltage taps (left) feed the battery management system; the rack voltage sensor (right) measures between the terminals. The placement of the fuse and contactors is schematic.

5.2 Conductors, protection and current measurement

This section takes the hardware of Figure 5.2 one piece at a time: the conductors that carry the current, the two devices that can interrupt it, and the shunt that measures it.

5.2.1 Busbars

A busbar is a heavy conductor, a bar or plate of copper or aluminum, sized to carry large currents with very little resistance. Busbars connect cells inside a module, modules inside a rack, racks to the DC bus, and the DC bus to the contactors, fuses and power converter. They

are used instead of cables wherever the current is large and the route is short and fixed, because a bar of large cross-section has a lower resistance and sheds heat better than a flexible cable of the same material.

A busbar's resistance is small but never zero. Whatever its value R, a current I through it produces a voltage drop and a heat generation,

$$V_{\text{drop}} = IR, \qquad P_{\text{heat}} = I^2 R, \tag{5.3}$$

by Ohm's law and Joule's law. The drop is lost from the voltage the cells produce before it reaches the rack terminals, and the heat must be carried away by the cooling system. The two quantities scale differently: the drop is proportional to the current, the heat to its square.

5.2.2 Resistance of a connection

A milliohm sounds negligible. Take a connection of $R = 1\,\text{m}\Omega$, which is a plausible value for a bolted joint between two busbars or for a short length of conductor. At $I = 100\,\text{A}$ equation (5.3) gives

$$V_{\text{drop}} = 100\,\text{A} \times 0.001\,\Omega = 0.1\,\text{V}, \qquad P_{\text{heat}} = (100\,\text{A})^2 \times 0.001\,\Omega = 10\,\text{W},$$

and at $I = 500\,\text{A}$

$$V_{\text{drop}} = 500 \times 0.001 = 0.5\,\text{V}, \qquad P_{\text{heat}} = 500^2 \times 0.001 = 250\,\text{W}.$$

Five times the current gives five times the drop but twenty-five times the heat. Figure 5.3 plots both relations: the left panel shows the voltage drop in volts against the current in amperes, the right panel the heat in watts against the same current, both for the same $1\,\text{m}\Omega$ connection. The left curve is a straight line; the right curve is a parabola that is flat at small current and steep at large current. Two hundred and fifty watts concentrated in one joint is the output of a small electric heater, delivered to a spot the size of a bolt head.

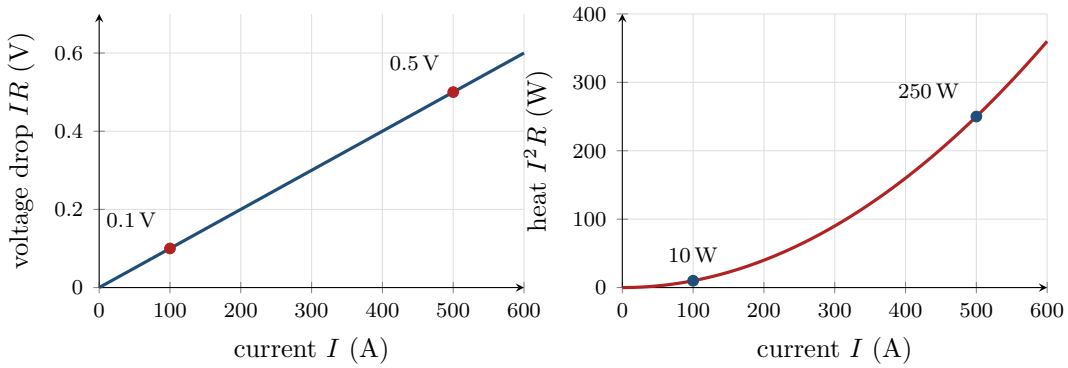


Figure 5.3: Voltage drop (left) and heat (right) of a single connection of $R = 1\,\text{m}\Omega$, from equation (5.3): $V_{\text{drop}} = IR$ and $P_{\text{heat}} = I^2 R$. The marked points are 100 A and 500 A.

Connections are where this matters most. A bolted joint whose torque has relaxed, a contact surface that has oxidized, or a crimp that has loosened under thermal cycling can rise from a fraction of a milliohm to several milliohms. Its heat rises in proportion, the heat raises its temperature, the temperature accelerates the oxidation, and the resistance rises further. A poor connection therefore becomes a thermal problem, and in the worst case a fire hazard, well before it becomes an electrical one. The same fact gives a diagnostic opportunity: if the resistance of a connection can be estimated from voltage and current measurements, a degraded connection reveals itself as a resistance that grows over months. Section 5.3 shows how the rack-level measurements provide exactly such an estimate.

5.2.3 The fuse

A fuse is a protective element: a calibrated metal link that melts and opens the circuit permanently if the current is dangerously large for long enough. The phrase "for long enough" matters. A fuse responds to the heat accumulated in its link, which grows with I^2 multiplied by time, so a modest overload is tolerated for a long time, a large overload for a short time, and a short circuit for only milliseconds. Fuses protect against short circuits (a direct low-resistance path between the two rack poles, through which the cells would drive thousands of amperes), catastrophic overcurrent, and equipment faults in the converter or the bus.

A fuse is a one-shot device. Once it has operated it must be found and replaced; it does not reset. It is therefore the last line of protection, intended to act only when every controlled means of stopping the current has failed or was too slow. In normal operation a fuse is simply one more small resistance in the series path of Figure 5.2, and its drop and heat are part of the connection drops of equation (5.2).

5.2.4 The contactor

A contactor is an electrically controlled high-current switch between a rack and the DC bus. A small control current energizes a coil, the coil pulls a set of heavy contacts together, and the rack is connected; removing the control current lets a spring pull the contacts apart, and the rack is isolated. Closed means connected; open means isolated. The battery management system or the higher-level control system commands it, for fault protection, for maintenance, for shutdown, when a voltage or a temperature is outside its allowed range, and for emergency isolation.

Unlike a fuse, a contactor opens and closes repeatedly, and that is its purpose: it is the controlled means of connecting and disconnecting a rack. A rack whose contactor is open contributes nothing to the container and draws nothing from the bus; it is electrically absent even though its cells are still charged and still dangerous. Like the fuse, a closed contactor is a small resistance in the series path, and the resistance of its contacts grows slowly with wear.

5.2.5 The current shunt

Measuring a large current directly is difficult; measuring a small voltage precisely is comparatively easy. A shunt converts the first problem into the second. It is a resistor of small, precisely known resistance R_{shunt} placed in the current path, so that the voltage across it is proportional to the current, and the current follows from Ohm's law,

$$I = \frac{V_{\text{shunt}}}{R_{\text{shunt}}}. \tag{5.4}$$

Example 5.1 (A shunt reading)**.** A shunt of $R_{\text{shunt}} = 100\,\mu\Omega = 0.0001\,\Omega$ shows a voltage of $V_{\text{shunt}} = 10\,\text{mV} = 0.010\,\text{V}$. By equation (5.4),

$$I = \frac{0.010\,\text{V}}{0.0001\,\Omega} = 100\,\text{A}.$$

At 500 A the same shunt shows 50 mV. The measurement electronics must therefore resolve tens of microvolts to resolve a fraction of an ampere: 0.1 A corresponds to $10\,\mu$V.

Figure 5.4 draws the arrangement. The heavy horizontal conductor carries the rack current I from left to right through the resistive element (the zigzag) between two massive current terminals. Two thin sense leads leave the element at points just inside the current terminals and run down to a voltmeter, which reads V_{shunt}. The separation of the two sets of terminals is deliberate: the large current flows through the heavy terminals, and the voltmeter, which

draws almost no current, measures only the drop across the calibrated element and not the drop across the bolted joints at its ends. This four-terminal arrangement is what makes the reading depend on R_{shunt} alone.

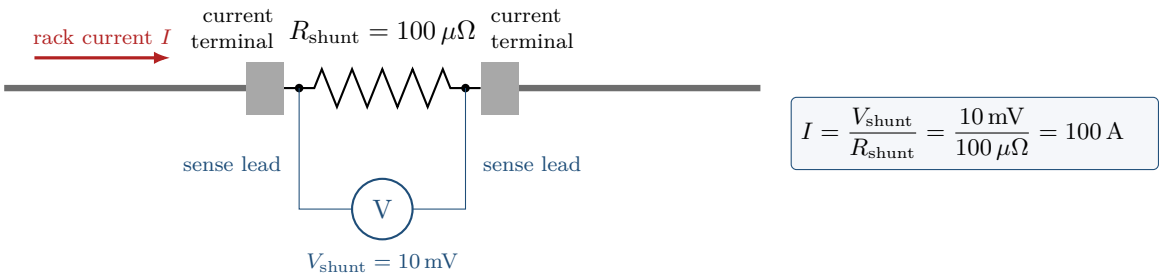


Figure 5.4: Current measurement with a shunt. The rack current passes through a calibrated resistance $R_{\text{shunt}} = 100\,\mu\Omega$; two sense leads carry the voltage across it to a voltmeter, and the current is computed by equation (5.4). The values are those of Example 5.1.

In the reference container the shunt is part of the connection hardware between the cell voltage taps and the rack terminals, as in Figure 5.2. Its reading is the rack current I of the reference record: one channel per rack, and the only direct measurement of the current that flows through all 396 cells of that rack.

The shunt's own voltage drop. A shunt measures current by producing a voltage drop, so the device that measures the current also slightly changes the voltage of the circuit it measures. Its drop is IR_{shunt}: 10 mV at 100 A and 50 mV at 500 A, against a rack voltage of about 1300 V, a fraction of under 0.004 %. Its heat is I^2R_{shunt}, 1 W at 100 A and 25 W at 500 A. Both are small and calibrated, but they are not zero, and because the shunt sits between the cell taps and the rack terminals, its drop is one of the terms that separates the sum of the cell voltages from the rack voltage. The next section collects all such terms into one parameter.

5.3 The string voltage identity

The rack voltage can be obtained in two ways in the reference record: read from the rack voltage sensor, or added up from the 396 cell voltage sensors. This section writes the relation that should hold between the two and shows why that relation is a check on the measurements rather than a definition.

5.3.1 One lumped series resistance

Between the cell voltage taps and the rack terminals the current passes busbars, the fuse, the contactors, the shunt and a number of bolted connectors. Each has its own resistance, and the connection drop of equation (5.2) is the sum of their drops. Because the same current I flows through all of them, that sum is the current times the sum of the resistances,

$$V_{\text{rack}} \approx \sum_{i=1}^{396} V_i - I\left(R_{\text{busbars}} + R_{\text{fuse}} + R_{\text{contactors}} + R_{\text{shunt}} + R_{\text{connectors}}\right) = \sum_{i=1}^{396} V_i - IR_{\text{series}}. \quad (5.5)$$

The lumped series resistance R_{series} is defined precisely as the total resistance of everything in the current path that lies outside the cell voltage taps. A resistance inside a cell, or between the two taps of one cell, is already contained in that cell's measured voltage V_i; a resistance outside every pair of taps is not, and belongs to R_{series}. With the sign convention of this book ($I > 0$ on discharge) the drop IR_{series} lowers the rack voltage on discharge and raises it on charge.

The individual resistances cannot be separated from the rack measurements, because they enter equation (5.5) only through their sum. One parameter summarizes several components, and that is all the measurements can support. As an illustration, busbars and cables of 1.2 mΩ, a fuse of 0.4 mΩ, two contactors of 0.2 mΩ each, a shunt of 0.1 mΩ and bolted connections totaling 0.4 mΩ give $R_{\text{series}} = 2.5\,\text{m}\Omega$; at the rack current of 85 A that is a drop of 0.21 V and a heat of $85^2 \times 0.0025 \approx 18\,\text{W}$ (illustrative values). This lumping is the logic of the string identity: the details of the hardware are replaced by one number that the data can determine.

5.3.2 The sensor offset

Equation (5.5) predicts that at zero current the rack voltage equals the sum of the cell voltages exactly, because the drop IR_{series} vanishes. Field data do not always agree. Suppose that with the rack at rest the rack voltage sensor reads 1301.0 V while the sum of the cell voltage sensors gives 1300.0 V. The one-volt difference at $I = 0$ cannot be a resistive drop. It is a persistent measurement offset V_{off}: a constant disagreement between the two measurement chains, caused by calibration differences in the sensors or their converters.

The model must therefore carry the offset as a second parameter. The difference between the two measurements is

$$V_{\text{rack}} - \sum_{i=1}^{396} V_i = V_{\text{off}} - IR_{\text{series}}, \tag{5.6}$$

or, written for the rack voltage,

$$V_{\text{rack}} = \sum_{i=1}^{396} V_i + V_{\text{off}} - IR_{\text{series}}. \tag{5.7}$$

Equation (5.7) is very close to the relation used in the reference analysis, and it is the string voltage identity of this book. In the example above, $V_{\text{off}} = 1301.0 - 1300.0 = +1.0\,\text{V}$.

The offset is one volt in thirteen hundred, about 0.077 % of the rack voltage, which is within the accuracy of an ordinary high-voltage sensor. It could equally come from the cell side: a bias of $1.0/396 \approx 2.5\,\text{mV}$ in each cell sensor, too small to notice on any single cell, adds up to one volt over the string. The identity determines the offset between the two chains; it does not say which chain carries it.

5.3.3 Two measurements of one voltage

The rack voltage is therefore represented twice in the reference record: once as the directly measured rack voltage, and once as the sum of 396 cell voltages. After the connection drop and the offset are accounted for, the two should agree. Figure 5.5 draws the two routes. The upper route is one sensor across the rack terminals; the lower route is 396 sensors across the individual cells, whose readings are added. Both arrive at the comparison box on the right, which also receives the rack current from the shunt, forms the difference of equation (5.6) and compares it with $V_{\text{off}} - IR_{\text{series}}$. The two outcomes on the far right are agreement, meaning the measurements are mutually consistent, and disagreement, meaning something is wrong.

This is measurement redundancy: two independent measurements of the same physical quantity. If they disagree, something is wrong, and the list of candidates is short: a sensor (a failed or drifting channel), the wiring (a loose voltage tap), the model (a resistance that has changed, or a term missing from equation (5.7)), the calibration (an offset that has moved), or the timing (the two routes sampled at different instants). The last is easy to overlook. If the cell voltages are sampled one second after the rack voltage and the current steps from zero to 100 A in between, the 396 cell voltages fall by their internal drops while the rack reading does not; at the illustrative 0.8 mΩ per cell of Chapter 4 that is $396 \times 100 \times 0.0008 \approx 32\,\text{V}$ of apparent disagreement for that one sample.

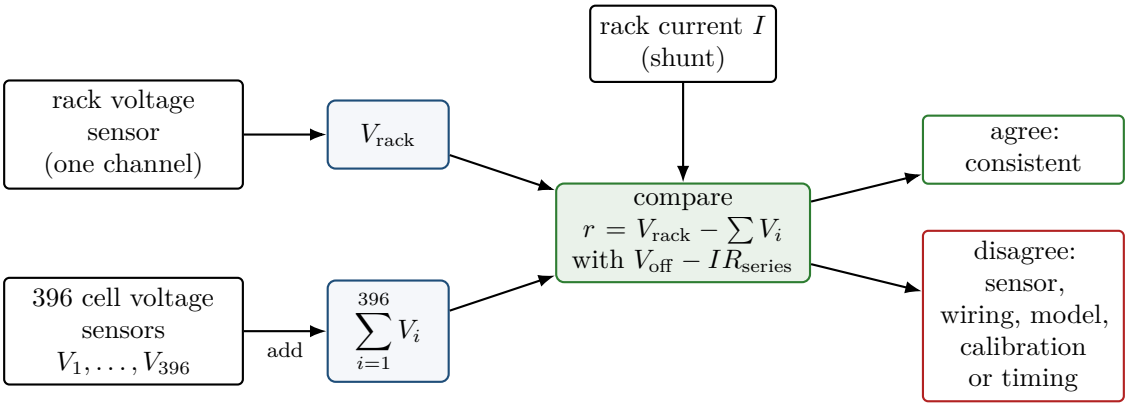


Figure 5.5: The two measurement routes to one rack voltage. The direct reading and the sum of 396 cell readings meet in the comparison of equation (5.6); agreement within noise means the measurements are consistent, a persistent disagreement locates an error.

Principle 5.1. *The string voltage identity $V_{\text{rack}} = \sum_i V_i + V_{\text{off}} - IR_{\text{series}}$ is a consistency relation between two independent measurements of one voltage. Its residual is near zero when sensors, wiring, model, calibration and timing are all sound, and a persistent residual points to one of them.*

5.3.4 Estimating the two parameters

Equation (5.6) is a straight line in the current. Figure 5.6 plots it: the horizontal axis is the rack current I in amperes, negative on charge and positive on discharge; the vertical axis is the residual $V_{\text{rack}} - \sum_i V_i$ in volts. The solid line has intercept V_{off} at $I = 0$ and slope $-R_{\text{series}}$; the dashed line is the same relation without an offset. With the illustrative $V_{\text{off}} = 1.0\,\text{V}$ and $R_{\text{series}} = 2.5\,\text{m}\Omega$, the residual is 1.5 V at a charging current of $-200\,\text{A}$ and 0.5 V at a discharging current of $+200\,\text{A}$. The figure makes the two parameters visible: the offset is the height of the line where it crosses zero current, and the connection resistance is its downward slope.

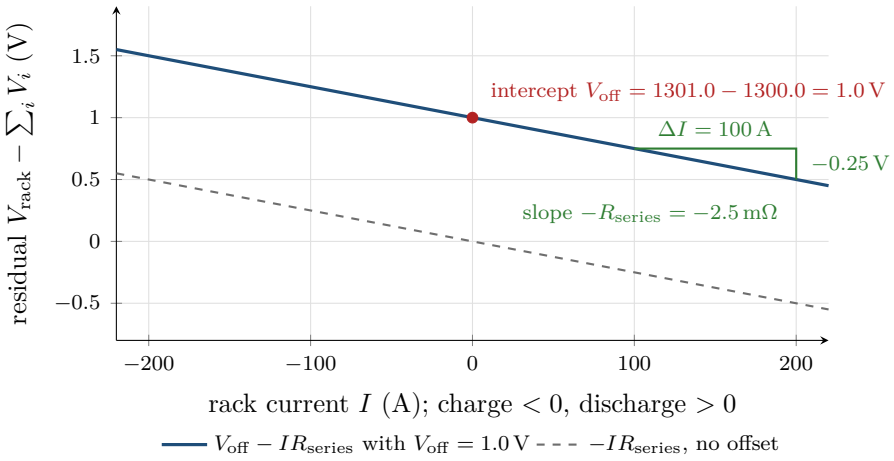


Figure 5.6: The string identity residual against rack current, equation (5.6): a straight line with intercept V_{off} and slope $-R_{\text{series}}$. Illustrative parameters $V_{\text{off}} = 1.0\,\text{V}$ and $R_{\text{series}} = 2.5\,\text{m}\Omega$; the dashed line is the same connection resistance with no sensor offset.

In practice the two parameters are estimated by fitting a straight line to many pairs $(I, V_{\text{rack}} - \sum_i V_i)$ from the record, the residual of the fit being the noise of the measurements.

The fit is well determined only if the current varies. If the record contained only one current value, say a constant 85 A, every observation would fall at the same point of the horizontal axis, and any line through that point would fit: a larger offset and a larger resistance would explain the data as well as a smaller offset and a smaller resistance. Rest periods fix the intercept, and charging and discharging at different currents fix the slope. Once both are estimated, a rise of R_{series} over months is the signature of the degrading connection of Section 5.2.2, and a shift of V_{off} is the signature of a sensor recalibration or drift.

5.4 The common bus and what is not measured

The nine racks of the reference container meet at the DC bus. This section states what the bus implies for the rack voltages and the rack currents, and what it does not allow to be checked because no sensor sits on the bus itself.

5.4.1 Equal rack voltages on a common bus

The nine racks are connected in parallel: the positive terminal of every rack is joined to one bus conductor and every negative terminal to the other. Parallel elements share their terminal voltage, so

$$V_1 \approx V_2 \approx \cdots \approx V_9 \approx V_{\text{bus}}, \tag{5.8}$$

where V_r is the terminal voltage of rack r. The approximation sign allows for the short conductor between each rack terminal and the bus, whose drop is the rack current times a resistance of the order of a milliohm.

This physical fact constrains how the rack voltage readings can be interpreted. If one rack reports 1300.0 V and another 1301.5 V, the difference of 1.5 V is suspicious, because the two terminals are tied to the same bus. For it to be a real voltage difference, the conductor between the two racks would have to drop 1.5 V; at 100 A that requires 15 mΩ, fifteen times the connection of Section 5.2.2, dissipating $100^2 \times 0.015 = 150$ W in a short piece of busbar. A difference that persists when the current is zero cannot be a drop at all. Persistent differences among the nine reported rack voltages must therefore be sensor disagreement, not a real difference of bus voltage. The physics of the connection constrains the interpretation of the sensors: nine channels that should read the same value are, in effect, nine measurements of one quantity, and their spread is a direct estimate of the sensor offsets.

5.4.2 The division of current among racks

Voltages are equal across parallel racks; currents are not. At the bus, charge is conserved (Kirchhoff's current law), so the bus current is the sum of the rack currents,

$$I_{\text{bus}} = I_1 + I_2 + \cdots + I_9 = \sum_{r=1}^{9} I_r. \tag{5.9}$$

How I_{bus} divides among the racks depends on the racks themselves. Each rack behaves approximately as a source equal to the sum of its cells' open-circuit voltages, $U_r = \sum_i U_i$, in series with its total resistance R_r (cells plus R_{series}). With the bus at V_{bus}, the current of rack r is

$$I_r = \frac{U_r - V_{\text{bus}}}{R_r}. \tag{5.10}$$

Racks with different states of charge (hence different U_r), different resistances, different temperatures (which change both) and different connection impedances carry different currents. Three racks of the same container might carry 110 A, 105 A and 118 A at the same moment.

Example 5.2 (Current split from resistance differences)**.** Take $V_{\text{bus}} = 1300\,\text{V}$ and three racks with the same open-circuit sum $U_r = 1335.2\,\text{V}$, so that each sees a driving difference of $35.2\,\text{V}$. With rack resistances of 0.320, 0.335 and $0.298\,\Omega$ (about $0.8\,\text{m}\Omega$ per cell, illustrative), equation (5.10) gives

$$I_1 = \frac{35.2}{0.320} = 110\,\text{A}, \quad I_2 = \frac{35.2}{0.335} \approx 105\,\text{A}, \quad I_3 = \frac{35.2}{0.298} \approx 118\,\text{A}.$$

A resistance difference of five percent produces a current difference of five percent. Equally, a rack whose open-circuit sum is 1.6 V higher, only $1.6/396 \approx 4\,\text{mV}$ per cell, carries $1.6/0.320 = 5\,\text{A}$ more.

The racks thus share the load unequally and the sharing changes with time: a rack that discharges faster than its neighbors lowers its own state of charge and open-circuit voltage, which reduces its share, which is a self-correcting tendency. The rack shunts measure each I_r, so the split is observed directly.

5.4.3 The missing bus-current sensor

The reference record has no current measurement at the DC bus or at the converter. There is therefore no independent measurement of the total current. The total can be computed from equation (5.9), but that computation is not a conservation check, because the same measurements build both sides of the equation. What the record supports is

$$I_{\text{bus,inferred}} = \sum_{r=1}^{9} I_r,$$

a definition, not

$$I_{\text{bus,measured}} \stackrel{?}{=} \sum_{r=1}^{9} I_r,$$

a test. The residual of the first, $I_{\text{bus,inferred}} - \sum_r I_r$, is zero by construction whatever errors the nine shunts contain. There is no redundancy, and the whole-container current balance is therefore declined in the reference analysis: writing it down would look like a validation while checking nothing.

Figure 5.7 shows where the missing sensor would sit. The thick horizontal line is the DC bus; the nine boxes below it are the racks, each connected through its own current sensor (the green circles, which are the shunts of the reference record); on the right the bus continues to the power conversion system and the grid. The dashed red circle on the bus between the last rack and the converter marks the place where a bus-current sensor would measure I_{bus}; in the reference container there is none. The nine rack currents enter the bus node; the one current that leaves it is unmeasured.

5.4.4 When a conservation equation is a test

The bus balance illustrates a general rule. A conservation equation says that what enters a boundary equals what leaves it plus what is stored inside. It becomes diagnostically useful only if enough of the quantities crossing its boundary are measured independently, or are otherwise constrained, so that the equation could fail. At the bus, nine flows are measured and one is not; the unmeasured one is then simply defined by the other nine, and the equation has no capacity to disagree with the data.

The value of the missing sensor can be stated in numbers. Suppose one of the nine shunts reads two percent high. At 85 A that is an error of 1.7 A in one rack current. With a bus-current sensor, the balance residual $I_{\text{bus,measured}} - \sum_r I_r$ would show $-1.7\,\text{A}$ and flag the shunt set;

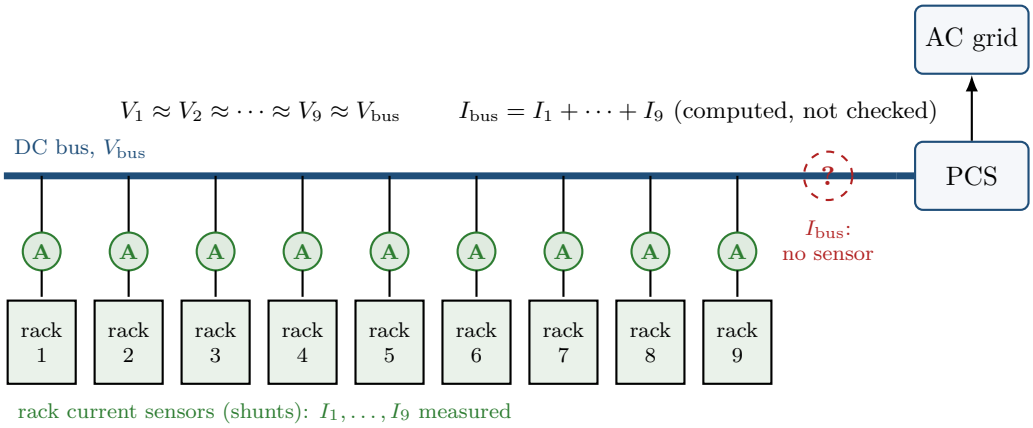


Figure 5.7: The common DC bus of the reference container. Nine racks in parallel share one bus voltage, equation (5.8), and each rack current is measured by its shunt. The total current I_{bus} of equation (5.9) has no sensor of its own, so the balance at the bus can be computed but not tested.

without one, the error passes into every quantity derived from that rack current, including its reported state of charge, and nothing at the bus can reveal it. A quantity can also be constrained without a direct sensor: if the converter reported its AC power P_{AC} and its efficiency η were known, the DC current could be estimated as $P_{\text{AC}}/(\eta\, V_{\text{bus}})$. Such a substitute is a test only if its uncertainty is small compared with the errors sought; with an efficiency known to $\pm 1\,\%$ at 769 A (Section 5.6.3) the substitute is uncertain by about 8 A, too coarse to detect the 1.7 A error of this illustration.

> **Principle 5.2.** *A conservation equation is diagnostically useful only if enough of the quantities crossing its boundary are independently measured or otherwise constrained. When every term is built from the same measurements, the equation is an identity, not a test.*

5.5 The battery management system

The sensors of the previous sections feed one supervisory system. This section describes what that system does, how it keeps the cells of a string together, what its temperature sensors can and cannot say, and why the state of charge it reports is a computation rather than a measurement.

5.5.1 What the BMS does

The battery management system (BMS) is not one sensor but a control and estimation system. It receives the cell voltages, the rack voltage, the rack current and the temperatures. From them it estimates the state of charge, detects when a voltage, current or temperature approaches its limit, balances the cells, raises alarms, opens and closes the contactors, and communicates with a higher-level controller, to which it reports what the rack can currently accept or deliver. In a container like the reference one the BMS is itself hierarchical: a board in each module measures that module's cells, a controller in each rack gathers the nine module boards and commands the rack's contactors, and a container-level unit gathers the nine racks.

The BMS has two jobs that pull in different directions. Its protective job is to keep every cell inside its safe voltage, current and temperature window, which it does by limiting the rack current and, as a last controlled resort, opening the contactor. Its estimating job is to tell

the rest of the system what the battery contains and what it can do. Both jobs rest on the same few measured channels, and in the reference record one of the BMS outputs, the reported state of charge, is derived from current integration rather than measured, a point developed in Section 5.5.5.

5.5.2 Cell balancing

Cells are never identical. Small differences in capacity, internal resistance, self-discharge and temperature make the states of charge of cells in one string drift apart, even though the same current passes through all of them. Suppose that in one rack cell A is at 80 %, cell B at 79 % and cell C at 72 %. On charge, A reaches its upper limit before C; on discharge, C reaches its lower limit first. Because the cells are in series, the whole rack must stop when any one cell reaches a limit.

Example 5.3 (The cost of an eight-point mismatch)**.** Take the three cells above, a capacity of 300 Ah each, and limits at 100 % and 0 % for simplicity (illustrative). Charging stops when A reaches 100 %, after 20 points or 60 Ah; C is then at 92 %. Discharging stops when C reaches 0 %, after 72 points; A is then at 8 %. The string can move between C empty and A full, a window of $100 - (80 - 72) = 92$ points, or $0.92 \times 300 = 276$ Ah instead of 300 Ah. The eight-point spread between the highest and the lowest cell removes eight percent of the rack's usable capacity, even though no cell has lost any capacity.

The BMS reduces the mismatch by balancing: it adjusts the charge of individual cells so that their states of charge converge. A balanced string uses the capacity of its cells; an unbalanced string is limited by its extremes.

5.5.3 Passive balancing and the side current

The simplest and most common method is passive balancing. A small resistor, switched by the BMS across a cell that is more charged than its neighbors, bleeds a small current around that cell, so that during charge it gains charge more slowly and the others catch up. The excess charge is not moved anywhere useful; it is dissipated as heat in the resistor. With an illustrative 33 Ω resistor across a 3.3 V cell the bleed current is 0.1 A and the heat 0.33 W. Removing one point of state of charge, 3 Ah of a 300 Ah cell, takes $3/0.1 = 30$ hours, and removing the eight points of Example 5.3 would take 240 hours. Passive balancing therefore corrects slow drift over many cycles; it cannot correct a large mismatch quickly. Active balancing, which moves charge from a high cell to a low one through a small converter instead of burning it, is faster and less wasteful but more complex.

Balancing has a consequence for the charge balance of the rack. The bleed current flows around a cell without passing the shunt, so the charge that leaves the cells is not exactly the charge that the shunt counts. The BMS electronics, if powered from the string, draw a further small current, and every cell loses a little charge to self-discharge. The reference analysis lumps these into one small, unmeasured side current I_{side}, treated as a constant for each rack, so that the charge change of the cells is

$$\Delta Q_{\text{cells}} = -\int I \,\mathrm{d}t - I_{\text{side}}\,\Delta t, \tag{5.11}$$

with I the shunt current, positive on discharge. If the main current says 100 Ah moved, the cells may have changed by slightly more or slightly less. With an illustrative $I_{\text{side}} = 0.05$ A over the 20 hours of a slow discharge of 100 Ah, the cells give up 101 Ah: the side term is one percent of the throughput, small but systematic.

5.5.4 Temperature sensors and their locations

Each rack of the reference container carries 216 temperature sensors, 24 per module, about one for every $396/216 \approx 1.8$ cells. The density is high because temperature inside a rack is spatially nonuniform. Hot spots arise from poor connections (Section 5.2.2), from cells of higher resistance, from uneven cooling, from edge effects (cells at the end of a module or the top of a rack see different airflow from those in the middle), from failed fans and from blocked airflow. Chapter 6 develops the heat balance that governs these temperatures.

In the reference record the physical locations of the temperature sensors are undocumented: each channel has an index but no stated position. The reference analysis therefore declines a module-level thermal model and uses a rack-level thermal average, $\bar{T} = \frac{1}{216}\sum_{j=1}^{216} T_j$. This is not a loss of rigor but its opposite: a module-level model would require assigning each sensor to a place, and any such assignment would be invented. The analysis does not pretend the data contain more than they do.

The reason is that location is what turns a reading into a physical statement. Suppose two sensors read $T_1 = 300\,\mathrm{K}$ and $T_2 = 306\,\mathrm{K}$ (27°C and 33°C). If T_1 is known to be on module 1 at the bottom of the rack and T_2 on module 9 at the top, the six-kelvin difference is a gradient of about $6/8 = 0.75\,\mathrm{K}$ per module pitch along the direction of airflow, and it can be compared with the heat generated and the air temperature rise. If T_2 is known to sit on a busbar joint, the difference may be the signature of that joint's resistance. Without locations the only statement available is that one sensor is hotter than another. Physical interpretation requires a physical mapping.

5.5.5 State of charge is computed, not measured

A BMS does not measure state of charge. It measures cell voltages, the rack current and temperatures; no sensor in the system outputs "SOC = 65 %". The BMS computes the state of charge from the measured channels. The simplest computation is coulomb counting: integrate the measured current and divide by an assumed capacity,

$$\mathrm{SOC}(t) = \mathrm{SOC}(t_0) - \frac{1}{C_{\mathrm{assumed}}} \int_{t_0}^{t} I(\tau)\,\mathrm{d}\tau, \tag{5.12}$$

with I positive on discharge, so that the chain is $I(t) \to \int I\,\mathrm{d}t \to \mathrm{SOC}(t)$. The picture of a water tank applies: the gauge on the tank is not a float that senses the water level but a meter on the outlet pipe, and the level shown is the starting level minus the volume that has flowed out.

Example 5.4 (A coulomb count)**.** A rack at SOC = 80 % with a nominal capacity of 300 Ah discharges 30 Ah. Equation (5.12) gives

$$\mathrm{SOC} = 0.80 - \frac{30}{300} = 0.80 - 0.10 = 0.70 = 70\,\%.$$

The current was measured; the 70 % was calculated.

An analysis of the reference record shows that its reported state of charge behaves exactly like such a count: it moves in proportion to the integrated rack current, and the proportion implies an assumed capacity of about 287 to 291 Ah. The reported state of charge is therefore a derived channel, a function of the current channel and of a constant chosen by the BMS designer, not a second sensor.

5.5.6 Drift of a coulomb count

Pure coulomb counting accumulates error, because equation (5.12) integrates every error in its inputs. A small bias in the current sensor adds up over time: an offset of 0.5 A (illustrative) over 24 hours is 12 Ah, four points of state of charge on a 300 Ah rack, although the current reading at any instant looks accurate. A wrong assumed capacity scales every step: the count is right at the start and wrong by a growing amount as charge is moved.

Example 5.5 (Drift from a wrong capacity)**.** A rack whose cells have aged to a true capacity of 270 Ah is managed by a BMS that still assumes 300 Ah. Starting from 100 %, 135 Ah is discharged. The true state of charge is

$$\mathrm{SOC}_{\mathrm{true}} = 1 - \frac{135}{270} = 0.50 = 50\,\%,$$

while the BMS computes

$$\mathrm{SOC}_{\mathrm{BMS}} = 1 - \frac{135}{300} = 0.55 = 55\,\%.$$

The reported value is five points too optimistic, and the error keeps growing: when the true state reaches 0 % after 270 Ah, the BMS still reports 10 %.

Figure 5.8 plots the two counts. The horizontal axis is the charge discharged from full, in ampere-hours; the vertical axis is the state of charge in percent. The steeper line is the true state for a 270 Ah rack, the shallower line the count of a BMS assuming 300 Ah. The two start together and separate linearly; the marked pair at 135 Ah is Example 5.5.

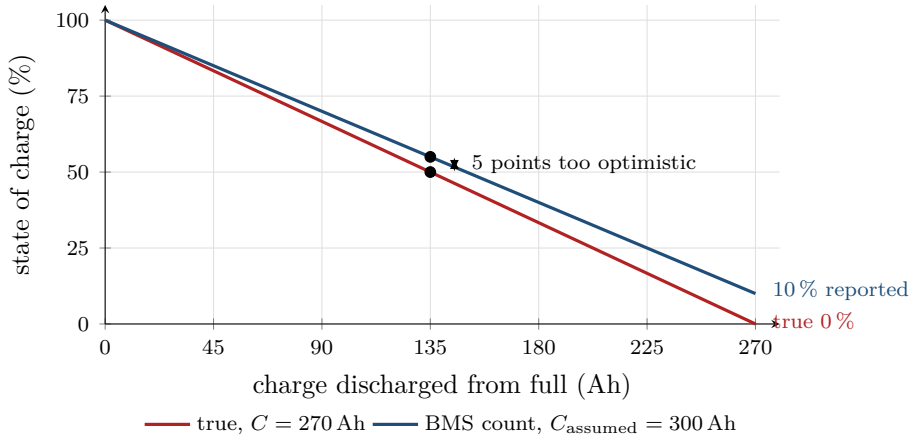


Figure 5.8: Coulomb counting with a wrong capacity, equation (5.12): true state of charge $100(1 - q/270)$ and the BMS count $100(1 - q/300)$ against the charge q discharged from full. At $q = 135$ Ah the true value is 50 % and the count 55 %.

More sophisticated battery management systems do not rely on the count alone. They combine current integration with the measured cell voltages, the temperature, an open-circuit-voltage curve $U(\mathrm{SOC})$ and a battery model such as that of Chapter 4, often inside an observer or a Kalman-type estimator. The count supplies a smooth short-term trajectory; the voltage, compared with the model's prediction, pulls the estimate back toward the true state whenever the two disagree. That correction is precisely what pure coulomb counting lacks.

5.5.7 Measured quantities and estimated states

The consequence for analysis is direct. Because the reported state of charge of the reference record is a coulomb count of the rack current, the current I and the reported SOC are not two

independent pieces of evidence. Using the state-of-charge channel in a charge balance largely tests the BMS's coulomb-counting arithmetic, the integration of equation (5.12), rather than independently measuring the true charge in the cells. Agreement between the two says that the BMS integrated correctly; it does not say that the shunt is accurate or that the assumed capacity is right.

Table 5.1 lists the channels of one rack of the reference record with the quantity each represents and whether it is measured by a sensor or computed from other channels. Current, voltage and temperature are measured; state of charge is estimated; state of health, the ratio of present to original capacity, is even less directly observable, since it must be inferred from how the estimated state of charge and the measured voltages behave over many cycles.

Table 5.1: Channels of one rack of the reference record: what each represents and whether it is measured or derived.

Channel	Count	Status	Source
cell voltage V_i	396	measured	module sensing boards, one tap pair per cell
rack voltage V_{rack}	1	measured	sensor across the rack terminals
rack current I	1	measured	shunt, equation (5.4)
temperature T_j	216	measured, location undocumented	module sensing boards
reported SOC	1	derived	coulomb count of I, implied capacity 287–291 Ah
SOH	none	not reported; inferred	from capacity estimated over many cycles
bus current I_{bus}	none	not measured	computable only as $\sum_r I_r$

Principle 5.3. *A measured quantity is not an estimated state. Current, voltage and temperature are measured; state of charge is computed from them, and a channel computed from another channel adds no independent evidence about the physics.*

5.6 Power conversion, control and the whole picture

The racks and their BMS produce and absorb direct current; the grid runs on alternating current and on dispatch commands. This section describes the converter and the controllers that connect the two, follows one megawatt from the grid command down to a single cell, and assembles the complete physical picture.

5.6.1 The power conversion system

The power conversion system (PCS) is the power-electronic converter between the DC bus and the AC grid. When the battery discharges it inverts direct current to alternating current; when the battery charges it rectifies alternating current to direct current. Beyond the conversion itself it controls the active power exchanged with the grid, and therefore the charge and discharge current drawn from the bus; it sets or follows the DC operating point; it often supplies or absorbs reactive power to support the grid voltage; it synchronizes with the grid's frequency and phase; and it implements the required responses to grid faults, such as riding through a voltage dip or disconnecting.

The PCS is the actuator of the BESS: the one device that acts on the power flow. The BMS can only permit, limit or interrupt the current; it is the PCS that decides, within those permissions, how much current flows. The bus voltage, in contrast, is set mainly by the racks:

with their contactors closed, the bus sits at the common terminal voltage of equation (5.8), and the PCS draws the current that its power command requires at that voltage.

5.6.2 The control hierarchy

The devices of this chapter form a control hierarchy:

grid or dispatch command → site controller (EMS) → PCS → DC bus → racks → BMS → cells.

The site controller decides what the asset does: it receives a dispatch command or a market schedule and turns it into a power set point. The PCS controls the power flow to meet that set point. The BMS protects the cells and estimates their state, and it reports upward the limits within which the PCS and the site controller must stay.

Figure 5.9 draws the hierarchy. The central column is the chain of command and power from the dispatch command at the top to the cells at the bottom; solid arrows are commands and power flow. The BMS sits beside the racks and cells rather than in the power path, because it measures and commands (through the contactors) without carrying the main current. The dashed arrow from the BMS to the site controller is the information that travels upward: limits, alarms and the estimated state of charge. The notes on the left follow the one-megawatt example of the next subsection; the notes on the right state the role of each controller.

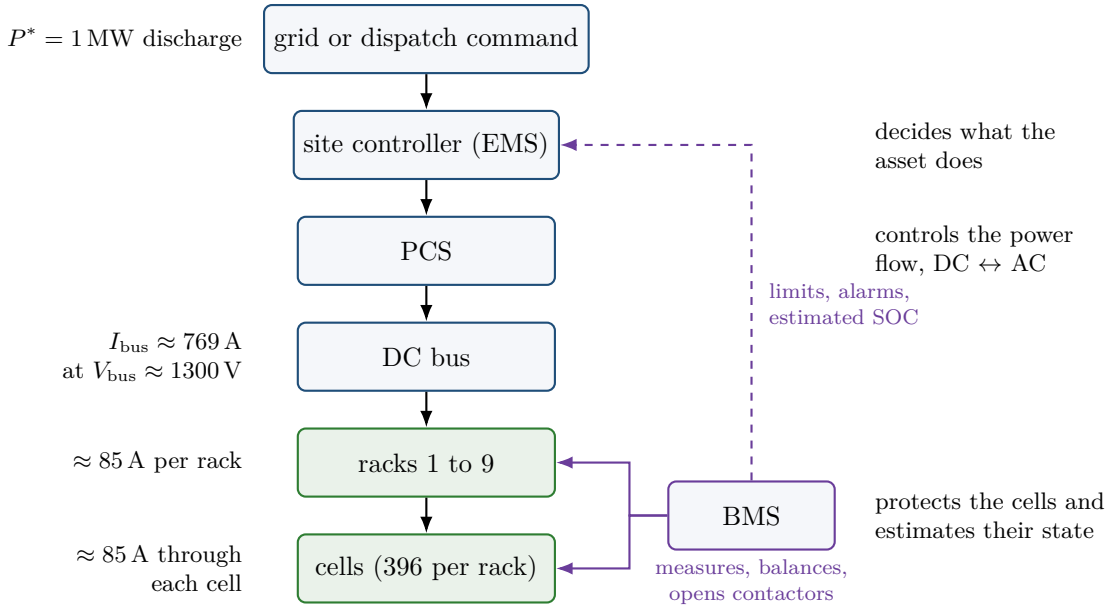


Figure 5.9: The control hierarchy of the reference container, from the dispatch command to the cells. Solid arrows are commands and power flow; the dashed arrow is information reported upward by the BMS. The left notes follow the one-megawatt discharge of Section 5.6.3.

5.6.3 A one-megawatt discharge

Suppose the site controller commands a discharge of $P^* = 1\,\text{MW}$ and the bus sits at $V_{\text{bus}} \approx 1300\,\text{V}$. Ignoring converter losses, the current the PCS must draw from the bus is

$$I_{\text{bus}} = \frac{P^*}{V_{\text{bus}}} = \frac{1{,}000{,}000\,\text{W}}{1300\,\text{V}} \approx 769\,\text{A}. \tag{5.13}$$

Shared over nine racks, that is about $769/9 \approx 85\,\text{A}$ per rack (equally shared, for the purpose of the example; Section 5.4.2 showed that the real split is uneven). Because each rack is a series

string, the same 85 A flows through every one of its 396 cells. At a cell capacity of 300 Ah that is a rate of $85/300 \approx 0.28$ per hour, a gentle discharge that would empty the battery in about three and a half hours.

The grid dispatch is thereby connected down to a single cell: a number typed by an operator in megawatts becomes a current of 85 A in a cell the size of a book. Converter losses only sharpen the point: with an illustrative conversion efficiency of 98 %, the DC side must deliver $1/0.98 \approx 1.02$ MW, or 785 A and 87 A per rack, for the grid to receive 1 MW.

5.6.4 One weak cell in a megawatt system

The link from grid to cell runs in the other direction too. Suppose one cell in one rack approaches its minimum voltage, because it is at a lower state of charge than its neighbors or has a higher resistance. By $V = U - IR$, reducing the current raises that cell's voltage, so the BMS first limits the rack's current and, if the cell still reaches its limit, opens the rack's contactor. In the second case eight racks remain, and they must carry more current to deliver the same 1 MW: $769/8 \approx 96$ A each, 12.5 % more than before. The resistive heat in each remaining rack grows with the square of its current, by $(96.2/85.5)^2 \approx 1.27$, and every remaining cell sees a larger IR drop, which moves it closer to its own lower voltage limit and closer to its temperature limit.

Table 5.2 works the three cases with an illustrative rack resistance of 0.32 Ω (about 0.8 mΩ per cell): all nine racks sharing equally, the weak rack limited to 40 A, and the weak rack disconnected. The columns give the current in each healthy rack, the resistive heat in each healthy rack, and the total resistive heat of the container.

Table 5.2: A 1 MW discharge at 1300 V with one weak rack. Rack resistance 0.32 Ω, illustrative; heat is I^2R per rack.

Case	Current per healthy rack	Heat per healthy rack	Total resistive heat
nine racks share equally	85.5 A	2.34 kW	21.0 kW
weak rack limited to 40 A	91.2 A	2.66 kW	21.8 kW
weak rack disconnected	96.2 A	2.96 kW	23.7 kW

The chain is therefore

weak cell → rack limitation → less current available → more load on the other racks → system power limitation.

If the remaining racks reach their own limits first, the site controller can no longer deliver 1 MW and must reduce the commanded power. One cell among 3,564 has limited a megawatt system. This is the series-string property of Section 5.1.3, the slowest runner setting the pace, carried up through every level of the hierarchy.

Principle 5.4. *In a series string the weakest cell sets the limits of its rack, and through the shared bus a limited rack shifts load onto the others; a cell-level limit can therefore become a system-level power limit.*

5.6.5 The complete physical picture

The pieces can now be assembled. From the grid inward: the AC grid connects to the PCS; the PCS connects to the DC bus; the DC bus connects racks 1 to 9 in parallel (Figure 5.7). Inside each rack the current passes a contactor, a fuse, a shunt and busbars and modules 1 to 9, which together contain 396 cells in series (Figure 5.2). Around the rack, voltage, current

and temperature sensors feed the BMS, which estimates the state of charge, protects the cells, balances them and controls the contactors; above it, the site controller and the PCS turn dispatch commands into current (Figure 5.9).

Every element of this picture has appeared in one of the equations of the chapter. The busbars, fuse, contactors and shunt are the series resistance R_{series} of the string identity. The rack and cell voltage sensors are its two measurement routes. The bus is the equality of the rack voltages and the current balance. The BMS is the source of the derived state-of-charge channel and of the side current. The PCS is the boundary of the DC system, the place where the missing bus-current measurement would sit. The hardware and the relations available for analysis are the same object seen from two sides.

5.6.6 Three relations to keep

Three relations from this chapter carry forward. First, the string voltage identity,

$$V_{\text{rack}} \approx \sum_i V_i - IR_{\text{connections}} + V_{\text{offset}},$$

equation (5.7), a consistency relation between two independent measurements. Second, the common-bus equality $V_1 \approx \cdots \approx V_9$, equation (5.8), a constraint that turns rack voltage disagreements into sensor offsets. Third, the bus current balance $I_{\text{bus}} = \sum_r I_r$, equation (5.9), which is physically true but cannot be independently checked without a bus-current sensor.

The three differ in exactly the property that matters for inference. The first is a test because both sides are measured independently. The second is a test because nine channels measure one quantity. The third is not a test, because one of its terms is not measured. Which equations provide independent information, and which only restate the measurements, becomes the central question of the later chapters, and of Chapter 9 in particular.

5.7 What to retain

Definition 5.1 (The hardware around the cells)**.** A grid battery is a hierarchy of series strings joined in parallel, whose connection hardware adds a lumped resistance and whose sensors measure current, voltage and temperature but not state of charge. A relation among the measurements informs the analysis only when its terms are measured independently: the string voltage identity and the common-bus equality do, the bus current balance without a bus sensor does not.

Table 5.3 lists the chapter's objects with the subsection that defines each.

5.8 Exercises

Exercise 5.1. A bolted joint between two busbars of one rack degrades from 0.2 mΩ to 1.5 mΩ. (a) Compute its voltage drop and heat before and after at the rack current of 85 A and at a short peak of 300 A. (b) By how much does R_{series}, as estimated from the string identity, rise, and what change in the slope of Figure 5.6 does that produce? (c) The rack shunt is $R_{\text{shunt}} = 75\,\mu\Omega$ and reads 6.0 mV. Compute the current and the shunt's heat. If the true shunt resistance is 1 % higher than the value the BMS uses, what current error results, and how many ampere-hours does a coulomb count accumulate in error over ten hours at that current?

Exercise 5.2. For one rack the residual $V_{\text{rack}} - \sum_i V_i$ averages 0.80 V at rest, 0.50 V at a discharge current of 100 A and 1.10 V at a charge current of −100 A. (a) Estimate V_{off} and R_{series} from equation (5.6). (b) Express the offset as a bias per cell sensor, and as a fraction of

Table 5.3: The objects of the hardware chapter.

Object	Definition	Section
cell, module, rack, container	$1 \to 44$ in series $\to 9$ in series $\to 9$ in parallel; 3,564 cells, 3.42 MWh	5.1.1
module voltage	$44 \times 3.2 = 140.8$ V	5.1.2
rack	series string of 396 cells carrying one current	5.1.3
IR, I^2R	drop and heat of a connection; 0.1 V, 10 W at 100 A and 1 mΩ	5.2.1, 5.2.2
fuse, contactor	one-shot protection; repeatable commanded switch	5.2.3, 5.2.4
$I = V_{\text{shunt}}/R_{\text{shunt}}$	shunt current measurement	5.2.5
R_{series}	lumped resistance outside the cell voltage taps	5.3.1
V_{off}	offset between rack sensor and summed cell sensors	5.3.2
$V_{\text{rack}} = \sum_i V_i + V_{\text{off}} - IR_{\text{series}}$	string voltage identity; a redundancy check	5.3.2, 5.3.3
$V_1 \approx \cdots \approx V_9 \approx V_{\text{bus}}$	common-bus voltage equality	5.4.1
$I_{\text{bus}} = \sum_r I_r$	bus current balance; an identity without a bus sensor	5.4.2, 5.4.3
balancing, I_{side}	mismatch reduction; unmeasured side current	5.5.2, 5.5.3
$\bar{T}$	rack-level thermal average, used because sensor locations are undocumented	5.5.4
coulomb-counted SOC	derived channel, equation (5.12); implied capacity 287–291 Ah	5.5.5
PCS, EMS, control hierarchy	actuator; site controller; dispatch to cell	5.6.1, 5.6.2

a 1300 V rack voltage. (c) A month later the residual at 100 A discharge is 0.20 V while the rest value is unchanged. Which parameter has moved, by how much, and which hardware would be inspected first? (d) Explain why a record in which the rack only ever discharged at a constant 100 A could not have separated the two parameters.

Exercise 5.3. A BMS assumes $C_{\text{assumed}} = 300$ Ah for a rack whose true capacity is 280 Ah. (a) Starting from 90 %, 168 Ah is discharged. Compute the true and the reported state of charge. (b) In addition, the shunt reads 0.3 A high on discharge. Over 48 hours of operation, what extra error in ampere-hours and in points of state of charge does the bias add? (c) Explain why comparing the reported SOC with the integrated rack current from the same record could not detect either error.

Exercise 5.4. The container is commanded to discharge 0.8 MW with the bus at 1280 V. (a) Compute I_{bus} and the current per rack with nine racks sharing equally, ignoring converter losses. (b) Two racks are disconnected by their BMS. Compute the new current per rack and the ratio of the per-rack heat to the nine-rack case. (c) With an illustrative rack resistance of 0.32 Ω, compute the total resistive heat in both cases. (d) Repeat (a) with a conversion efficiency of 97.5 %.

5.9 Connection to the TERRA framework

The hardware of this chapter is the graph on which the later inference runs, and the measurement layout decides which of its relations carry information.

- **A rack is an edge with one flow.** The series string of Figure 5.2 is an edge whose flow, the current I, is common to 396 cells. The shunt measures that flow once, and every cell component on the edge inherits it; the per-cell voltage model of Chapter 4 is a component relation evaluated at that shared flow.
- **The string identity is a redundancy relation.** Equation (5.7) relates two measured routes to one potential difference through two parameters, R_{series} and V_{off}. Its residual is

a test statistic, and the two parameters are identifiable only if the record contains both rest and varied current (Section 5.3.4).

- **The DC bus is a conservation node with an unmeasured edge.** Equation (5.9) is charge conservation on the bus node, with nine measured rack edges and one unmeasured PCS edge. With no independent measurement crossing that boundary the node yields an identity, not a test, which is why the container current balance is declined.
- **The bus potential ties nine sensors to one state.** Equation (5.8) makes the nine rack voltage channels nine measurements of one node potential, so their persistent spread is assigned to sensor offset parameters rather than to physics.
- **Reported SOC is a derived channel.** The coulomb count of equation (5.12) is a function of the current channel; the charge balance on a rack node uses the measured current, the capacity as an unknown parameter and the side current I_{side} of equation (5.11) as a lumped unmeasured term, and treats the reported SOC as a check on the BMS arithmetic, not as independent evidence.
- **Unmapped sensors limit the representation.** Temperature channels without locations cannot be attached to module nodes, so the heat balance is written on one rack node with the average $\bar{T}$; the graph is no finer than the measurement map that supports it.

In summary, the hardware around the cells turns one cell equation into a network: series edges carrying shared currents, a parallel node sharing one voltage, a lumped connection resistance, sensor offsets, and a supervisory system whose state-of-charge output is computed from the same current it reports. Of the three relations this network supplies, two are tests and one is an identity, and the difference is fixed by where the sensors are, not by the physics.

The next chapter follows the I^2R heat generated in the cells and connections into the thermal balance of the rack.

Chapter 6

Thermal behavior and cooling: heat generation, thermal mass and the rack heat balance

6.1 Heat as a conserved quantity

The previous chapters treated a battery as a device that stores and releases charge. It is also a device that generates heat and a system that must remove it: every ampere that passes through a cell dissipates some energy inside it, and a container of thousands of cells needs fans, ducts or liquid loops to carry that energy away. The question of this chapter is

how heat generated by current, stored in thermal mass and removed by cooling sets the temperature of a battery, and what a set of temperature readings can determine about that process.

The starting point is a conservation law, written for heat instead of charge:

$$\text{rate of heat stored} = \text{heat generated} - \text{heat removed}.$$

Everything in the chapter is a reading of this one balance: what generates the heat, what stores it, what removes it, how fast the temperature responds, and which of its coefficients a telemetry record can recover. For the reference container the balance is written at rack level, one temperature per rack, because the 216 temperature sensors of each rack have no documented positions; Section 6.4 explains why that choice is forced rather than convenient.

Jargon: Joule heating, thermal capacitance, thermal inertia, conductance, time constant, structural unidentifiability, thermal runaway

Joule heating (also ohmic or resistive heating) is the conversion of electrical energy into heat when a current flows through a resistance; its rate is I^2R. *Thermal capacitance* C_{th}, in joules per kelvin, is the heat needed to raise the temperature of a body by one kelvin; for a body of mass m and specific heat c_p it is about $m\,c_p$. *Thermal inertia* is the resulting reluctance of a large thermal capacitance to change temperature quickly: the temperature lags the heating. A thermal *conductance* h, in watts per kelvin, is the heat flow that a one-kelvin temperature difference drives across a path, here from the battery to its cooling air; its inverse $1/h$ is a thermal resistance. The thermal *time constant* $\tau = C_{\text{th}}/h$ is the time a temperature takes to cover about 63 % of a step toward its new equilibrium. A model is *structurally unidentifiable* when two or more different parameter values produce exactly the same predicted measurements for every input, so that no quantity of data of that kind can tell them apart. *Thermal runaway* is a self-accelerating failure in which rising temperature speeds up heat-releasing reactions inside a cell, which raise the temperature further, faster than any cooling can remove the heat.

Key idea

Battery temperature is the state of a heat balance: $C_{\text{th}}\,\mathrm{d}T/\mathrm{d}t = \text{generation} - \text{removal}$, with generation growing as the square of the current and removal growing with the temperature difference to the cooling air. From rack-mean temperatures and rack currents alone, only the coefficients divided by the thermal capacitance can be recovered.

6.1.1 Joule heating

A battery is not an ideal source. Chapter 4 modeled its non-ideality as an internal resistance R in series with the open-circuit voltage: when a current I flows, a voltage IR is lost across that resistance. The power dissipated in a resistance is the voltage across it times the current through it,

$$P_{\text{heat}} = (IR)\,I = I^2R, \tag{6.1}$$

and that power appears as heat inside the cell. Two properties follow from the square. First, the sign of the current does not matter: in the book's convention $I > 0$ is discharge and $I < 0$ is charge, and I^2R is positive in both cases, so a battery heats while charging as well as while discharging. Second, heat grows faster than current. For an illustrative battery block with lumped resistance $R = 10\,\text{m}\Omega$,

$$I = 100\,\text{A}:\ P_{\text{heat}} = 100^2 \times 0.010 = 100\,\text{W}, \qquad I = 200\,\text{A}:\ P_{\text{heat}} = 200^2 \times 0.010 = 400\,\text{W}.$$

Doubling the current quadruples the heat.

Figure 6.1 plots equation (6.1). The horizontal axis is the current magnitude in amperes; the vertical axis is the heat generated in watts. The solid curve is the 10 mΩ block of the example, with its two operating points marked; the dashed curve is a block of half that resistance, 5 mΩ, with the operating point 150 A that Section 6.2.5 uses. Both curves are parabolas: flat near zero current and steep at high current. A battery that spends most of its time at low current and occasionally runs hard generates most of its heat in the hard intervals, and an average current understates the heat because the mean of I^2 exceeds the square of the mean of I.

The quadratic also sets the fraction of the power that is lost. If the illustrative block is a module of 16 cells in series at a nominal 3.2 V each, its voltage is 51.2 V and at 100 A it delivers about 5.1 kW, of which 100 W, about 2 %, is lost as heat. At 200 A it delivers about 10.2 kW and loses 400 W, about 4 %. The loss fraction is IR/V, which grows linearly with the current, so fast operation costs efficiency as well as temperature.

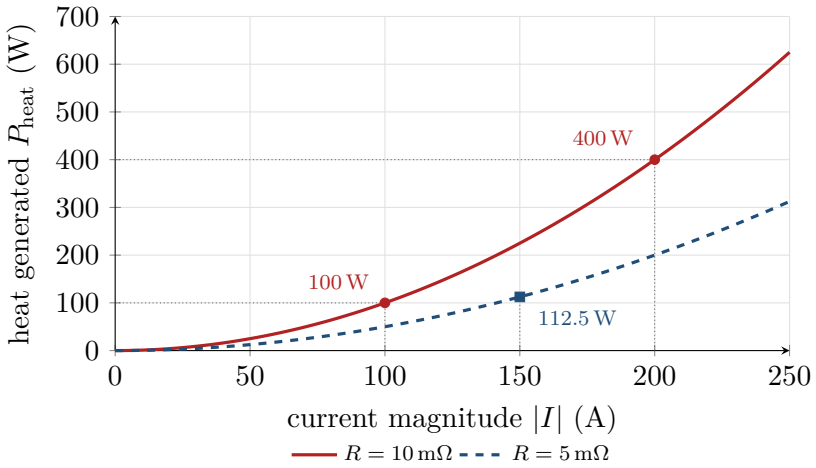


Figure 6.1: Joule heating against current, $P_{\text{heat}} = I^2R$ (equation (6.1)), for two illustrative lumped resistances. Doubling the current from 100 A to 200 A quadruples the heat from 100 to 400 W; the square marks the 150 A, 5 mΩ point of the worked example.

Principle 6.1. *Resistive heat grows as the square of the current and does not depend on its direction: doubling the current quadruples the heat, whether the battery is charging or discharging.*

6.1.2 Other sources of heat

I^2R is not the whole story. The voltage a cell loses under load is not all ohmic: Chapter 4 split it into an instantaneous ohmic part IR_0 and a slow polarization part V_1, which represents the energy spent driving the electrochemical reactions at the electrode surfaces and the energy spent moving lithium ions through the electrolyte and the electrode particles against concentration gradients (transport limitations). All of this lost voltage turns into heat. With current positive on discharge, the irreversible heat is the lost voltage times the current,

$$\dot{Q}_{\text{irr}} = I\,(U - V) = I^2R_0 + I\,V_1, \tag{6.2}$$

and when the current has been steady long enough for the polarization voltage to settle at $V_1 = IR_1$, it becomes $I^2(R_0 + R_1)$: still proportional to I^2, with a larger effective resistance than the ohmic one alone.

A third term, the *reversible* or *entropic* heat, comes from the entropy change of the cell reaction. It is proportional to I rather than I^2, to the absolute temperature, and to the slope $\mathrm{d}U/\mathrm{d}T$ of the open-circuit voltage with temperature; with the book's sign convention it is $-I\,T\,\mathrm{d}U/\mathrm{d}T$. Its sign depends on the direction of the current and on the state of charge, so over part of the range it cools the cell rather than heating it. For lithium iron phosphate cells such as those of the reference container it is usually small compared with the irreversible terms at high current, but it is not always negligible at low current, where I^2 is small.

The reference thermal model does not separate these mechanisms. It keeps one lumped generation term proportional to I^2, whose coefficient absorbs the ohmic resistance, the settled polarization resistance and whatever part of the other losses scales the same way. That is a modeling decision forced by the data: none of the separate terms is measured, and Section 6.5.2 returns to what the lumped coefficient means.

6.2 Storage, cooling and the heat balance

Heat generated inside a battery does one of two things: it stays and raises the temperature, or it leaves through the cooling system. This section quantifies each and combines them into the balance equation and its equilibrium.

6.2.1 Thermal mass and thermal capacitance

Heat raises temperature, but not instantly. The amount of heat needed to raise a body's temperature by one kelvin is its thermal capacitance,

$$C_{\mathrm{th}} \approx m\, c_p, \tag{6.3}$$

where m is the mass in kilograms and c_p the specific heat in joules per kilogram per kelvin. If no heat leaves, all the generated heat $\dot{Q}$ is stored, and the temperature rises at the rate

$$C_{\mathrm{th}} \frac{\mathrm{d}T}{\mathrm{d}t} = \dot{Q}, \qquad \text{that is} \qquad \frac{\mathrm{d}T}{\mathrm{d}t} = \frac{\dot{Q}}{C_{\mathrm{th}}}. \tag{6.4}$$

The same heat warms a light system quickly and a massive one slowly. With illustrative values typical of a large prismatic lithium iron phosphate cell, $m = 5.4\,\mathrm{kg}$ and $c_p = 1000\,\mathrm{J/(kg\,K)}$, the cell's thermal capacitance is $C_{\mathrm{th}} = 5400\,\mathrm{J/K}$; a 112.5 W heat input with no cooling would raise it at $112.5/5400 = 0.021\,\mathrm{K}$ per second, about 1.25 K per minute. A block ten times heavier, $C_{\mathrm{th}} = 54\,\mathrm{kJ/K}$, warms ten times more slowly under the same heat, 0.125 K per minute. A thermal capacitance of 54 kJ/K is the illustrative thermal mass used in the rest of the chapter, paired with heat inputs of about a hundred watts and a conductance of about 15 W/K so that the worked times come out in hours. A whole rack of 396 such cells has a thermal capacitance near $396 \times 5400 \approx 2 \times 10^6\,\mathrm{J/K}$ and generates kilowatts rather than a hundred watts at full current, so its heat inputs and its conductance are larger in the same proportion and the time scales are of the same order.

Thermal inertia. A small metal block on a heater warms in seconds; a large battery rack under the same power warms over hours. The difference is the thermal capacitance, and its consequence is that temperature is a *state* with memory. Equation (6.4) fixes the rate of change of temperature, not the temperature itself, so the present temperature is the result of the whole history of heating and cooling up to now, not of the present current. Two racks carrying the same current at the same moment can be at different temperatures because one has been running hard for an hour and the other has just started.

6.2.2 Cooling as a conductance to the surroundings

A battery system removes heat with forced air, fans, a heating, ventilation and air-conditioning (HVAC) unit, and in some designs liquid cooling plates. The simplest law that describes all of them is proportional:

$$\dot{Q}_{\mathrm{cool}} = h\,(T - T_{\mathrm{env}}), \tag{6.5}$$

where h is an effective thermal conductance in W/K and T_{env} is the temperature of the cooling air or of the ambient surroundings. The law says that heat flows from hot to cold at a rate proportional to the temperature difference: when the battery is barely warmer than the air, little heat is removed; when it is much warmer, more is removed. The conductance h lumps together everything about the cooling path, the air flow rate, the surface area, the contact between cells and the air or the plates, and it is a property of the installation as much as of the battery. A faster fan raises h; a clogged filter lowers it.

6.2.3 The heat balance equation

Combining generation, equation (6.1), storage, equation (6.4), and removal, equation (6.5), gives the heat balance

$$C_{\text{th}} \frac{\mathrm{d}T}{\mathrm{d}t} = I^2 R - h\,(T - T_{\text{env}}), \tag{6.6}$$

which reads term by term as

$$\text{heat storage rate} = \text{heat generation} - \text{heat removal}.$$

It is the thermal form of a conservation law: energy entering the battery as heat either stays, raising T, or leaves through the cooling path. Nothing else happens to it.

Figure 6.2 draws the balance as a node. The dashed outline is the boundary of the control volume, the battery whose heat is being counted. The circle inside is the stored thermal energy, $C_{\text{th}}T$, whose rate of change is the storage term. The arrow entering from the left is the generation, I^2R plus the other sources of Section 6.1.2; the arrow leaving to the right is the cooling flow $h(T - T_{\text{env}})$ into the cooling air, drawn as a second box at temperature T_{env}. The current enters the figure only through the generation arrow: the thermal node does not see the current itself, only the heat it produces.

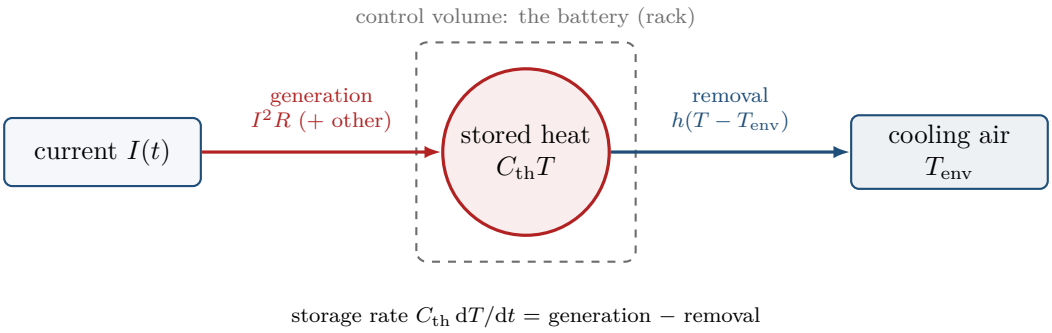


Figure 6.2: The heat balance of equation (6.6) drawn as a node. Generation driven by the current enters the battery, cooling proportional to the temperature difference leaves it, and the difference accumulates as stored heat and appears as a change in temperature.

6.2.4 Warming, cooling and equilibrium

The sign of the right-hand side of equation (6.6) decides the direction of the temperature:

- if generation exceeds removal, $I^2R > h(T - T_{\text{env}})$, then $\mathrm{d}T/\mathrm{d}t > 0$ and the battery warms;
- if removal exceeds generation, $\mathrm{d}T/\mathrm{d}t < 0$ and the battery cools;
- if they are equal, $\mathrm{d}T/\mathrm{d}t = 0$ and the temperature holds.

The balance is self-correcting. As the battery warms, the difference $T - T_{\text{env}}$ grows and so does the removal, until removal catches up with generation. Setting $\mathrm{d}T/\mathrm{d}t = 0$ gives the equilibrium,

$$T - T_{\text{env}} = \frac{I^2 R}{h}. \tag{6.7}$$

The steady temperature rise above the cooling air is the heat generated divided by the cooling conductance. It is proportional to I^2, so a doubled current gives four times the rise, and inversely proportional to h, so a halved conductance gives twice the rise. The thermal capacitance does not appear: thermal mass decides how long the battery takes to reach equilibrium, not where the equilibrium is.

Principle 6.2. *At equilibrium the temperature rise above the cooling air equals the heat generated divided by the cooling conductance, I^2R/h; thermal mass sets only how quickly that equilibrium is reached.*

6.2.5 A worked equilibrium

Take an illustrative block carrying $I = 150\,\mathrm{A}$ with $R = 5\,\mathrm{m\Omega}$. It generates

$$P_{\mathrm{heat}} = 150^2 \times 0.005 = 22\,500 \times 0.005 = 112.5\,\mathrm{W}.$$

With a cooling conductance $h = 15\,\mathrm{W/K}$, equation (6.7) gives the steady rise

$$T - T_{\mathrm{env}} = \frac{112.5\,\mathrm{W}}{15\,\mathrm{W/K}} = 7.5\,\mathrm{K},$$

and with cooling air at 25°C the block settles at $T \approx 25 + 7.5 = 32.5$°C. At that temperature the cooling removes $15 \times 7.5 = 112.5\,\mathrm{W}$, exactly the generation. The same block at 75 A would generate a quarter of the heat, 28.1 W, and settle 1.9 K above the air; at 300 A it would generate 450 W and settle 30 K above the air, at 55°C, a temperature most lithium-ion manufacturers place at or beyond the upper limit for operation. The quadratic of Figure 6.1 becomes, through equation (6.7), a quadratic in temperature.

6.3 Thermal dynamics and time scales

The equilibrium says where the temperature goes; this section says how fast it gets there, gives the electrical circuit that obeys the same equation, and follows the temperature after the current changes.

6.3.1 The thermal time constant

Dividing equation (6.6) by h shows its time scale. Write the temperature rise as $\theta = T - T_{\mathrm{env}}$ with T_{env} constant; then

$$\frac{C_{\mathrm{th}}}{h}\frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{I^2R}{h} - \theta, \qquad \tau = \frac{C_{\mathrm{th}}}{h}. \tag{6.8}$$

The quantity τ has units of $(\mathrm{J/K})/(\mathrm{W/K}) = \mathrm{s}$: it is the thermal time constant. For a current switched on at $t = 0$ and then held, starting from the air temperature, the solution is

$$\theta(t) = \theta_\infty(1 - e^{-t/\tau}), \qquad \theta_\infty = \frac{I^2R}{h}. \tag{6.9}$$

After one time constant the rise has covered $1 - e^{-1} \approx 63\,\%$ of the way to equilibrium, after three time constants 95 %, after five more than 99 %. A large thermal capacitance gives a slow response; a large conductance gives a fast one, because strong cooling both lowers the equilibrium and reaches it sooner.

For the illustrative block of Section 6.2.5, with $C_{\mathrm{th}} = 54\,\mathrm{kJ/K}$ and $h = 15\,\mathrm{W/K}$,

$$\tau = \frac{54\,000\,\mathrm{J/K}}{15\,\mathrm{W/K}} = 3600\,\mathrm{s} = 1\,\mathrm{h}.$$

After one hour at 150 A the block is $7.5 \times 0.632 = 4.74\,\mathrm{K}$ above the air; after three hours, 7.13 K. Tripling the thermal mass to 162 kJ/K with the same cooling gives $\tau = 3\,\mathrm{h}$: the same 7.5 K equilibrium, reached three times more slowly, with only 2.13 K of rise after the first hour.

Figure 6.3 draws both responses. The horizontal axis is hours since the current was switched on; the vertical axis is the temperature rise above the cooling air. The dashed horizontal line is

the common equilibrium $\theta_\infty = 7.5\,\mathrm{K}$. The solid curve is the one-hour block, the dashed curve the three-hour block, and the dots mark the 63 % point of each, reached at $t = \tau$. The dotted straight line is the initial slope of the one-hour curve; it reaches the equilibrium exactly at $t = \tau$, which is a second way to read the time constant off a measured curve.

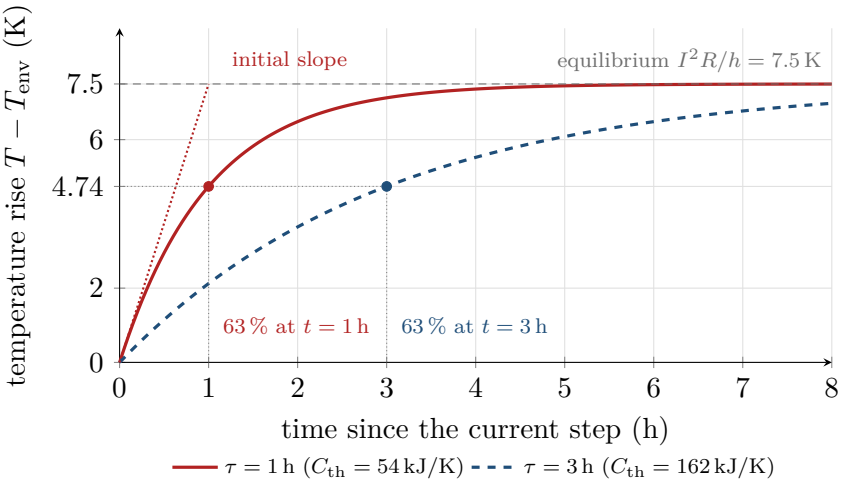


Figure 6.3: Temperature rise after a current step, $\theta(t) = \theta_\infty(1 - e^{-t/\tau})$ (equation (6.9)), for the illustrative block with $I = 150\,\mathrm{A}$, $R = 5\,\mathrm{m\Omega}$ and $h = 15\,\mathrm{W/K}$, so $\theta_\infty = 7.5\,\mathrm{K}$, and two thermal capacitances, $54\,\mathrm{kJ/K}$ ($\tau = 1\,\mathrm{h}$) and $162\,\mathrm{kJ/K}$ ($\tau = 3\,\mathrm{h}$). Thermal mass changes the speed, not the destination.

Mismatch of time scales. Current in a battery system changes in seconds: a power converter can reverse a rack from full charge to full discharge in well under a second. Temperature responds over minutes or hours. The mismatch matters for field telemetry in two ways. The temperature recorded at a given moment reflects the current over the preceding few time constants, not the current at that moment, so a regression of temperature on present current misreads the physics; the model has to be dynamic. And the sampling of temperature can be much coarser than that of current without losing the thermal dynamics: the reference record reports temperature about once every 213 s, which is $3600/213 \approx 17$ readings per hour, enough to trace a response with a time constant of an hour, but not enough to see anything that happens in seconds.

6.3.2 The thermal RC analogy

Equation (6.6) has exactly the form of the equation of an electrical circuit that Chapter 4 used for polarization: a capacitor C in parallel with a resistor R, fed by a current source I. Kirchhoff's current law at the top node says that the source current splits between charging the capacitor and flowing through the resistor,

$$C\,\frac{\mathrm{d}V}{\mathrm{d}t} = I - \frac{V}{R}, \qquad \text{against} \qquad C_{\mathrm{th}}\,\frac{\mathrm{d}T}{\mathrm{d}t} = \dot{Q}_{\mathrm{gen}} - h\,(T - T_{\mathrm{env}}). \tag{6.10}$$

Term by term, temperature plays the role of voltage, heat flow the role of current, thermal capacitance the role of capacitance, and the thermal resistance $1/h$ the role of electrical resistance. The cooling-air temperature T_{env} plays the role of the ground, the reference against which the node voltage is measured. The time constants match as well: RC for the circuit and $(1/h)C_{\mathrm{th}} = C_{\mathrm{th}}/h$ for the battery.

Figure 6.4 draws the two circuits side by side. On the left is the electrical circuit: a current source (circle with arrow), a capacitor (two plates) and a resistor (zigzag) all connected between

a top node at voltage V and the ground rail. On the right is the thermal circuit with the same topology: the heat source $\dot{Q}_{\text{gen}}$, the thermal capacitance C_{th} and the thermal resistance $1/h$ connected between a node at temperature T and a rail at the cooling-air temperature T_{env}. The correspondence is useful in practice: every result about RC circuits, including the step response of Figure 6.3, carries over to the battery's temperature, and a battery with several thermal masses connected by several heat paths is a network of such circuits.

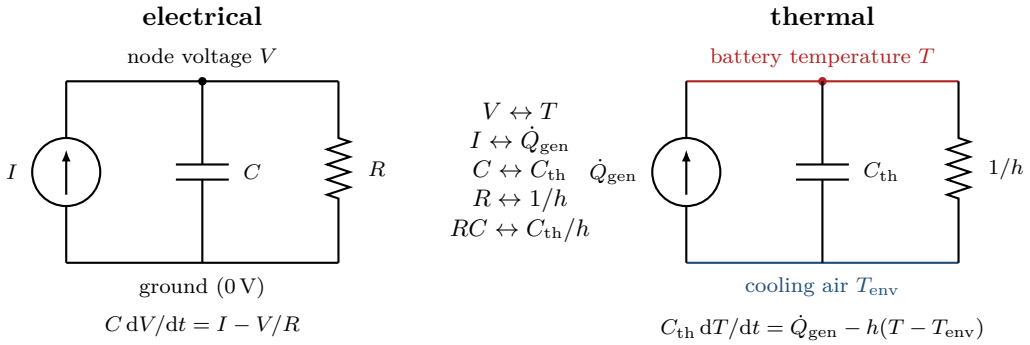


Figure 6.4: The electrical RC circuit (left) and the thermal circuit of the heat balance (right), equation (6.10). The two have the same topology and the same equation; the middle column lists the correspondence. The cooling-air temperature plays the role of ground.

6.3.3 Temperature after a current change

Consider a battery after a period of hard discharge, when the current drops. Generation drops at once, because it is a function of the present current. The stored thermal energy does not: it is still there, and it leaves only as fast as the cooling can carry it. In the illustrative block of Section 6.3.1, one hour at 150 A leaves the block 4.74 K above the air, holding $54\,000 \times 4.74 \approx 256$ kJ more heat than at the start. If the current then drops to 50 A, generation falls from 112.5 to 12.5 W while cooling is still removing $15 \times 4.74 = 71.1$ W, so the block loses heat at about 58.6 W and its temperature starts down, but only gradually, with the one-hour time constant, toward the new equilibrium $12.5/15 = 0.83$ K above the air.

A battery is not a single uniform lump, and that lag has a second form. Heat is generated inside the cells and must travel outward through the cell casing and the module structure before it reaches the cooling air; a temperature sensor on a casing or a busbar sits partway along that path. When the current drops, the inside of the cell is still hotter than the casing, so heat keeps flowing toward the sensor, and the sensor reading can keep rising for a while after the current has dropped. It then flattens, when the inside and the casing reach the same temperature, and only then declines.

Figure 6.5 shows both effects for an illustrative current profile, 150 A for one hour followed by 50 A. The upper panel is the heat generated in watts; the lower panel is the temperature rise above the cooling air in kelvin, both against hours. The gray dashed curve is the single-lump model of equation (6.6) with $\tau = 1$ h; it turns down the moment the current drops. The other two curves are a two-stage version of the same physics: an inner stage, the cell interior, that relaxes toward the generation-driven equilibrium with a time constant of half an hour (thin red), and an outer stage, the sensor location, that follows the inner stage with a time constant of one hour (thick green). The sensor reading is 3.00 K above the air when the current drops, keeps rising for about 22 minutes to a peak of 3.54 K, and then declines. Temperature does not follow current; it follows the heat stored along the path, and a reading taken shortly after a current change can move in the opposite direction from the change.

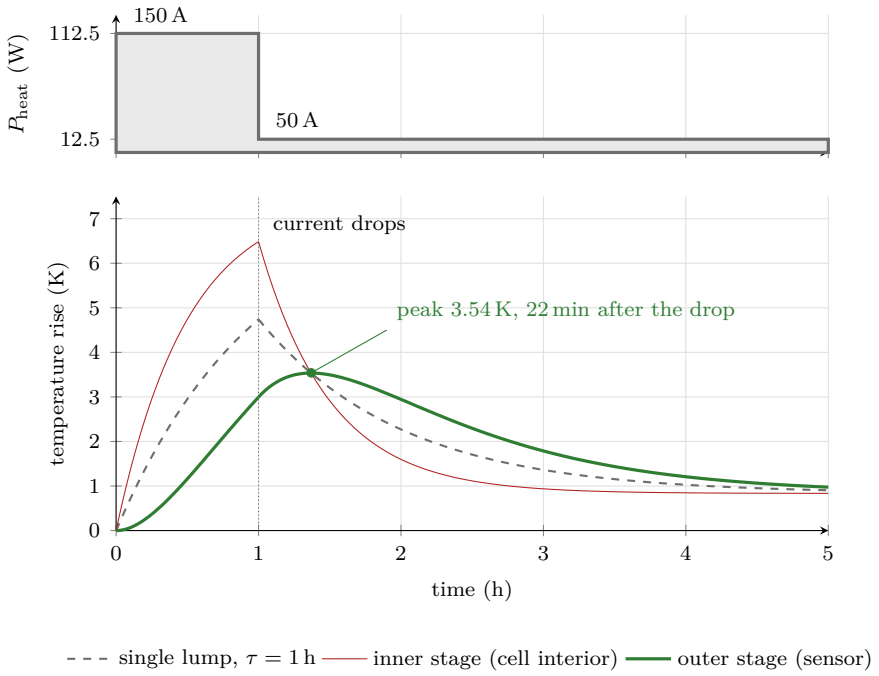


Figure 6.5: Temperature after a current drop, illustrative. Upper panel: heat generation for 150 A then 50 A through $R = 5\,\text{m}\Omega$. Lower panel: rise above the cooling air ($h = 15\,\text{W/K}$, equilibria 7.5 and 0.83 K). Single lump: $\theta = 7.5(1 - e^{-t})$, then $0.83 + 3.91\,e^{-(t-1)}$. Two-stage chain with time constants 0.5 h (inner) and 1 h (outer): the outer-stage step response is $S(t) = 1 + e^{-2t} - 2e^{-t}$, and the sensor reads $7.5\,S(t) - 6.67\,S(t-1)$ after the drop. The sensor keeps rising after the current has fallen.

6.4 Temperature sensors and the rack-mean boundary

The heat balance has so far been written for one temperature. A rack has many, and this section describes what the reference record's temperature channels contain, why their number does not translate into spatial information, and which single temperature the balance can be written for.

6.4.1 Local hot spots

Temperature inside a rack is not uniform. Cells near the cooling inlet run cooler than cells near the outlet, cells in the middle of a module run warmer than those at its edges, and connections carrying the full rack current heat locally. A rack whose sensors average 30°C may have individual locations reading 28, 29, 31 and 38°C. The 38°C point stands out by 8 K, and it may indicate any of several things: a poor electrical connection, whose contact resistance generates its own I^2R heat; locally weak cooling, such as a blocked air path; a cell or module with higher internal resistance than its neighbors; a faulty or miscalibrated sensor; or an abnormal cell beginning to fail. Some of these are harmless and some are early signs of a serious fault.

The average hides all of them. One sensor 8 K above the rest raises the mean of 216 sensors by only $8/216 \approx 0.04$ K, far below the 1 K resolution of any single reading. A rack-mean model is therefore blind to local structure by construction, and local anomalies must be looked for in the individual channels, relative to each other, rather than in the mean.

6.4.2 The temperature channels of the reference record

The reference record carries 216 temperature channels per rack, nine racks, 1944 channels in all. Each reports with a resolution of about 1 K and about one reading every 213 s, so each channel produces $86\,400/213 \approx 406$ readings per day, and each rack about $216 \times 406 \approx 87\,600$ temperature values per day.

What the record does not carry is the physical placement of the sensors. It is not documented which sensor sits near which cell, which module, which part of the air path, which busbar, or how far from the cooling inlet. A channel is a number and a time series; it is not a location. That single gap severely limits spatial inference: without positions, the readings cannot be assembled into a temperature field, cannot be compared with the geometry of the air flow, and cannot be assigned to the cells whose voltages the record also reports.

6.4.3 Why location matters

Suppose sensor 47 always reads warmer than sensor 12. The observation is real and repeatable, but its meaning depends entirely on where the two sensors are. Sensor 47 may be near a hot module, which would be a property of the cells; near a busbar, which would be a property of the connections; in a pocket of stagnant air, which would be a property of the cooling; or it may simply carry a calibration bias, which would be a property of the sensor. Each explanation calls for a different action, and without positions the record cannot choose between them.

The general statement is that the number of measurements is not the amount of identifiable information. Two hundred and sixteen channels whose positions are unknown support far fewer physical conclusions than a dozen channels whose positions are known, because the physics of heat flow, the heat balance of each module and the conduction between neighbors, is written in terms of locations. A measurement that cannot be attached to a location cannot enter a balance written for that location.

Principle 6.3. *The number of measurements is not the amount of identifiable information. Temperature channels whose physical placement is unknown cannot support a balance for any region smaller than the one that certainly contains them all.*

6.4.4 The rack mean as the thermal boundary

Without placement, no module-level heat balance can be written: there is no way to say which channels belong to which module's control volume. The one control volume that certainly contains all 216 sensors is the rack. The finest thermal boundary the reference record supports is therefore the rack, and the temperature used in its balance is the mean of the rack's channels at each reading,

$$T_r = \frac{1}{216} \sum_{j=1}^{216} T_{r,j}, \tag{6.11}$$

where $T_{r,j}$ is channel j of rack r. The rack mean T_r is then modeled dynamically by the heat balance, with the rack current as the input.

Figure 6.6 draws the reduction. The grid on the left represents the 216 channels of one rack, shaded from cool (light) to warm (dark), with one hot channel in red; the grid's arrangement is schematic, because the true arrangement is exactly what is unknown. The arrow is the average of equation (6.11), and the circle on the right is the single rack-mean temperature that enters the heat balance. Everything to the left of the arrow is information the rack balance does not use.

Averaging has one benefit. A single channel reports in 1 K steps, so its rounding error is up to 0.5 K, with a standard deviation of $1/\sqrt{12} \approx 0.29$ K if the error is spread evenly. When the true temperatures across the rack differ by several kelvin, the rounding errors of different channels are close to independent, and the standard deviation of the rounding error of their mean is about $0.29/\sqrt{216} \approx 0.02$ K. The rack mean can therefore follow changes far smaller than one reading's resolution, such as the 0.83 K steady rise of the 50 A example, which a single 1 K channel cannot resolve.

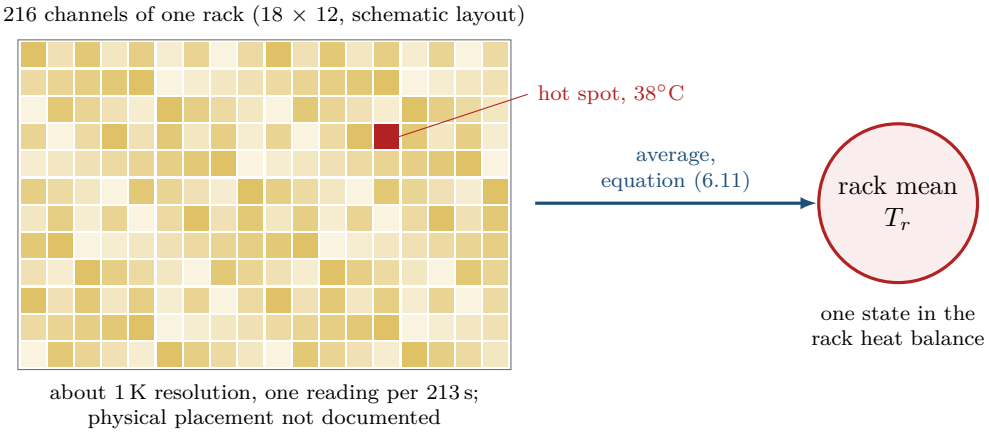


Figure 6.6: From 216 temperature channels to one rack-mean temperature. Shading is schematic (darker is warmer); one channel is a hot spot. Because sensor positions are undocumented, the rack is the finest control volume known to contain all channels, and its mean is the temperature the heat balance is written for.

6.5 The reference thermal equation and its identifiability

This section writes the heat balance in the form the reference record supports, names its three lumped coefficients, and shows that the record can determine only their ratios to the thermal capacitance, together with the one additional measurement that would separate them.

6.5.1 The block equation

The reference record is sampled, not continuous, so the derivative in equation (6.6) is replaced by a difference over a time block of length Δt, such as the interval between two temperature readings. For rack r, with rack current I_r and rack-mean temperature T_r, the reference thermal model is

$$\frac{\Delta T_r}{\Delta t} - \frac{k}{C_{\text{th}}} I_r^2 + \frac{h}{C_{\text{th}}} (T_r - T_{\text{ref}}) + \frac{q_0}{C_{\text{th}}} = 0, \tag{6.12}$$

or, solved for the rate of change,

$$\frac{\Delta T_r}{\Delta t} = \frac{k}{C_{\text{th}}} I_r^2 - \frac{h}{C_{\text{th}}} (T_r - T_{\text{ref}}) - \frac{q_0}{C_{\text{th}}}. \tag{6.13}$$

The right-hand side has three terms, each divided by the thermal mass: an ohmic heating term kI_r^2, a cooling term $h(T_r - T_{\text{ref}})$, and a steady unaccounted heat term q_0. It is equation (6.6) divided by C_{th}, with three changes that the following subsection explains: a lumped coefficient k in place of a resistance, a reference temperature T_{ref} in place of the cooling-air temperature, and an extra constant.

Two details of the discretization matter in practice. The temperature T_r in the cooling term is taken at the start of the block (or as the block average), and the heating term needs the average of I_r^2 over the block, not the square of the average current: if the current alternates within a block between $+100\,\text{A}$ and $-100\,\text{A}$, the average current is zero but the average of I_r^2 is $10\,000\,\text{A}^2$, and the heat is real. A record whose current is sampled more often than its temperature allows that average to be computed; a record sampled at the temperature rate does not.

6.5.2 The coefficients k, h and q_0

The heating coefficient k. The model writes $k\,I_r^2$ rather than $I_r^2 R$ because k is a lumped effective heating coefficient, not a single resistor. The rack current passes through 396 cells in series and through the busbars, connectors and fuses of the rack, so the resistive heat is $I_r^2(\sum_{i=1}^{396} R_i + R_{\text{series}})$, plus the settled polarization losses of Section 6.1.2, all of which scale as I_r^2. With an illustrative $0.25\,\text{m}\Omega$ per cell, the cells alone would contribute $99\,\text{m}\Omega$ and about $2.2\,\text{kW}$ of heat at $150\,\text{A}$. The internal heat paths are not directly known, and k absorbs them all; its units are W/A^2, which is ohms, but it is not the resistance of any one component.

The cooling conductance h. As in equation (6.5), h is an effective conductance and $h(T_r - T_{\text{ref}})$ is the heat lost to the cooling system. A large h means strong cooling: a small temperature excess drives a large heat flow. The reference temperature T_{ref} is a fixed value standing in for the cooling-air temperature. If the true cooling-air temperature T_{env} is constant but different from T_{ref}, then $h(T_r - T_{\text{env}}) = h(T_r - T_{\text{ref}}) - h(T_{\text{env}} - T_{\text{ref}})$, and the second term is a constant that the model cannot distinguish from q_0.

The steady unaccounted heat q_0. q_0 represents a persistent heat input or loss that neither the I^2 heating nor the simple cooling law explains. Possible contributors are the heat of the rack's auxiliary electronics (the battery management boards and their power supplies), a cooling-air temperature that differs from T_{ref} as just shown, or plain model bias. The reference model treats q_0 as one lumped term and does not assign it to a named mechanism. In the sign convention of equation (6.12), a positive q_0 is a steady loss and a negative q_0 a steady gain; written as a source, $+q_0$, the fitted value simply changes sign.

6.5.3 Only ratios are identifiable

Equation (6.13) contains the four parameters k, h, q_0 and C_{th} only through three combinations: k/C_{th}, h/C_{th} and q_0/C_{th}. The thermal capacitance never appears on its own. Multiply all four parameters by 10: each ratio is unchanged, the right-hand side is unchanged, and the predicted temperature trajectory is unchanged for every current profile. The same holds for any positive factor λ,

$$(k,\ h,\ q_0,\ C_{\text{th}}) \longrightarrow (\lambda k,\ \lambda h,\ \lambda q_0,\ \lambda C_{\text{th}}) \quad \Longrightarrow \quad \text{identical } T_r(t) \text{ for every } I_r(t). \tag{6.14}$$

The data therefore cannot determine k, h and C_{th} separately, and the thermal mass cannot be recovered from this equation alone.

This is *structural* unidentifiability. It is not a matter of noise, of short records or of poor excitation: more data of the same kind, at any precision, leaves the ambiguity exactly where it is, because the ambiguity is a property of the model's equations and not of the measurements. It differs from *practical* unidentifiability, where the parameters are determined in principle but the data at hand are too noisy or too uninformative to pin them down, and which more or better data can cure. Chapter 9 develops both.

6.5.4 A numerical demonstration

Rename the three ratios

$$\dot{T} = a\,I^2 - b\,(T - T_{\text{ref}}) - c, \qquad a = \frac{k}{C_{\text{th}}}, \quad b = \frac{h}{C_{\text{th}}}, \quad c = \frac{q_0}{C_{\text{th}}}. \tag{6.15}$$

The data identify a, b and c. Infinitely many quadruples $(k, h, q_0, C_{\text{th}})$ map to the same triple. In arbitrary consistent units, $k = 10$ with $C_{\text{th}} = 100$ gives $a = 0.1$, and so does $k = 20$ with $C_{\text{th}} = 200$. Table 6.1 lists three complete parameter sets: set A, set B with every parameter doubled, and set C with only the thermal capacitance doubled. A and B share all three ratios; C does not.

Table 6.1: Three parameter sets for the reference thermal equation, in arbitrary consistent units. Sets A and B differ by the common factor $\lambda = 2$ and give identical ratios; set C changes C_{th} alone.

Set	k	h	q_0	C_{th}	$a = k/C_{\text{th}}$	$b = h/C_{\text{th}}$	$c = q_0/C_{\text{th}}$	$\tau = 1/b$	$(a-c)/b$
A	10	5	2	100	0.10	0.050	0.020	20	1.6
B	20	10	4	200	0.10	0.050	0.020	20	1.6
C	10	5	2	200	0.05	0.025	0.010	40	1.6

Figure 6.7 simulates all three sets under the same current profile: a unit current from $t = 0$ to $t = 40$, then zero. The horizontal axis is time and the vertical axis is the temperature excess $T - T_{\text{ref}}$, both in the same arbitrary units, starting from zero. Set A is the solid line and set B the open circles placed on it; they coincide at every point, so no record of T and I could distinguish them. Set C is the dashed line. It reaches the same equilibrium excess, 1.6 with the current on, because the equilibrium depends on $(a - c)/b$, in which C_{th} cancels; but it gets there half as fast, because its time constant $1/b = C_{\text{th}}/h$ is doubled, and after the current stops it settles toward $-c/b = -0.4$ along a different path. Changing a ratio changes the trajectory; changing all four parameters together does not.

Principle 6.4. *In the rack heat balance only the ratios k/C_{th}, h/C_{th} and q_0/C_{th} are identifiable. Scaling all four parameters by a common factor leaves every predicted temperature unchanged.*

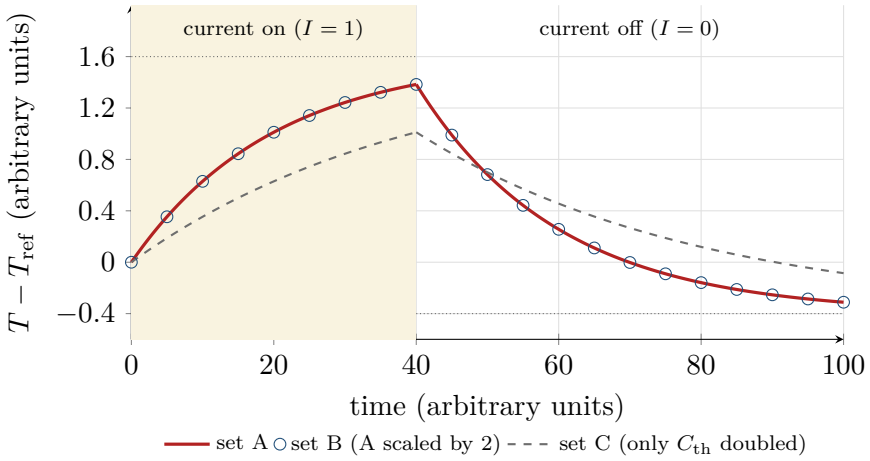


Figure 6.7: The scaling ambiguity of equation (6.14). Trajectories of $\dot{T} = aI^2 - b(T - T_{\mathrm{ref}}) - c$ for the parameter sets of Table 6.1: with the current on, $T - T_{\mathrm{ref}} = 1.6(1 - e^{-bt})$; after $t = 40$ it relaxes toward $-c/b = -0.4$. Sets A and B produce one trajectory; set C, with only C_{th} changed, produces another. Dotted lines mark the two equilibria.

6.5.5 What an independent thermal capacitance would allow

The ambiguity is one-dimensional: fix any one of the four parameters and the other three follow. The natural one to fix is C_{th}, because it can be obtained from sources independent of the operating record: the masses of the components (cells, module housings, racks, busbars) and their specific heats, as in equation (6.3); a calorimetric test, in which a known heat is supplied and the temperature rise is measured; or manufacturer data. With C_{th} in hand, the identified ratios convert back to physical coefficients:

$$k = \Big(\frac{k}{C_{\mathrm{th}}}\Big)C_{\mathrm{th}} = a\,C_{\mathrm{th}}, \qquad h = b\,C_{\mathrm{th}}, \qquad q_0 = c\,C_{\mathrm{th}}. \tag{6.16}$$

The uncertainty of the independent C_{th} then passes into k, h and q_0 in proportion: a C_{th} known to $\pm 10\,\%$ gives k, h and q_0 to about $\pm 10\,\%$ plus their own estimation error.

Some useful quantities need no C_{th} at all. The time constant $\tau = 1/b$ and the steady rise per unit I^2, $a/b = k/h$, are ratios in which C_{th} cancels, and they are what an operator needs to predict how hot a rack will get at a given current and how long it will take. The general strategy is worth stating separately: when only a combination of parameters is identifiable, measure one member of the combination independently, or report only the combinations.

6.6 Thermal faults, electro-thermal coupling and the system so far

The heat balance is also a diagnostic tool: a temperature that departs from what the balance predicts is evidence of a change in one of its terms. This section shows how a cooling fault appears, why heating and cooling faults can be confused and how current separates them, then couples the thermal model to the electrical one and assembles the three-layer system.

6.6.1 Cooling faults

Suppose the normal cooling conductance of the illustrative block is $h = 20\,\mathrm{W/K}$ and a degraded fan or blocked air path lowers it to $h = 10\,\mathrm{W/K}$. For the same $150\,\mathrm{A}$ through $5\,\mathrm{m\Omega}$, the

generation is still 112.5 W, and the steady rise of equation (6.7) doubles:

$$\text{normal: } \frac{112.5}{20} = 5.6\,\text{K}, \qquad \text{degraded: } \frac{112.5}{10} = 11.3\,\text{K}.$$

With cooling air at 25°C, the block settles at 36.3°C after the fault instead of 30.6°C before it. The time constant C_{th}/h also doubles, so the block heats and cools more slowly than before. The signature of a cooling fault is a temperature higher than expected for the observed I^2: the current is the same, the heat generated is the same, and the temperature is not.

6.6.2 Competing explanations of a high temperature

A temperature higher than the model expects is evidence that something changed, but not of what. Each term of the balance offers an explanation: increased internal resistance generates more heat for the same current; weaker cooling removes less heat for the same excess; a warmer cooling-air temperature raises the level everything is measured from; an unmodeled heat source adds heat; and a drifting temperature sensor reports a higher number without any change in the battery. These are competing hypotheses, and the temperature alone does not choose among them.

Take a concrete case: the rack temperature is 5 K above what the model predicts. Five hypotheses are on the table:

- H_1: internal resistance has increased;
- H_2: cooling conductance has decreased;
- H_3: the cooling-air temperature has increased;
- H_4: the temperature sensor is biased;
- H_5: an auxiliary heat load has increased.

The question is not which of them is possible (all are) but which is most consistent with all the measurements available: the temperatures at every current level, the voltages, the time course of the excess, and the other channels. That is model-based diagnosis: each hypothesis is a specific change to the model, each change predicts a specific pattern in the data, and the data are compared with each prediction. Chapter 8 develops the method.

6.6.3 How current separates the hypotheses

The five hypotheses enter the steady balance in two different ways. H_1 and H_2 change the generation per unit I^2 or the removal per kelvin, so their excess scales with the heat load. H_3, H_4 and H_5 add a constant: a warmer air temperature shifts the equilibrium by the air-temperature change, a biased sensor shifts the reading by the bias, and an extra heat load δq shifts the equilibrium by $\delta q/h$, none of which depends on the current. The shape of the excess against I^2 therefore separates the two groups. An anomaly present only at high current supports a resistive cause; an offset present at low current as well points to the air temperature, the sensor or a persistent external heat source.

Table 6.2 works this out for the illustrative block of Section 6.2.5 ($k = 5\,\text{m}\Omega$, $h = 15\,\text{W/K}$), with each hypothesis sized to produce the observed 5 K excess at 150 A. H_1 requires the lumped resistance to rise from 5 to 8.33 mΩ, so that the equilibrium rise is $8.33\times10^{-3}\times22\,500/15 = 12.5\,\text{K}$ instead of 7.5 K. H_2 requires the conductance to fall to $112.5/12.5 = 9\,\text{W/K}$. H_3 and H_4 require 5 K of air-temperature change or of sensor bias, and H_5 requires $\delta q = 5 \times 15 = 75\,\text{W}$ of extra heat. At 50 A, where the normal rise is 0.83 K, the first group predicts an excess of only 0.56 K and the second still predicts 5 K; at rest, the first predicts none.

Table 6.2: Predicted steady temperature excess under each hypothesis, illustrative block ($k = 5\,\mathrm{m\Omega}$, $h = 15\,\mathrm{W/K}$, $C_{\mathrm{th}} = 54\,\mathrm{kJ/K}$), each hypothesis sized to a 5 K excess at 150 A. The last column is the evidence that separates hypotheses with the same current signature.

Hypothesis	Change	150 A	50 A	0 A	Other evidence
H_1 resistance up	k: 5 → 8.33 mΩ	5 K	0.56 K	0	larger voltage drop IR at the same current; time constant unchanged
H_2 cooling down	h: 15 → 9 W/K	5 K	0.56 K	0	no voltage change; time constant 1 → 1.67 h
H_3 air warmer	T_{env}: +5 K	5 K	5 K	5 K	all channels shift together, with the thermal lag
H_4 sensor bias	reading +5 K	5 K	5 K	5 K	appears at once, without lag, in the affected channels only
H_5 extra heat	$\delta q = 75\,\mathrm{W}$	5 K	5 K	5 K	thermal lag; auxiliary power, if metered

Figure 6.8 plots the same predictions as a function of load. The horizontal axis is I^2 in thousands of A^2; the vertical axis is the steady temperature excess over the normal model. The solid line through the origin is the prediction of H_1 and H_2, an excess proportional to I^2; the dashed horizontal line is the prediction of H_3, H_4 and H_5, a constant offset. The two lines meet at the 150 A observation, which is why one observation at one current cannot separate them, and they are far apart at 50 A and at rest. A record that spans several current levels traces out one shape or the other: an excess that grows as I^2 suggests resistive heating, an excess that stays roughly constant suggests an offset or a persistent external source. The shape of the relationship matters more than any single value.

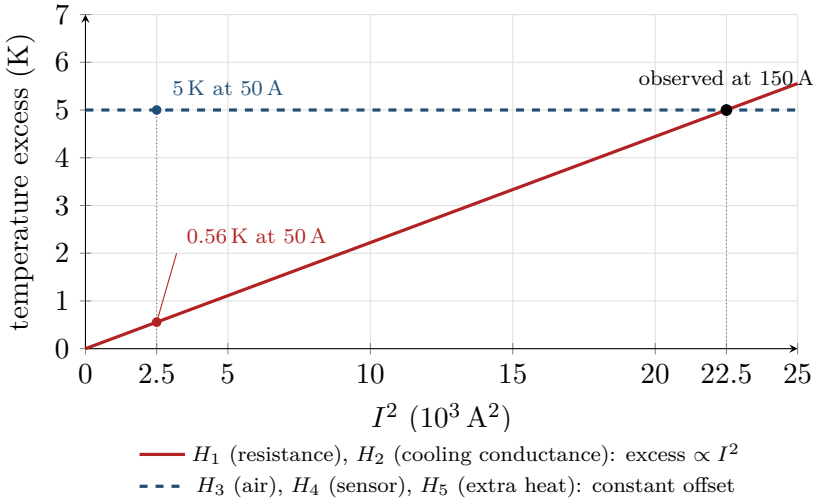


Figure 6.8: Steady temperature excess against I^2 under the five hypotheses of Table 6.2. Proportional line: excess $= 5\,\mathrm{K} \times I^2/22\,500\,\mathrm{A}^2$; constant line: 5 K. A single observation at 150 A lies on both; observations at other currents separate the groups.

Within each group, the current shape is no help and other evidence is needed, which is the last column of Table 6.2. H_1 and H_2 give the same steady temperatures at every current, but H_1 also raises the voltage drop, from $150 \times 0.005 = 0.75\,\mathrm{V}$ to $150 \times 0.00833 = 1.25\,\mathrm{V}$ in the example, which the voltage channels see, and it leaves the time constant unchanged; H_2 leaves

the voltage alone and lengthens the time constant from $54\,000/15 = 3600\,\mathrm{s}$ to $54\,000/9 = 6000\,\mathrm{s}$, which the cooling curve after a current drop shows. H_4 is a property of the measurement and not of the battery, so it appears as a step without thermal lag, and usually in one or a few channels rather than in all 216; H_3 and H_5 are real changes of heat flow and appear with the lag of the time constant. Separating H_3 from H_5 needs a measurement outside the thermal record: the cooling-air temperature or the auxiliary power.

Principle 6.5. *A temperature excess that grows as I^2 points to the generation or removal coefficients; one that is constant in current points to an offset, either in the air temperature, the sensor or a persistent heat source. Hypotheses with the same current signature are separated only by other evidence: voltage, time constant, lag, or channel-to-channel pattern.*

6.6.4 Electro-thermal coupling

The thermal model and the electrical model are not separate. Current produces heat through the resistance, and temperature in turn changes almost every property of the cell: the internal resistance, the usable capacity, the rate of aging, the power the cell can deliver, and the limits within which it may safely operate. The loop is

$$\text{current} \to I^2R \text{ heating} \to \text{temperature} \to \text{resistance changes} \to \text{voltage behavior changes} \to \text{heating changes.}$$

In the normal operating range the internal resistance of a lithium-ion cell falls as temperature rises, so this loop is stabilizing: a warmer cell has a lower resistance and generates less heat for the same current. It runs the other way in the cold, where a high resistance means large voltage drops and large heat, which is one reason battery systems are derated at low temperature. Aging closes a slower loop, which Chapter 7 takes up: high temperature accelerates the side reactions that raise resistance and fade capacity, and higher resistance means more heat.

6.6.5 Thermal runaway as a different regime

Thermal runaway is not an extreme case of the heat balance studied here; it is a different regime with different physics. It is a self-accelerating failure: temperature raises the rate of chemical reactions inside the cell, those reactions are exothermic, the heat they release raises the temperature further, and the cycle accelerates. Reaction rates grow roughly exponentially with temperature, whereas the cooling of equation (6.5) grows only linearly. Below a critical temperature the linear removal can match the generation and an equilibrium exists; beyond it the exponential generation outruns any linear removal, there is no equilibrium, and the temperature rises until the reactants are consumed.

The model of this chapter, I^2 heating plus proportional cooling, is a balance for the normal operating regime. It has a single stable equilibrium for every current and cannot produce runaway, because its generation does not depend on temperature. It is useful for detecting the conditions that precede runaway, such as weak cooling or a hot spot, but it is not a model of runaway itself, and its predictions must not be extrapolated into that regime.

6.6.6 The state-space form and the system so far

In state-space form the thermal model is

$$\frac{\mathrm{d}T}{\mathrm{d}t} = a\,I^2 - b\,(T - T_{\mathrm{env}}) + c, \qquad a = \frac{k}{C_{\mathrm{th}}}, \quad b = \frac{h}{C_{\mathrm{th}}}, \tag{6.17}$$

where c is a lumped constant whose sign is a matter of convention; in the convention of equation (6.12) it is $c = -q_0/C_{\text{th}}$. Temperature is one more dynamic state. The battery model now holds three: the state of charge, the polarization voltage V_1, and the temperature T.

Collecting the layers built so far, with current positive on discharge,

$$\text{charge:} \quad \frac{\mathrm{dSOC}}{\mathrm{d}t} = -\frac{I}{C}, \tag{6.18}$$

$$\text{voltage:} \quad V = U(\mathrm{SOC}) - IR - V_1, \qquad \frac{\mathrm{d}V_1}{\mathrm{d}t} = -\frac{V_1}{R_1 C_1} + \frac{I}{C_1}, \tag{6.19}$$

$$\text{thermal:} \quad C_{\text{th}} \frac{\mathrm{d}T}{\mathrm{d}t} = k\, I^2 - h\,(T - T_{\text{env}}) + q_0. \tag{6.20}$$

The thermal line is written here, as is common, with q_0 as a source; in the convention of equation (6.12) its sign is reversed. The state vector is $(\mathrm{SOC}, V_1, T)^{\mathsf{T}}$ and the input is the current $I(t)$.

Figure 6.9 draws the structure. The box on the left is the single input, the current. The three boxes on the right are the three layers, each labeled with its equation. Solid arrows carry the current into every layer; the arrow from the charge layer to the voltage layer is the open-circuit voltage $U(\mathrm{SOC})$; the dashed arrow from the thermal layer back to the voltage layer is the electro-thermal coupling of Section 6.6.4, the dependence of resistance on temperature, which the reference model leaves out but the physics does not. The right-hand column lists what the reference record observes of each layer: a reported state of charge, which is derived from the current rather than measured; cell and rack voltages; and the 216 temperature channels, reduced to the rack mean. One current drives all three layers: it moves the state of charge and through it the open-circuit voltage, it produces the resistive voltage drop, and it generates the heat that sets the temperature.

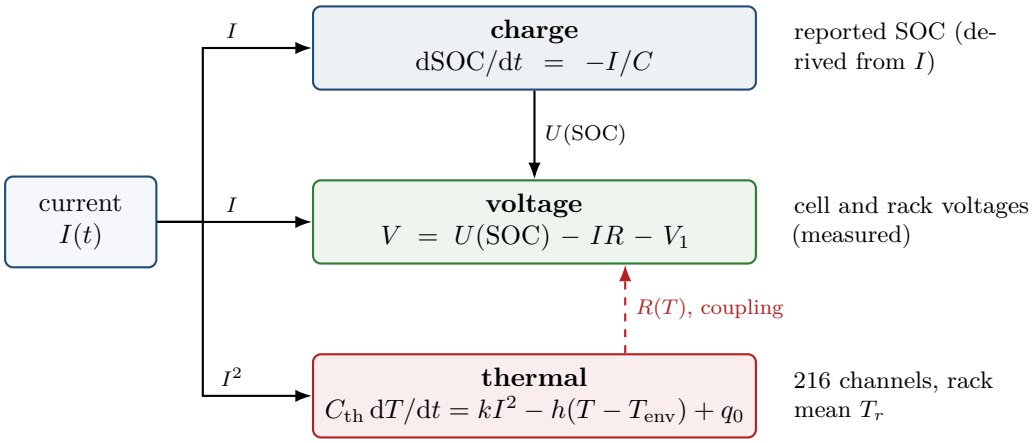


Figure 6.9: The battery model after this chapter: three layers driven by one current, equations (6.18)–(6.20). The state of charge feeds the voltage through the open-circuit voltage; temperature feeds back on resistance (dashed). The right column names what the reference record observes of each layer.

6.7 What to retain

Definition 6.1 (Rack heat balance)**.** A battery is a heat-generating and heat-removing system whose temperature is the state of a conservation law, $C_{\text{th}}\,\mathrm{d}T/\mathrm{d}t = kI^2 - h(T - T_{\text{env}}) - q_0$. From rack currents and rack-mean temperatures, only the ratios of k, h and q_0 to the thermal capacitance can be identified; separating them requires an independent measurement of C_{th}.

Table 6.3 lists the chapter's objects and the subsection that defines each.

Table 6.3: The objects of the thermal model and where each is defined.

Object	Expression	Role	Defined in
Joule heating	$P_{\text{heat}} = I^2 R$	generation, quadratic in current	6.1.1
Thermal capacitance	$C_{\text{th}} \approx m c_p$	storage; sets the speed of response	6.2.1
Cooling law	$\dot{Q}_{\text{cool}} = h(T - T_{\text{env}})$	removal, linear in temperature excess	6.2.2
Heat balance	equation (6.6)	conservation of heat	6.2.3
Equilibrium rise	$I^2 R / h$	steady temperature above the air	6.2.4
Time constant	$\tau = C_{\text{th}} / h$	time to 63 % of a step	6.3.1
Rack mean	T_r, equation (6.11)	finest boundary the record supports	6.4.4
Block equation	equation (6.12)	the reference thermal model	6.5.1
k, h, q_0	lumped coefficients	heating, cooling, steady offset	6.5.2
Identifiable ratios	a, b, c of equation (6.15)	what the data determine	6.5.3
Hypotheses H_1–H_5	Table 6.2	explanations of a temperature excess	6.6.2
Three-layer system	equations (6.18)–(6.20)	charge, voltage and heat from one current	6.6.6

6.8 Exercises

Exercise 6.1. A battery block with $R = 6\,\text{m}\Omega$ and cooling conductance $h = 12\,\text{W/K}$ carries $I = 120\,\text{A}$ with cooling air at 22°C. Compute the heat generated, the steady temperature rise and the steady temperature. Then compute the steady temperature if (a) the current rises to 180 A, and (b) the current stays at 120 A but the conductance falls to 8 W/K. Which change has the larger effect, and why?

Exercise 6.2. A block has $C_{\text{th}} = 72\,\text{kJ/K}$ and $h = 20\,\text{W/K}$ and starts at the air temperature. A constant 150 W of heat is switched on at $t = 0$. Compute the time constant, the equilibrium rise, and the rise after 30 min and after 2 h. How many temperature readings at the reference record's 213 s interval fall within the first time constant, and how long does the block take to reach 95 % of its equilibrium rise?

Exercise 6.3. Fitting equation (6.15) to one rack of a record gives $a = 5 \times 10^{-8}\,\text{K/(s A}^2\text{)}$, $b = 1.25 \times 10^{-4}\,\text{s}^{-1}$ and $c = 5 \times 10^{-5}\,\text{K/s}$. (a) Without any further information, compute the time constant and the steady rise above T_{ref} at 150 A. (b) An independent estimate gives $C_{\text{th}} = 2.0\,\text{MJ/K}$; compute k, h and q_0. (c) Repeat (b) with $C_{\text{th}} = 1.0\,\text{MJ/K}$ and show that the answers to (a) do not change.

Exercise 6.4. The normal model of a rack predicts a steady rise of 12 K at 200 A. The rack-mean temperature is found to exceed the model by 4.8 K at 200 A, by 1.2 K at 100 A and by nothing at rest. (a) Which group of hypotheses in Table 6.2 does the pattern support? (b) The cooling time constant measured after the current stops is unchanged, and the rack voltage drop at 200 A is larger than before. Which hypothesis remains, and by what fraction has the heating coefficient k changed? (c) A second rack shows an excess of 2 K at 0, 100 and 200 A, in all 216 channels, appearing gradually over about an hour. Which hypotheses remain, and what measurement would separate them? If $h = 250\,\text{W/K}$, how much extra heat would H_5 require?

6.9 Connection to the TERRA framework

The rack heat balance is the second conservation law in the battery model, after the charge balance, and it maps onto a graph as directly.

- **The rack heat balance is an energy conservation node.** Equation (6.12) is the balance of a node whose stored quantity is $C_{\mathrm{th}}T_r$, with a generation term kI_r^2 flowing in, a cooling edge $h(T_r - T_{\mathrm{ref}})$ flowing out to an ambient node, and a constant q_0. Figure 6.2 is that node drawn by hand.

- **One edge flow feeds two conservation laws.** The rack current is the flow on the series-string edge common to 396 cells; it enters the charge balance linearly and the heat balance through I^2. The three-layer structure of Figure 6.9 is one edge flow shared by a charge node, a voltage relation and a heat node, and the dependence of R on T is a component relation that couples them.

- **The measurement boundary sets the node.** With sensor positions undocumented, no measurement can be attached to a module-level control volume, so the rack is the finest thermal node the reference record can represent, and the 216 channels are measurements of one node state through equation (6.11). Finer nodes could be written, but no measurement would cross their boundaries; their balances would be identities, not tests.

- **Structural identifiability is checked before estimation.** The scaling of equation (6.14) leaves the likelihood unchanged, so the posterior over C_{th} alone is exactly its prior, whatever the record length. The framework's second question, whether there is enough independent information to determine the unknowns, is answered here in advance: estimate the ratios a, b, c, or add an independent C_{th} as a prior and recover k, h, q_0 by equation (6.16).

- **Faults are hypotheses with signatures.** H_1 to H_5 are parameter shifts of the same node, and Figure 6.8 is their predicted signature against load. A signature is detectable only if the normal model cannot absorb it: a constant offset present from the start of a record is absorbed by a freely fitted q_0, so only a *change* in the offset relative to a reference period can be detected, and H_1 is separated from H_2 only by bringing the voltage relation into the same inference.

In summary, a battery's temperature is the state of a heat balance with quadratic generation, linear removal and a thermal capacitance that sets the speed of response; the reference record supports that balance only at rack level and only up to a common scale, and a temperature anomaly becomes a diagnosis only when its shape against current and its agreement with the voltages are examined together. The next chapter turns to the slow consequences of current, temperature and time: how a battery ages and how its state of health is defined.

Chapter 7

Degradation and state of health: capacity fade, resistance growth and cell imbalance

7.1 Two kinds of degradation

The previous chapters described how a battery behaves now: how its voltage responds to current, how its state of charge moves, how its hardware is arranged and how it heats. This chapter asks how that behavior changes as the battery ages. The question of the chapter is

How does a battery change as it ages, and which of those changes can be determined from its telemetry?

The quantity usually reported for aging is the state of health (SOH). Like the state of charge, it is not measured by any sensor; it is estimated. It is also more complicated than the state of charge, because there is no single quantity called health that is sufficient for every purpose. A battery can lose the ability to store charge, lose the ability to deliver power, or have its cells drift apart, and these are different losses with different causes and different consequences.

Jargon: state of health, capacity fade, resistance growth, calendar and cycle aging, depth of discharge, equivalent full cycle, self-discharge, state of power, remaining useful life

The *state of health* (SOH) is an estimate of how degraded a cell or battery is compared with when it was new; most often it is the ratio of present capacity to rated capacity. *Capacity fade* is the loss of charge capacity C with age. *Resistance growth* is the rise of internal resistance R with age. *Calendar aging* is the degradation that accumulates with elapsed time, even when the battery is not used; *cycle aging* is the degradation caused by charging and discharging. The *depth of discharge* (DoD) of a cycle is the state-of-charge swing between its highest and lowest points. An *equivalent full cycle* counts partial cycles by the charge they move: two half-depth cycles make one equivalent full cycle. *Self-discharge* is the slow loss of stored charge with no external current. The *state of power* (SOP) is the charge or discharge power the battery can safely deliver now. The *remaining useful life* (RUL) is a forecast of the time until an end-of-life criterion is reached.

Key idea

Health is not one directly observable number. A battery ages along at least three separate axes, capacity fade, resistance growth and cell imbalance, and each axis has to be estimated from information that actually depends on it.

7.1.1 The new cell

Take a new cell with charge capacity $C_{\text{new}} = 300\,\text{Ah}$ and internal resistance $R_{\text{new}} = 0.5\,\text{m}\Omega$. The capacity is the charge the cell can move between its upper and lower voltage limits; the resistance sets how far the terminal voltage V falls below the open-circuit voltage U when current flows, and how much heat the cell generates. At a nominal 3.2 V, a 300 Ah lithium iron phosphate cell stores about $3.2 \times 300 = 960\,\text{Wh}$, which is consistent with the reference container, whose 3,564 cells store about 3.42 MWh.

After some years of service the same cell might measure $C = 270\,\text{Ah}$ and $R = 0.8\,\text{m}\Omega$ (illustrative values). Its capacity has fallen and its resistance has risen. These are two different kinds of degradation: the first reduces how much charge the cell can hold, the second reduces how easily it can deliver that charge. Figure 7.1 draws both against age. The left panel has age in years on the horizontal axis and capacity in ampere-hours on the vertical axis, with a dashed line at 80 % of the rated 300 Ah; the right panel has the same horizontal axis and resistance in milliohms on the vertical axis. The curves are illustrative. The capacity curve falls quickly at first and then more steadily, a shape often seen in practice; the resistance curve rises steadily and passes through the values used later in this chapter. At year 6 the cell has reached the state $C = 270\,\text{Ah}$, $R = 0.8\,\text{m}\Omega$.

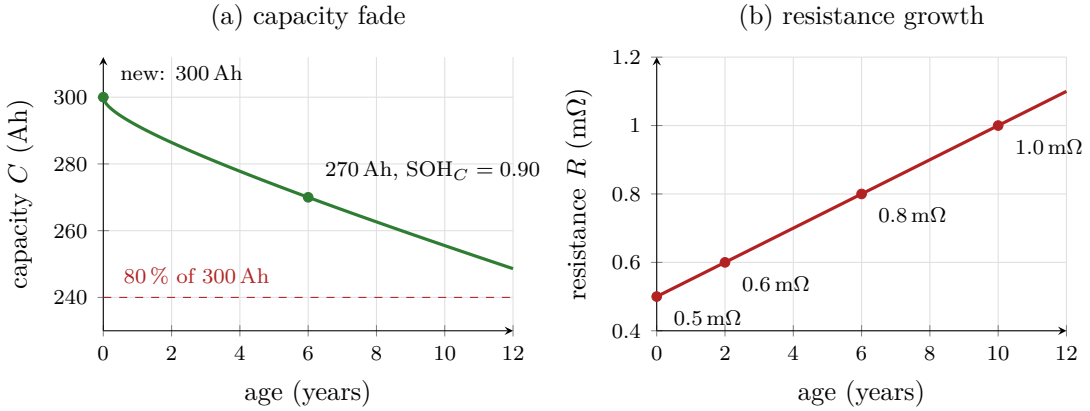


Figure 7.1: Capacity fade and resistance growth of one cell with age. Illustrative curves: (a) $C(t) = 300 - 6\sqrt{t} - 2.55\,t\,\text{Ah}$ with t in years, a square-root term for the early fade and a linear term for the steady fade; (b) $R(t) = 0.5 + 0.05\,t\,\text{m}\Omega$. At year 6 the cell holds 270 Ah ($\text{SOH}_C = 0.90$) and has $R = 0.8\,\text{m}\Omega$.

7.1.2 Capacity fade

The most common definition of state of health is the capacity ratio

$$\text{SOH}_C = \frac{C_{\text{current}}}{C_{\text{rated,new}}}, \tag{7.1}$$

where C_{current} is the capacity the cell has now and $C_{\text{rated,new}}$ the capacity it was rated for when new. The subscript C records that this is a capacity-based health. For the aged cell of Section 7.1.1,

$$\text{SOH}_C = \frac{270\,\text{Ah}}{300\,\text{Ah}} = 0.90,$$

so its capacity-based state of health is 90 %.

The same ratio applies to a string of cells, with one caution that Section 7.4 develops: a series string can only move as much charge as its weakest cell allows, so a rack's capacity-based health

is not the average of its cells' health. If every cell of the reference container had faded uniformly to 90 %, its stored energy would fall from about 3.42 MWh to about $0.90 \times 3.42 = 3.08$ MWh. A common end-of-life criterion for stationary storage is $\mathrm{SOH}_C = 80\,\%$, the dashed line of Figure 7.1(a).

7.1.3 What 90 percent does and does not mean

A capacity-based SOH of 90 % means that the cell holds 90 % of the usable charge it held when new. It means nothing more. In particular it does not mean that the cell is 90 % safe: safety depends on internal defects, temperature and abuse, none of which the capacity ratio measures. It does not mean that the cell retains 90 % of its power capability: power is limited by resistance and voltage limits, which Section 7.1.4 shows can degrade much faster than capacity. And it does not mean that 90 % of the cell's life remains.

The last misreading is the most common. With an end-of-life criterion at 80 %, a cell at 90 % has used half of its allowed capacity loss, not one tenth of its life. How long the remaining 10 points take depends on the future duty, temperature and storage conditions, and the fade rate is generally not constant: in Figure 7.1(a) the first 10 points of fade take six years and the next 10 take about eight and a half. The time to the criterion is a forecast, treated in Section 7.6.3.

7.1.4 Resistance growth

Suppose the resistance of the cell doubles from 0.5 mΩ to 1.0 mΩ. At the same current I the ohmic voltage drop IR doubles and the resistive heating $P_{\mathrm{heat}} = I^2R$ doubles. At an illustrative discharge current of 150 A (half the capacity per hour, a rate of 0.5C) the drop grows from $150 \times 0.0005 = 0.075$ V to 0.150 V per cell, and the heat from $150^2 \times 0.0005 = 11.25$ W to 22.5 W per cell. A rack of 396 cells in series carries the same current through every cell, so the rack loses $396 \times 0.150 = 59.4$ V instead of 29.7 V to resistance, and generates 8.9 kW of heat instead of 4.5 kW.

The consequences are more voltage sag, more heat, and lower power capability, because the terminal voltage reaches its lower limit at a smaller current. All three appear even if the capacity has hardly changed. A cell can therefore pass a capacity test comfortably and still be unable to deliver the power it once delivered, which is the reason the chapter treats capacity fade and resistance growth as two separate kinds of degradation rather than as two views of one.

7.2 Health as a vector

This section shows that two cells can be degraded in ways that cannot be ranked by a single number, collects the separate aspects of health into a vector, and states the physical reasons capacity falls and resistance rises.

7.2.1 Two cells with different kinds of health

Consider two aged cells. Cell A has $C = 285$ Ah and $R = 1.2$ mΩ. Cell B has $C = 250$ Ah and $R = 0.6$ mΩ. Which is healthier has no single answer. Cell A stores more charge: $285/250 = 1.14$ times as much as cell B, so it sustains a given discharge 14 % longer. Cell B delivers power with less sag: at 150 A its ohmic drop is 0.09 V against 0.18 V for cell A, and it generates 13.5 W of heat against 27 W.

Figure 7.2 places the two cells on a plane. The horizontal axis is internal resistance in milliohms, the vertical axis is capacity in ampere-hours; up is more stored charge and left is less voltage sag and heat, so the best cell sits in the upper left corner. The new cell of Section 7.1.1 sits there, at (0.5 mΩ, 300 Ah). Cell A lies to the upper right and cell B to the lower left: each

is better than the other along one axis and worse along the other. For an application that discharges slowly over four hours, cell A is the better cell; for one that delivers short, high-power pulses, cell B is. State of health is multidimensional, and any single number that ranks A against B has made a choice of application without saying so.

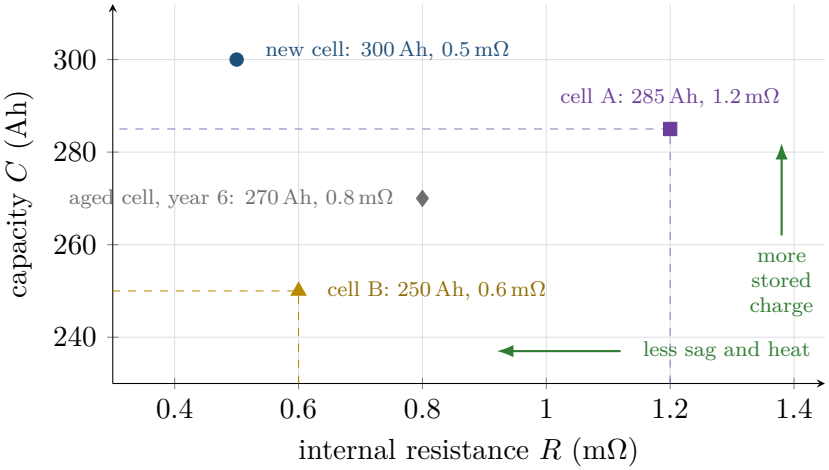


Figure 7.2: Cells on the capacity–resistance plane. Up is more capacity, left is less resistance. Cell A (285 Ah, 1.2 mΩ) and cell B (250 Ah, 0.6 mΩ) do not dominate each other: the dashed lines from each cell bound the region that would be better than it in both respects, and the other cell lies outside that region. Values are illustrative.

7.2.2 The health vector

The natural description of health is therefore a vector,

$$H = \left[\, C,\ R,\ \text{self-discharge rate},\ \text{imbalance},\ \text{thermal behavior},\ \ldots \,\right]^{\mathsf{T}}, \tag{7.2}$$

whose components are the capacity, the internal resistance, the rate at which the cell loses charge at rest (Section 7.4.5), the spread of states across the cells of a string (Section 7.4), and how the cell heats and cools. The list is open: a particular application may add the polarization parameters of the equivalent circuit, or the rate at which each component is changing.

A battery management system (BMS) may collapse this vector into one reported number, most often the capacity ratio of equation (7.1) or a weighted mixture of capacity and resistance. That is a reporting convenience. The physics is richer, and a reader who sees only the collapsed number cannot recover the components from it: cells A and B of Figure 7.2 would receive the same weighted score for many choices of weights.

> **Principle 7.1.** *State of health is a vector. Capacity, resistance, self-discharge and imbalance degrade separately, have different consequences, and must be estimated and reported separately.*

7.2.3 Why capacity degrades

A lithium-ion cell stores charge by moving lithium between two electrodes. The capacity is set by how much lithium is available to move and how much electrode material is available to receive it. Aging removes both. Unwanted side reactions at the electrode surfaces consume lithium and lock it into surface films, where it can no longer take part in charge storage; this is called loss of lithium inventory. Mechanical stress from repeated expansion and contraction

cracks electrode particles and breaks their electrical contact, so part of the active material no longer participates; this is called loss of active material.

Either mechanism reduces the usable capacity C_{usable}, the charge the cell can move between its voltage limits. Neither is visible directly in a voltage or current reading at one instant: the cell still has the same open-circuit voltage curve over most of its range, but it reaches the ends of that range after less charge has moved. That is why capacity fade is only revealed by a measurement that spans a large part of the state-of-charge range, a point Section 7.5 returns to.

7.2.4 Why resistance increases

The same processes make the transport of lithium and electrons harder. Surface films thicken and slow the transfer of lithium into the electrode; the electrolyte degrades and conducts ions less well; particles lose contact and the electronic paths become longer. Each of these raises the internal resistance R, so the ohmic drop IR and the heating I^2R rise at every current.

The effect compounds when the battery is asked for a fixed power rather than a fixed current. Power is $P = VI$ with $V = U - IR$, so the current needed for power P solves $I(U - IR) = P$. For an illustrative per-cell power of 450 W at $U = 3.25$ V, the new cell (0.5 mΩ) needs 141.5 A and its terminal voltage is 3.18 V; the cell at 1.0 mΩ needs 144.9 A at 3.11 V. The aged cell draws more current to make up for its lower voltage, and its heat rises from 10.0 W to 21.0 W, slightly more than double. The same requested power is harder on an older cell, and the extra heat accelerates the temperature-driven aging of Section 7.3.4.

7.3 Aging mechanisms and stress factors

This section separates the two ways a battery ages, with time and with use, and takes the stress factors that set their rates one by one: depth of discharge, the cycle count, temperature, high state of charge and high current.

7.3.1 Calendar aging and cycle aging

Calendar aging happens with elapsed time, even when the battery is not cycled: a cell stored on a shelf for a year loses capacity. Its rate depends on the temperature, the state of charge at which the cell sits, the chemistry, and the time spent at high voltage. Cycle aging happens because the battery is charged and discharged. Its severity depends on the depth of discharge, the charge and discharge rate (C-rate), the temperature during cycling, and the state-of-charge window in which the cycles take place.

Figure 7.3 draws the two as separate inputs to the same degradation. The upper boxes are the two kinds of aging, the lists under them are their stress factors, and the lower box is what both produce: capacity fade and resistance growth. Temperature and state of charge appear in both lists, so a battery that is held hot and full ages faster whether or not it is cycled. A common first model adds the two contributions,

$$\Delta C \approx \Delta C_{\text{cal}}(t,\ T,\ \text{SOC}) + \Delta C_{\text{cyc}}(\text{throughput},\ \text{DoD},\ \text{C-rate},\ T), \tag{7.3}$$

which says that a stationary battery that idles for most of the day is dominated by its calendar term and one that cycles hard by its cycle term. The additive form is an approximation: the two mechanisms share the same surface reactions and are not strictly independent.

7.3.2 Depth of discharge

The depth of discharge of a cycle is the state-of-charge swing it covers. A cycle from SOC = 90 % down to 20 % and back has

$$\text{DoD} = 0.90 - 0.20 = 0.70 = 70\,\%.$$

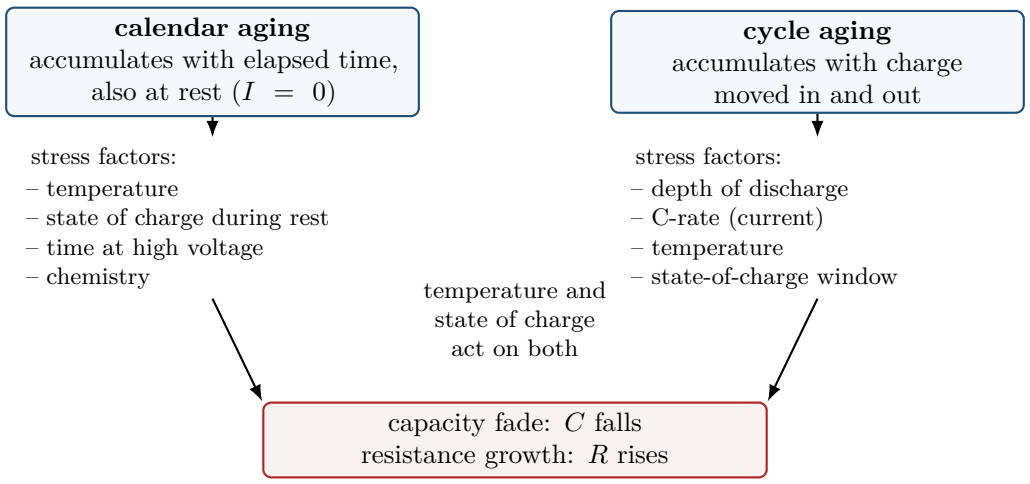


Figure 7.3: Calendar aging and cycle aging as two inputs to the same degradation. Each has its own stress factors; temperature and state of charge appear in both. The outputs are the two kinds of degradation of Section 7.1.

A shallow cycle from 60 % to 50 % and back to 60 % has a swing of about 10 %. The first moves $0.70\,C$ of charge out of the cell and back; the second moves $0.10\,C$.

Different depths age a cell differently, and not in simple proportion to the charge moved. A deep cycle drives the electrodes through a large range of lithium content, with correspondingly large expansion and contraction of the active material; a shallow cycle around the middle of the range keeps that strain small. For many chemistries seven shallow 10 % cycles cause less wear than one 70 % cycle, even though they move the same charge. The depth of each cycle is therefore part of the duty record a degradation model needs, not just the total.

7.3.3 Equivalent full cycles

Because real duty is a mixture of partial cycles of many depths, a common accounting metric is the number of equivalent full cycles,

$$N_{\mathrm{EFC}} = \frac{1}{C}\int \max(I, 0)\,\mathrm{d}t, \tag{7.4}$$

the total charge discharged divided by the capacity (current is positive on discharge). For closed cycles, which return to their starting state of charge, this equals the sum of the depths of discharge. Two 50 % cycles give $N_{\mathrm{EFC}} = 0.5 + 0.5 = 1$, approximately one full equivalent cycle.

Consider an illustrative day with one $90 \to 20$ % cycle and two shallow $60 \to 50 \to 60$ % cycles. It counts $0.70 + 0.10 + 0.10 = 0.90$ equivalent full cycles, or $0.90 \times 365 \approx 329$ per year. Equation (7.4) is useful because it reduces any duty to one number, but by the argument of Section 7.3.2 that number is incomplete: two days with the same count but different mixes of depths do not age the cell equally. Equivalent full cycles are an accounting metric, and a degradation model weights them by depth, rate and temperature.

7.3.4 Temperature

Temperature changes both the immediate electrical behavior of a cell and its long-term aging rate. In the short term a warm cell has lower resistance and delivers power more easily. In the long term high temperature accelerates the side reactions that consume lithium and grow surface films. Reaction rates typically follow an exponential law in temperature; an illustrative form is

$$k(T) = k_{\mathrm{ref}}\,\exp\!\Big[\frac{E_a}{k_B}\Big(\frac{1}{T_{\mathrm{ref}}} - \frac{1}{T}\Big)\Big], \tag{7.5}$$

where k is the aging rate, k_{ref} its value at the reference temperature T_{ref}, E_a an activation energy and k_B Boltzmann's constant. With an illustrative activation energy that makes the rate double for every 10 K, a container held at 35°C instead of 25°C ages twice as fast on its calendar term.

Very low temperature causes problems of a different kind, especially on charge. In a cold cell lithium enters the negative electrode slowly, and at high charging current some of it deposits as metallic lithium on the electrode surface instead; this lithium is lost to storage and can damage the cell. The thermal management of the previous chapter therefore does more than prevent overheating today: by keeping the cells in a moderate band it preserves their life.

7.3.5 High state of charge

Long periods at very high state of charge accelerate some degradation mechanisms, because the electrodes sit at the extreme of their lithium content and the cell sits at its highest voltage, where side reactions are fastest. A battery that sits at 100 % for weeks ages differently from one stored at a moderate state of charge such as 50 %, even if neither is cycled.

This makes the resting state of charge an operating decision with a cost. A stationary battery kept full so that it is always ready to discharge earns readiness at the price of calendar aging; one parked in the middle of its range ages more slowly but needs time to charge before a long discharge. The operating policy therefore weighs more than dispatch revenue: it also weighs the aging that each resting state of charge causes.

7.3.6 High current

Heating grows with the square of the current. For a cell at $R = 0.8\,\mathrm{m\Omega}$, a current of 150 A (0.5C for a 300 Ah cell) generates $150^2 \times 0.0008 = 18\,\mathrm{W}$, and a current of 300 A (1C) generates 72 W: twice the current, four times the heat. For a 396-cell rack these are 7.1 kW and 28.5 kW.

High current also stresses the electrochemistry directly: lithium has to cross the electrode surfaces faster than it can spread inside the particles, which creates large local gradients and, on charge at low temperature, favors the lithium deposition of Section 7.3.4. High C-rate operation therefore contributes to degradation both through heat and through this direct stress, by an amount that depends on chemistry and conditions.

7.4 Imbalance and the weakest cell

A rack is not one cell but a string of cells that are never identical. This section shows how differences in capacity and self-discharge make the cells' states drift apart, why the weakest cell sets the usable window of the whole string, and why the spread of health matters as much as its average.

7.4.1 Unequal capacities in a series string

A rack of the reference container is a series string of 396 cells, and the 396 cells are not identical. Suppose three of them have $C_1 = 300\,\mathrm{Ah}$, $C_2 = 295\,\mathrm{Ah}$ and $C_3 = 270\,\mathrm{Ah}$. Because the cells are in series, the same current I flows through every one of them. Each cell's state of charge nevertheless obeys its own charge balance,

$$\frac{\mathrm{dSOC}_i}{\mathrm{d}t} = -\frac{I}{C_i}, \tag{7.6}$$

with I positive on discharge. The same current divided by a smaller capacity moves the state of charge faster.

The third cell's state of charge moves $300/270 = 1.11$ times as fast as the first cell's. On every discharge it reaches its lower limit sooner, and on every charge it reaches its upper limit sooner. A current that is exactly right for the average cell is slightly too much for the smallest one at both ends of the range.

7.4.2 A numerical example of capacity-driven imbalance

Take two cells, both at SOC = 100 %, with $C_A = 300$ Ah and $C_B = 270$ Ah, and discharge 135 Ah through the string. Then

$$\mathrm{SOC}_A = 1 - \frac{135}{300} = 0.55, \qquad \mathrm{SOC}_B = 1 - \frac{135}{270} = 0.50.$$

The cells started together and are now five percentage points apart. Nothing failed and no current bypassed either cell; the gap is created entirely by the difference in capacity. This is capacity-driven imbalance.

Figure 7.4 draws the two trajectories. The horizontal axis is time in hours at a constant discharge current of 135 A, the vertical axis is state of charge; both lines start at 100 %. After one hour (135 Ah) they are at 55 % and 50 %. After two hours (270 Ah) cell B is empty while cell A still holds 10 %, or 30 Ah. If the string is stopped when its first cell reaches empty, those 30 Ah in cell A cannot be used.

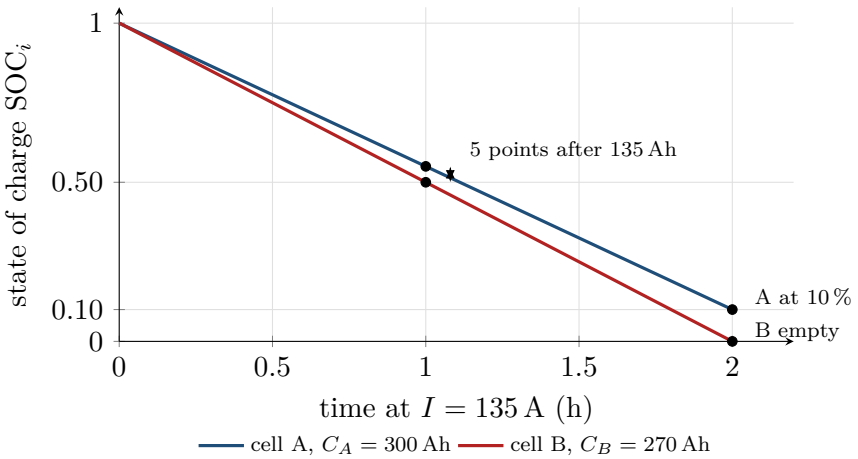


Figure 7.4: Two cells in series discharged by the same current diverge. Closed form $\mathrm{SOC}_i(t) = 1 - It/C_i$ from equation (7.6) with $I = 135$ A, $C_A = 300$ Ah and $C_B = 270$ Ah. After 135 Ah the cells are at 55 % and 50 %; after 270 Ah cell B is empty and cell A still holds 30 Ah.

7.4.3 Why imbalance grows

Capacity is one of several differences that make cells drift apart. Differences in self-discharge make some cells lose charge faster at rest (Section 7.4.5). Balancing circuits, which bleed small currents from the fullest cells to bring them back in line, act unevenly and correct only offsets in state of charge, not differences in capacity. Differences in resistance make cells sit at different terminal voltages under the same current and generate different heat. Differences in temperature across a rack, which the reference record's 216 temperature sensors per rack can reveal, make cells age at different rates by equation (7.5).

These sources reinforce each other. A cell with higher resistance runs hotter, a hotter cell ages faster, and a cell that ages faster loses capacity and gains resistance sooner than its neighbors. Over time one cell reaches its voltage limit first on discharge or on charge, and the

protection system stops the whole rack when it does. Variation in health across cells therefore matters as much as average health.

7.4.4 The weakest cell governs the rack

The usable window of a series string can be written down exactly. Let cell i hold charge $C_i\,\mathrm{SOC}_i$ above its empty point and have room $C_i(1-\mathrm{SOC}_i)$ below its full point. Because every cell carries the same charge, the string can discharge only until the first cell is empty and charge only until the first cell is full:

$$Q_{\mathrm{dis}} = \min_i\ C_i\,\mathrm{SOC}_i, \qquad Q_{\mathrm{ch}} = \min_i\ C_i\,(1-\mathrm{SOC}_i), \qquad Q_{\mathrm{dis}} + Q_{\mathrm{ch}} \le \min_i\ C_i. \tag{7.7}$$

The total window is at most the smallest capacity in the string, and it reaches that bound only when the smallest cell is also perfectly aligned in state of charge with the others at one end of the range.

Suppose most cells of a rack are at $\mathrm{SOH}_C = 90\,\%$ and one is at 70 %. The weak cell may define the usable window of the whole rack. Figure 7.5 shows twelve illustrative cells from such a string. The horizontal axis is the cell index, the vertical axis the cell's capacity in ampere-hours. Eleven cells lie between 267 and 273 Ah; cell 8 has 210 Ah and is drawn in red. The dashed gray line is the mean, 265.2 Ah; the red line is the minimum. A string whose cells all start full can deliver 210 Ah, not 265 Ah: about 21 % less than the average suggests. Rack performance is governed by the tail of the distribution, not its mean, and averaging across cells hides the cell that limits the rack. This is why per-cell behavior is emphasized throughout this book and why the reference record keeps all 396 cell voltages per rack.

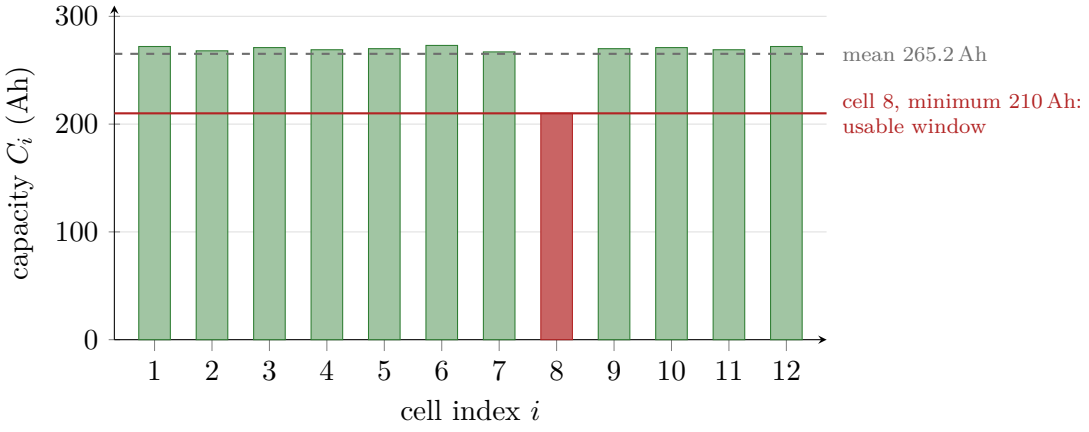


Figure 7.5: The usable window of a series string is set by its weakest cell. Twelve illustrative cells: eleven at 267–273 Ah (about 90 % of 300 Ah) and one at 210 Ah (70 %). By equation (7.7) the string can move at most 210 Ah, although the mean capacity is 265.2 Ah.

> **Principle 7.2.** *A series string is limited by its weakest cell. The usable window of a rack is bounded by the minimum cell capacity, not the mean, so the question of which cell limits the rack can matter more than the average health.*

7.4.5 Self-discharge and the side current

A cell slowly loses charge even with no external current: small internal reactions and leakage paths drain it. With the external current $I = 0$, its state of charge still falls. In a charge balance

self-discharge appears as a small side current $I_{\mathrm{sd},i}$ that adds to the external current,

$$\frac{\mathrm{dSOC}_i}{\mathrm{d}t} = -\frac{I + I_{\mathrm{sd},i}}{C_i}. \tag{7.8}$$

At rack level the same idea appears as one lumped term. The rack-level charge model fitted to the reference record carries a lumped side current, I_{side}, which absorbs self-discharge, balancing currents and the consumption of the rack's own controller: every path by which charge leaves the cells without passing through the measured current sensor. It is small, but because it is integrated over long periods it is not negligible.

Self-discharge also creates imbalance without any capacity fade. Take two cells with identical capacities of 300 Ah, both resting at 50 %, and let cell A self-discharge at an illustrative 3 % of capacity per month and cell B at 1 %. After one month at rest cell A is at 47 % and cell B at 49 %. On the next discharge cell A holds 141 Ah and cell B 147 Ah, so cell A reaches its lower voltage limit 6 Ah earlier, and the string stops there. The cells have the same capacity and the same resistance, and the string has still lost part of its window. Imbalance and capacity fade are separate components of the health vector of equation (7.2).

7.4.6 Health as variability

Consider two racks with the same average cell capacity, 290 Ah. In rack A every cell lies between 285 and 295 Ah. In rack B most cells are at 300 Ah and a minority are near 230 Ah. The averages are similar; the operational risk is very different. Figure 7.6 draws the two distributions as histograms of 396 cells each. The horizontal axis is cell capacity in ampere-hours, the same scale for both panels, and the vertical axis is the number of cells in each 2 Ah bin. The dashed line marks the mean of 290 Ah in both.

Rack A's smallest cell is at about 285 Ah, so by equation (7.7) the rack can move close to 285 Ah. Rack B's smallest cells are at about 228 Ah, so the rack can move only about 228 Ah, some 20 % less, with the same mean. A health report that gives the mean capacity would rate the two racks equal. The question "which cell is limiting?" separates them, and it can matter more than "what is the average capacity?".

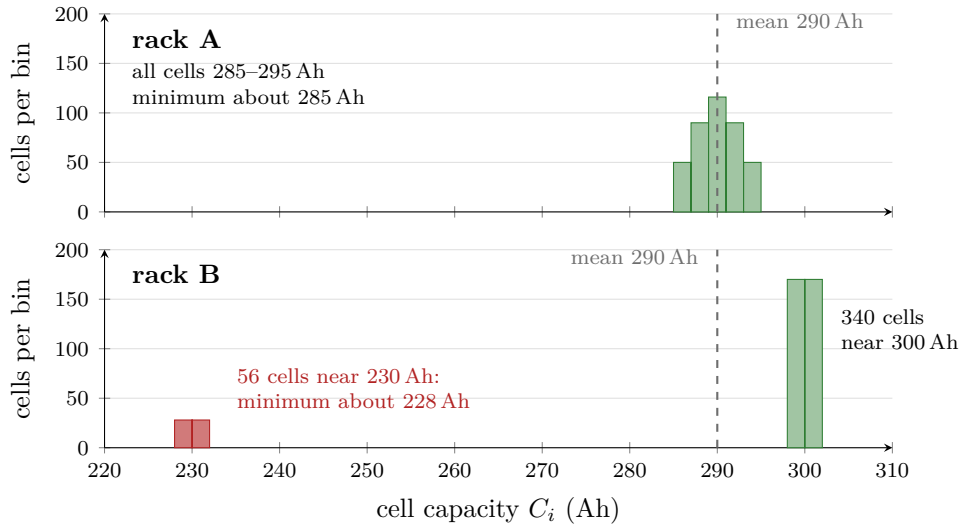


Figure 7.6: Two racks of 396 cells with the same mean capacity, 290 Ah. Rack A (top) has a narrow distribution, 285–295 Ah; rack B (bottom) has 340 cells near 300 Ah and 56 near 230 Ah. By equation (7.7) rack B's usable window is set by its low tail and is about 20 % smaller. Distributions are illustrative.

7.5 Estimating health

Health is not measured; it is estimated from voltage, current and temperature. This section shows why capacity is hard to estimate without independent information, and why resistance is easier.

7.5.1 Capacity from charge and state of charge

Integrating equation (7.6) over a discharge from t_1 to t_2 gives the obvious capacity estimate,

$$\hat{C} = \frac{\Delta Q}{\Delta \mathrm{SOC}} = \frac{\int_{t_1}^{t_2} I \,\mathrm{d}t}{\mathrm{SOC}(t_1) - \mathrm{SOC}(t_2)}. \tag{7.9}$$

If 270 Ah is discharged while the state of charge goes from 100 % to 10 %, then $\Delta\mathrm{SOC} = 0.90$ and

$$\hat{C} = \frac{270\,\mathrm{Ah}}{0.90} = 300\,\mathrm{Ah}.$$

The numerator is a measured quantity: the integral of the current sensor's reading. The denominator is not. The state of charge is not measured by any sensor, so the calculation is only as good as the answer to one question: where did the two state-of-charge values come from? If they were themselves estimated with an assumed capacity, the calculation is circular.

7.5.2 The circularity

A battery management system typically estimates state of charge by coulomb counting,

$$\mathrm{SOC}_{\mathrm{BMS}}(t) = \mathrm{SOC}_0 - \frac{1}{C_{\mathrm{assumed}}} \int_0^t I \,\mathrm{d}\tau, \tag{7.10}$$

with C_{assumed} a capacity stored in its configuration. Substituting equation (7.10) into the denominator of equation (7.9) gives

$$\Delta\mathrm{SOC}_{\mathrm{BMS}} = \frac{\Delta Q}{C_{\mathrm{assumed}}} \quad \Longrightarrow \quad \hat{C} = \frac{\Delta Q}{\Delta Q / C_{\mathrm{assumed}}} = C_{\mathrm{assumed}}.$$

The estimate returns the assumed capacity identically, whatever the true capacity is and however much data is used. The calculation uses an output that was generated from the quantity being sought, and it rediscovers that quantity.

Figure 7.7 draws the loop. The boxes are quantities and calculations, and the arrows show what each is computed from. The measured current and the assumed capacity enter the coulomb count, which produces the reported state of charge; the reported state of charge and the integrated current enter the capacity estimate; and the estimate comes back, along the dashed red path, equal to the assumed capacity. The only way out is a quantity that does not pass through the loop, drawn as the green independent anchor.

The analysis of the reference record finds exactly this. The reported state of charge in the reference record is a coulomb count, and the rack-level charge calculation recovers the capacity assumed by the plant's state-of-charge estimator rather than establishing the true capacity independently. The agreement between the two numbers is an identity, not a confirmation.

The error this hides can be large. Suppose the true capacity of a cell has faded to 270 Ah but the BMS still assumes 300 Ah. Discharging 270 Ah from full empties the cell, yet the coulomb count moves the reported state of charge only by $270/300 = 0.90$, to 10 %. The system believes 30 Ah remain in a cell that has none, and the capacity estimate from equation (7.9) still reads 300 Ah.

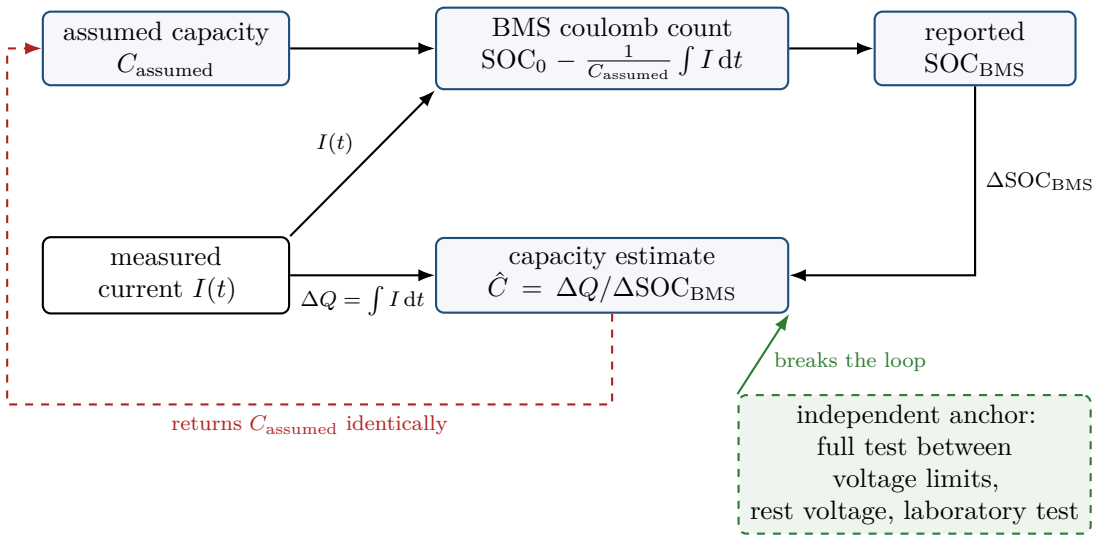


Figure 7.7: The circular path of a capacity estimate built from a coulomb-counted state of charge. The assumed capacity sets the reported state of charge through equation (7.10); dividing the integrated current by the change in that state of charge returns the assumed capacity. Only information that does not pass through the assumed capacity (green) can establish the true capacity.

7.5.3 Independent anchors for capacity

To learn the true capacity, the estimate needs an independent anchor: information about the state of charge that does not come from the assumed capacity. Five sources are common. A full charge and discharge test between the voltage limits fixes both endpoints physically, so the charge moved between them is the capacity. Trusted state-of-charge endpoints, such as a full charge that ends at the upper voltage limit with the current tapered to near zero, fix one end of the range. Rest voltage information, read from a cell that has rested long enough for the polarization voltage V_1 to decay, gives the state of charge through the open-circuit voltage curve $U(\text{SOC})$; for lithium iron phosphate that curve is flat over most of the range, so the rest voltage is informative mainly near the ends. Laboratory characterization of sample cells gives capacity directly under controlled conditions. Finally, long operational records can determine capacity when the duty excites a wide range of state of charge and when independent constraints, such as voltage limits reached at known points, tie the charge count to the physics.

What these have in common is independence. Ten years of coulomb-counted state of charge contain no more information about the true capacity than one day of it, because every sample depends on the same assumed number. What capacity estimation needs is independent information, not more samples of the same dependent quantity.

> **Principle 7.3.** *A capacity estimate is only as independent as its state-of-charge input. Dividing integrated current by a coulomb-counted state of charge returns the assumed capacity; establishing the true capacity requires information that does not pass through that assumption.*

7.5.4 Resistance from current transients

Resistance is often easier to estimate than capacity. When the current changes quickly, the open-circuit voltage and the polarization voltage have no time to move, and the terminal voltage

changes by the ohmic part alone. From $V = U(\mathrm{SOC}) - IR - V_1$,

$$\Delta V \approx -R\,\Delta I \qquad \Longrightarrow \qquad \hat{R} = -\frac{\Delta V}{\Delta I}. \tag{7.11}$$

If an increase in discharge current of $\Delta I = 100\,\mathrm{A}$ causes the cell voltage to fall by $\Delta V = -0.07\,\mathrm{V}$, then

$$\hat{R} = \frac{0.07\,\mathrm{V}}{100\,\mathrm{A}} = 0.0007\,\Omega = 0.7\,\mathrm{m\Omega}$$

(illustrative values). The minus sign follows from the convention that current is positive on discharge: more discharge current lowers the voltage.

This estimate needs no full cycle and no state-of-charge value. Any operational record with current steps provides it, and because the reference record keeps every cell voltage, it gives one resistance per cell rather than one per rack. Two cautions apply. The voltage change must be read over a short window: over longer windows the polarization voltage V_1 responds as well and the ratio grows toward $R_0 + R_1$, the total resistance of the equivalent circuit. And resistance depends on temperature and state of charge, so estimates are comparable only at similar conditions.

7.5.5 Resistance as a trend indicator

Because it can be estimated from routine operation, resistance makes a good health indicator to track over time. Table 7.1 lists the illustrative resistance history of Figure 7.1(b) together with the voltage drop and heat it implies at 150 A, per cell and per 396-cell rack. Over ten years the resistance rises from 0.5 through 0.6 and 0.8 to 1.0 mΩ, and with it the rack's resistive heat rises from 4.5 kW to 8.9 kW at the same current.

Table 7.1: Illustrative resistance history and its consequences at a discharge current of 150 A.

Year	R (mΩ)	cell drop IR (V)	cell heat I^2R (W)	rack heat, 396 cells (kW)
0	0.5	0.075	11.3	4.5
2	0.6	0.090	13.5	5.3
6	0.8	0.120	18.0	7.1
10	1.0	0.150	22.5	8.9

A rising resistance indicates increasing internal loss, and it often moves at a different rate from capacity: in Figure 7.1 resistance has doubled by year 10 while capacity has fallen by only 15 %. A health system that tracks capacity alone misses the change that limits power; one that tracks resistance alone misses the change that limits energy. Practical health systems track more than one parameter.

7.6 State of health, state of power, remaining useful life and the time-varying model

This section separates the four questions asked of a battery, how full, how degraded, how much power and how much life, and then places degradation inside the model equations of the earlier chapters, where capacity and resistance become slowly evolving hidden states.

7.6.1 State of health as an estimate

There is no state-of-health sensor. What exists is a chain from the measured voltage, current and temperature, together with the cycle history, through a model to an estimate:

$$\text{voltage, current, temperature, history} \longrightarrow \text{degradation model} \longrightarrow \text{SOH estimate.}$$

Every number at the end of the chain depends on the model in the middle, and different algorithms give different estimates from the same telemetry. One algorithm may fit capacity to full charges, another may count equivalent full cycles and apply a fade curve, a third may combine resistance and capacity; all three are legitimate and they need not agree.

A reported value such as "90 % SOH" therefore needs a definition before it can be used. It may be the estimated capacity divided by the rated capacity, as in equation (7.1); it may be a mixture of capacity and resistance; it may be a measure of power capability; or it may be a manufacturer-specific metric whose formula is not published. The useful first question about any reported state of health is how it was defined, and the second is what information it was estimated from, which Section 7.5.2 showed can make a capacity-based value an echo of its own assumption.

7.6.2 State of power

The state of power is how much charge or discharge power can safely be delivered now. It depends on the state of charge (through U), on the resistance, on the temperature (through the resistance and through temperature limits), and on the voltage and current limits of the cell. Ignoring the slow polarization voltage, the largest discharge current that keeps the terminal voltage above its lower limit $V_{\min}$ is $(U - V_{\min})/R$, and with a current limit $I_{\max}$ the discharge state of power is

$$\mathrm{SOP}_{\mathrm{dis}} = V\, I^{\star}, \qquad I^{\star} = \min\Big(I_{\max},\ \frac{U(\mathrm{SOC}) - V_{\min}}{R}\Big), \qquad V = U(\mathrm{SOC}) - I^{\star}R. \tag{7.12}$$

Example 7.1. Take illustrative values $U = 3.0\,\mathrm{V}$ at a low state of charge, $V_{\min} = 2.8\,\mathrm{V}$ and $I_{\max} = 300\,\mathrm{A}$. For the new cell at $0.5\,\mathrm{m\Omega}$ the voltage limit would allow $0.2/0.0005 = 400\,\mathrm{A}$, so the current limit binds: $I^{\star} = 300\,\mathrm{A}$, $V = 3.0 - 0.15 = 2.85\,\mathrm{V}$, and $\mathrm{SOP}_{\mathrm{dis}} = 855\,\mathrm{W}$. For the same cell at $1.0\,\mathrm{m\Omega}$ the voltage limit binds first: $I^{\star} = 0.2/0.001 = 200\,\mathrm{A}$, $V = 2.8\,\mathrm{V}$, and $\mathrm{SOP}_{\mathrm{dis}} = 560\,\mathrm{W}$, 65 % of the new value. If cold doubles the resistance again to $2.0\,\mathrm{m\Omega}$, the state of power falls to $2.8 \times 100 = 280\,\mathrm{W}$.

Example 7.1 shows why state of health and state of power are different quantities. An old battery may keep most of its energy capacity and lose much of its power capability, because the two depend on different components of the health vector. A cell at $\mathrm{SOH}_C = 90\,\%$ with doubled resistance has 90 % of its energy and 65 % of its low-state-of-charge power.

7.6.3 Remaining useful life and the four questions

The remaining useful life is how long until an end-of-life criterion is reached, for example $\mathrm{SOH}_C = 80\,\%$. For the cell of Figure 7.1(a) at year 6, with 270 Ah and a criterion at 240 Ah, the present fade rate is the slope of the curve, $3/\sqrt{6} + 2.55 = 3.77\,\mathrm{Ah}$ per year, and a linear extrapolation gives

$$\mathrm{RUL} \approx \frac{270 - 240}{3.77} \approx 8.0 \text{ years;}$$

following the curve itself, whose square-root term slows, gives about 8.6 years. A fade rate of 3.0 or 4.5 Ah per year, the kind of difference a warmer site or a heavier duty makes, moves the answer to 10 or 6.7 years. The remaining useful life is a forecast: it depends on assumptions about

future duty, temperature and degradation mechanisms. The state of health is a present-state estimate. They answer different questions and carry different kinds of uncertainty.

Table 7.2 sets the four questions side by side. The state of charge asks how full the battery is now; the state of health asks how degraded it is compared with new; the state of power asks how much power it can safely provide now; the remaining useful life asks how much useful life remains. Three are present-state estimates on different time scales; the fourth is a forecast.

Table 7.2: The four questions asked of a battery.

Quantity	Question	Kind	Depends mainly on
SOC	How full is it now?	present estimate	integrated current, capacity
SOH	How degraded is it compared with new?	present estimate	capacity, resistance (by definition)
SOP	How much power can it safely provide now?	present estimate	SOC, resistance, temperature, limits
RUL	How much useful life remains?	forecast	present health, future duty and temperature

7.6.4 Degradation in the model equations

The equivalent circuit model of Chapter 4 treated capacity and resistance as constants. With aging they become functions of time, $C = C(t)$ and $R = R(t)$, and the model equations become

$$\frac{\mathrm{dSOC}}{\mathrm{d}t} = -\frac{I}{C(t)}, \qquad V = U(\mathrm{SOC}) - I\,R(t) - V_1. \tag{7.13}$$

The model is now slowly time-varying. Its parameters are no longer fixed numbers to be looked up but hidden, evolving states to be estimated, each with its own slow dynamics. A simple description of those dynamics is a drift with a small random step,

$$C_{k+1} = C_k + \delta_{C,k}, \qquad R_{k+1} = R_k + \delta_{R,k},$$

where the steps δ are small compared with the precision of any single measurement and carry the aging of Section 7.3.

The time-varying form makes the circularity of Section 7.5.2 precise. In equation (7.13) the current is measured and the state of charge is hidden; the capacity appears only as the divisor of the current. A record in which the state of charge is never anchored independently determines only the ratio of integrated current to state-of-charge change, so it cannot separate a change in $C(t)$ from a change in the state of charge it was supposed to explain.

7.6.5 Time scales

The states of equation (7.13) move on very different time scales. Figure 7.8 places them on a logarithmic axis of time from one second to ten years. Each bar spans the range of times over which the state changes appreciably. The polarization voltage V_1 moves in seconds to minutes; the state of charge in minutes to hours; the temperature T also in minutes to hours; capacity and resistance in months to years.

The separation is wide: about four decades of time lie between the temperature's hours and the months over which capacity changes measurably. The battery is a multi-timescale dynamical system, and the separation is useful. Over a window of hours or days, capacity and resistance can be treated as constants and estimated as parameters; over months, the fast states average out and the slow states can be tracked from a sequence of such windows.

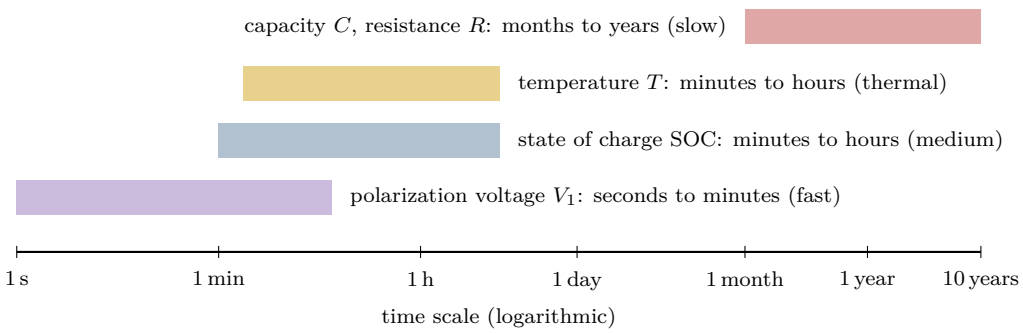


Figure 7.8: The time scales of the battery's states, from one second to ten years on a logarithmic axis. The fast and medium states change within a day; capacity and resistance change over months and years. Bar extents are indicative.

7.6.6 The compact state vector

Collecting the hidden quantities of the chapter gives the state vector and the measurement vector

$$x = \left[\,\mathrm{SOC},\ T,\ V_1,\ C,\ R\,\right]^{\mathsf{T}}, \qquad y = \left[\,V,\ I,\ T_{\mathrm{sensor}}\,\right]^{\mathsf{T}}. \tag{7.14}$$

The state vector contains the fast polarization voltage, the medium state of charge and temperature, and the slow capacity and resistance; the measurement vector contains what the reference record actually keeps for each cell or rack: voltage, current and the temperature read by a sensor, which is not the same as the cell's internal temperature. The degradation problem becomes an estimation problem: given the measurements over time, what can be inferred about the hidden states and parameters, and how precisely?

That framing replaces a single reported number with a set of questions. Which cells limit the rack? Is the reduced performance due to capacity loss, resistance growth, temperature or imbalance? How certain is each answer? What measurement would distinguish the possibilities? These are richer questions than "SOH = 87 %", and they are the questions the last two chapters of this book take up.

7.7 What to retain

Definition 7.1 (Degradation and state of health)**.** Health is not one directly observable quantity. A battery ages by capacity fade, resistance growth and cell imbalance, which must be separated; a reported state of health is an estimate whose definition and information sources determine what it can mean.

Table 7.3 lists the chapter's objects and the subsection that defines each.

7.8 Exercises

Exercise 7.1. A cell rated at 300 Ah now measures 255 Ah, and its resistance has grown from 0.5 mΩ to 0.9 mΩ. (a) Compute SOH_C. (b) At a discharge current of 150 A, compute the ohmic drop and the resistive heat before and after, and the ratio of each. (c) Using equation (7.12) with $U = 3.0\,\mathrm{V}$, $V_{\min} = 2.8\,\mathrm{V}$ and $I_{\max} = 300\,\mathrm{A}$, compute the discharge state of power before and after and express the aged value as a fraction of the new one. Compare that fraction with SOH_C.

Exercise 7.2. Three cells in series have capacities 300, 290 and 255 Ah and all start at SOC = 100 %. (a) The string discharges at 120 A for two hours. Compute each cell's state of

Table 7.3: The objects of this chapter.

Object	Meaning	Defined in
$\mathrm{SOH}_C = C_{\mathrm{current}}/C_{\mathrm{rated,new}}$	capacity-based state of health; 270/300 = 90 %	7.1.2
resistance growth	rise of R; doubles IR and I^2R at fixed current	7.1.4
health vector H	capacity, resistance, self-discharge, imbalance, thermal behavior	7.2.2
calendar and cycle aging	aging with time and with use; equation (7.3)	7.3.1
DoD	state-of-charge swing of a cycle	7.3.2
N_{EFC}	discharged charge divided by capacity	7.3.3
$\mathrm{dSOC}_i/\mathrm{d}t = -I/C_i$	per-cell charge balance; capacity-driven imbalance	7.4.1
usable window	bounded by $\min_i C_i$; the weakest cell governs	7.4.4
$I_{\mathrm{sd},i}$, I_{side}	self-discharge and the lumped rack side current	7.4.5
$\hat{C} = \Delta Q/\Delta\mathrm{SOC}$	capacity estimate; circular with a coulomb-counted SOC	7.5.2
independent anchor	state-of-charge information not derived from C_{assumed}	7.5.3
$\hat{R} = -\Delta V/\Delta I$	resistance from a current transient	7.5.4
SOP	power deliverable now, equation (7.12)	7.6.2
RUL	forecast time to an end-of-life criterion	7.6.3
$C(t)$, $R(t)$	slowly time-varying parameters as hidden states	7.6.4
$x = [\mathrm{SOC}, T, V_1, C, R]^{\mathsf{T}}$	the multi-timescale state vector	7.6.6

charge from equation (7.6). (b) Using equation (7.7), compute the charge the string can deliver from full before its first cell is empty, and the charge it could deliver if every cell had the mean capacity. By what percentage does the mean overstate the usable charge? (c) Repeat (b) when each cell must stay above SOC = 10 %.

Exercise 7.3. (a) A rack current steps from 50 A to 170 A of discharge, and one cell's voltage falls from 3.262 V to 3.178 V within the first second. Estimate the cell's resistance with equation (7.11) and compare it with the new value of 0.5 mΩ. If all 396 cells of the rack had this resistance, what would the step in rack voltage be, before adding the connection hardware? (b) A BMS assumes $C_{\mathrm{assumed}} = 300$ Ah for a cell whose true capacity is 270 Ah. Starting from full, 216 Ah is discharged. Compute the reported state of charge, the true state of charge, the capacity estimate of equation (7.9) from the reported values, and the charge the BMS believes remains compared with what actually remains.

Exercise 7.4. A rack follows the illustrative daily duty of Section 7.3.3: one cycle from 90 % to 20 % and two cycles from 60 % to 50 % and back. (a) Compute the equivalent full cycles per day and per year. (b) Suppose each equivalent full cycle removes 0.02 Ah of capacity and calendar aging removes 1.5 Ah per year at 25°C. Starting from 270 Ah, estimate the remaining useful life to a criterion of 240 Ah assuming constant rates. (c) Using the illustrative doubling per 10 K of Section 7.3.4 applied to the calendar term only, repeat (b) for a site at 35°C. Explain why the answer is a forecast and not a measurement.

7.9 Connection to the TERRA framework

Degradation adds slow states to the battery model and, with them, the first estimation questions whose answer can be that the data do not determine the quantity sought.

- **Capacity and resistance are slow component states.** In the per-cell voltage model, equation (7.13), $C(t)$ and $R(t)$ are states of the cell component with drift dynamics, not constants. The time-scale separation of Figure 7.8 lets the framework hold them fixed within an estimation window and track them across windows.

- **The side current is an unmeasured flow on the charge balance.** The rack node's charge conservation, $\mathrm{d}Q/\mathrm{d}t = -(I + I_{\mathrm{side}})$, has one measured edge flow, the rack current, and one unmeasured leak, the lumped side current of Section 7.4.5. The leak is estimated as a parameter of that node, and self-discharge, balancing and controller consumption cannot be separated within it unless another measurement crosses the boundary.

- **A coulomb-counted state of charge is a derived channel.** Reported state of charge is computed from current by equation (7.10). Fitting capacity to it uses a conservation boundary that no independent measurement crosses, so the result of Figure 7.7 is an identity, not a test: capacity is structurally unidentifiable from current and reported state of charge alone, and an independent anchor is the additional measurement that makes it identifiable.

- **The series string is one edge with 396 components.** The common current flows through every cell, and the per-cell charge balances of equation (7.6) share it. The usable window of equation (7.7) is a bound evaluated cell by cell, which is why the framework carries the distribution of Figure 7.6 rather than its mean.

- **Resistance is practically identifiable only under excitation.** Equation (7.11) needs current steps. A record with flat current leaves the posterior on R wide however long it is, while one with frequent steps narrows it for every cell whose voltage is measured.

In summary, degradation turns the battery's parameters into slowly evolving hidden states, and the chapter's main results (the weakest cell, the circular capacity estimate, the easier resistance estimate) are statements about which of those states the measured channels can determine. The next chapter turns from slow, normal aging to faults: departures whose signatures must lie outside what a model of normal degradation can absorb.

Chapter 8

Faults, anomalies, detection, diagnosis and isolation

8.1 Anomaly, fault, detection, diagnosis and isolation

Chapters 3 to 7 described a battery that behaves normally: a cell whose voltage follows its open-circuit curve and its internal resistance, a rack whose 396 cells carry one current, a heat balance that sets the temperature, and an aging process that slowly lowers capacity and raises resistance. This chapter turns from normal behavior to its departures. The question is

what does it mean for something to be wrong in a battery energy storage system, and how would an operator know?

Answering it requires five terms that are often used as if they meant the same thing: anomaly, fault, detection, diagnosis and isolation. They do not, and most confused reasoning about battery faults comes from treating a symptom as a cause or an alarm as an explanation.

Jargon: anomaly, fault, residual, detection, diagnosis, isolation, signature, detectability, isolability

An *anomaly* is observed behavior that differs from what is expected. A *fault* is a physical or measurement problem that can cause abnormal behavior. A *residual* r is a measured value minus the value a model predicts for it. *Detection* decides whether something is wrong at all; *diagnosis* decides what is wrong; *isolation* decides where. A fault's *signature* is the pattern its presence adds to the measurements, as a function of current, time or charge moved. *Detectability* is the property that a fault can be told apart from normal behavior with the available measurements and model; *isolability* is the property that it can be told apart from the other candidate faults.

Key idea

An anomaly is a symptom and a fault is a cause. A fault is detectable only if its signature produces a pattern that the normal model cannot already reproduce by adjusting its own parameters, and isolable only if that pattern differs from the pattern of every competing fault.

This section defines the five terms, places them in the order in which they are used, and applies them to one low cell voltage.

8.1.1 Anomaly and fault

An *anomaly* is something observed that differs from expected behavior. Four examples from a container like the reference container are: one cell voltage that is consistently lower than the others; a rack temperature that rises faster than usual for the same current; one rack current that differs from the currents of the other eight racks; and a reported rack voltage that disagrees with the sum of that rack's 396 cell voltages. Each of these is a statement about the data. None of them says why the data look that way. An anomaly does not give its cause.

A *fault* is a physical or measurement problem that can cause abnormal behavior. Typical faults in a battery energy storage system are a degraded cell, a bad voltage sensor, a loose connection, a failed cooling fan, a bias in a current sensor, a stuck contactor (a contactor is the switch that connects a rack to the common bus) and an imbalance of state of charge between cells. A fault is a cause; an anomaly is a symptom. The relation between the two lists is many-to-many: one fault can produce several anomalies, and one anomaly can be produced by several faults.

Figure 8.1 draws that relation. The boxes on the left are faults, the boxes on the right are anomalies, and each gray arrow means "can produce". The row of three boxes at the bottom is the sequence of decisions developed in the next two subsections, with an example answer under each. Two features of the upper part matter. A loose connection reaches three different anomalies, so seeing any one of them does not single it out. And the low cell voltage is reached from three different faults, so seeing it does not single out any one of them. Moving from the right column to the left column is the work of the chapter.

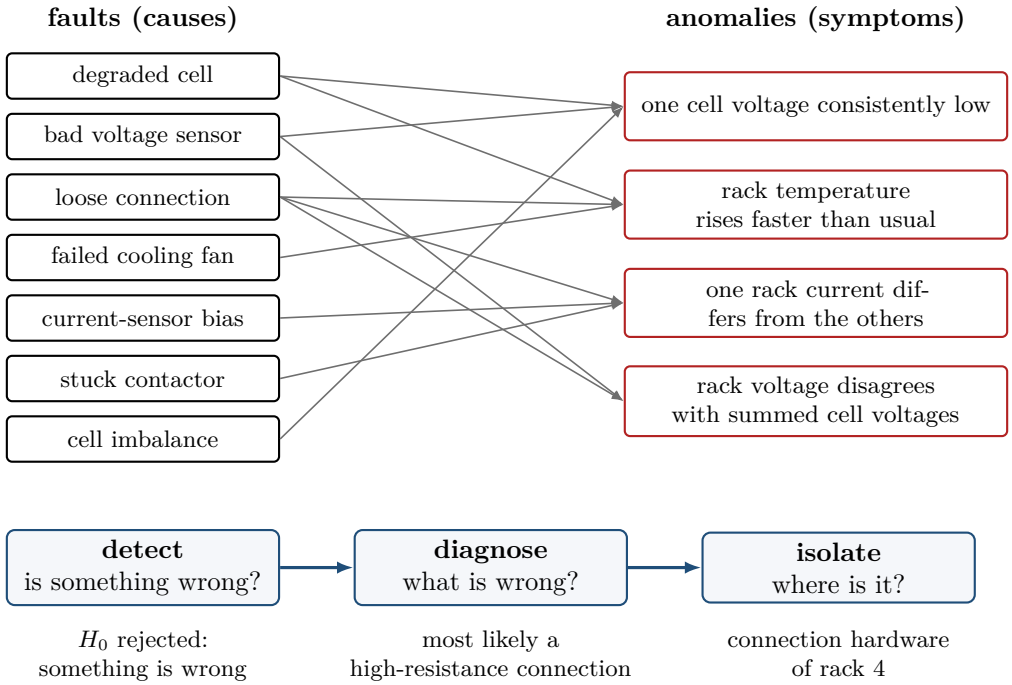


Figure 8.1: Symptoms and causes. Top: faults (left) and the anomalies each can produce (right); the relation is many-to-many, so an anomaly does not identify its fault. Bottom: the three decisions taken in order, with an example answer under each.

8.1.2 Detection and diagnosis

Detection asks whether something is wrong. It is a choice between two hypotheses,

$$H_0: \text{ the system behaves normally}, \qquad H_1: \text{ the system behaves abnormally}, \tag{8.1}$$

and its output is roughly binary: an alarm or no alarm. A hypothesis is a candidate statement about the system that the data can make more or less plausible; H_0 is called the normal or null hypothesis.

Diagnosis asks what is wrong. It is a choice among several competing explanations of the same abnormal behavior, for example

H_1 : sensor bias, H_2 : high connection resistance, H_3 : weak cell, H_4 : cooling problem.

Detection compares normal against not normal; diagnosis compares one abnormal explanation against another. The two are different problems with different outputs. A detector can work well while knowing nothing about the possible causes: it needs only a description of normal behavior and a rule for deciding when the data are too far from it. A diagnosis needs, in addition, a description of what each fault would do to the data, so that the data can favor one explanation over another. Section 8.2 builds the detector and Section 8.4 builds the diagnosis. The subscript of a hypothesis numbers the candidates of one particular problem; H_1 means sensor bias here and something else in a later example.

8.1.3 Isolation

Isolation asks where the fault is: in which cell, rack, sensor, connection or cooling zone. For the reference container the candidate locations are many: 3,564 cells, nine racks, nine rack current sensors, 3,564 cell voltage channels, $9 \times 216 = 1{,}944$ temperature sensors, and the connection hardware and cooling equipment of every rack. The three decisions are taken in order,

$$\text{detect} \longrightarrow \text{diagnose} \longrightarrow \text{isolate},$$

and the bottom row of Figure 8.1 shows one pass through it: something is wrong; the most likely explanation is a high-resistance connection; the connection is in the hardware of rack 4. Diagnosis and isolation often proceed together, because the location of an anomaly is itself evidence about its cause: a departure common to all 396 cells of one rack points to something the rack shares, such as its connection hardware, while a departure confined to one cell points to that cell or its sensor. Section 8.4.3 works this rack 4 example through in full.

8.1.4 One low cell voltage

Suppose most cells of a rack report 3.30 V and one reports 3.10 V. That is an anomaly: a cell 0.20 V below its neighbors. It is not yet a diagnosis, because at least six different causes can produce it. Table 8.1 lists them with the mechanism by which each lowers the reading and the observation that would distinguish it. Each column of the table is used later in the chapter: the mechanism becomes a term in a model, and the distinguishing observation becomes a signature.

The six causes fall into three groups. Two are physical properties of the cell's charge. The cell may sit at a *lower state of charge*, so its open-circuit voltage $U(\mathrm{SOC})$ is lower; on the flat middle of the lithium iron phosphate curve a 0.20 V gap requires the cell to be near the steep low end. Or it may have a *lower capacity*, so that the same current empties it faster and it reaches the low end of its curve before its neighbors. Two are properties of how the voltage responds to current. A *higher internal resistance* R increases the drop IR; at an illustrative 150 A the gap of 0.20 V corresponds to an extra $0.20/150 = 1.33\,\mathrm{m\Omega}$. A *colder cell* sags more through the same mechanism, because resistance rises as temperature falls. Two are properties of the measurement. A *biased sensor* reports 0.20 V less than the cell's true voltage, and *faulty wiring or sensing*, such as a corroded sense lead, corrupts the reading without anything being wrong with the cell.

The right-hand column of Table 8.1 contains the whole method of this chapter in compressed form. The causes differ in *how* the gap depends on current, on charge moved, on temperature

Table 8.1: Six causes of one cell reading 3.10 V while its neighbors read 3.30 V, and what would distinguish each.

Cause	Mechanism	Distinguishing observation
lower state of charge	lower $U(\text{SOC})$	gap persists at rest; disappears after the cells are balanced
lower capacity	SOC falls faster for the same current	gap grows with charge moved; largest near empty
higher internal resistance	larger drop IR	gap scales with current; vanishes at rest; reverses sign on charge
colder cell	resistance rises as temperature falls	gap scales with current and tracks the local temperature sensors
biased sensor	reading = true voltage + b	gap constant at rest and under load, on the flat and the steep parts of the curve
faulty wiring or sensing	extra resistance or poor contact in the sense lead	gap erratic or intermittent; may change when balancing circuits switch

and on time. A single reading of 3.10 V contains none of that information; a record of the cell over a period in which current, charge and temperature vary contains all of it.

8.2 Residuals and thresholds

Detection needs a definition of normal and a rule for deciding when the data have left it. This section builds both: the residual against a physical model, the uncertainty that decides whether a residual is large, and the threshold that turns it into an alarm.

8.2.1 Model-based detection

Model-based detection compares each measurement with what the physics of the normal system predicts for it. For a cell, the simplest prediction is the lumped-resistance model of Chapter 4,

$$V_{\text{predicted}} = U(\text{SOC}) - IR, \tag{8.2}$$

with current positive on discharge. The residual is the difference between what was measured and what was predicted,

$$r = V_{\text{measured}} - V_{\text{predicted}}. \tag{8.3}$$

If the cell is healthy and the model is good, $r \approx 0$ within the uncertainty of the sensor and the model. A large $|r|$ raises suspicion. This is the core of residual-based fault detection: the model removes everything about the measurement that normal physics explains, and what remains is examined.

An illustrative example: a cell at a state of charge where $U = 3.30$ V, with $R = 0.4$ mΩ, carries $I = 150$ A. The predicted voltage is $3.30 - 150 \times 0.0004 = 3.24$ V. A measurement of 3.23 V gives $r = -0.01$ V. The value of the residual is that it removes the large, expected effect of the current (a 60 mV drop) and leaves the small, unexpected part (10 mV). Without the model, a reading of 3.23 V at 150 A and a reading of 3.30 V at rest could not be compared; with it, both reduce to residuals near zero.

Figure 8.2 shows what a residual looks like over time. The horizontal axis is two hours of operation, the vertical axis is the residual in millivolts, the shaded horizontal band is ±15 mV (three standard deviations of the normal residual, explained in Section 8.2.3), and the shaded vertical strip marks a fifteen-minute interval during which a fault adds a departure of 22 mV. Outside the strip the residual scatters around zero and stays inside the band almost everywhere; inside the strip it lifts out of the band and stays out. Two isolated samples, circled, cross the band at minutes 29 and 54 with no fault present. They are the subject of Section 8.2.4.

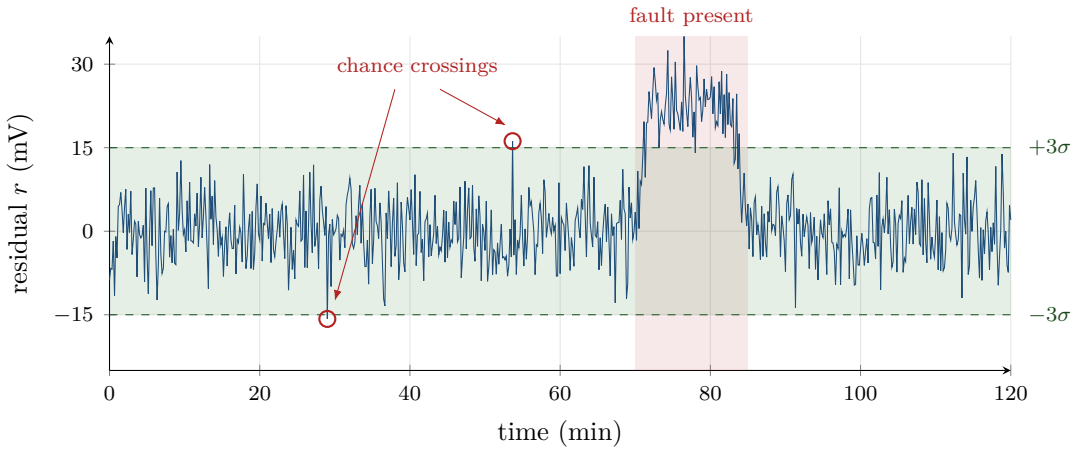


Figure 8.2: A cell-voltage residual over two hours, sampled every 10 s. Under the normal model the residual is Gaussian noise with $\sigma = 5\,\text{mV}$; between minutes 70 and 85 a fault adds 22 mV. The band is $\pm 3\sigma$. Inside the fault window 66 of the 67 samples on the plateau lie outside the band; outside it, 2 of 630 samples do. Synthetic data from `scripts/ch08_data.py` (seed 8); values are illustrative.

8.2.2 Residuals are not automatically faults

A residual of $r = 0.08\,\text{V}$ is not by itself evidence of a fault. At least five sources can produce it, and four of them are not faults. Each is worked here with illustrative numbers that produce exactly 0.08 V.

- *Measurement noise.* If the channel is a rack voltage sensor with a stated accuracy of 0.1 % of a 1,300 V reading, its uncertainty is 1.3 V and 0.08 V is far inside it. If the channel is a cell voltage with a noise of 5 mV, the same 0.08 V is 16 standard deviations.
- *Model error.* Equation (8.2) omits the polarization voltage V_1 of Chapter 4. With an illustrative polarization resistance $R_1 = 0.53\,\text{m}\Omega$, after a long discharge at 150 A the omitted term settles near $150 \times 0.00053 = 0.08\,\text{V}$.
- *Parameter error.* If the resistance used in the model is wrong by $0.5\,\text{m}\Omega$, the predicted drop at 160 A is wrong by $160 \times 0.0005 = 0.08\,\text{V}$.
- *Timing mismatch.* If the current steps by 200 A and the voltage sample is time-stamped one sample before the current sample, the model applies the old current to the new voltage, and with $R = 0.4\,\text{m}\Omega$ the residual during the step is $200 \times 0.0004 = 0.08\,\text{V}$.
- *A true fault.*

The question is therefore never "is r nonzero?" but

is r too large to arise plausibly under the normal model?

That question requires an uncertainty. Write the residual under H_0 as a random variable with standard deviation σ that combines every non-fault source,

$$r \mid H_0 \sim \mathcal{N}(0, \sigma^2), \qquad \sigma^2 = \sigma_{\text{sensor}}^2 + \sigma_{\text{model}}^2, \tag{8.4}$$

where σ_{model} collects model error, parameter error and timing error, each expressed in volts. A residual is large or small only relative to σ. The same 0.08 V is noise on one channel and a strong signal on another.

Principle 8.1. *A residual is evidence, not a verdict. Its size means something only relative to the uncertainty of the normal model, and a model that omits a real effect produces residuals that look like faults.*

8.2.3 Thresholds

The simplest detector raises an alarm when the residual exceeds a threshold τ,

$$\text{alarm if } |r| > \tau, \qquad \text{for example } \tau = 3\sigma. \tag{8.5}$$

Under the Gaussian model of equation (8.4) the probability that a normal residual exceeds 3σ in either direction is $\Pr(|r| > 3\sigma \mid H_0) = 2\,(1 - \Phi(3)) = 0.0027$, where Φ is the standard normal cumulative distribution function. That is the source of the two chance crossings in Figure 8.2: the 630 fault-free samples are expected to contain $630 \times 0.0027 \approx 1.7$ of them.

The choice of τ is a compromise. If τ is too low, ordinary noise crosses it often and the detector produces many false alarms. If τ is too high, real faults whose departure is smaller than τ never cross it and are missed. No value of τ removes both errors at once; the next subsection makes the compromise quantitative.

8.2.4 False positives and false negatives

A *false positive* is an alarm when nothing is wrong. A *false negative* is no alarm when a fault is present. Their rates are

$$\alpha = \Pr(|r| > \tau \mid H_0), \qquad \beta = \Pr(|r| \le \tau \mid H_1), \tag{8.6}$$

and $1 - \beta$ is the *sensitivity*: the probability of catching a fault that is present. The trade-off is between sensitivity and false-alarm rate.

Figure 8.3 draws it. The horizontal axis is the residual; the two curves are its probability density under H_0 (centered at zero) and under a fault H_1 that shifts it by 30 mV, both with $\sigma = 10$ mV. The vertical dashed lines are the thresholds $\pm\tau$, here drawn at $2\sigma = 20$ mV so that both areas are visible. The red areas are the parts of the H_0 curve beyond the thresholds: their total is α. The blue area is the part of the H_1 curve inside the thresholds: its total is β. Moving the threshold to the right shrinks the red areas and grows the blue one; moving it to the left does the opposite. With these numbers,

$$\begin{aligned} \tau = 2\sigma = 20\,\text{mV}: &\quad \alpha = 0.0455, \quad \beta = \Phi\!\left(\tfrac{20-30}{10}\right) = 0.159, \\ \tau = 3\sigma = 30\,\text{mV}: &\quad \alpha = 0.0027, \quad \beta = \Phi(0) = 0.500, \end{aligned}$$

neglecting the negligible part of the H_1 curve below $-\tau$. Raising the threshold from 2σ to 3σ divides the false-alarm rate by 17 and triples the miss rate.

The trade-off becomes decisive when many channels are monitored at once. The reference container has 3,564 cell voltage channels, 1,944 temperature channels and nine rack currents and nine rack voltages: 5,526 channels. If each is tested against $\tau = 3\sigma$ at every scan and the residuals are independent, the expected number of false alarms per scan is

$$5{,}526 \times 0.0027 \approx 14.9.$$

At one scan per second that is about 54,000 false alarms an hour, which no operator can use. A threshold that is conservative for one channel is not conservative for five thousand. Two remedies are standard. The first raises the threshold: at 5σ the per-test rate falls to 5.7×10^{-7}, but for the fault of Figure 8.3 the miss rate rises to $\Phi(2) = 0.98$. The second requires *persistence*: an alarm only after, say, three consecutive crossings of the same sign. If the residuals are

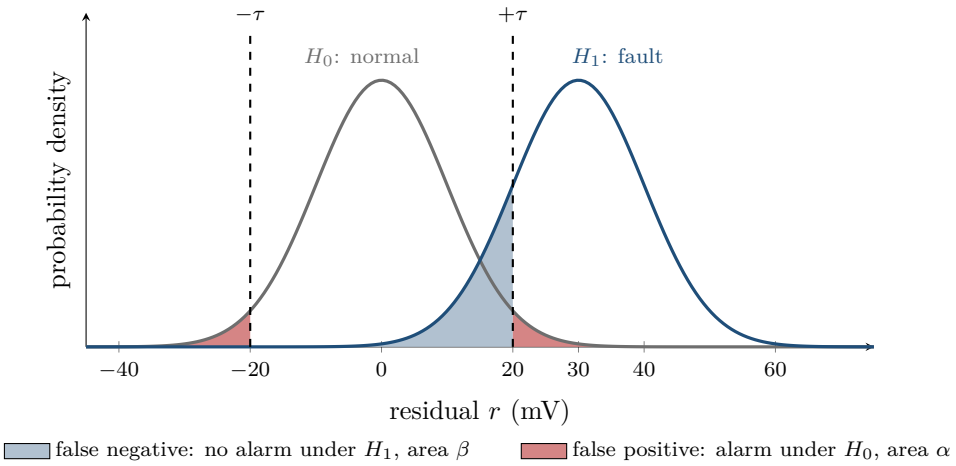


Figure 8.3: The threshold trade-off. Densities of the residual under the normal hypothesis, $\mathcal{N}(0, 10^2)$, and under a fault, $\mathcal{N}(30, 10^2)$, in millivolts, with thresholds at $\pm\tau = \pm 20\,\text{mV}$. Red: false-positive probability $\alpha = 0.0455$. Blue: false-negative probability $\beta = 0.159$. Values are illustrative.

independent, the chance of that under H_0 is $2 \times 0.00135^3 \approx 4.9 \times 10^{-9}$ per channel per scan, about one false alarm every ten hours across the whole container, while a fault that persists, like the one in Figure 8.2, still crosses the band on almost every sample and is caught within seconds. Persistence works because faults last and noise does not; it fails when the noise is itself correlated in time, which is the usual case for model error.

8.3 Fault types and their signatures

A detector says that a residual is too large; it does not say why. This section goes through the main fault types of a battery energy storage system and shows that each one leaves a characteristic pattern, its signature, in how the residual depends on current, charge and time.

8.3.1 Sensor bias and redundancy

A sensor with a bias b reports

$$V_{\text{measured}} = V_{\text{true}} + b. \tag{8.7}$$

With $b = 0.05\,\text{V}$ the sensor reads 50 mV high always: at rest, under load, on charge and on discharge. Seen through that one channel, the bias looks like a physical change in the cell, for example a slightly higher state of charge. Without a second, independent view of the same quantity there is nothing to compare it with, and a bias cannot be told apart from the physics it imitates.

Redundancy supplies the second view. The reference record contains both a rack voltage and the 396 cell voltages of that rack. The rack terminals sit at the ends of the series string, so the rack voltage should equal the sum of the cell voltages up to the drop across the connection hardware between the cells and the terminals and a constant offset between the two measurement systems. Define the discrepancy

$$\Delta V \equiv V_{\text{rack}} - \sum_{i=1}^{396} V_i = V_{\text{off}} - I R_{\text{series}}, \tag{8.8}$$

where R_{series} is the lumped resistance of busbars, contactors and fuses and V_{off} the constant offset. Equation (8.8) is the *rack string voltage identity*. It is a redundancy relation: two

independent sets of sensors measure the same physical voltage, so their disagreement is a residual with a known physical form.

If one sensor drifts while the others stay consistent, the identity exposes it. A bias b that appears in the rack voltage sensor shifts ΔV by $+b$; a bias that appears in one cell's sensor shifts the sum by b and ΔV by $-b$. Either way the discrepancy steps by $|b|$ at the moment the bias appears. A drifting cell sensor is also exposed by its neighbors: the 396 cells of a rack carry the same current, so their voltages should agree to within their state-of-charge spread. A biased cell sensor departs from the median of its 395 neighbors by the same 50 mV on the flat middle of the open-circuit curve and on its steep ends, while a real state-of-charge difference produces a gap that is small on the flat middle and large at the ends.

The word *drifts* matters. A bias present from the first day is a different matter: it adds a constant to ΔV that is indistinguishable from V_{off} itself, a point developed in Section 8.5.2. Redundancy detects a change in a sensor relative to the others; it cannot detect an error that was there before the comparison began unless the offset is known independently.

8.3.2 High-resistance connection

A degraded busbar joint or contactor raises R_{series}. The extra voltage drop it causes is $I\,\delta R$, where δR is the increase, so the discrepancy is small at low current and large at high current. For an illustrative healthy $R_{\text{series}} = 10\,\text{m}\Omega$ and an increase of $\delta R = 5\,\text{m}\Omega$:

Current I	20 A	150 A	300 A
extra drop $I\,\delta R$	0.10 V	0.75 V	1.50 V
extra heat $I^2\,\delta R$ at the joint	2 W	113 W	450 W

(The reference container stores 3.42 MWh in 3,564 cells, about 0.96 kWh per cell, which at a nominal 3.2 V is about 300 Ah; so 300 A is roughly a one-hour discharge rate.) The drop grows in proportion to current and the heat in proportion to its square. A residual that grows with current points to resistance, unlike a constant offset, which is the same at every current. The heat is concentrated at one joint rather than spread over 396 cells, so it can raise the local temperature noticeably even when the voltage effect is a small fraction of a 1,300 V rack.

8.3.3 Offset versus resistance

Equation (8.8) has two unknowns with different roles: V_{off} is constant in current and R_{series} sets the slope. Plotted against I, the discrepancy is a straight line with intercept V_{off} and slope $-R_{\text{series}}$. That is a linear regression with a design matrix X whose rows are $(1, -I_k)$ for samples $k = 1, \dots, n$ and a parameter vector $\theta = (V_{\text{off}}, R_{\text{series}})$,

$$\Delta V_k = V_{\text{off}} - I_k R_{\text{series}} + e_k, \qquad \hat{\theta} = (X^{\mathsf{T}}X)^{-1}X^{\mathsf{T}}\Delta V, \tag{8.9}$$

with e_k the noise. The least-squares estimate of the slope has variance

$$\text{Var}(\hat{R}_{\text{series}}) = \frac{\sigma^2}{\sum_k (I_k - \bar{I})^2}, \tag{8.10}$$

which depends on how widely the current varies, not merely on how many samples there are. With an illustrative per-sample noise of 0.3 V on ΔV and 100 samples spread uniformly over 0 to 300 A, $\sum_k (I_k - \bar{I})^2 = 100 \times 300^2/12 = 750{,}000\,\text{A}^2$ and the standard deviation of $\hat{R}_{\text{series}}$ is $0.3/\sqrt{750{,}000} = 0.35\,\text{m}\Omega$, so a 5 mΩ increase stands out at about 14 standard deviations. If all 100 samples were taken at the same current, the denominator would be zero and the slope would not be estimable at all.

Figure 8.4 shows the two faults as they appear in the residual of a model calibrated on the healthy rack. The horizontal axis is current; the vertical axis is the change in ΔV caused by the fault. The dashed line at zero is the healthy rack. The offset fault, $-0.75\,\mathrm{V}$, is a horizontal line. The resistance fault, $\delta R = 5\,\mathrm{m\Omega}$, is a line through the origin with slope $-5\,\mathrm{m\Omega}$. The two lines cross at 150 A. A record taken only at 150 A sees the same residual under both faults and cannot tell them apart; a record in which the current varies sees two different lines. Varying current separates offset from resistance.

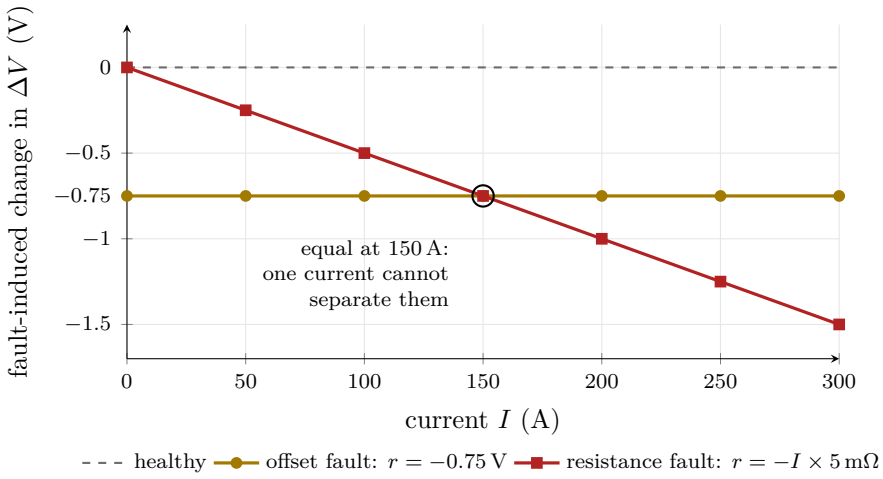


Figure 8.4: Residual of the string voltage identity against current for an offset fault (constant, $-0.75\,\mathrm{V}$) and a resistance fault ($-I\,\delta R$ with $\delta R = 5\,\mathrm{m\Omega}$). The offset sets the intercept and the resistance the slope; the two coincide only at 150 A. Values are illustrative.

8.3.4 Weak cell

A cell can be weak in two different ways, and they leave different signatures. A cell with *low capacity* C_i changes its state of charge faster than its neighbors for the same current,

$$\frac{\mathrm{d}\,\mathrm{SOC}_i}{\mathrm{d}t} = -\frac{I}{C_i}, \tag{8.11}$$

with I in amperes, C_i in ampere-hours and t in hours. The effect is cumulative and depends on history. An illustrative example: a cell of 270 Ah in a string of 300 Ah cells, starting with all cells at SOC $= 0.90$ and discharging at 150 A. After one hour the healthy cells are at $0.90 - 150/300 = 0.40$ and the weak cell at $0.90 - 150/270 = 0.344$. On the flat middle of the lithium iron phosphate curve the voltage gap is only a few millivolts; it becomes large only as the weak cell approaches empty, which it reaches when the healthy cells still hold $270 - 243 = 27$ Ah, nine percent of their capacity. The gap grows with the charge moved, $\int I\,\mathrm{d}t$, and persists after the current stops.

A cell with *high resistance* R_i sags more under current,

$$\Delta V_i \approx I\,\delta R_i, \tag{8.12}$$

with an immediate effect proportional to the present current. For an illustrative $\delta R_i = 0.2\,\mathrm{m\Omega}$ at 150 A the cell reads 30 mV below its neighbors from the first second of the discharge, and the gap vanishes the moment the current stops.

The difference is diagnostically valuable. Capacity loss is a slow, integrating effect that is invisible at the start of a discharge and present at rest; resistance increase is an instantaneous, current-proportional effect that is present at the start of a discharge and absent at rest. A record containing both rest periods and current steps separates them.

8.3.5 Cooling fault

A cooling fault, such as a failed fan or a blocked air path, lowers the effective cooling conductance h of the rack heat balance of Chapter 6,

$$C_{\mathrm{th}}\frac{\mathrm{d}T}{\mathrm{d}t} = P_{\mathrm{heat}} - h\,(T - T_{\mathrm{env}}), \qquad P_{\mathrm{heat}} = kI^2 + q_0. \tag{8.13}$$

For the same current and the same environment, a lower h produces three effects: a larger temperature rise, a longer thermal recovery after the load ends, and a hotter steady state if the load is held. With illustrative values $k = 0.158\,\mathrm{W/A^2}$ (396 cells at $0.4\,\mathrm{m\Omega}$), $q_0 = 0$, $I = 150\,\mathrm{A}$ and $C_{\mathrm{th}} = 2.0\,\mathrm{MJ/K}$, the heat is $P_{\mathrm{heat}} \approx 3.6\,\mathrm{kW}$, and halving h by a fan failure gives

	healthy, $h = 400\,\mathrm{W/K}$	fan failed, $h = 200\,\mathrm{W/K}$
steady rise P_{heat}/h	8.9 K	17.8 K
time constant C_{th}/h	5,000 s (1.4 h)	10,000 s (2.8 h)
recovery to 10 % of the rise, $(C_{\mathrm{th}}/h)\ln 10$	3.2 h	6.4 h

A hotter rack is one symptom with several causes. More heat (a higher k, from cell resistance growth or a bad joint), less cooling (a lower h), a warmer environment (a higher T_{env}) and a biased temperature sensor all raise the reading. They differ in signature: more heat scales with I^2 and vanishes at rest; less cooling changes both the rise and the recovery time; a warmer environment shifts the temperature at rest as well as under load; a sensor bias shifts it by a constant.

8.3.6 Sensor fault versus real thermal fault

Suppose one temperature sensor reads 10 K hotter than expected. Two hypotheses compete: H_1, a real hot spot; and H_2, a sensor bias. The reference record has 216 temperature sensors per rack, so the reading can be checked against several kinds of evidence, listed in Table 8.2 with what each hypothesis predicts for it.

Table 8.2: Evidence that separates a real hot spot from a biased temperature sensor reading 10 K high.

Evidence	H_1: real hot spot	H_2: sensor bias
neighboring sensors	elevated, less so with distance	normal
correlation with current	excess grows with I^2, lagging by the thermal time constant	excess constant, independent of current
persistence during rest	excess decays toward zero over hours	excess stays at 10 K
voltage or resistance of nearby cells	changed: extra sag if the heat comes from resistance, lower resistance if the cells are warmer	unchanged
spatial consistency	fits the airflow pattern and the cell layout	isolated point that no heat path explains

No single row of Table 8.2 decides the question. A hot spot small enough could fail to warm its neighbors measurably; a bias could coincide with a time of day when current is high. The rows are combined. This combination of independent kinds of evidence is called *evidence fusion*, and Section 8.4 gives it a formal rule.

8.3.7 Fault signatures

The previous subsections share one idea: each fault has a signature, a characteristic dependence of its effect on the operating variables. Table 8.3 collects the four basic ones.

Table 8.3: Fault signatures: how each fault's effect depends on current, charge and time.

Fault	Signature	Behavior over an operating cycle
sensor offset	$\Delta V \approx b$, constant	same at rest and under load, on charge and on discharge
resistance increase	$\Delta V \propto I$	zero at rest, follows the current, reverses sign on charge
heating increase	$\Delta T \propto I^2$ (approximately)	follows the square of the current with the thermal lag, same sign on charge and discharge
capacity reduction	effect accumulates with $\int I \, dt$	grows with charge moved, persists at rest, resets only on rebalancing

Figure 8.5 draws the four signatures against a common operating cycle. The top panel is the current: rest, a discharge at one level, rest, a discharge at twice that level, rest. The four panels below are the effects of the four faults on the same time axis, each in its own arbitrary units. The offset is a constant line, blind to the current. The resistance signature is a copy of the current. The heating signature is the current squared, so the second pulse is four times the first, not twice. The capacity signature is the running integral of the current: it ramps during each pulse and holds its level through the rests. The four shapes are different functions of the same input, and that is what makes a diagnosis possible.

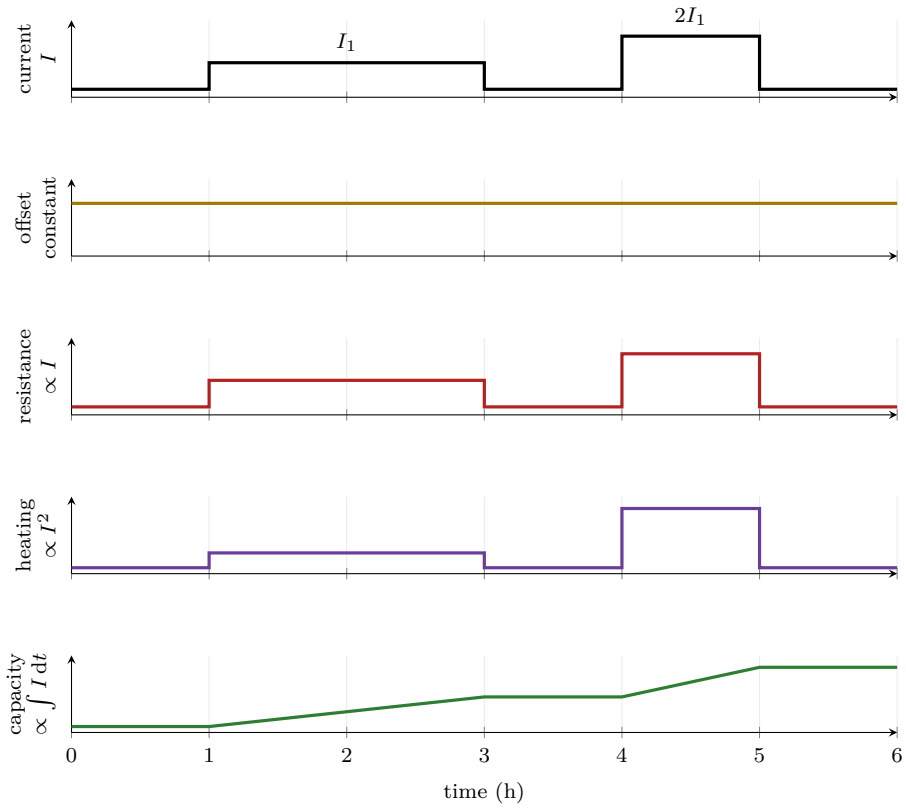


Figure 8.5: The four basic signatures over one operating cycle. Top: a current profile with a pulse I_1 from hour 1 to 3 and a pulse $2I_1$ from hour 4 to 5. Below: the effect of a sensor offset (constant), a resistance increase ($\propto I$), a heating increase ($\propto I^2$, drawn without the thermal lag) and a capacity reduction ($\propto \int I \, dt$), each in arbitrary units.

8.4 Diagnosis as competing hypotheses

Signatures say what each fault would do; diagnosis asks which fault the data actually support. This section states diagnosis as a comparison of competing hypotheses by their posterior probabilities, works one complete example, and defines what a model must provide for diagnosis to be possible.

8.4.1 Competing hypotheses

Consider an abnormally high rack temperature with three candidate explanations: H_1, high internal resistance (more heat); H_2, cooling degradation (less heat removed); H_3, sensor bias (no thermal change at all). Diagnosis asks for the probability of each given the data D, $\Pr(H_i \mid D)$. Bayes' rule gives it from two ingredients,

$$\Pr(H_i \mid D) = \frac{p(D \mid H_i)\ \Pr(H_i)}{\sum_j p(D \mid H_j)\ \Pr(H_j)}. \tag{8.14}$$

The *prior* $\Pr(H_i)$ is how plausible the explanation was before the data, for example from the failure history of similar equipment. The *likelihood* $p(D \mid H_i)$ is how probable the observed data would be if H_i were true; it is computed from the signature of H_i and the noise model of equation (8.4). The *posterior* $\Pr(H_i \mid D)$ is the plausibility after the data. The denominator only rescales the numerators so that the posteriors sum to one over the hypotheses considered.

The difference from a threshold is in the output. A threshold says abnormal or not abnormal. Probabilistic diagnosis says which explanation is most consistent with the evidence and how uncertain that conclusion is. Two cautions follow from the formula. The posteriors are relative to the list of hypotheses: a cause left off the list receives probability zero however well it would fit, so a practical hypothesis set includes an "other" entry. And independent pieces of evidence multiply in the likelihood, which is the formal version of the evidence fusion of Section 8.3.6.

8.4.2 Threshold versus probability

A detector applied to the abnormal temperature says

$$|r| > 3\sigma \implies \text{alarm}.$$

A probabilistic diagnosis of the same data might say

$$\Pr(H_1 \mid D) = 0.60, \qquad \Pr(H_2 \mid D) = 0.30, \qquad \Pr(H_3 \mid D) = 0.10.$$

These numbers arise, for example, from equal priors of $1/3$ and likelihoods in the ratio $6 : 3 : 1$: each posterior is its likelihood divided by $6 + 3 + 1 = 10$. The output carries information the alarm does not. High internal resistance is the leading explanation, but the ratio of its posterior to that of cooling degradation, $0.60/0.30 = 2$, means the data favor it only twofold; one maintenance visit in three directed at the cell would find the cooling at fault instead. The sensor bias is unlikely but not excluded.

The two outputs answer different problems. The detector answers "should someone look?"; the diagnosis answers "what should they look at first, and how sure is that?". A diagnosis also says which measurement would help next: here, one that separates H_1 from H_2, such as the thermal recovery time after the load ends, which under H_2 is twice as long (Section 8.3.5) and under H_1 is unchanged.

8.4.3 A complete example

Rack 4 shows a larger-than-normal voltage drop at high current. Four observations describe it. (E1) The discrepancy ΔV of equation (8.8) departs from its commissioning line by almost nothing at low current. (E2) The departure grows linearly with current. (E3) The rack temperature is slightly elevated. (E4) The 396 cell voltages of rack 4 are mutually consistent: no cell stands out from its neighbors. Three hypotheses compete: H_1, an offset in the rack voltage sensor; H_2, an increased connection resistance; H_3, cell degradation.

The reasoning is qualitative first. The current dependence (E1, E2) argues against a pure offset, which would be constant in current. The rack-wide character of the effect, with consistent cells (E4), argues against degradation of one cell, which would make that cell stand out. The extra heat (E3) supports resistance, since $I^2\,\delta R$ is dissipated at the joint, and is not predicted by a sensor offset. The diagnosis favors H_2.

Table 8.4 makes the same reasoning quantitative with illustrative priors and likelihoods. Each likelihood is the probability of the observation under the hypothesis; E1 and E2 are taken together because they are one observation of the shape of the residual. The observations are treated as conditionally independent given the hypothesis, so the likelihoods multiply.

Table 8.4: Diagnosis of rack 4. Priors and likelihoods are illustrative; the product column is prior times the three likelihoods, and the posterior is the product divided by its column sum, 0.1611.

Hypothesis	prior	E1+E2	E3	E4	product	posterior
H_1: rack sensor offset	0.40	0.05	0.30	0.90	0.0054	0.034
H_2: connection resistance	0.30	0.80	0.70	0.90	0.1512	0.939
H_3: cell degradation	0.30	0.30	0.50	0.10	0.0045	0.028

Figure 8.6 shows the result. Each hypothesis has two bars: the prior (gray) and the posterior (blue). The prior favored the sensor offset, because offsets are common; the evidence moves almost all of the probability to the connection resistance. The posterior of H_1 falls from 0.40 to 0.034 mainly through E1+E2, whose likelihood under an offset is 0.05; the posterior of H_3 falls from 0.30 to 0.028 mainly through E4, whose likelihood under a degraded cell is 0.10. Each competing hypothesis is eliminated by a different observation, which is why the four observations together are decisive when no one of them is. Isolation follows from the same evidence: the effect is in rack 4 and in no single cell, so the fault is in the connection hardware of rack 4.

8.4.4 What makes a good diagnostic model

A diagnosis is possible only if different faults produce distinguishable signatures: fault A produces pattern A, fault B produces pattern B, fault C produces pattern C, and the patterns differ. If two faults produce the same pattern, no amount of reasoning on the data separates them; they can be distinguished only with more information. That is a lack of *isolability*.

A compact way to check this is a table of which residual features each fault affects. Table 8.5 uses four features drawn from this chapter: the constant part of the string-identity residual, its current-proportional part, a cell standing out from its neighbors, and a residual in the rack heat balance. Each row is the pattern of one fault. Two faults are isolable from each other when their rows differ. The last two rows are identical: a failed cooling fan and an unusually warm environment both raise the rack temperature and touch nothing else. If the ambient temperature T_{env} is not measured, the two cannot be told apart from the record. If it is measured, a warm environment becomes part of the normal model's input, its row empties, and the remaining fan failure is isolated.

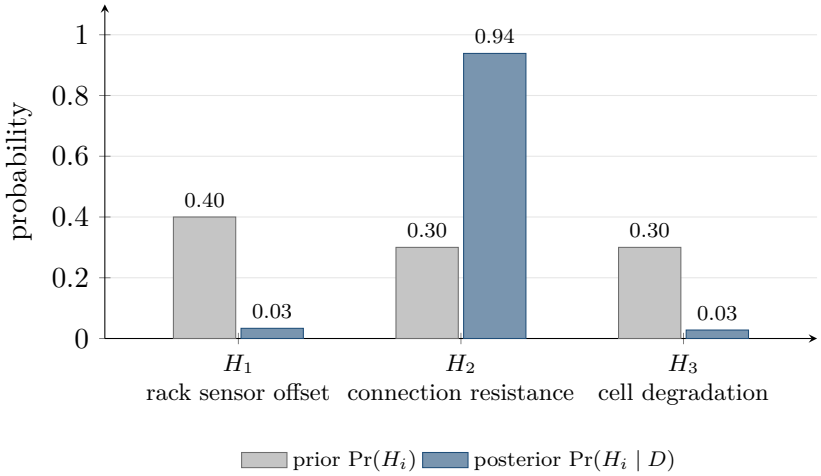


Figure 8.6: Prior and posterior probabilities of the three hypotheses for rack 4, from Table 8.4. The evidence moves the probability from the a priori most common explanation to the one whose signature matches the data. Values are illustrative.

Table 8.5: Which residual features each fault affects. "Weak" means a change that may lie below the noise of that feature. Identical rows are not isolable from each other.

Fault	ΔV constant part	ΔV part $\propto I$	cell vs neighbors	heat balance
rack voltage sensor drift	yes	no	no	no
connection resistance	no	yes	no	yes
cell voltage sensor bias	weak	no	yes	no
weak cell (high resistance)	no	no	yes	weak
cooling fan failure	no	no	no	yes
warm environment (T_{env} unmeasured)	no	no	no	yes

8.4.5 Detectability versus isolability

Two properties must be kept apart. *Detectability* asks whether a departure from normal can be told at all: whether the fault's signature leaves anything in the residual. *Isolability* asks whether the causing fault can be told: whether its signature differs from that of every competing fault. Detectability is a property of one fault against the normal model; isolability is a property of a pair of faults against each other.

A fault can be detectable but not isolable. Return to Figure 8.4 and suppose the rack has operated at a steady 150 A throughout the record. Under both the offset fault and the resistance fault the residual is -0.75 V at every sample. Something is clearly wrong with the voltage behavior, so the fault is detected; but sensor bias and connection resistance remain equally plausible, because their signatures coincide at the only current observed. The cure is not a better detector but a record with a second current level, which gives the two hypotheses different predictions. In Table 8.5 a record at one current merges the constant and the current-proportional columns into one, and the rows of a rack voltage sensor drift and a connection resistance then differ only through the heat balance, whose evidence may be weak.

Principle 8.2. *Detection needs a fault's signature to differ from normal behavior; diagnosis needs it to differ from the signatures of the other faults. Both are properties of the measurements and the model together, not of the fault alone.*

8.5 Faults the model can absorb

The previous sections assumed a normal model with known parameters. In practice the normal model has unknown parameters that are fitted from the same data, and that freedom can hide a fault completely. This section states when it does, in algebra and in geometry.

8.5.1 Departures with declared shapes

An analysis of the reference record defines faults as departures with *declared shapes* and asks, for each, whether the available measurements could distinguish it from the parameters already being fitted. Write the n samples of a channel as a vector y. A departure is a vector a, its shape, multiplied by an unknown scalar magnitude ϕ. The four declared shapes are

$$\begin{aligned} \text{offset:} \quad & a_k = 1, \\ \text{proportional to a column:} \quad & a_k = x_{k,j} \text{ (for example } a_k = -I_k), \\ \text{step at sample } k_0\text{:} \quad & a_k = 1 \text{ if } k \geq k_0, \text{ else } 0, \\ \text{drift:} \quad & a_k = t_k - t_0. \end{aligned} \tag{8.15}$$

The offset is a sensor bias, the column-proportional shape is a resistance-like fault, the step is something that changed at one moment (a joint that loosened, a sensor that was disturbed), and the drift is something that changes steadily. The question asked of each is not whether the fault would change the measurements; every one of them would. It is whether the change creates something the normal model cannot already explain away. A fault is detectable only if its signature creates something the normal model, with its parameters free, cannot reproduce.

8.5.2 The absorbed offset

The string voltage identity (8.8) contains an unknown offset V_{off}, fitted from the data. Suppose the fault of interest is also a constant voltage offset b, a rack sensor that reads high. The data are then

$$\Delta V_k = (V_{\text{off}} + b) - I_k R_{\text{series}} + e_k,$$

and the fit of equation (8.9) returns $\hat{V}_{\text{off}} = V_{\text{off}} + b$ with residuals identical to those of a healthy rack. With illustrative numbers: the true offset is 0.20 V, the fault adds $b = 0.30$ V, the fit reports an offset of 0.50 V, and nothing in the residual reveals that 0.30 V of it is a fault. A fault offset is the same thing, as far as the data can tell, as a different fitted offset.

The fault is invisible. That is not because the sensor is bad or the noise is high; with noise-free data the result would be the same. It is because the fault and a model parameter are mathematically indistinguishable: they enter the measurement in exactly the same way. The same holds for a resistance fault when R_{series} is fitted: the extra $-I_k\,\delta R$ is absorbed into $\hat{R}_{\text{series}} = R_{\text{series}} + \delta R$, and the residuals again look healthy.

8.5.3 The span argument

The general statement is a fact about linear algebra. Let the normal model be $y = X\theta + e$, with X the $n \times p$ design matrix and θ the p unknown parameters. A fault adds $a\phi$,

$$y = X\theta + a\phi + e. \tag{8.16}$$

If the signature a lies in the span of the columns of X, that is, if $a = Xc$ for some vector c, then

$$X\theta + a\phi = X\theta + Xc\,\phi = X(\theta + c\,\phi), \tag{8.17}$$

and the faulty data are exactly the data of a healthy system with the parameter vector $\theta + c\phi$. The fault is absorbed into a different parameter vector, and it is undetectable from that combination of model and dataset. For the offset fault of the previous subsection X has rows $(1, -I_k)$, a is the column of ones, and $c = (1, 0)$; for the resistance fault a is the column $-I$ and $c = (0, 1)$.

Most signatures are neither wholly inside nor wholly outside the span. The least-squares fit removes the part of y that lies in the span, by the projection $P_X = X(X^{\mathsf{T}}X)^{-1}X^{\mathsf{T}}$, so the residual after fitting is $r = y - P_X y$. Substituting equation (8.16),

$$r = (e - P_X e) + \phi\, a_\perp, \qquad a_\perp = a - P_X a. \tag{8.18}$$

Only $a_\perp$, the part of the signature orthogonal to the span, reaches the residual. The part $P_X a$ is absorbed into the fitted parameters. The strength of the evidence for the fault is $\phi\,\|a_\perp\|/\sigma$: under H_0 the sum of squared residuals divided by σ^2 follows a chi-squared distribution with $n - p$ degrees of freedom, and the fault shifts it by $\phi^2\|a_\perp\|^2/\sigma^2$.

A four-sample example shows that the split depends on the operating history, not only on the fault. Take X with rows $(1, -I_k)$ and a step fault $a = (0, 0, 1, 1)$ that begins at the third sample. If the currents were $I = (0, 100, 200, 300)$ A, rising with time, the regression of a on the columns of X gives $P_X a = 0.5 + 0.004\,(I - 150) = (-0.1, 0.3, 0.7, 1.1)$ and

$$a_\perp = (0.1, -0.3, 0.3, -0.1), \qquad \|a_\perp\|^2 = 0.2 = 10\,\%\text{ of }\|a\|^2 = 2.$$

Because the current rose steadily, a step in time looks almost like a dependence on current, and ninety percent of the step is absorbed: a voltage step of $\phi = -0.5$ V is reported mostly as an increase of $0.004 \times 0.5 = 2\,\text{m}\Omega$ in the fitted R_{series}. If the same four currents had been applied in the order $I = (0, 300, 100, 200)$ A, the current would be uncorrelated with the step, $P_X a$ would be the constant 0.5, and

$$a_\perp = (-0.5, -0.5, 0.5, 0.5), \qquad \|a_\perp\|^2 = 1 = 50\,\%\text{ of }\|a\|^2 :$$

the mean of the step is still absorbed by V_{off}, but its change in time is not. The same fault, the same sensors and the same model are five times more detectable under the second operating history.

8.5.4 The geometric picture

Equations (8.17) and (8.18) have a geometric reading. Every possible record of n samples is a point in an n-dimensional space. As θ varies, the predictions $X\theta$ of the normal model sweep out a subspace, the column space of X. A fault moves the record in the direction of its signature a.

Figure 8.7 draws this in perspective. The shaded parallelogram is the column space of X: every record the normal model can produce by changing its parameters. The green arrow is a fault direction lying inside it, $a = Xc$. The red arrow is a fault direction pointing out of it. The dashed segment in the plane is the projection $P_X a$ of the red arrow onto the plane, and the dashed vertical segment is the remainder $a_\perp$, at right angles to the plane. A fault along the green direction moves the record to another point the normal model can reach, so it is indistinguishable from ordinary parameter variation. A fault along the red direction moves the record off the plane; the fit can follow it only as far as its projection, and the vertical component is left in the residual, where it can be detected.

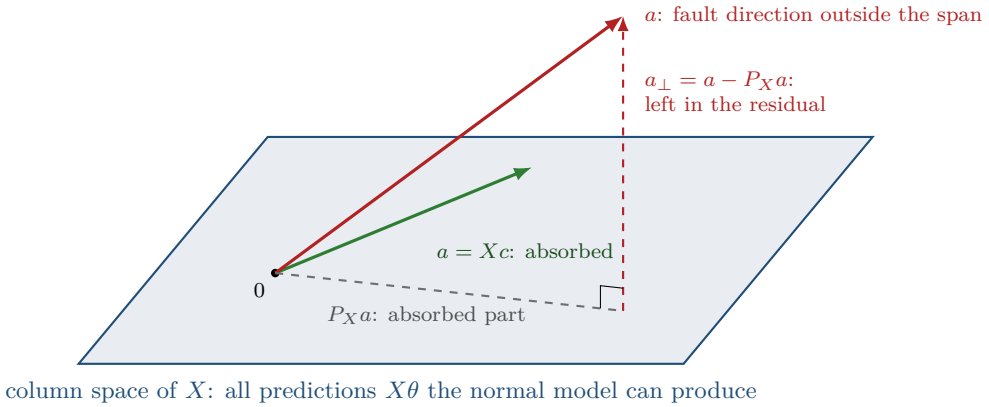


Figure 8.7: Absorbed and detectable fault directions. The plane is the set of records the normal model can produce; a fault direction in the plane (green) is absorbed by a change of parameters, while a direction leaving the plane (red) keeps its orthogonal component $a_\perp$ in the residual. Drawn in three dimensions for a space of any dimension.

> **Principle 8.3.** *A fault is detectable from a given model and dataset only if its signature has a component outside the span of the model's free parameters. The size of that component, not the size of the fault, sets how detectable it is.*

8.6 Instrumentation and what makes a fault visible

If a fault is invisible because the model is free to absorb it, the remedy is information that removes the freedom. This section shows how a single commissioning measurement, an extra sensor, or a different kind of record makes such faults visible, and why more samples of the same record do not.

8.6.1 Commissioning measurements

If R_{series} is unknown and fitted from live data, a resistance-like fault is absorbed by a changed fitted R_{series}, as in Section 8.5.2. If instead R_{series} is measured independently once, at commissioning, and treated as known thereafter, the model loses that freedom. The term $I_k R_{\text{series}}$ moves to the known side of the equation, the design matrix keeps only its column of ones, and

the resistance signature $a = -I$ is no longer in its span; its orthogonal part $-(I - \bar{I})$ is nonzero whenever the current varies. A later resistance change then stands out.

Figure 8.8 shows the effect with the same illustrative rack as before: commissioning values $V_{\text{off}} = 0.20\,\text{V}$ and $R_{\text{series}} = 10\,\text{m}\Omega$, and a later fault that raises R_{series} to $15\,\text{m}\Omega$. Both panels plot the measured discrepancy ΔV against current for the faulty rack (black points). In the left panel R_{series} is a free parameter: the fitted line passes through the points with slope $-15\,\text{m}\Omega$, the residuals are near zero, and the fault has been absorbed into the parameter. In the right panel R_{series} is fixed at its commissioning value: the model line keeps the slope $-10\,\text{m}\Omega$, the points fall progressively below it, and the residual (red segments) grows to 1.5 V at 300 A. The same data, read against a model with one parameter fewer, reveal the fault.

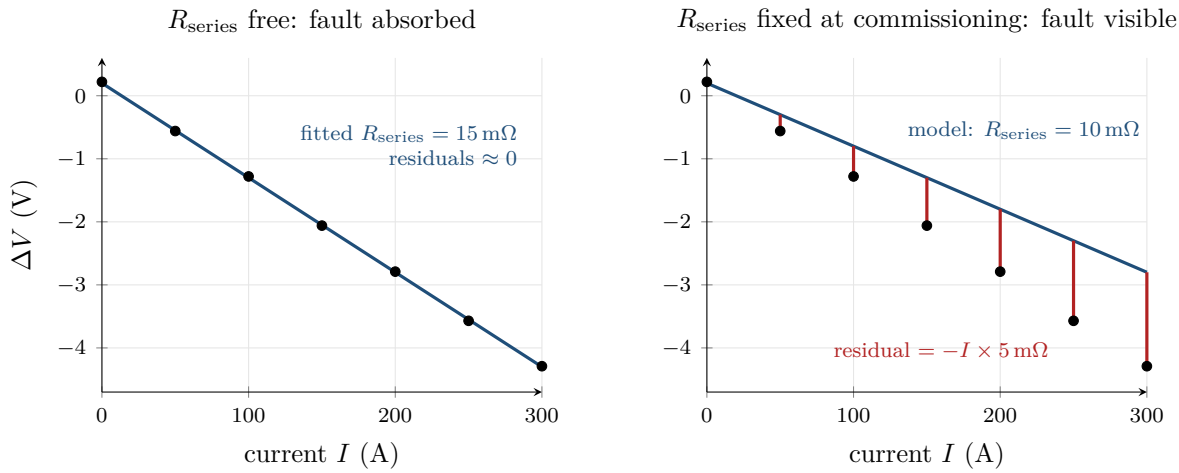


Figure 8.8: The effect of a commissioning measurement. The points are the discrepancy $\Delta V = V_{\text{off}} - IR_{\text{series}}$ of a rack whose R_{series} has risen from 10 to $15\,\text{m}\Omega$, with $V_{\text{off}} = 0.20\,\text{V}$ and small noise. Left: with R_{series} fitted, the model follows the fault. Right: with R_{series} fixed at its commissioning value, the fault appears as a residual proportional to current. Values are illustrative.

The same argument applies to offsets. If the voltage offsets V_{off} between the rack sensor and the cell sensors are measured once at commissioning, the column of ones also leaves the design matrix, and the absorbed offset of Section 8.5.2 becomes a visible constant residual. Measuring the string resistance and the voltage offsets once at commissioning therefore makes otherwise invisible faults detectable. One caution applies: a parameter that legitimately changes, as resistance does with aging (Chapter 7), cannot be fixed forever without eventually producing false alarms. The commissioning value is better treated as a known starting point with a stated rate of normal change, so that a change faster than aging stands out while aging itself does not.

8.6.2 Instrumentation limits

Some faults are invisible not because of the model but because of the sensors. The nine racks of the reference container connect in parallel to a common bus, so the bus current is the sum of the rack currents,

$$I_{\text{bus}} = \sum_{r=1}^{9} I_r. \tag{8.19}$$

If the only current sensors are the nine rack sensors, as in the reference record, then I_{bus} is computed from them, and the residual $I_{\text{bus}} - \sum_r I_r$ is zero by construction for any data at all. The relation is an identity, not a test. Suppose the rack 4 current sensor reads 5 A high. The computed bus current is also 5 A high, and nothing in the record disagrees. The consequence is

not small: rack 4's coulomb-counted state of charge drifts by 5 Ah per hour of operation, so over a day of continuous coulomb counting the error would reach 120 Ah, two-fifths of a 300 Ah rack, unless a full charge resets it.

An independent bus-current meter turns the identity into a test,

$$r = I_{\text{bus,measured}} - \sum_{r=1}^{9} I_r = -5\,\text{A},$$

and the bias is detected. It is still not isolated: a bias in any of the nine rack sensors enters the sum with the same coefficient, so the nine candidate signatures are identical, and the residual says that some rack sensor is wrong but not which one. Isolating it requires a relation that involves rack 4 alone, such as its state of charge at the next full charge. None of this is a failure of an algorithm. It is an instrumentation limitation: the information is not in the record.

8.6.3 More of the same data does not help

When a fault is undetectable because its signature lies in the span of the model, adding samples of the same measurement does not change that. Ten thousand samples of an insufficient measurement are still insufficient. In the language of Section 8.5.3, the augmented design matrix $[X\ a]$ has linearly dependent columns for every n when $a = Xc$; appending rows of the same kind preserves the dependence, so the fault magnitude ϕ and the parameters it is confused with remain undetermined, with infinite variance along the direction $(c, -1)$, at $n = 10$ and at $n = 10{,}000$ alike.

Two kinds of missing information must be distinguished. When the limitation is noise, more samples help: the standard deviation of a mean falls as $\sigma/\sqrt{n}$, so a 5 mV noise averaged over 10,000 samples leaves 0.05 mV. When the limitation is structural, more samples do not help, because the missing information is not a matter of precision. What is needed is a different measurement. Table 8.6 lists the different measurements that this chapter has shown to matter and the faults each one makes visible.

Table 8.6: Measurements that remove structural limits, and the faults each makes detectable or isolable.

Measurement	Freedom or identity it removes	Fault made visible
independent bus current	$I_{\text{bus}} = \sum_r I_r$ is no longer an identity	rack current-sensor bias (detected, not isolated)
known sensor offsets	V_{off} is no longer fitted	constant sensor bias present from the start
known string resistance	R_{series} is no longer fitted	connection-resistance increase
mapped temperature locations	spatial pattern becomes interpretable	hot spot versus temperature-sensor bias
rest-voltage measurements	IR terms vanish at $I = 0$	capacity and state-of-charge effects separated from resistance
ambient temperature	T_{env} becomes an input, not a cause	cooling fault versus warm environment

Principle 8.4. *When a fault is hidden by the structure of the model or of the sensor set, the remedy is a different measurement, not more of the same one. Sample size reduces noise; it does not create identifiability.*

8.6.4 The conceptual ladder

The chapter's terms form a ladder, each rung built on the one below it:

(1) *measurement*: what a sensor reports;

(2) *residual*: observation minus model;

(3) *anomaly*: behavior that departs from what was expected;

(4) *fault*: the underlying cause;

(5) *detection*: deciding that something is wrong;

(6) *diagnosis*: deciding what is wrong;

(7) *isolation*: deciding where it is.

A measurement becomes a residual only through a model, a residual becomes an anomaly only through an uncertainty, and an anomaly points to a fault only through the signatures of the candidate faults. Each step can fail on its own. The last lesson of the chapter is the strongest: a fault can exist and still be fundamentally undetectable from the available telemetry, because its effect is indistinguishable from the ordinary freedom of the normal model. Whether that is so is a property of the model and the sensor set, and it can be decided before any fault occurs.

8.7 What to retain

Definition 8.1 (Fault detectability)**.** An anomaly is a symptom and a fault is its cause; detection, diagnosis and isolation decide, in order, whether, what and where. A fault is detectable from a given model and record only if its signature has a component outside what the model's free parameters can absorb, and isolable only if that component differs from the signature of every competing fault.

Table 8.7 collects the chapter's objects with the subsection that defines each.

8.8 Exercises

Exercise 8.1. The 3,564 cell voltage channels of the reference container are scanned once per second, and each residual is Gaussian with $\sigma = 8\,\mathrm{mV}$ under H_0 and independent across channels and scans. (a) Compute the expected number of false alarms per hour for the two-sided thresholds $\tau = 3\sigma$ and $\tau = 5\sigma$. (b) A fault shifts one cell's residual by $30\,\mathrm{mV}$. Compute the probability that a single scan misses it at each threshold. (c) Using the persistence rule of Section 8.2.4 with three consecutive same-sign crossings at $\tau = 3\sigma$, compute the false-alarm rate per hour and the probability that three consecutive scans of the faulty cell all cross.

Exercise 8.2. The discrepancy of the rack string voltage identity is measured at four currents: $(I, \Delta V) = (0\,\mathrm{A}, 0.20\,\mathrm{V})$, $(100\,\mathrm{A}, -1.10\,\mathrm{V})$, $(200\,\mathrm{A}, -2.40\,\mathrm{V})$ and $(300\,\mathrm{A}, -3.70\,\mathrm{V})$. The commissioning values are $V_{\mathrm{off}} = 0.20\,\mathrm{V}$ and $R_{\mathrm{series}} = 10\,\mathrm{m\Omega}$. (a) Fit equation (8.9) and state $\hat{V}_{\mathrm{off}}$ and $\hat{R}_{\mathrm{series}}$. (b) Compute the residuals against the commissioning model and say which signature of Table 8.3 they match. (c) Suppose the record had contained only the sample at $200\,\mathrm{A}$. Show that an offset fault of $-0.60\,\mathrm{V}$ explains it as well as the resistance fault found in (a), and state which property, detectability or isolability, is lost.

Exercise 8.3. One temperature sensor reads $10\,\mathrm{K}$ hot. The prior probabilities are 0.3 for a real hot spot (H_1) and 0.7 for a sensor bias (H_2). Two observations are made: the neighboring sensors read normally, with likelihood 0.2 under H_1 and 0.9 under H_2; and the excess stays at

Table 8.7: The objects of fault detection and diagnosis.

Object	Definition	Section
anomaly, fault	observed departure; its physical or measurement cause	8.1.1
H_0, H_1	normal and abnormal hypotheses of detection	8.1.2
detect, diagnose, isolate	whether, what and where	8.1.3
$r = V_{\text{measured}} - V_{\text{predicted}}$	residual against the normal model	8.2.1
σ	combined sensor, model, parameter and timing uncertainty	8.2.2
$\lvert r\rvert > \tau$, $\tau = 3\sigma$	threshold detector	8.2.3
α, β	false-positive and false-negative rates	8.2.4
b	sensor bias	8.3.1
$\Delta V = V_{\text{off}} - IR_{\text{series}}$	rack string voltage identity, a redundancy relation	8.3.1
intercept and slope	offset versus resistance, separated by varying current	8.3.3
$\mathrm{dSOC}_i/\mathrm{d}t = -I/C_i$, $I\,\delta R_i$	capacity (cumulative) versus resistance (immediate)	8.3.4
h	cooling conductance; a cooling fault lowers it	8.3.5
signature	dependence of a fault's effect on I, I^2, $\int I\,\mathrm{d}t$, time	8.3.7
$\Pr(H_i \mid D)$	posterior probability of a competing explanation	8.4.1
detectability, isolability	differs from normal; differs from other faults	8.4.5
declared shapes	offset, column-proportional, step, drift	8.5.1
$y = X\theta + a\phi$, $a = Xc$	a fault absorbed into the parameters	8.5.3
$a_\perp = a - P_X a$	the detectable part of a signature	8.5.3
commissioning measurement	removes a free parameter from the model	8.6.1
$I_{\text{bus}} = \sum_r I_r$	an identity unless measured independently	8.6.2

10 K through a three-hour rest, with likelihood 0.05 under H_1 and 0.95 under H_2. (a) Compute the posterior of each hypothesis using each observation alone and both together, assuming conditional independence. (b) Using the healthy time constant of Section 8.3.5, 1.4 h, compute the fraction of a real excess that would remain after three hours of rest, and explain why the rest observation is the more informative of the two.

Exercise 8.4. Take the model of Section 8.5.3 with rows $(1, -I_k)$ and a drift fault $a = (0, 1, 2, 3)$ over four equally spaced samples. (a) With currents $I = (0, 100, 200, 300)$ A, show that a lies in the span of X, find c, and state what the fit reports instead of the drift. (b) With currents $I = (0, 300, 100, 200)$ A, compute $P_X a$, $a_\perp$ and the fraction $\|a_\perp\|^2/\|a\|^2$. (c) Explain, using Section 8.6.3, why repeating the first current sequence a thousand times does not make the drift detectable, and name one measurement from Table 8.6 that would.

8.9 Connection to the TERRA framework

Before estimating anything, the TERRA framework asks which departures from normal would be detectable. This chapter supplies the objects that question is asked about.

- **The string voltage identity is a redundancy relation on the series-string edge.** The rack is an edge whose flow, the current, is common to 396 cells, and equation (8.8) states that the voltage across the edge measured at its terminals equals the sum measured cell by cell, up to $V_{\text{off}} - IR_{\text{series}}$. Two independent sensor sets cross the same boundary, so the relation is a test and its residual is the input to every rack-level diagnosis of Section 8.4.3.

- **A fault is a hypothesis with a declared signature.** Each shape of equation (8.15) is an extra column a with an unknown magnitude ϕ attached to a component relation. The framework's detectability check is equation (8.18): compute $a_\perp$ against the parameters

left free, and report a fault whose $a_\perp$ is zero as undetectable before any data are fitted, as with the offset fault against a free V_{off}.

- **A commissioning measurement is a tight prior.** Fixing R_{series} in Figure 8.8 corresponds to a prior on a component parameter whose width is the measurement uncertainty plus the expected aging. The narrower the prior, the less of a resistance fault the posterior can absorb into R_{series}, which is the probabilistic version of removing a column from X.

- **Charge conservation at the bus node is an identity without an independent meter.** Equation (8.19) is the conservation law of the node where the nine rack edges meet. With no independent measurement crossing that node's boundary it constrains nothing; with one, it becomes a test that detects but cannot isolate a rack current-sensor bias.

- **The thermal balance is an energy-conservation node with competing causes.** Equation (8.13) places heat generation, cooling and environment on one node, so more heat, less cooling and a warm environment are three hypotheses on the same node, isolable only when T_{env} is measured (Table 8.5).

- **Diagnosis is a posterior over hypotheses, and the next measurement is chosen to sharpen it.** With hidden component states, physical relations, measurements, fault hypotheses and uncertainty in one model, the framework computes quantities such as $\Pr(\text{fault} \mid \text{measurements})$ or $\Pr(\text{cell degraded} \mid V, I, T)$ rather than a threshold crossing, and asks which next measurement (a rest period, a second current level, a recovery-time observation) would most reduce the uncertainty between competing diagnoses.

In summary, fault detection and diagnosis in a battery energy storage system is a question about information rather than about alarms. A residual means something only against a model and its uncertainty; a diagnosis means something only when the candidate faults have different signatures; and a fault whose signature the model can absorb is invisible however long the record. The decisive step is therefore taken before any data are examined: listing the faults of interest with their shapes, and checking each against the model's free parameters and the sensors actually installed.

The next chapter applies that check systematically to the reference record and asks what its measurements can and cannot determine.

Chapter 9

Sensors, observability, identifiability and information content

9.1 What can be known

The eight chapters before this one built the models of a battery: the cell and its open-circuit voltage (Chapter 3), the equivalent circuit (Chapter 4), the hardware and its sensors (Chapter 5), the heat balance (Chapter 6), aging (Chapter 7) and faults (Chapter 8). Each model has states that are not measured and parameters that are not known. This chapter asks the question that comes before any of those models is fitted to field data:

given the sensors installed, what can actually be known about the states and parameters of the system?

The analysis of the reference record is built around this question, and it takes three steps in a fixed order. First, it asks whether the available readings can represent a physical balance at all. Second, it asks whether the record holds enough independent information to determine the unknowns of that balance. Only then does it fit. The reason for the order is that a large amount of data does not necessarily carry a large amount of information. The reference record holds, per rack, one current, one rack voltage, one reported state of charge, 396 cell voltages and 216 temperatures: 615 channels per rack and $9 \times 615 = 5{,}535$ for the container. At one sample per minute the cell voltages alone produce $3{,}564 \times 1{,}440 \approx 5.1$ million readings a day. Whether those millions determine a single resistance is a separate question, and it is the subject of this chapter.

The chapter stops at that question. How a parameter is estimated once it is known to be determinable (least squares, the weighting of measurements by their noise, the standard errors of the estimates) is outside this book.

Jargon: observability, identifiability, excitation, redundancy, structural and practical identifiability, degrees of freedom, boundary

A state is *observable* if its value can be reconstructed from the measurements taken over time; a parameter is *identifiable* if the measurements determine its value uniquely. *Excitation* is variation in the inputs (current, ambient temperature) large enough to make a parameter's effect visible in the outputs. *Redundancy* means that two or more independent measurement paths constrain the same quantity, so that they can be compared. A parameter is *structurally* unidentifiable when the model equations themselves cannot separate it from other parameters, even with perfect noiseless data; it is *practically* unidentifiable when the equations could separate it but the data are too noisy, too coarse or too still to do so. The *degrees of freedom* of a problem are the number of independent equations minus the number of unknowns. A *boundary* is the surface around a part of the system across which a conservation law is written; a balance on a boundary tests something only if a measurement crosses it independently.

Key idea

Data volume is not information. A state or parameter is known from a record only if independent measurements move in the direction it controls by more than their noise and resolution, and a fitted number does not establish that by itself.

The four subsections below state the core of the idea with two equations, then separate the two kinds of unknown a battery model carries: states that evolve in time and parameters that are fixed.

9.1.1 Equations and unknowns

The simplest example is one equation in two unknowns,

$$x + y = 10. \tag{9.1}$$

It has infinitely many solutions: $(1, 9)$, $(2, 8)$, $(5, 5)$ and every other pair on the line $y = 10 - x$. The equation is true; it simply does not contain enough information to fix both unknowns. The problem is *underdetermined*. A second, independent equation,

$$x - y = 2, \tag{9.2}$$

removes the ambiguity. Adding the two equations gives $2x = 12$, so $x = 6$, and then $y = 10 - 6 = 4$.

Figure 9.1 draws this. Each panel has x on the horizontal axis and y on the vertical axis, both from 0 to 10, and each line is the set of points satisfying one equation. Panel (a) holds equation (9.1) alone: a whole line of solutions, three of which are marked. Panel (b) adds equation (9.2): the two lines cross at one point, $(6, 4)$, and the solution is unique. Panel (c) is used in Section 9.5.4. The word *independent* matters: a second equation $2x + 2y = 20$ is the first one multiplied by two, draws the same line, and adds nothing. In matrix form the two equations of panel (b) have the coefficient matrix $\left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$, whose determinant is $-2 \neq 0$; the dependent pair has $\left(\begin{smallmatrix} \frac{1}{2} & \frac{1}{2} \end{smallmatrix}\right)$, whose determinant is zero. Everything in this chapter is a version of the difference between panels (a) and (b): whether the measurements supply enough independent equations to pin the unknowns to a point.

9.1.2 Observable states

A *state* is a quantity that evolves in time and summarizes what the system remembers of its past. For a rack the two most important states are the state of charge and the temperature,

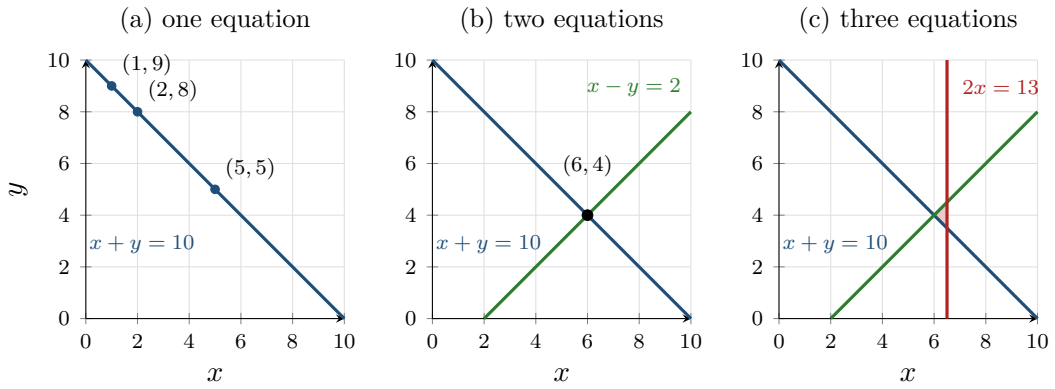


Figure 9.1: Equations as lines in the plane of the unknowns. (a) One equation, $x + y = 10$: a line of solutions, of which $(1, 9)$, $(2, 8)$ and $(5, 5)$ are three. (b) A second independent equation, $x - y = 2$: the lines cross at the unique solution $(6, 4)$. (c) A third equation, $2x = 13$ (red), is inconsistent with the first two (colored as in panel (b)): the three lines enclose a small triangle (shaded) and share no common point.

collected in the state vector

$$\mathbf{x}(t) = \begin{bmatrix} \mathrm{SOC}(t) \\ T(t) \end{bmatrix}. \tag{9.3}$$

Neither is a sensor reading in the sense that matters here: the state of charge is a property of the electrode chemistry, and the temperature of the cells is not the temperature at the tip of a sensor. What is measured is the terminal voltage V, the current I and a sensor temperature T_{sensor}. The states are *observable* if those measurements, taken over a window of time, carry enough independent information to reconstruct $\mathrm{SOC}(t)$ and $T(t)$.

The test is whether two different state histories could produce the same measurements. Write the dynamics as $\mathbf{x}(t + \Delta t) = f(\mathbf{x}(t), I(t))$ and the measurements as $\mathbf{y}(t) = g(\mathbf{x}(t), I(t))$, with $\mathbf{y} = [V, T_{\mathrm{sensor}}]^{\mathsf{T}}$. If the map from the initial state to the whole measured sequence $\mathbf{y}(t_0), \mathbf{y}(t_1), \ldots$ is one-to-one, the state is observable: each sequence points back to exactly one starting state, and the dynamics then carry it forward. If several internal states produce the same measurements, the state is not fully observable, and no amount of processing of those measurements will tell them apart. For the state of charge, the voltage depends on it through the open-circuit voltage $U(\mathrm{SOC})$ of Chapter 3; where that curve is steep, two different states of charge give clearly different voltages, and where it is flat they do not. Section 9.3.4 returns to this.

9.1.3 Identifiable parameters

Identifiability is the same question asked about parameters, the quantities a model treats as fixed. Take the simplest voltage model of Chapter 4,

$$V = U(\mathrm{SOC}) - I\,R, \tag{9.4}$$

with the internal resistance R unknown and current positive on discharge. If the current varies, the term IR varies with it, and the resistance can be estimated: two samples at the same state of charge with currents $I_1 \neq I_2$ give

$$R = -\frac{V_2 - V_1}{I_2 - I_1}. \tag{9.5}$$

If instead $I = 0$ throughout the record, equation (9.4) reduces to $V = U(\mathrm{SOC})$, and R has disappeared from the data. Every value of R explains the record equally well.

The resistance still exists physically: the cell has the same resistance whether or not current flows. What fails is the experiment. The sensitivity of the measured voltage to the parameter is $\partial V/\partial R = -I$, and it is zero at every sample of a record with no current. A parameter is identifiable from a dataset only if the dataset excites it, and *not identifiable from this dataset* is a statement about the data, not about the battery.

9.1.4 Observability and identifiability compared

The two words answer the same kind of question about two kinds of unknown. Observability concerns hidden *states*, which change in time: $\mathrm{SOC}(t)$, the cell temperature $T(t)$, the polarization voltage $V_1(t)$ of Chapter 4. Identifiability concerns unknown *parameters*, which a model treats as constant: the capacity C, the resistance R, the cooling conductance h. A state is reconstructed at every instant; a parameter is determined once from the whole record.

The two overlap when parameters vary with time. Aging (Chapter 7) makes the capacity fall and the resistance rise over months, and a model that tracks this treats C and R as slowly varying states, appended to the state vector of equation (9.3). Their estimation then becomes an observability question with a long time scale, and the conditions of the following sections (excitation, independent channels, resolution) apply to both in the same way.

9.2 Excitation and redundancy

A parameter is seen only when the system moves in the direction it controls, and a measurement is checked only when a second, independent path reaches the same quantity. This section treats the two conditions, excitation and redundancy, and the two ways a record can appear to have more information than it does: repeated points and derived channels.

9.2.1 Excitation

Consider again estimating R from $V = U - IR$, now with U also unknown over a short window in which the state of charge barely changes. If the record contains currents of 20, 50, 100 and 150 A, the term IR takes four clearly different values and the effect of R appears strongly: with an illustrative $R = 1\,\mathrm{m\Omega}$ it ranges from 20 mV to 150 mV. If the current is about 50 A all the time, every sample satisfies $V \approx U - 50\,R$, and only that one combination is determined. A voltage of 3.250 V is explained equally well by $U = 3.300\,\mathrm{V}$ with $R = 1\,\mathrm{m\Omega}$ and by $U = 3.275\,\mathrm{V}$ with $R = 0.5\,\mathrm{m\Omega}$, since $3.275 - 50 \times 0.0005 = 3.250$. This is equation (9.1) again: one independent equation in two unknowns, a line of solutions.

The system must move enough to expose the parameter. To make "enough" precise, write the model as known signals multiplied by unknown parameters; the signal that multiplies a parameter, taken over all samples, is that parameter's *design column*. For R the design column is $-I$. The analysis of the reference record checks, before accepting a fit, that each design column moves sufficiently relative to the measurement scatter: the spread of the column times a plausible value of the parameter must clearly exceed the noise. With the four currents above the standard deviation of I is about 49.5 A, and with $R = 1\,\mathrm{m\Omega}$ the resistance moves the voltage by about 49.5 mV, fifty times an illustrative scatter of 1 mV. With $I = 50 \pm 0.5\,\mathrm{A}$ the same product is 0.5 mV, half the scatter, and the column does not move enough to count.

Principle 9.1. *A parameter is informed only by variation in the signal that multiplies it. A record in which that signal is constant determines a combination of parameters, not the parameter.*

9.2.2 A three-point and a repeated-point record

Write the voltage response as a drop from the rest value, $\Delta V = -IR$, and take three measurements at different currents: (50 A, −0.05 V), (100 A, −0.10 V) and (150 A, −0.15 V). The slope between the first and last is

$$\frac{-0.15 - (-0.05)}{150 - 50} = \frac{-0.10\,\text{V}}{100\,\text{A}} = -0.001\,\Omega, \qquad R = 1\,\text{m}\Omega, \tag{9.6}$$

and the middle point lies on the same line. Even if the intercept is not assumed to be zero, three points at three currents determine both the slope and the intercept (which comes out as zero), and the third point is a surplus equation that checks the straight-line model itself.

Now take three samples all at (100 A, −0.10 V). There are as many samples as before, but no independent excitation: the record contains the same point three times. Figure 9.2 compares the two records. Both panels have the current on the horizontal axis and the voltage drop in millivolts on the vertical axis. In panel (a) the three spread points fix a single line, of slope −1 mV/A. In panel (b) the three samples sit on top of each other, and three of the infinitely many lines through that point are drawn: $R = 0.5$, 1 and 1.5 mΩ, with intercepts of −50, 0 and +50 mV. Each fits the repeated point exactly. If the intercept is known to be zero the repeated point does give $R = 1\,\text{m}\Omega$, but the two extra copies only average the noise down by $\sqrt{3}$; they add no new direction and cannot check linearity. More samples are not more parameter information.

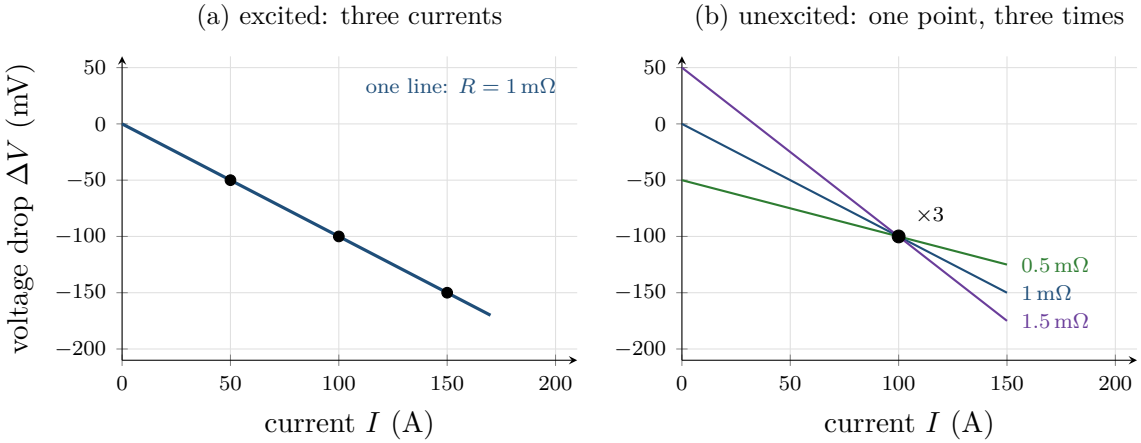


Figure 9.2: Excited and unexcited records of the voltage drop $\Delta V = c - IR$, with the intercept c unknown. (a) Points at 50, 100 and 150 A determine a single line, $R = 1\,\text{m}\Omega$, $c = 0$. (b) Three samples at 100 A, −100 mV are one point; every line with $c = -100\,\text{mV} + (100\,\text{A})\,R$ fits it, of which $R = 0.5$, 1 and 1.5 mΩ are drawn. Values are illustrative.

9.2.3 Redundancy in the string voltage identity

The reference container measures each rack voltage V_r directly and also measures all 396 cell voltages of the rack. Chapter 5 showed that the rack is a series string, so the two are related by

$$V_r \approx \sum_{i=1}^{396} V_i \;-\; I\,R_{\text{series}} \;+\; V_{\text{off}}, \tag{9.7}$$

where R_{series} is the lumped resistance of the connection hardware between the cells and the rack terminals and V_{off} is a constant offset between the two measurement systems. Two measurement paths now constrain one quantity. Figure 9.3 draws them: the string of cells along the top, each

cell with its own voltage sensor; the blue path sums the 396 cell readings and corrects for the connection drop and the offset; the green path is the single rack voltage sensor spanning the whole string. Both arrive at a comparison box, whose output is the residual

$$r_{\text{string}} = V_r - \Big(\sum_{i=1}^{396} V_i - I\,R_{\text{series}} + V_{\text{off}}\Big). \tag{9.8}$$

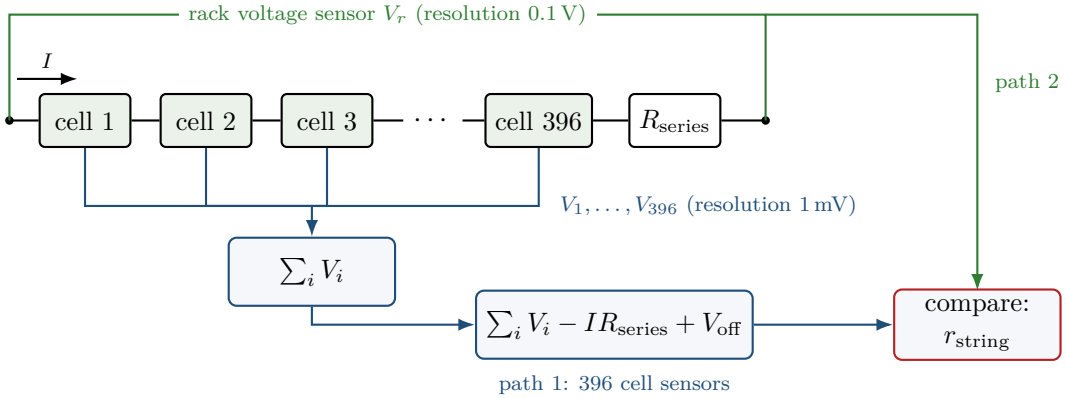


Figure 9.3: The string voltage identity, equation (9.7), as two measurement paths to one quantity. Path 1 (blue) sums the 396 cell voltage readings and corrects for the connection resistance R_{series} and the offset V_{off}; path 2 (green) is the rack voltage sensor across the whole string. Their difference is the residual r_{string} of equation (9.8).

Redundancy is what makes fault detection possible. If every sensor is correct and the model holds, r_{string} stays within the noise. If one cell sensor reads 50 mV high, the blue path rises by 50 mV and the residual moves by -50 mV; if the rack sensor drifts, the residual moves the other way. Without the second path, a bad sensor may be undetectable: a lone rack voltage that reads 0.3 V high looks exactly like a rack whose cells are 0.3 V higher. The comparison is limited by the coarser path. The rack voltage is resolved in steps of 0.1 V, a rounding error of up to ± 50 mV, while the sum of 396 cell readings each rounded to 1 mV has a rounding spread of only about $0.29\,\text{mV} \times \sqrt{396} \approx 6\,\text{mV}$ if the rounding errors are independent. A single-cell sensor error of 50 mV is therefore at the edge of what one sample of the residual can show, and becomes clear only over many samples in which the rack rounding varies. The same residual, regressed on the current, also carries the connection resistance: with an illustrative $R_{\text{series}} = 2.5\,\text{m}\Omega$, the term IR_{series} is 0.25 V at 100 A, two and a half rack resolution steps, but only 0.025 V at 10 A, a quarter of a step.

9.2.4 Independent and derived channels

The reference record contains a rack current and a reported state of charge, and they look like two measurements. They are not. Chapter 2 described coulomb counting: the battery management system (BMS) produces the reported state of charge by integrating the same current and dividing by a capacity it assumes,

$$\text{SOC}_{\text{BMS}}(t) = \text{SOC}_{\text{BMS}}(t_0) - \frac{1}{C_{\text{assumed}}} \int_{t_0}^{t} I\,\mathrm{d}\tau. \tag{9.9}$$

Figure 9.4 shows the chain as boxes from left to right: the current sensor, an integrator, a division by the assumed capacity C_{assumed}, and the reported state of charge. The current channel and the reported state of charge both enter the record, but they share a single physical

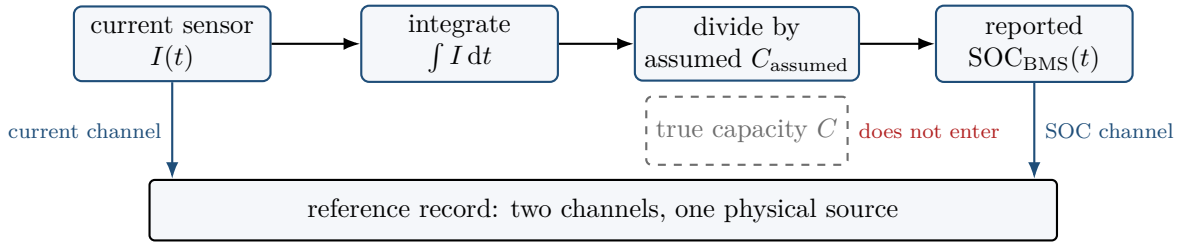


Figure 9.4: The reported state of charge as a derived channel. It is computed from the current channel by integration and division by an assumed capacity, equation (9.9), so the record's two channels carry one measurement. The true capacity of the cells (dashed) plays no part in the chain.

source. The true capacity C of the cells, drawn dashed below the chain, does not enter it anywhere.

Using both channels as independent evidence overstates the information in the record. An error in the current sensor appears in both: a bias of $+0.5\,\text{A}$ over a 10-hour discharge adds $5\,\text{Ah}$ to the integrated charge, and with $C_{\text{assumed}} = 300\,\text{Ah}$ the reported state of charge falls by an extra $5/300 \approx 1.7\,\%$. The two channels then agree with each other perfectly, because one is computed from the other, and the agreement creates false confidence. An analysis that counts them as two measurements counts the current sensor twice and the battery not at all.

9.2.5 The charge balance as an accounting identity

The physical charge balance of a rack is

$$\text{SOC}(t) = \text{SOC}_0 - \frac{1}{C}\int_{t_0}^{t} I\,\mathrm{d}\tau, \tag{9.10}$$

and it looks like an equation from which the capacity C can be estimated: measure the state of charge and the current, and find the C that makes the two sides agree. If the state of charge used is the reported one, equation (9.9), then the left side was computed from the right side with $C = C_{\text{assumed}}$. Substituting gives zero residual exactly when $C = C_{\text{assumed}}$, whatever the true capacity of the cells. Fitting the capacity rediscovers the BMS assumption.

The charge balance written on reported data is therefore an accounting identity, not a physical measurement: it largely tests the controller's arithmetic rather than the battery. Its residual is not zero in practice, but what makes it nonzero is the rounding of the reported value to $1\,\%$ (that is $3\,\text{Ah}$ of a $300\,\text{Ah}$ string), resets and recalibrations performed by the BMS, and timing differences between the two channels. Those are facts about the BMS. A capacity estimate that reflects the cells requires an independent state-of-charge reference, such as the open-circuit voltage at rest at two well-separated states of charge, which is the subject of Section 9.6.1.

9.3 Structural and practical identifiability

A parameter can fail to be determined for two different reasons: the equations may not separate it from other parameters at all, or they may separate it in principle while the data are too weak to do so. This section takes the two in turn, with the rack thermal model as the example of the first and the flat voltage curve of the cell chemistry as the example of the second.

9.3.1 Structural identifiability: the thermal ratios

A parameter set is *structurally unidentifiable* when the model itself does not allow the parameters to be separated, even with perfect noiseless data. The rack heat balance of Chapter 6 is

$$C_{\mathrm{th}}\,\frac{\mathrm{d}T}{\mathrm{d}t} = k\,I^2 + q_0 - h\,(T - T_{\mathrm{env}}), \tag{9.11}$$

with heating coefficient k, steady unaccounted heat q_0, cooling conductance h and thermal capacitance C_{th}. Dividing by C_{th} gives the form that the temperature trajectory actually obeys,

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{k}{C_{\mathrm{th}}}\,I^2 + \frac{q_0}{C_{\mathrm{th}}} - \frac{h}{C_{\mathrm{th}}}\,(T - T_{\mathrm{env}}). \tag{9.12}$$

Only three combinations appear: k/C_{th}, q_0/C_{th} and h/C_{th}. Four parameters, three determinable ratios.

Multiply k, h, q_0 and C_{th} all by 10. Every ratio is unchanged, so equation (9.12) is unchanged, and so is every temperature trajectory it produces, for every current and ambient history. Table 9.1 shows two illustrative parameter sets related this way; the first corresponds to a rack with a time constant $C_{\mathrm{th}}/h = 2 \times 10^4\,\mathrm{s} \approx 5.6$ hours and a steady rise of $(0.1 \times 100^2 + 200)/100 = 12\,\mathrm{K}$ above ambient at 100 A. The second gives the same time constant, the same rise and the same trace. The four parameters cannot all be determined: the solutions form the line $(\lambda k, \lambda q_0, \lambda h, \lambda C_{\mathrm{th}})$, $\lambda > 0$, in the four-dimensional parameter space, the same shape as the line of solutions of equation (9.1).

Table 9.1: Two parameter sets for the rack heat balance, equation (9.11), that produce identical temperature trajectories. Values are illustrative.

Quantity	Set A	Set B (×10)	Ratio to C_{th} (both sets)
k (W/A^2)	0.1	1	5×10^{-8} K/(A^2 s)
q_0 (W)	200	2,000	1×10^{-4} K/s
h (W/K)	100	1,000	$5 \times 10^{-5}\,\mathrm{s}^{-1}$
C_{th} (J/K)	2×10^6	2×10^7	1

9.3.2 Why more data does not resolve a structural ambiguity

One day of the reference record, one month or ten years of the same kind of data: the ambiguity of Table 9.1 remains, because every one of those records is generated by equation (9.12), and that equation depends on the parameters only through their ratios. More data of the same kind sharpens the three ratios; it never separates the four parameters. This distinguishes a structural problem from a noise problem, which more data does reduce.

The solution is different information, not more of the same. If the thermal capacitance is measured independently (for example from the mass of the rack and the specific heat of its materials), the remaining three follow at once from the ratios: $k = (k/C_{\mathrm{th}})\,C_{\mathrm{th}}$, $h = (h/C_{\mathrm{th}})\,C_{\mathrm{th}}$ and $q_0 = (q_0/C_{\mathrm{th}})\,C_{\mathrm{th}}$. Fixing any one of the four in this way fixes the scale λ and therefore all of them. What breaks a structural ambiguity is a measurement that enters the equations in a way the existing ones do not.

9.3.3 Practical identifiability

A parameter can be identifiable in theory and still not be determined by the data, because its effect is tiny compared with the noise. Take $V = U(\mathrm{SOC}) - IR$ again with $R = 0.5\,\mathrm{m\Omega}$, and a

record in which the current varies by only ±1 A. The resistance-related variation of the voltage is

$$\Delta I \cdot R = 1\,\mathrm{A} \times 0.0005\,\Omega = 0.0005\,\mathrm{V} = 0.5\,\mathrm{mV}. \tag{9.13}$$

With a sensor resolution of 100 mV the effect is two hundred times smaller than one reading step and is buried: the reported value does not change when the true voltage moves by half a millivolt, except on the rare samples that sit at a rounding boundary. For the effect to reach even one resolution step, the current would have to swing by $0.1\,\mathrm{V}/0.0005\,\Omega = 200\,\mathrm{A}$.

Two failure modes. The two kinds of failure are different in kind and have different remedies. In *structural* unidentifiability the equations cannot distinguish the parameters; only new information of a different kind helps, as in Section 9.3.2. In *practical* unidentifiability the equations can distinguish them in principle, but the data quality or the excitation is too weak; better sensors, more variation in the inputs, or a deliberate experiment help. Both produce the same symptom in a fit, a parameter that can move a long way without changing the fit much, and the diagnosis of which one is present decides what to do.

9.3.4 The flat voltage plateau of LFP cells

The cell chemistry of the reference container, lithium iron phosphate (LFP), has the open-circuit voltage curve described in Chapter 3: steep near empty and near full, and almost flat over a wide middle range. On that plateau $\mathrm{d}U/\mathrm{dSOC} \approx 0$: substantial changes in the state of charge produce tiny changes in voltage. The state of charge is then observable in principle and practically weak, the state analog of Section 9.3.3.

The size of the problem follows from the slope. An error δV in the voltage maps into an error in the state of charge inferred from it of

$$\delta\mathrm{SOC} \approx \frac{\delta V}{\mathrm{d}U/\mathrm{dSOC}}. \tag{9.14}$$

Illustratively, if the open-circuit voltage changes by about 30 mV between 30 % and 70 %, the slope is about 0.75 mV per percent. One step of the 1 mV cell-voltage resolution then corresponds to about 1.3 % of state of charge, and a 10 mV error from an imperfectly modeled polarization voltage corresponds to about 13 %. Much of the reference record lies on this flat portion, so the voltage there says little about the state of charge, and an independent check of the reported state of charge has to come from the steep ends of the curve.

9.4 What the sensors resolve

A channel being present in the record does not mean it carries the effect of interest. This section takes the four properties of a sensor channel that decide whether it does: its resolution, its sampling rate, the alignment of its clock with the other channels, and the location of the sensor.

9.4.1 Resolution

The reference record resolves cell voltage to 1 mV, rack voltage to 0.1 V, the reported state of charge to 1 % and temperature to 1 K. An effect smaller than the resolution of the channel that must show it may carry no information. A thermal fault that raises a temperature by 0.2 K, read by sensors that report whole kelvins, changes most readings not at all and a few by one step, and is hard to resolve.

Figure 9.5 sets five such comparisons side by side. Each bar is one effect sought, labeled with the channel that must show it; its length is the ratio of the size of the effect to the resolution of that channel, on a logarithmic axis. The dashed line marks a ratio of one: effects to its left are

smaller than one reading step, effects to its right span several. The bars start at the line, so their direction says which side of it they fall on. Four of the five fall to the left: the 0.5 mV resistance effect of equation (9.13) read on a 100 mV rack channel, the 0.2 K thermal fault, an illustrative 1 Ah charge discrepancy (0.33 % of 300 Ah) read on a 1 % state-of-charge channel, and a 1 % change of state of charge on the plateau (0.75 mV) read on a 1 mV cell channel. Only the deliberate one, a 100 A current pulse through 0.5 mΩ read on a cell channel, is fifty steps wide.

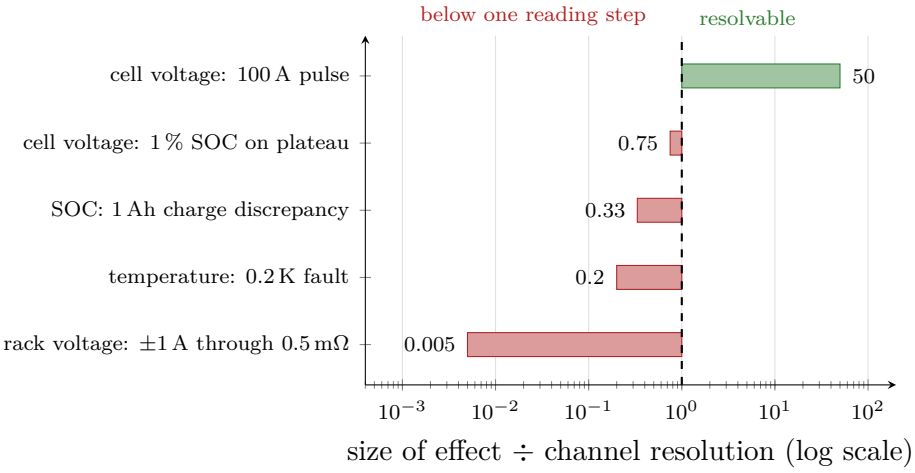


Figure 9.5: Size of five effects relative to the resolution of the channel that must show them, with the resolutions of the reference record: cell voltage 1 mV, rack voltage 0.1 V, state of charge 1 %, temperature 1 K. Bars start at a ratio of one (dashed); red bars are effects smaller than one reading step. The effect sizes are illustrative.

9.4.2 Sampling rate

A process that changes in 2 s and is sampled every 60 s is missed. The rule of thumb is that a change lasting τ needs a sampling interval of at most about $\tau/2$ to be seen reliably; with a 60 s interval, a 2 s event is caught by a sample only if a sample happens to fall inside it, which for independently timed events has probability $2/60 \approx 3.3\,\%$.

The channels of the reference record are sampled at different rates. The rack current is sampled about every second, the cell voltages about once a minute, and the temperatures much more slowly. A current pulse of a few seconds is therefore visible in the current channel and almost never in the cell voltages, and the short-time-constant part of the polarization response of Chapter 4, which settles in seconds to tens of seconds, is largely invisible at one-minute sampling. A channel present in the dataset is not temporally equivalent to every other channel present in it: the effective resolution in time of any relation between two channels is set by the slower one.

9.4.3 Timing alignment

Different channels are recorded by different devices, each with its own clock. Suppose the current changes sharply at $t = 10.0$ s, from 20 A to 120 A, and the one-second current record captures the change. A cell-voltage sample falls at $t = 10.6$ s on the voltage logger's clock. Figure 9.6 draws the situation on a common time axis from 8 to 13 s. The top panel is the current, with its one-second samples as dots. The bottom panel is the true voltage of one cell, which drops by $100\,\text{A} \times 0.5\,\text{m}\Omega = 50\,\text{mV}$ at the step (illustrative values), with the single voltage

sample marked. The shaded band is where that sample may truly have been taken if the two clocks disagree by up to one second.

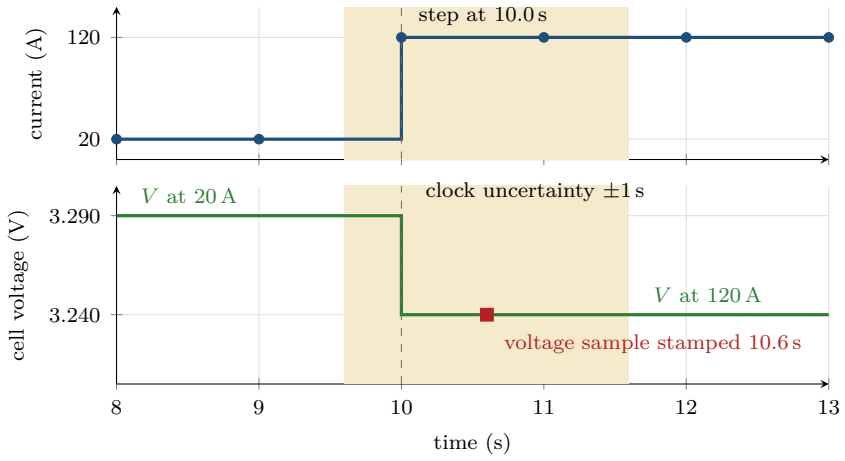


Figure 9.6: Timing misalignment between a one-second current record (top) and a cell-voltage sample (bottom). The current steps from 20 to 120 A at 10.0 s; the true cell voltage drops by 50 mV at the same instant ($R = 0.5\,\mathrm{m\Omega}$, illustrative). The voltage sample is stamped 10.6 s, but with clocks that may disagree by one second it may truly lie anywhere in the shaded band, which straddles the step.

Treating the two samples as simultaneous manufactures a voltage–current relation the battery does not have. If the voltage sample was in truth taken just before the step, it shows the voltage at 20 A but is paired with a current of 120 A: the analysis sees a 100 A rise in current with no voltage drop and infers $R \approx 0$. If it was taken after, it gives the true 0.5 mΩ. The same record can therefore yield anything from zero to the correct value, depending on an offset between clocks that is not itself recorded.

The analysis of the reference record guards against this in two ways. It interpolates all channels onto a common time grid, and it excludes windows in which the current moves too much for asynchronous clocks to be compared safely. The size of "too much" follows from the numbers: if the current changes by at most $\Delta I_{\max}$ within the clock uncertainty, the pairing error in voltage is at most $\Delta I_{\max} R$, and keeping that below the 1 mV cell resolution with $R = 0.5\,\mathrm{m\Omega}$ requires $\Delta I_{\max} \leq 2\,\mathrm{A}$. This is not just data cleaning. It protects identifiability: without it, the excitation that a resistance estimate needs (Section 9.2.1) is exactly the condition that corrupts the pairing.

9.4.4 Spatial mapping

Each rack of the reference container carries 216 temperature sensors, 1,944 in the container, and their physical locations are not documented. The record therefore contains $T_1, \ldots, T_{216}$ for each rack but not $T(x, y, z)$, the temperature as a function of position. A statement such as "module 4 is hotter than module 5" cannot be made, because no reading is known to belong to module 4. For the same reason the analysis of the reference record builds no thermal boundary finer than the rack: a heat balance around a module needs the temperatures on that module, and the record cannot say which they are.

What remains possible is everything that does not need the map: the rack mean, the spread of the 216 readings, the maximum, and the behavior of one sensor relative to the distribution of the others over time. A sensor that reads persistently higher than its peers can be flagged; whether it sits on a hot cell, near a fan outlet or on a faulty mounting cannot be decided from the record. Location is information, and a channel without it supports fewer questions than the

same channel with it.

9.5 Boundaries, degrees of freedom and parameter trade-offs

Conservation laws are the strongest relations in a battery model, but a conservation law is a test only where independent measurements meet it. This section places the boundary, counts equations against unknowns, and then turns to the subtler failure in which a model fits well with the wrong parameters.

9.5.1 The container boundary

Draw a boundary around the whole container. Charge conservation says that the current leaving it on the DC bus equals the sum of the nine rack currents,

$$I_{\text{bus}} = \sum_{r=1}^{9} I_r. \tag{9.15}$$

Figure 9.7 draws the configuration: the dashed rectangle is the boundary, the nine racks sit inside it each with its own current sensor, the rack leads join a common bus, and the bus leaves the container to the power conversion system (PCS). Equation (9.15) is an independent check only if a bus current or an inverter current is measured where the bus crosses the boundary. The reference record lacks that measurement. Defining $I_{\text{bus}} = \sum_r I_r$ in the analysis then creates a quantity that satisfies the balance by construction: an identity, not a test.

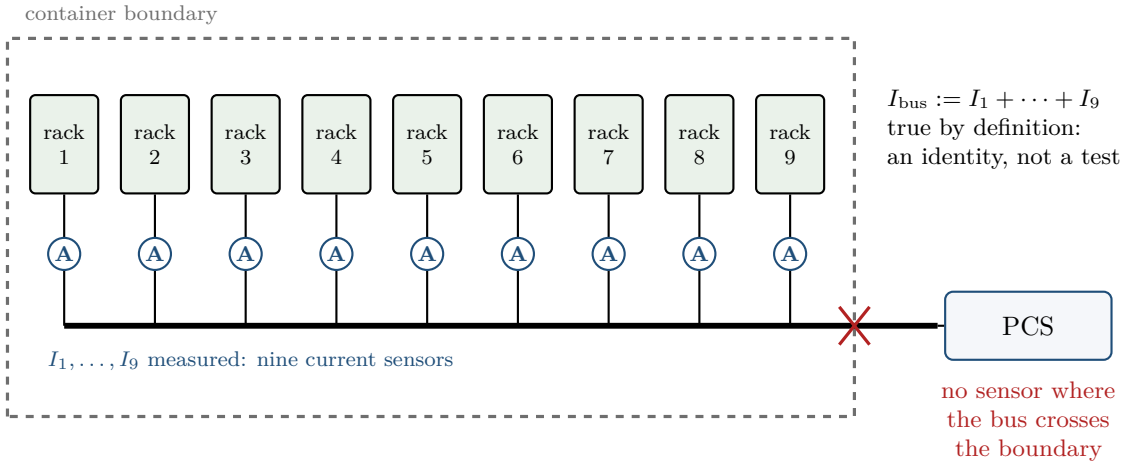


Figure 9.7: The container boundary of the reference container. Nine rack current sensors (circles marked A) lie inside the boundary; no measurement crosses it where the bus leaves for the power conversion system. The container balance, equation (9.15), can then only define the bus current, not check the rack sensors.

9.5.2 A conservation law that is not a test

The relation $I_{\text{bus}} = I_1 + I_2 + \cdots + I_9$ is true. Without an independent I_{bus}, though, it cannot reveal a wrong rack sensor. Suppose the sensor of rack 4 reads 3 A high while every rack carries 100 A. The sum is then 903 A, the defined bus current is 903 A, and the residual of equation (9.15) is zero, as it is for any error in any of the nine sensors. A conservation law alone does not guarantee diagnostic information.

With an independent measurement I_{PCS} of the current entering the conversion system, the residual $r_{\mathrm{bus}} = I_{\mathrm{PCS}} - \sum_r I_r$ becomes $900 - 903 = -3\,\mathrm{A}$, and the error is detected if it exceeds the noise of the comparison. Even then, one residual cannot say which of the nine sensors is wrong: it is one equation in nine possible sensor errors. For a conservation law to carry diagnostic information, independent measurements must cross the boundary; for it to locate a fault, more independent relations are needed, which is the subject of the next subsection.

9.5.3 Counting equations and unknowns

The degrees of freedom of a problem are the number of independent equations minus the number of unknowns. With fewer equations than unknowns the problem is *underdetermined* (panel (a) of Figure 9.1); with as many, it is *just determined* (panel (b)); with more, it is *overdetermined*. An overdetermined problem is what makes fault detection possible, because arbitrary errors in the data cannot then be fitted perfectly: the surplus equations leave somewhere for an error to show.

The analysis of the reference record counts the surplus of relations beyond the fitted constants before it accepts a balance. Applied to the objects of this chapter: the string identity of equation (9.7) gives one equation per time stamp and has two constant unknowns, R_{series} and V_{off}, so N time stamps give a surplus of $N - 2$, but only if the current varies, since at constant current $V_{\mathrm{off}} - IR_{\mathrm{series}}$ is one combination and the surplus counts one unknown fewer than it should. The container balance of equation (9.15) without a bus measurement has one equation and one new unknown, I_{bus}, at every time stamp: a surplus of zero, and nothing to test. The thermal balance of equation (9.11) has four parameters of which three combinations are determinable; counting four unknowns there would overstate what the record can deliver.

9.5.4 Overdetermination and the inconsistent third equation

Return to the two equations $x + y = 10$ and $x - y = 2$, with the solution $x = 6$, $y = 4$. Suppose one of the right-hand sides is wrong, say the first should have read 11. The two equations then give $x = 6.5$, $y = 4.5$ and fit the wrong data perfectly: a just-determined problem absorbs any error into its solution and gives no sign that anything is amiss.

Add a third equation, $2x = 13$. The first two imply $2x = 12$, so the three equations are inconsistent by one unit, and the inconsistency is revealed. Panel (c) of Figure 9.1 shows it: the third line, $x = 6.5$, misses the intersection of the other two, and the three lines enclose a small triangle instead of meeting at a point. The extra equation is what creates fault-detection capability. The size of the inconsistency is the residual of Chapter 8, and which equation is responsible is a further question that one surplus equation cannot answer alone.

9.5.5 Parameter trade-offs

Consider a model $y = a\,x_1 + b\,x_2$ in which the two inputs x_1 and x_2 always move together. Raising a and lowering b by a compensating amount gives almost the same prediction. In the extreme $x_2 = x_1$ the model is $y = (a + b)\,x_1$ and only the sum $a + b$ is determined. With x_2 close to but not equal to x_1, the two are poorly separated: illustratively, at $x_1 = 4$ and $x_2 = 4.2$ the pair $a = 2$, $b = 1$ predicts $2 \times 4 + 1 \times 4.2 = 12.2$, and the very different pair $a = 2.5$, $b = 0.5$ predicts $2.5 \times 4 + 0.5 \times 4.2 = 12.1$. A difference of 0.1 is easily within the noise.

This is the design-column condition of Section 9.2.1 applied to two parameters at once: each column must move, and the two columns must also move differently from each other. The analysis of the reference record checks this after fitting, by examining how strongly the estimates of pairs of parameters are tied to each other. A strong tie means that the data determine a combination and not the two parameters separately.

9.5.6 Heating versus cooling

A rack temperature higher than expected fits two different explanations: more heating or less cooling. From equation (9.11) at steady state, the temperature rise is

$$T - T_{\text{env}} = \frac{k\,I^2 + q_0}{h}. \tag{9.16}$$

Raising the total heat input $kI^2 + q_0$ by 10 % and lowering h by $1 - 1/1.1 \approx 9.1\,\%$ give exactly the same rise. With the values of Table 9.1 at 100 A, both turn the 12 K rise into 13.2 K. If the current and the ambient temperature do not vary enough, the two explanations trade off against each other along this direction.

The two are separated by variation. The heating term changes with I^2; the cooling term changes with $T - T_{\text{env}}$. A record in which the current takes several levels while the ambient temperature also moves gives the two design columns different shapes, and the trade-off narrows. In a record without that variation, the fit is good while the interpretation is uncertain. A good prediction is not a correct parameter explanation.

9.5.7 Fitting for the wrong reason

A perfectly reproduced temperature trace does not mean that the cooling coefficient has been estimated. Too much heating, too much cooling, the wrong thermal mass and the wrong offset can all compensate for each other. The structural family of Table 9.1 is the exact case; the trade-off of Section 9.5.6 is the approximate one. A third arises from the sensors. If every temperature sensor reads β too high, the measured temperature is $T + \beta$, and substituting into equation (9.11) shows that the model fits the biased record exactly with q_0 replaced by $q_0 + h\beta$. A sensor offset of an illustrative 1 K with $h = 100$ W/K is indistinguishable from 100 W of phantom heat.

Parameter identifiability must therefore be checked separately from fit quality. The quality of a fit measures how well some parameter set reproduces the record; identifiability asks whether only one parameter set does. Each of the three compensations above leaves the first untouched and destroys the second.

> **Principle 9.2.** *A good fit establishes that the model can reproduce the data. It does not establish that the parameters which reproduce it are the true ones; that requires the identifiability checks, made separately.*

9.6 Matching the dataset to the question

The previous sections give the conditions under which a record determines an unknown. This section turns them around: it asks which variation each parameter needs, which experiment supplies it, and how to state the question so that a record can be judged against it.

9.6.1 Information comes from variation

A parameter becomes visible when the data move in the direction it controls. Its sensitivity, the derivative of the measured quantity with respect to the parameter, is the signal that has to vary:

- the resistance needs variation in current, since $\partial V/\partial R = -I$;
- the capacity needs a meaningful accumulated charge, since by equation (9.10) $\partial \text{SOC}/\partial C = C^{-2} \int I \,\mathrm{d}t$;

- the cooling coefficient needs temperature differences, since the cooling term is $h\,(T - T_{\mathrm{env}})$;
- the thermal mass needs transients, since C_{th} multiplies $\mathrm{d}T/\mathrm{d}t$ and is invisible when the temperature is steady.

The capacity shows how the size of the variation sets the precision. With an independent state of charge known at two rests to within δSOC for their difference, and a swing ΔSOC between them, the relative uncertainty of the capacity is

$$\frac{\delta C}{C} \approx \frac{\delta \mathrm{SOC}}{\Delta \mathrm{SOC}}. \tag{9.17}$$

With $\delta\mathrm{SOC} = 1\,\%$, a swing of $80\,\%$ gives the capacity to about $1.25\,\%$, and a swing of $10\,\%$ to only $10\,\%$. The state of charge used here must be independent of the current, as Section 9.2.5 showed; the reported one is not.

9.6.2 Experiments and the parameters they reveal

Because each parameter needs its own variation, different experiments reveal different parameters. Table 9.2 lists the four standard ones. A current pulse supplies sharp variation in current and estimates the resistance. A long charge or discharge cycle supplies a large accumulated charge and estimates the capacity. A thermal cool-down, in which heating stops and the rack relaxes toward ambient, supplies a transient with a known zero heat input and estimates the ratio h/C_{th} from its time constant. A rest period lets the polarization decay so that the terminal voltage approaches the open-circuit voltage, and relates V to $U(\mathrm{SOC})$.

Table 9.2: Experiments, the variation each supplies, and the parameter it reveals.

Experiment	Variation supplied	Parameter revealed	What it does not reveal
Current pulse	sharp change in I over seconds	resistance R (and R_0)	capacity: the charge moved is small
Long charge or discharge cycle	large accumulated $\int I\,\mathrm{d}t$	capacity C	resistance, if the current is constant
Thermal cool-down	transient $\mathrm{d}T/\mathrm{d}t$ with heating off	ratio h/C_{th}	h and C_{th} separately (Section 9.3.2)
Rest period	current zero, polarization decaying	relation of V to $U(\mathrm{SOC})$	state of charge on the flat plateau

There is therefore no single best dataset. A record of steady operation at constant power is poor for resistance and may be good for a heat balance at steady state; a record of frequent short pulses is good for resistance and poor for capacity. A dataset is good or bad relative to a question.

9.6.3 Asking a specific question

The question “is this BESS dataset good?” has no answer. The question “can this dataset answer this specific question?” does, and each specific question may have a different answer. Table 9.3 applies the considerations of this chapter to five questions about the reference record and names the factor that decides each.

The answers differ even though the record is the same. Detecting a bus-current sensor error is impossible without a new measurement, while the connection resistance may be estimable from periods of strong current variation. Stating the question is the step that turns a general unease about data quality into a decision about what can and cannot be concluded.

Table 9.3: Five specific questions and the property of the reference record that decides each.

Question	What it needs	Deciding property of the record
Estimate the rack connection resistance	redundancy (string identity) and variation in I	rack voltage resolution 0.1 V against IR_{series} (Section 9.2.3)
Estimate the true cell capacity	a state of charge independent of the current	reported SOC is derived (Section 9.2.5); rest voltages lie mostly on the plateau
Isolate a thermal hot spot	temperatures tied to positions	sensor locations undocumented (Section 9.4.4)
Detect a bus-current sensor error	a measurement crossing the container boundary	no bus or inverter current in the record (Section 9.5.1)
Identify cooling degradation	variation in $T - T_{\text{env}}$ and an independent C_{th}	only h/C_{th} is structurally determinable (Section 9.3.1)

9.6.4 The capability sequence

Dataset capability is assessed in a fixed sequence of four questions:

(1) can the measurements represent the physical balance (are the terms of the balance measured, and does an independent measurement cross its boundary);

(2) is there enough independent information to estimate the unknowns (excitation, independent channels, surplus equations, structural identifiability);

(3) are the estimates credible (practical identifiability against resolution and noise, trade-offs between parameters, consistency with physics);

(4) which faults would actually be detectable (which departures from normal leave a signature the normal model cannot absorb).

Each question is a precondition for the next: an estimate from an unrepresentable balance is meaningless however credible it looks, and a fault is detectable only if the parameters of the normal model are pinned down well enough that a departure from them shows.

This is stronger than "fit a model and report the parameters". That procedure answers none of the four questions; it produces numbers whose meaning depends on the answers, and leaves the answers unexamined.

9.6.5 Can fit versus can know

An optimizer almost always returns a number. Run on the reference record, it may report $R = 0.82\,\text{m}\Omega$ and $C = 278\,\text{Ah}$. The question is whether the dataset constrains those numbers. A number can be mathematically fitted and physically plausible and still not be independently determined. The capacity fitted to the reported state of charge is the clearest case: it returns the BMS assumption to every digit, with a residual near zero that looks like high precision (Section 9.2.5). A resistance fitted to a record whose current barely moves returns some value on the line of Figure 9.2(b), chosen by the noise.

The existence of a fit is not knowledge of the parameter. The same holds for balances: a balance can produce a number even when the representation is invalid (the container balance defines a bus current that it then cannot test) or the identifiability is absent (the thermal balance returns four parameters of which only three combinations are determined). Every reported parameter therefore needs the companion statement of what in the record determines it.

9.6.6 The eight-step mental model

The chapter's conditions combine into a sequence of eight questions to ask of any record before fitting anything to it. Figure 9.8 draws it as a flow diagram: eight boxes read left to right in two rows, each holding one question and, beneath it, the subsection of this chapter that treats it. The first four establish what the record is and what it is asked to determine; the next three establish whether it can; only the eighth fits.

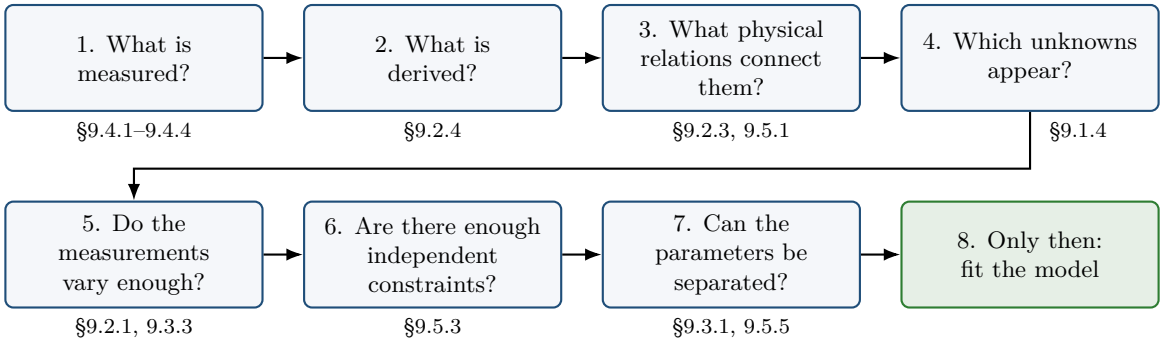


Figure 9.8: The eight-step mental model. The questions are asked in order; a negative answer at any step stops the sequence before the fit and says which information is missing. The section references point to where this chapter treats each step.

Each step can stop the sequence. If a channel is derived (step 2), it does not count as a second measurement; if the relations have no boundary crossed by an independent measurement (step 3), they define rather than test; if the unknowns outnumber the independent constraints (step 6) or appear only in combinations (step 7), the fit will return numbers that the record does not determine. A negative answer is a result: it says which experiment, sensor or independent reference would make the question answerable.

9.7 What to retain

Definition 9.1 (Information content of a record)**.** A record determines a state or parameter only when independent measurements vary in the direction it controls by more than their resolution and noise, and supply more independent constraints than there are unknowns. Lots of data does not mean lots of information, and a fitted number is not, by itself, a known one.

Table 9.4 collects the chapter's objects with the subsection that defines each.

9.8 Exercises

Exercise 9.1. A cell is measured at three currents, giving voltage drops of $(40\,\mathrm{A}, -20\,\mathrm{mV})$, $(80\,\mathrm{A}, -40\,\mathrm{mV})$ and $(120\,\mathrm{A}, -60\,\mathrm{mV})$. (a) Find R and the intercept of $\Delta V = c - IR$, and state the number of surplus equations. (b) The record is replaced by three samples at $(80\,\mathrm{A}, -40\,\mathrm{mV})$ with c unknown. Give the intercept that accompanies $R = 0.25\,\mathrm{m\Omega}$ and $R = 0.75\,\mathrm{m\Omega}$, and explain why both fit. (c) With $R = 0.5\,\mathrm{m\Omega}$, find the smallest current swing that changes the cell voltage by one step of the 1 mV cell resolution, and the swing that changes a 396-cell string voltage by one step of the 0.1 V rack resolution. (d) A 2 s current pulse occurs at a random time. What is the probability that a one-minute cell-voltage sample falls inside it?

Exercise 9.2. Take the rack heat balance of equation (9.11) with set A of Table 9.1 and a constant current of 100 A. (a) Compute the three ratios k/C_{th}, q_0/C_{th} and h/C_{th}, the time constant C_{th}/h in hours and the steady temperature rise. (b) Multiply all four parameters by 5

Table 9.4: The objects of observability and identifiability.

Object	Meaning	Subsection
independent equation	an equation not implied by the others	9.1.1
observability	states reconstructable from the measurements over time	9.1.2
identifiability	parameters uniquely determined by the measurements	9.1.3
excitation, design column	variation in the signal that multiplies a parameter	9.2.1
repeated point	many samples, one independent equation	9.2.2
string identity, r_{string}	two measurement paths to the rack voltage; their residual	9.2.3
derived channel	a channel computed from another, such as SOC_{BMS} from I	9.2.4
accounting identity	a balance that recovers an assumption, not a property	9.2.5
structural unidentifiability	parameters entering only in combinations, such as k/C_{th}	9.3.1
practical unidentifiability	effect smaller than noise or resolution	9.3.3
plateau sensitivity $\text{d}U/\text{dSOC}$	voltage information about the state of charge	9.3.4
resolution, sampling, alignment, location	the four properties of a sensor channel	9.4.1–9.4.4
boundary	a balance that tests only if a measurement crosses it	9.5.1
degrees of freedom	independent equations minus unknowns	9.5.3
parameter trade-off	a combination determined, its parts not	9.5.5
experiment	a source of the variation one parameter needs	9.6.2
capability sequence	represent, determine, trust, detect	9.6.4
eight-step mental model	the questions asked before any fit	9.6.6

and show that the ratios, the time constant and the rise are unchanged. (c) Increase k alone by 10 % and compute the new rise; then find the value of h that, with the original k, gives the same rise. (d) An independent measurement gives $C_{\text{th}} = 2.5 \times 10^6\,\text{J/K}$ and the ratios are as in (a). Find k, q_0 and h.

Exercise 9.3. The BMS of a rack assumes $C_{\text{assumed}} = 300\,\text{Ah}$; the true capacity is 278 Ah. A discharge moves 150 Ah, measured exactly. (a) Compute the change in reported state of charge and the true change. (b) Show that fitting C to the reported state of charge and the current with equation (9.10) returns 300 Ah, and state what the residual of that fit tests. (c) The current sensor now has a bias of $+0.5\,\text{A}$ over the 10-hour discharge. By how much do the integrated charge and the reported state of charge change, and why does the agreement between the two channels not reveal the bias? (d) An independent state of charge is available from two rests, with an uncertainty of 1.4 % on their difference. Using equation (9.17), what swing is needed to determine the capacity to within 2 %?

Exercise 9.4. Each of the nine racks of the reference container carries 100 A on discharge, and the sensor of rack 4 reads 3 A high. (a) Compute the bus current defined by equation (9.15) and the residual of that balance. (b) A DC current measurement I_{PCS} with independent noise of standard deviation 2 A per sample is added at the boundary. Compute the expected residual r_{bus}, and the number of independent samples whose average brings its standard deviation low enough that the residual is five standard deviations from zero. (c) Count equations and unknowns per time stamp with the new measurement and explain why the fault is detected but not located. (d) Suppose instead each rack sensor may have an unknown gain error, so that the reading is $(1 + g_r)I_r$, and the rack currents vary independently over time. What is the smallest number of time stamps for which the nine gains can be determined, and how many more are needed before the residual tests anything?

9.9 Connection to the TERRA framework

This chapter's question is the one TERRA asks before it estimates anything, in the order of Section 9.6.4: whether the measurements can represent the physical balance, whether there is enough independent information to determine the unknowns, whether the estimates are credible, and which departures from normal would be detectable.

- **The rack charge balance is a conservation node, and the reported state of charge is a derived channel on it.** Equation (9.10) is charge conservation on a rack node with the rack current as the flow on its edge. The reported state of charge of Figure 9.4 is attached to that edge's current by the BMS arithmetic, so the framework treats it as a function of the current channel, not as a second measurement, and the balance it closes carries no weight as evidence about C.
- **The series string is an edge, and the string identity is a redundancy relation.** The rack current is one flow common to the 396 cells of the string. Equation (9.7), with the per-cell voltage model as the component relation for each cell, ties two measurement paths to one quantity, and its residual r_{string} is a surplus equation of the kind counted in Section 9.5.3: the place where a sensor error or a change in R_{series} has somewhere to show.
- **A boundary with no independent measurement crossing it gives an identity.** The container node of Figure 9.7 has nine measured inflows and one unmeasured outflow, so its conservation law only defines I_{bus}. Adding a measured PCS current to the outflow edge turns the same node into a test with one surplus equation per time stamp, which detects a rack sensor error but cannot locate it.
- **The rack heat balance is an energy conservation node whose posterior stays wide along a line.** Equation (9.11) is the rack's energy node. Because the data depend only on the ratios of equation (9.12), the posterior over $(k, q_0, h, C_{\text{th}})$ remains as wide as the prior along the scaling direction of Table 9.1, however long the record; an independent C_{th} collapses it. Practical unidentifiability (weak excitation, the LFP plateau, the resolution ratios of Figure 9.5) widens the posterior in the same way, but more variation narrows it.
- **A fault is a hypothesis whose signature must lie outside what the normal model can absorb.** The temperature sensor offset of Section 9.5.7 is absorbed into q_0 and is therefore undetectable as a fault; the cell sensor error of Section 9.2.3 moves a residual that no normal parameter can cancel and is detectable. Detectability is a property of the model and the sensors together, settled before any data are fitted.

In summary, a battery record is a set of channels attached to a graph of conservation laws, and what it can determine is decided by the shape of that graph: which channels are independent, which relations have surplus equations, which boundaries are crossed by a measurement, and which parameters enter only in combination. The volume of data enters last, and only to sharpen what the graph already allows to be known. This book ends here, and hands the reader the eight questions of Figure 9.8 to carry to any record: what is measured, what is derived, which physical relations connect them, which unknowns appear, whether the measurements vary enough, whether there are enough independent constraints, whether the parameters can be separated, and only then, to fit the model.